JTCG/AS Joint Accreditation Support Activity

Credible Models for Credible Analysis . . .

"HOW TO VV&A WITHOUT REALLY TRYING"

SMART VV&A LESSONS LEARNED

November 1997

Joint Accreditation Support Activity (JASA)
Naval Air Warfare Center, Weapons Division (418100D)
China Lake, CA

JTCG/AS

JTCG/AS-97-M-00 9

EXECUTIVE SUMMARY

The Susceptibility Model Assessment and Range Test (SMART) Project was a five-year effort which developed and tested a comprehensive, cost-effective verification, validation and accreditation (VV&A) process for models and simulations (M&S) used in support of the system acquisition process. In developing and testing the VV&A process, SMART Project personnel collected a number of "lessons learned". Those lessons are documented here in a series of semi-independent articles which describe practical advice for planning and executing various elements of the process.

This Lessons Learned document is intended to provide advice to users of M&S in the acquisition community on how to accomplish effective M&S accreditation at minimum cost. It is organized around the same table of contents as the Joint Accreditation Support Activity (JASA) VV&A process description, and it is intended to be a companion to that document. It is hoped that the articles in this document will give the reader valuable insights into practical VV&A applications, and how to put together a cost-effective VV&A program.

The "lessons learned" described in the articles in each section of this document are organized around the top-level work breakdown structure (WBS) elements of the VV&A process. This WBS was developed over the five years of the SMART Project, and its elements are the result of a number of demonstrations of the process for actual acquisition program customers. The articles in this document are the lessons gleaned while performing VV&A activities for those real customers.

TABLE OF CONTENTS

MEET THE AUTHORS

David H. Hall - NAWCWPNS. Mr. Hall was the SMART Joint Project Manager, responsible for all aspects of the program; based on the overall SMART project experiences and DoD community response, he edited this document (all of the italicized comments are his) and authored: *"The Benefits of Standard Reporting", "Statistical vs Analytical Significance: How Much V&V is Enough", and "Integrating V&V Into Development: The Importance of Configuration Management"*

Michelle Kilikauskas - NAWCWPNS. V&V Manager for SMART, Ms. Kilikauskas also was responsible for assisting acquisition programs with use of SMART developed V&V data to support their accreditation requirements. Based on that experience, she has authored: *"The SMART Road to Accreditation: Lessons from the Front"*

Dennis K. Laack - Computer Sciences Corporation. Mr. Laack developed the bulk of the VV&A training course produced by the SMART project, which grew out of his experiences in applying V&V techniques to M&S accreditation requirements for various programs. Based on that experience, he has authored: *"Defining Accreditation Requirements: A Common Sense Approach", "Translating Accreditation Requirements into V&V Requirements", "Getting Expert Results from Expert Reviews", and "Accreditation Support Proces: More Lessons from the Front"*

Dr. Paul R. Muessig - NAWCWPNS. Dr. Muessig, the SMART Navy Deputy Project Manager, was responsible for all technical products developed under the project. Based on his oversight of the overall aspects of VV&A activities, he has authored: *"The Cost of Credibility" and "The Functional Approach to V&V: Why Bother?"*

Barry O'Neal - ASI Systems International. Contract team leader for the SMART project, Mr. O'Neal was responsible for the successful completion of all technical products . Based on his extensive V&V experience, he has authored: *"Everybody Else Does It Better", "Verification or Validation: Which One When?", "Building A Functional Area Template", "The Functional Approach to V&V: Why Bother?", and "The I in IV&V: Independence or Intelligence?"*

Chester Richardson - NAWCWPNS. T&E Manager for SMART, Mr. Richardson was responsible for coordinating test data collection from a number of tri-service test activities in support of model validation efforts. Based on the experiences gained in that effort, he has compiled and edited the article: *"Data Specification and Collection for Model Validation"*

Karl Simecka - Arrowhead Technologies, Inc. Mr. Simecka was responsible for coordination of most of the test data collection activities conducted for SMART. He is a co-author of the article: *"Data Specification and Collection for Model Validation"*

Willie Stewart - SIMSUM, Inc. Mr. Stewart was responsible for actual collection of much of the SMART test data. He also is a co-author of: *"Data Specification and Collection for Model Validation"*

Stewart W. Turner - SWTI, Inc. Mr. Turner was responsible for most of the rest of the test data collected for SMART. He also is a co-author of: *"Data Specification and Collection for Model Validation"*

I. INTRODUCTION

The Susceptibility Model Assessment and Range Test (SMART) project was a five-year effort which developed and tested a comprehensive, cost-effective verification, validation and accreditation (VV&A) process for models and simulations (M&S) used in support of the system acquisition process. In developing and testing that VV&A process, SMART project personnel collected a number of “lessons learned”. Those lessons are documented here in a series of semi-independent articles which describe practical advice for planning and executing various elements of the process.

The use of models and simulations (M&S) is growing throughout the Department of Defense (DoD) to support acquisition programs, to analyze strategic and tactical alternatives in conflict situations, and to train personnel from the unit to the command levels. This increased usage is being spurred in part by the increasing capabilities of computer systems, which in turn results in larger and more complex simulations of systems, operations, and environments.

Recognizing that simulations have been misused, especially in acquisition support applications, the DoD has established a requirement that all M&S used for significant purposes be accredited. (Accreditation is defined by DoD as the official determination that a model or simulation is suitable for a particular application.) Obviously, accreditation involves comparing the strengths and weaknesses of the selected M&S with the requirements of the application. This is done to determine if any unacceptable analytical results or operational impacts would result due to one or more M&S weaknesses or limitations.

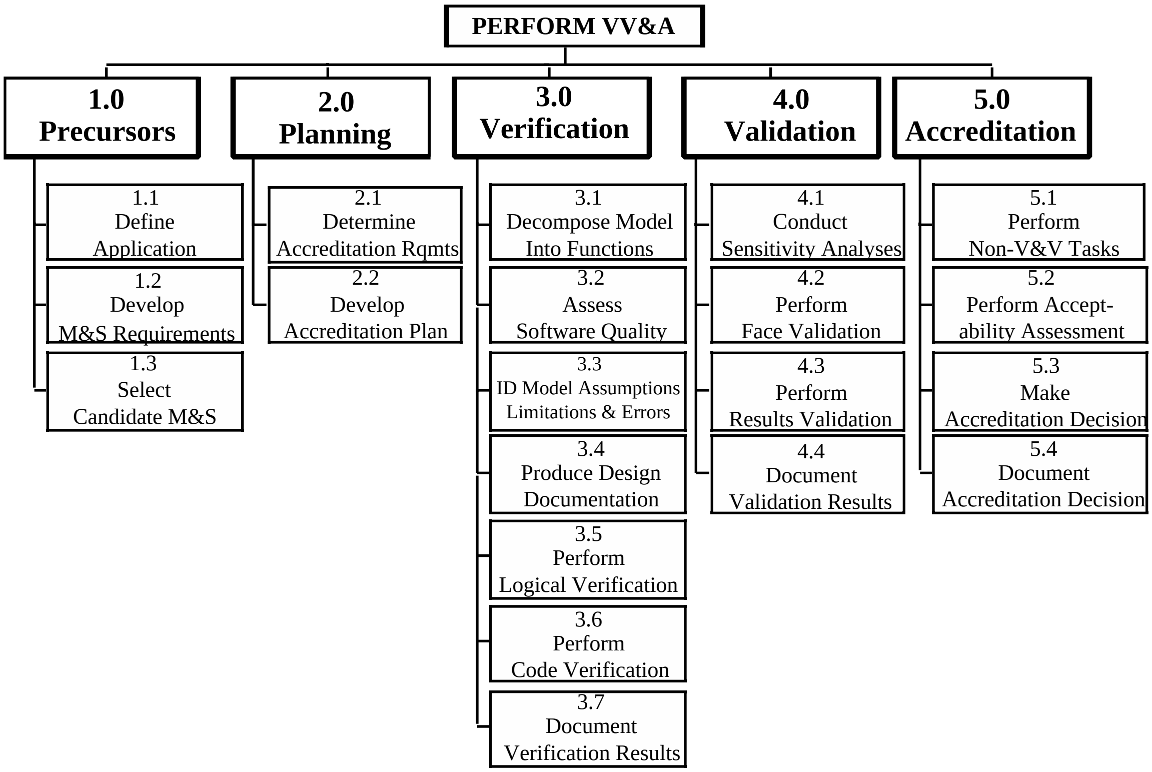
This document is intended to provide to users of M&S in the acquisition community advice on how to accomplish effective M&S accreditation at minimum cost. Its structure parallels that of the VV&A process description[1], and it is intended to be a companion to that document. It is hoped that the articles which follow will give the reader valuable insights into practical VV&A applications, and how to put together a cost-effective VV&A program.

---

[1] V&V from A to Z: A SMART Approach to Accreditation for Acquisition M&S, October 1997

2. THE VV&A WBS

The work-breakdown structure (WBS) for VV&A activities is shown in the figure. The "lessons learned" described in the articles in each section of this document are organized around the top-level WBS elements shown. This WBS was developed over the five years of the SMART project, and its elements are the result of a number of demonstrations of the process for actual acquisition program customers. The articles which follow are the lessons gleaned from performing VV&A activities for those real customers.



VV&A Work Breakdown Structure

3. WBS DETAILS

The VV&A process description, the companion document to this report, contains detailed descriptions of each of the tasks identified under the work breakdown structure. We don't intend to repeat those descriptions here; rather, this document is intended to describe some of the philosophy behind the generation of those WBS tasks, some practical lessons learned in trying to implement the WBS, and some traps and pitfalls to try and avoid in doing VV&A.

3.1 PRECURSORS

Precursors are those things that should be done in setting up the basic analysis or training problem before you even start to plan VV&A tasking. Ideally, these things will be done for you, the VV&A agent. Practically, they seldom will be done well, if at all, and you'll have to coax and cajole your customer into helping you get them done. But most of the time you'll have to do some of this yourself, and hope that your customer will agree that what you did adequately represents the M&S application under study.

DEFINING M&S REQUIREMENTS: A COMMON-SENSE APPROACH

by Dennis Laack

Perhaps the most important aspect of planning for VV&A activities is deciding exactly what you need a model to do. If you don't lay out exactly what questions need to be answered, what measures of effectiveness (MOE's) are going to be used to answer them, and how model outputs are going to help develop those MOE's, then you have no hope of deciding whether a model meets the program's needs or not. The following article describes some of the lessons learned while trying to help programs describe their M&S requirements. (ed.)

Since accreditation is defined as the official determination that a model or simulation is suitable for a particular application, it involves a comparison of the M&S requirements with the attributes of the selected model. This comparison of the M&S strengths and weaknesses with the M&S requirements is typically done in one of two ways. The first way is to explicitly define the requirements of the intended application as a basis for evaluating the M&S assumptions and limitations. The second way is to have a group of experts who understand the intended application conduct a review of the selected M&S to determine if any limitations or deficiencies might cause an unacceptable outcome. Lessons learned in using the first approach for development of accreditation rationale are the subject of this article.

M&S REQUIREMENTS - WHAT ARE THEY?

M&S requirements (sometimes called modeling requirements or acceptance criteria) are statements of what the model or simulation is required to do in order to analyze the problem at hand. These requirements also address whether the model is operationally compatible with the intended operating organization. These statements of the required model capabilities can be grouped into three primary categories: operating requirements, functional requirements, and fidelity requirements. Sometimes data requirements are also included. Each of these requirements is explained more fully in the following paragraphs.

Operating requirements identify the hardware and software (HW/SW) that are available to operate the model, any networking that is planned to facilitate data exchange either within a single simulation or between multiple simulations, any planned processing of input/output data, etc. These plans that address how the model will be operated are really statements of required model operating capabilities. If the model possesses these capabilities it can be used to fulfill the plans.

Functional requirements are a listing of the important attributes of a problem that must be represented or simulated by the model(s) that will be used in the application. This listing addresses system features and/or functions, physical phenomena, political and environmental conditions, and personnel or unit actions that have an important impact on the ultimate solution of the problem. Any interactions between these represented entities that are important to the problem must also be represented.

Fidelity requirements are, quite simply, the requirements that specify how well or how accurately each function must be represented within the M&S. Fidelity requirements are typically stated in terms of permissible tolerances on functional parameters that are calculated by the model. They may also be stated in other terms (e.g. a specification of certain sub-functions that should be included in the functional representation to accurately portray the overall function).

Data requirements typically address both model input and output parameters. The input data requirements are specifications that describe what data inputs are needed for the application as well as the accuracy and reliability requirements to make the model outputs sufficiently credible. Output data requirements are the list of parameters that must be produced by the model to provide necessary information that will be used to resolve the problem. (The definition of data requirements is included herein for completeness. However, these requirements will not be addressed further in this article.)

HOW ARE M&S REQUIREMENTS USED?

M&S requirements are primarily used for two purposes. They provide a basis for initially selecting a model that will be suitable for the intended purpose or application. They are also the basis for accrediting the model (this is where the term "acceptance criteria" originates). M&S requirements are derived through a detailed analysis of the problem that is to be solved. This analysis is very similar to the analysis that is done to define the requirements for a new model. In fact, if no model matches these M&S requirements, they essentially form the "conceptual model' which is the starting point for a model development effort.

In analyzing a problem to develop the M&S requirements, the first step is to develop a clearly documented problem statement that articulates all the essential concepts and ideas. Following that step, the overall problem requirements are defined and the approach for problem resolution is determined. For that part of the problem which will be addressed with M&S, the application parameters are defined. These lead to development of the M&S requirements.

To help clarify the relationship of the problem statement and application parameters to the resulting M&S requirements, and to show how these items affect the problem resolution, it is necessary to understand the terms *problem* and *application*. To help explain these terms, it is useful to look at a diagram of the problem life cycle shown in Figure 1. The entire diagram is the problem resolution process, and the questions or issues being addressed constitute the *problem*. Certain parts of the problem or selected pieces of information might be generated by using M&S. These parts of the problem are the *application* for the M&S.

The first two steps in this process, "develop problem statement" and "establish problem requirements", are the basis for the entire problem resolution process. These two factors determine whether M&S or some other approach will be used to gather data to support the decision or final outcome. Within the M&S application, the problem requirements are the basis for deriving the M&S requirements which, in turn, drive the selection and accreditation of the most appropriate model. The diagram also shows that the M&S requirements are the basis for determining verification, validation, and accreditation (VV&A) requirements which initiate the entire VV&A process, represented by the lighter shaded portion of the diagram. The problem statement, problem requirements and M&S requirements are critical seminal steps to successful problem resolution and, from a VV&A perspective, can be considered as preparatory steps.

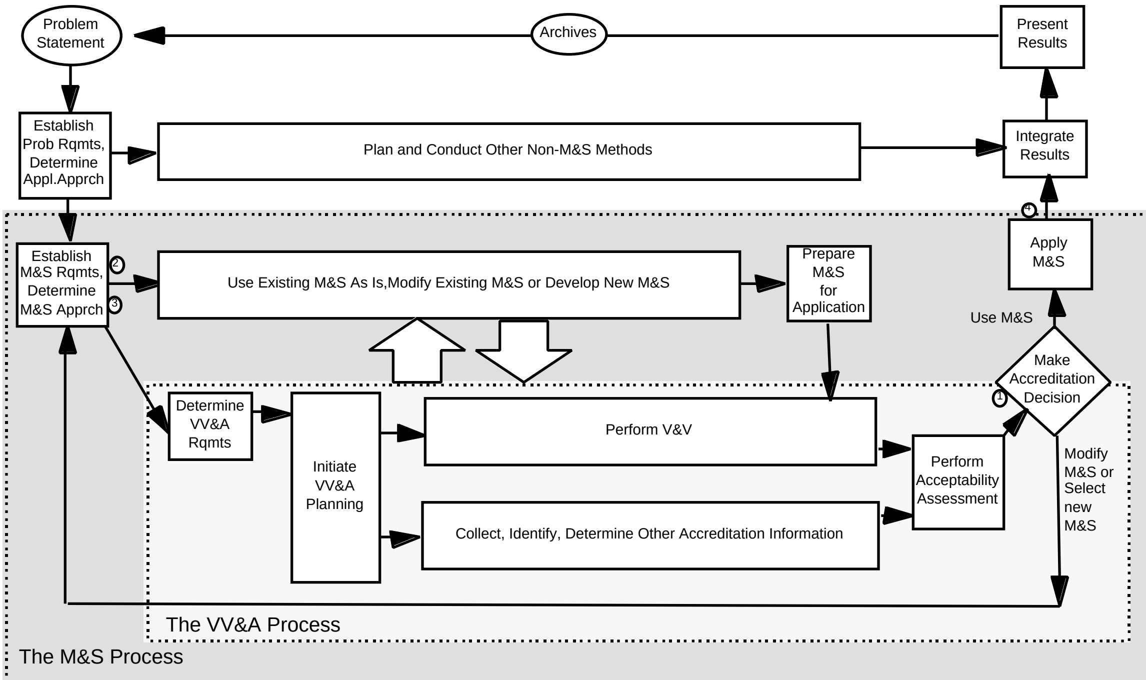


Figure 1 Problem Life Cycle

These preparatory steps are theoretically done prior to initiation of the VV&A process. In practice, M&S selection and usage seldom follow the ordered, sequential process depicted in Figure 1. Instead, a decision to use the M&S process, selection of the M&S, and initiation of the VV&A process are frequently all done simultaneously and without the benefit of a formally documented problem statement or any clearly specified requirements. M&S selection is usually based on the sponsor's general understanding of the problem and its requirements. Often these problem requirements are either poorly documented or not documented at all, and usually they are not well understood by the analysts who, in turn, must define the application, run the model, and interpret its results.

The difficulties that arise from this simultaneity are often compounded by the fact that the people managing the problem are often uninformed about VV&A issues and details. In many cases, the managers assign VV&A responsibilities to independent government organizations or contractors who are unaware and uninvolved in the basic problem. These typical actions exacerbate inherent VV&A difficulties and can result in wasted resources. The pitfalls related to VV&A preparations and the impacts of using independent organizations to plan V&V without adequate definition and guidance are part of the motivation that led to the process recommendations explained in the remaining sections of this article.

VV&A preparation and planning is made more difficult by the lack of common definitions for many V&V terms and techniques. Terminology and the actual tasks performed as part of any technique vary from one application community to the next. Managers often do not even know the accepted definitions of "verification", "validation", or "accreditation". This lack of common terminology makes clear communication between the managers, modelers, VV&A proponents, and V&V practitioners extremely prone to misunderstanding.

One clear solution to these difficulties is to establish and follow a logically structured process to develop the foundations for a successful VV&A effort. This process would be enhanced if managers had a basic VV&A orientation and all supporting organizations became fully acquainted with the problem, its objectives and constraints, and its metrics. Such steps, although bothersome, are critical to avoid false starts and poorly focused activities.

APPLICATION DEFINITION - A PREREQUISITE TO IDENTIFYING M&S REQUIREMENTS

The basic difficulty usually encountered in planning VV&A activities is that the problem being addressed and the role of M&S within the problem have not been well defined. This is not to say that the program managers have not defined their objectives; rather, they have usually not defined them in terms that can be readily applied to the question of M&S credibility: "How good do the M&S need to be to resolve my problem?" Very often the problem definition has not been documented, nor has the use of M&S outputs been rigorously correlated with the MOEs that address the questions to be answered.

Problem Description

The first step in defining the application is to clearly specify the problem along with the its purpose and the desired objectives. In a typical acquisition problem the purpose is usually to determine the best alternative design or to decide whether to proceed with a particular program. The objectives of such an acquisition problem are to achieve the best performance at minimum cost (acquisition and/or life cycle) or to determine if the performance of a system is above specified thresholds. These goals will drive the criteria that will be used for selecting and approving a model.

For non-acquisition related problems, the purpose and objectives may be specified in a different manner. Typical operations analysis problems can be posed as questions that deal with selection of investment or warfighting alternatives. The purpose of a typical analysis problem might be to define tri-Service force structure requirements for the 21st century. The objectives might include a list of warfighting goals or capabilities that are desired to be met with this force structure.

The problem purpose and objectives for a training application would not normally be stated in the form of a question. In these cases, the statements would normally describe the training goals and objectives that are to be achieved. For example, the purpose of a training exercise might be to train corps level commanders in how to deal with communications blackouts caused by enemy information warfare (IW). Note that it is not sufficient simply to identify the purpose as "training".

In any of these cases, the length of the description is not as important as its technical accuracy, and the inclusion of important factors that uniquely describe the problem, and which can be used to derive M&S application requirements. The description of the problem purpose and its objectives serves as the touchstone against which objective statements of model capabilities and model accreditation requirements can be defined. Poorly defined problems lead to poorly defined (or, at least, objectively unjustifiable) requirements for accreditation.

Note that the description of the purpose and objectives differs for the different application types. Therefore the type of application (acquisition support, operations analysis, or training) must also be specified along with any specific questions or issues that are explicitly identified.

Another element of the problem description is a definition of the metrics that are to be used to quantify the pertinent information, the threshold values that will control decisions or actions, and the critical parameters that must be calculated or estimated as a basis for solving the problem. One of the definitions of a *metric* is "a standard of measurement". In the context of problem solving, metrics are the means of measuring or quantifying the information that will be used to resolve the issues and answer the basic questions. Most non-training problems deal with making a choice between competing alternatives based on the tradeoff between some measure of cost, risk, and benefit. In this context, metrics are used to quantify the costs, risks, and benefits and to provide some means of ranking the different alternatives. These metrics are generically termed “measures of merit”, or MOMs. The ultimate decision or choice between different alternatives is usually determined by the specific values of critical metric(s). If the value of a particular decision metric, as quantified by the appropriate MOM(s), is above a predetermined level, one choice is made. If it is below that level, a different choice is made. This predetermined level is referred to as a “threshold”.

The term “measure of merit” is actually a generic term for the metrics that constitute a hierarchy of three reasonably well-defined and understood levels of evaluation. The lowest (most detailed) level deals with individual systems or components of individual systems. The metric typically used to quantify this lowest level is the Measure of Performance, or MOP. MOPs quantify parameters that characterize a system or its functionality. Typical MOPs that might be used to characterize a person might be: height, weight, visual acuity, speed, IQ, etc. Typical MOPs for an aircraft might be: maximum altitude, fuel consumption, maximum range, turning radius, maximum load, radar cross section, rate of climb, etc.

The next higher MOM is the Measure of Effectiveness, or MOE, which quantifies the parameters used to evaluate the interactions between two or more systems. For example, in an engagement between one (or more) aircraft and one (or more) anti-air defensive systems, some typical MOEs might be: probability of kill, probability of survival, weapon miss distance, threat missile leakage rate, etc. Frequently (albeit not necessarily) the integrated effects of several MOPs might be used to determine an MOE. Using the previous example, the probability of survival of a combat aircraft is influenced by the aircraft MOPs of speed and maneuverability as well as MOPs related to the capabilities of the defensive systems.

Finally, the highest (least detailed) MOM is the Measure of Outcome, or MOO. A MOO is used to quantify the effects or outcomes of large scale, multiple interactions between numerous different groups of systems and organizations. Some MOOs might be: the time required to capture specific campaign objectives, numbers of tanks destroyed, number of friendly casualties, amount of territory recaptured, etc.. Some of these MOOs, such as time to achieve objectives, can be influenced by the MOEs for aircraft engagements with air defenses, such as probability of survival or probability of target destruction.

To determine the most appropriate MOMs for a problem, the first step is to determine what information is critical to the solution of the problem. That is, what information will be used to make the necessary intermediate judgment(s) that support the ultimate decision? Identifying this necessary information requires a clear understanding of the overall problem’s purpose and objectives. The parameters that are normally used to quantify this type of information should be reviewed, and those most appropriate to the specific purpose selected as the metrics. The units in which these metrics will be quantified should also be selected if multiple choices exist.

Having specified the “units” in which each MOM will be expressed and measured, the next task is to identify how each MOM will be determined: either calculated directly by a model or from other parameters that are produced by models or from other sources of data. In some cases, a MOM will be a single parameter (e.g., probability of kill, speed, etc.). In other cases, a MOM may be some mathematical (or other algorithmic) combination of a number of different parameters (e.g., cost per operating day, number of logistics flights per day, etc.). In all cases, it is important to be judicious (and even parsimonious) in the parameterization of problem MOMs. It must be kept in mind that the evaluation of each MOM parameter will require a data source (either a model output, a combination of model outputs, a subject matter expert’s opinion, etc.). Also, the relationship between MOMs, MOM parameters and parameter data sources can easily add a level of complexity that may obscure rather than facilitate the analysis process.

Having defined the MOMs, the thresholds can now be defined. Depending on the particular problem, MOM thresholds may be more or less rigid. COEAs associated with acquisition programs tend to have rigid thresholds for selected performance parameters (e.g., weight, range, life cycle cost, etc.). Other problems (such as evaluation of alternative strategies or tactics) might have less rigid thresholds. In some cases, a choice might hinge on the relative size of a parameter rather than on its absolute value. Thus, there is usually an acceptable "tolerance" in the value of each MOM about the threshold that will not cause the final answer to the problem to change. Acceptable tolerances in a problem's MOMs should be defined based on a judgment of how much error is acceptable in the MOMs without affecting the final (or intermediate) decision(s). In some cases there might not be any threshold. Instead, alternatives might be compared on the basis of their relative ordering in terms of one or more MOM values.[2] The only requirement on the MOM tolerance would be that the MOM sensitivity to its constituent parameters was the same for both alternatives.

Definition of MOM thresholds has an important bearing on the specification of model fidelity requirements. Where possible, the acceptable MOM tolerances should be estimated so that analysts can have some idea of the accuracy needed in their work. Tight MOM tolerances can lead to stringent fidelity requirements for those model outputs to which they are related, which, in turn, can significantly increase the requirements for detailed V&V. On the other hand, undefined or arbitrarily loose MOM tolerances permit no causal relationship to be defined between model credibility and problem solutions. In that case, there is no objective way to determine whether or not model outputs are sufficiently credible for the application at hand, a situation which may well call into question not only the entire V&V effort, but the conclusions derived from M&S use. It is, therefore, well worth the time spent in defining MOM tolerances and thresholds.

Tolerances are typically overlooked in problem definition efforts since any decision thresholds are normally discrete values without a specified error tolerance. Thus, it may be necessary to raise the issue of acceptable tolerances with the accreditation authority early in this phase of the process; this will ensure that adequate consideration can be given to this question prior to expending extensive resources on an inappropriate model or on excessive V&V efforts.

In addition to the problem statement and the metrics, another critical part of the problem that must be articulated is the problem importance. The term *problem importance* refers to the relative weight attached to the problem outcome or decision. It does not refer to the importance of the model outputs that are used to support the problem decision. The importance of any problem can usually be described by defining the benefits expected to result from a "good" decision or outcome, the impacts of a "wrong" outcome, who or what will be impacted (a segment of society, individuals, equipment, and/or the environment), and how widespread these effects will be felt.

Potential problem impacts are the major effects that might be realized due to a choice or problem answer. Some of the different types of impacts are: cost (e.g., acquisition, maintenance, logistics, life-cycle, etc.), schedule, safety, public reaction or opinion, warfighting capability, likelihood of conflict, budgetary impacts, environmental impacts, etc. Each application may have several types of possible impacts that could result from each particular outcome or answer. The greater the impact, the greater the need to ensure that the information used to make the decision is accurate.

As indicated previously, the organizations or entities that are affected by the decision or problem outcome must be identified. These organizations might be either within or outside the DoD and might be any size from an entire country to a small military unit. The broader the scope of the impact in terms of who is affected, the greater the risk associated with accepting model results uncritically (i.e., with little, inadequate, or no attention to V&V).

---

[2] For example, in comparing two different tactics for employment of an on-board radar jammer, one MOE might be the miss distance of a threat missile. Evaluation of the different tactics would then be evaluated on the basis of which tactic yielded the greatest miss distance.

Understanding problem importance is necessary because it will be to determine the credibility requirements for the models that are selected and ultimately to determine the necessary level of V&V to be done. For example, if V&V activities identify what appear to be serious deficiencies in a model, it would be helpful to know how to assess the relative importance of those deficiencies with respect to concrete criteria related to the overall importance of the problem.

Consider a hypothetical missile acquisition program where some analysis must be done to select the most appropriate sets of test conditions from the entire conceptual operating envelope. If a particular model has been selected for use in planning the tests, and it is discovered on the basis of validation results that the model does not represent some missile function accurately enough, an assessment of the impact of this deficiency on the final testing conditions must be made. If this deficiency results in the missile range being underestimated, and the location of the test has limited geographical space, the deficiency can impact safety, is probably critical, and must be corrected. On the other hand, if the missile range is overestimated as a result of the deficiency, or if the test location is not geographically limited, the impact may be negligible and it may not be necessary to address the problem in as rigorous a fashion as it would be if human safety were at stake. Deficiencies in realism in this case might be mitigated with verbal caveats given to the test participants.

Another aspect of problem importance is its visibility. This visibility can vary in scope (from local to national or international) and in the degree of influence that can be exercised by the personnel who become aware of the outcome. The effects of this visibility can be positive or negative, depending on the nature of the outcome and the perceptions of those who become aware of it. Using the test planning example, the only visibility that might be expected would be to the next higher organization over the testing organization, and to the program manager who sponsored the tests. However, if there are serious problems with the tests, especially safety violations affecting personnel outside the military, greater visibility will be likely. The higher the visibility of the problem outcome, the more risk there is with the use of models whose credibility cannot be demonstrated.

The sponsor’s statements or concerns regarding the importance of the problem outcome are necessary elements of any statement of importance. It is true that an application sponsor’s specific concerns might be captured implicitly by the other statements regarding problem importance. However, it is highly desirable to ensure that these sponsor’s concerns are captured explicitly so that they can be factored into the level of risk that should be associated with model deficiencies uncovered during V&V. Some possible sponsor concerns related to the importance of the test planning problem might be: 1) possible program cancellation due to test failures or tests that do not obtain adequate information, 2) impacts of test failures on career prospects, 3) possible international interest in a system if tests are successful, etc. These concerns are related to the effects of the problem outcome and not to the direct outputs of the model.

Application Parameters

The problem statement should be accompanied by a description of the application (that part of the problem that will be addressed using M&S). The application description includes a statement of the specific goals and objectives that are to be achieved with M&S. It also includes a summary description of the scenarios and elements that are to be addressed in the analysis. The term “scenarios” includes: geographic areas and conditions (type terrain, weather, political climate, etc.); participants (systems, units, environmental operators); and boundaries (conditions or limits established for the problem). Finally, the application description identifies all the specific questions or issues that are specifically called out in program documentation as questions to be resolved through modeling.

A scenario description includes two components; the physical environment and the operating environment. The physical environment encompasses the natural and man-made surroundings in which the system operates or performs. It includes the terrain, construction (roads, buildings, bridges, etc.), oceanographic conditions, atmospheric and meteorological conditions, electromagnetic environment, etc. The operating environment is the set of non-physical conditions that affect system operation. For a theater level engagement simulation, the operating environment includes the rules of engagement, political considerations, strategic and tactical doctrine, etc. For an engineering level model, the operating conditions might include procedures that are imposed by system design or by operating manuals.

The organizations, forces, and systems that are involved in a scenario definition are readily apparent for higher level models. For example, the elements of a theater level engagement scenario (that might be used to evaluate the militarily significant contributions of a new aircraft) would include all the operating forces or organizations (corps, regiments, battalions, platoons, airborne strike units, logistics units, amphibious forces, etc.) and the systems that are employed (aircraft, weapons, communications, ships, tanks, etc.) which have unique properties that impact the problem under study. Applications wherein engineering level models are used also have “scenarios” in the form of a set of conditions in which the system performs. A scenario for an engineering level model might include the system being modeled (man-made or organic system); any other systems that affect the performance or operation of the system being modeled; any human interfaces or controls; and the environment in which the system will operate. Scenario descriptions for most other types of models would fall somewhere between these two (a theater engagement and an engineering model) and the scenario elements would include some combination of those indicated for these two types of models.

Another element of a scenario description is the scenario boundaries, which include all the limitations on the scenario imposed by the application. They include the geographical limitations, time constraints, operational and safety limitations as well as limitations or constraints that might be imposed due to the nature of the problem or environmental impacts (e.g. a missile test might be constrained as to the operating modes that can be used due to potential electromagnetic interference with nearby facilities.

In some applications there may be a critical combination of scenario conditions that will significantly affect the problem outcome. One example of such a problem is found in the liberty ships of World War II. The problem of fatigue cracking was accelerated due to the extreme cold of the North Atlantic. This combination of scenario conditions (cold temperature and vibration stresses) caused a unique problem that would not have been noted in other scenarios. In any application, the potential influence of different combinations of conditions should be postulated and, if any significant consequences seem possible, that combination should be identified as part of the scenario definition.

Finally, the scenario description should include a list of all simplifying assumptions. Any and all assumptions that are made either at the program level (to constrain the problem) or at the analytical level (to facilitate a problem solution) should be identified and documented. Simplifying assumptions at the program level are typically top level assumptions (e.g. nuclear weapons will not be used, strategic intelligence sources will or will not be available, etc.). Assumptions at the analytical level are typically focused on more detailed aspects of the problem, and are made to eliminate unnecessary complexity (e.g., personnel will be assumed to be trained and fully capable of performing their assigned functions).

Preparing a problem statement initially appears almost trivial. However, in trying to prepare such a document, the most difficult aspect is often determining that it is complete and comprehensive. The traditional, and apparently the “best” approach to defining the problem statement is to use some sort of checklist that lists numerous different types of factors that should be addressed when preparing the problem statement. The Work Breakdown Structure (WBS) developed by SMART is a good starting checklist for developing a set of M&S requirements. During the course of using the problem statement to define modeling requirements and to tailor the V&V efforts, questions may arise which are not answered by the written statement. In these cases, it is often necessary to review the problem and add to the statement any missing or overlooked information.

The whole purpose of explicitly defining the problem is to provide a common, well-understood starting point for the analysts who must define the application details, run the model, and analyze the resulting data to arrive at a recommendation. Without such an explicit statement, there is a significant possibility that analysts, in a sincere attempt to avoid errors, will make conservative assumptions that can lead either to gold plating the model or to unnecessary V&V. A clear problem definition, explicit application description, and specific M&S requirements are the manager's best tools for controlling M&S (and V&V) costs.

Independent organizations who are detached from the primary problem solving organization cannot define the problem. If such organizations are assigned VV&A responsibilities, their reasonable and proper role would be to facilitate problem definition by developing basic questions, synthesizing the answers into a comprehensive problem statement, and documenting the final problem statement. Application sponsors and problem analysts are the only ones who have a broad enough understanding of all the problem ramifications to adequately develop a problem statement that would provide a sufficient basis to develop M&S requirements. When preparing statements of work or work agreements for personnel or organizations who will perform V&V tasks, the objectives, tasks, and division of responsibilities should be clearly spelled out. Otherwise, organizations that do not have the requisite background may be assigned total responsibility and may ultimately produce an inadequate product.

Motivating sponsors and problem analysts to participate in problem definition is often difficult. Briefings on the concepts and the need for this activity are seldom sufficient motivation. Preparation of a VV&A plan with questions replacing sections may be helpful. Also, preparation of the accreditation report outline, which includes specific statements of needed items or questions relating to the problem definition, can help motivate sponsors to clearly define the problem. Wherever possible, any VV&A efforts that have already been undertaken should be recognized and incorporated into these draft plans and reports. One sure way to discourage people is to ignore or belittle what has already been done.

HOW ARE M&S REQUIREMENTS DERIVED?

If modeling requirements tailored to the problem at hand are defined before any model is selected for an application, there is a greater chance that they will be developed objectively, and that any possible bias towards a particular model will be minimized. The use of objectively developed requirements also simplifies the review of V&V (and other) data used to support an accreditation decision, because the comparison between available model information and clearly defined criteria leaves little room for misinterpretation. Modeling requirements can be grouped into three categories: Functional Requirements, Fidelity Requirements and Operating Requirements.

Operating Concepts

As mentioned previously, operating requirements are necessary attributes of the model that make it usable by a particular group of analysts. These attributes include hardware and software compatibility features, documentation that explains how to input data and run the model, training in model use that might be available, and any other features that allow the user to intelligently use the model. There is no magic technique for defining operating requirements. A simple and straight-forward approach is to identify all the different sets of hardware systems and operating software that are available to the program's analysts. Also, considering their experience with different models, any training requirements can be specified for models other than those with which they are familiar. If there are any plans for automated data pre- or post-processing, they lead to other requirements for model compatibility with available data processing software and hardware.

A checklist of operational considerations can be prepared over time as additional analyses are performed using a variety of models. This checklist would not be a listing of requirements, but would be a listing of the different types and categories of requirements. Such a checklist can then be used in future VV&A efforts to serve as a guide in selecting the minimal set of operating requirements that will allow the model to be run and produce credible results.

Often, the selection of M&S is made based on what is available, rather than on what is required. Many times analysts select those M&S that are already running on their computers and with which they are familiar. While it is true that operating requirements need to be considered in the selection of M&S, and that schedule pressures may force programs into making do with what they already have in the way of M&S and hardware, it is also true that using operational requirements as the principal basis for M&S selection is not the best way to do business. M&S selection should be based primarily on the analytical (functional and fidelity) requirements of the problem definition, and operating requirements should come second. By identifying problem requirements in detail up front, a program should have the time to plan an operating concept which allows for acquisition of the necessary operating resources that satisfy the primary M&S requirements.

Functional Requirements

Functional requirements are a listing of those aspects of reality (e.g., individual system performance characteristics, natural phenomena, interactions between systems and the environment, interactions between systems, operator roles or decisions, etc.) that a model must represent or address in order to be of use in the problem at hand. Depending on the fidelity requirements and the effects of dynamic interactions, these representations can be in the form of a look up table, a single parameter, an pre-set input parameter, or a calculated output using a number of inputs and a set of mathematical operations.

There are basically three different approaches to defining these functional requirements: a formal approach, a semi-formal approach, and an informal approach. The formal approach uses the Knowledge Acquisition/Engineering process described in the Naval Synthetic Force Simulation System V&V Plan dated 1 April 1996. This process involves a series of well defined steps starting with a domain analysis, functional decomposition and task analyses, development of a list of tasks and systems, preparation of system descriptions and task descriptions, preparation of logical diagrams for each system and task, definition of metrics and thresholds, and preparation of functional requirements. The end product is a validated conceptual model which is, in essence, a statement of functional requirements. Fidelity requirements may also be developed with this approach.

This formal approach is most appropriate for large, complex problems that involve numerous systems and a variety of forces all interacting with each other. It is also appropriate in those applications where some third or fourth level function can have a significant effect on the outcome. Since this approach will entail the greatest expenditure of resources, it is most appropriate for only the most important (and well funded) applications.

The semi-formal approach bypasses the above formal steps. Instead, a team of subject matter experts is convened to review the application and collectively identify the functional requirements based on their expert understanding of the problem. The process is enhanced by the use of a checklist of possible functions that should be developed prior to even looking at the application. This informal approach is well suited to problems or applications that are not quite as complex and/or do not have the same level of importance (or funding).

The informal approach to developing functional requirements is similar to the semi-formal approach. A small group of analysts or even a single analyst engaged in resolving the problem compile a list of functions that are necessary to adequately model the problem. This approach is only appropriate for very simple applications and ones where the level of importance does not warrant greater expenditures.

Fidelity Requirements

Fidelity requirements define how well or how accurately each function must be represented. The general approach to defining these requirements is to identify the acceptable tolerances on the parameters that affect the outcome of the problem or application and flow them down to the model and functional level. The process is similar to that used in determining an error budget for a system. The difference is that an error budget is created using the principle that allowed tolerances are calculated for a given parameter assuming that other parameters vary to their maximum or minimum allowed values. In calculating fidelity requirements, all other parameters, except the one for which the allowed tolerance is being calculated, are expected to be free of any errors. This assumption is made out of necessity. Considering the large number of functions in most models and the relatively stringent tolerances that are expected in most applications, it would be impossible to use a strict error budget approach. Such an approach would yield tolerances that could not be met with any model that is built using today's technology.

There are three approaches to determining fidelity requirements. The first is purely subjective and uses expert judgment to translate allowed tolerances at the decision information level to acceptable tolerances at the functional level. This approach is certainly the least costly and most appropriate for simple simulations and where VV&A efforts are underfunded.

The second, or "semi-analytical" approach uses a formal sensitivity analysis to relate tolerances at the overall problem decision level to acceptable tolerances on each model output. These tolerances on model outputs are essentially the model level fidelity requirements. To determine the functional level fidelity requirements, model experts use their judgment to identify critical internal functions within the model and to assign appropriate fidelity requirements to these functions. This approach is most appropriate in those applications where the M&S outputs are used either in a model or a mathematical equation to produce information that is directly applicable to the problem outcome. In these cases, a classical sensitivity analysis can be used to relate problem level parameters and tolerances to model output tolerances.

Explicit error budgets generally are not possible because of the large number of functions and the consequent very small tolerances that would be allowable at the functional level. Some subjective judgment is necessary in flowing the tolerances down to the functional level.

The third, or "analytic" approach employs a two-tiered formal sensitivity analysis to the application to determine: first, the acceptable tolerances on the model outputs; and second, acceptable tolerances at the functional level. Because of the large number of functions within most models it is not economically feasible to translate model level tolerances to the functional level using this analytic approach, except in the very simplest of models. As model complexity increases, the level of difficulty of this approach increases significantly.

SUMMARY

A clear definition of M&S requirements is needed for several reasons, including: deciding upon the problem analysis methods, selecting the appropriate M&S (if modeling is to be used in the analysis), planning the VV&A efforts, and accrediting the selected M&S. A complete definition of requirements depends on proper preparation. The preparation steps include development of a clear and comprehensive problem statement, identification of the overarching problem requirements, and determination of the approach to problem analysis.

The VV&A process is part of the application life cycle and is initiated with a set of accreditation requirements that are derived from the M&S requirements. M&S requirements are also the basis for model selection and for model accreditation. Because of their importance and widespread use throughout the application life cycle, the time and effort invested in developing these requirements is well worthwhile.

3.2 PLANNING

After problem definition, the single most important task to perform is planning the VV&A program. It is the planning phase of VV&A that makes or breaks the program. What are the elements of the M&S that are truly critical to this particular application, how many and what kind of tasks are required to address those critical elements, and how do those relate to funding and schedule realities? These questions must be answered in detail before any V&V work begins, or resources and time are wasted. Too many programs jump right into detailed validation of a few engineering details of a model, without determining first what is really needed; the result usually is very expensive V&V which does not address the true needs of the program.

<u>TRANSLATING ACCREDITATION REQUIREMENTS INTO V&V REQUIREMENTS</u>

by Dennis Laack

Once you've agreed on what problems M&S are going to be used to solve, what MOE's will address those problems, and how M&S outputs will generate those MOE's, and once you've gotten by whatever institutional barriers are in the way of doing V&V, then you need to figure out just what V&V tasks you need to do. But the program manager won't give you carte blanche - you need to do only those V&V tasks that directly support accreditation requirements. If you try to do <u>everything</u> you'll break the bank. The following article describes an approach to minimizing required V&V tasking, while still meeting the requirements. The program manager wants to know, "How little V&V can I get away with and still make an informed accreditation decision?" The article answers this question and allows you, the VV&A agent, to relate your V&V requirements to the sponsors need for M&S credibility. (ed.)

A long standing problem within the modeling and simulation (M&S) community has been the determination of how much verification and validation (V&V) is sufficient to support an accreditation decision. In practice, analysts use their best judgment to select the V&V techniques that are used, and the amount of funding available often controls the extent of the effort. Any V&V results that are produced are captured in some form (e.g., a report or a briefing), which is then used to justify an accreditation decision. Rarely are V&V requirements more rigorously defined by basing them on an analysis of the intended application.

This article presents an intuitive, but objective approach to determining the type and level of V&V that should be done to support accreditation for a particular application. The approach consists of two steps:

1) Establish the level of confidence in M&S results that is needed, based on a risk and benefit analysis of the application at hand.

2) Using look-up tables developed for the application (examples are included in this article), select and tailor the most appropriate V&V activities.

This approach is an adaptation of risk analysis techniques that are described in MIL-STD-882C, "System Safety Program Requirements".

FUNDAMENTAL CONCEPTS - TYING CREDIBILITY TO M&S PURPOSE

M&S are tools that provide information used to reduce risks or increase benefits in decision-making. Figure 1 is a conceptual flow diagram illustrating this concept. The information produced by M&S is often combined with that generated from other sources (e.g. handbooks, test data, expert opinion, etc.) to make some type of decision. These decisions result in (or contribute to) a variety of potential benefits and risks. The primary purpose of getting information prior to making a decision is to maximize the benefits while minimizing the risks.

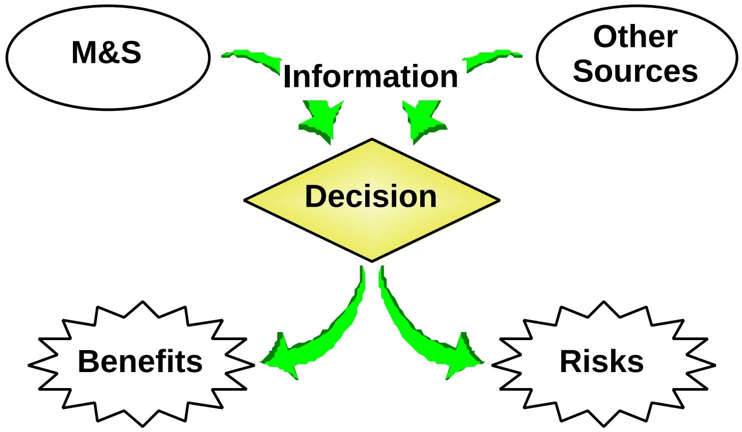


Figure 1: Generic Problem Paradigm

It stands to reason that the information used to make a decision must have a higher level of credibility as the potential benefits and/or risks increase. The credibility of M&S (and hence, of the information that they provide) can be improved by V&V and by corroborating information from other sources. Thus, the level of V&V required to support M&S accreditation for a given application depends on two factors: the degree of potential benefits and risks, and the availability of other information to corroborate M&S results.

RISK ASSESSMENT

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

$$\text{Risk} = (\text{Impact Level}) \times (\text{Probability of Occurrence})$$

In most cases the factors in this equation cannot be absolutely quantified. Instead, different levels are subjectively assigned based on some accepted criteria. One set of accepted criteria is presented in MIL-STD-882C, wherein the impact levels are divided into four levels of severity: catastrophic, critical, marginal, and negligible. The impact categories that are discussed in 882C are personnel and equipment safety, environmental damage, and occupational illness. Considering the broader uses of M&S, some additional categories could be added including: user capabilities or effectiveness, cost, schedule, and political or public adverse reaction. Some guidelines for determining the level of impact for each of the different impact categories is given in Table 1. This table uses the guidelines found in MIL-STD-882C, and adds similar criteria for the additional categories.

Table 1: Impact Severity Levels

IMPACT CATEGORIES	IMPACT LEVELS			
	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
ENVIRONMENT 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
SCHEDULE	Slip reduces DoD capabilities	Slip causes cost impacts	Slip causes internal turmoil	Republish schedules
POLITICAL	Widespread (Watergate)	Significant (Tailhook '91)	Embarrassment ($200 hammer)	Local

The other factor affecting risk is the probability that the impact will occur or be experienced. The probability of occurrence can be described as the expected number of occurrences per unit of time, over the life of a system, per number of events, or per items in a population. 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 the user to other types of impacts that would be experienced over time or over a number of events.

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

** The number of items should be specified

MIL-STD-882C also presents several sample tables that relate the level of risk to the different levels of impact and probabilities. These tables are termed "risk assessment matrices". Each application can have a unique risk assessment matrix. Table 3 is a typical risk assessment matrix.

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

The point of this discussion is that, for each M&S application, if risks associated with the use of M&S can be identified and associated with probabilities of occurrence and impact (given occurrence), it is possible to determine the level of risk. High levels of risk will require high levels of confidence in M&S results, leading to more stringent V&V requirements. Lower levels of risk will lead to fewer V&V requirements. We will discuss the actual relationship between the level of risk and the level of V&V later.

BENEFITS ASSESSMENT

"Benefits" are sort of the flip side of "Risks"; positive risks, if you will. Sometimes, the use of M&S will lead to very definite benefits, which can be evaluated and "quantified" with techniques that are similar to the risk assessment techniques addressed 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, etc.). One other significant benefit type is a reduction in operating costs. 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 sample criteria for assigning a beneficial impact level to the different types of benefits.

Table 4 Beneficial Impact Levels

BENEFIT CATEGORIES	BENEFICIAL IMPACT			
	REVOLUTIONARY	SIGNIFICANT	MARGINAL	NEGLIGIBLE
USER EFFECTIVENESS	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%

The other factor that is important in quantifying benefits is the breadth of their effects. This factor depends on two other factors: the scope of the benefit and the probability of its occurrence. The term "scope" refers to the size of organization that might experience the benefit. The probability of occurrence is a factor that is generally not considered when discussing benefits. However, just as in a risk analysis, there is some probability that the benefit will or will not materialize.

Ideally, a three dimensional matrix could be developed that would identify the required level of confidence for the information used in making a decision against these three factors that quantify benefits. Since such a matrix is difficult to portray, the factors of probability and scope can be combined into a single factor termed the percentage of a population that is affected in a given period of time. There is no objective and analytical process by which these two factors can be combined. Instead the merger of these factors must be based on informed judgment.

For our purposes, namely development of a V&V tailoring technique for use within the DoD acquisition community, the population base that seems most appropriate is the total armed forces, and the expected period of time wherein benefits are realized may be arbitrarily set at one year. (This means that for a combat system the one year period would be during a time when the armed forces are engaged in combat for a year). Table 5 is a subjectively developed table that provides some guideline for quantifying the percentage factor. Individual programs faced with the requirement to accredit M&S for use in support of an acquisition decision may wish to modify this table. (Although this methodology is designed for use in acquisition programs, there is nothing that would prevent its use in operations analysis or training applications.) The important issue is that the criteria are explicitly defined so that the rationale for determining what V&V techniques should be used can be clearly understood by anyone wishing to review the underlying logic.

Table 5: Percentage of Armed Forces Affected

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 or 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

Knowing the percentage of a population affected and the degree of impact, a table can be developed that identifies the level of benefits. Table 6 is a benefits assessment matrix. Although the benefit levels shown in Table 6 are reasonable, individual programs may wish to construct their own benefits matrix using similar principles.

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

Once again, the point of this discussion is that for each M&S application, if benefits associated with the use of M&S can be identified and associated with probabilities of occurrence and impact (given occurrence), it is possible to determine the level of benefit. High levels of benefit may demand high levels of confidence in M&S results, leading to more stringent V&V requirements. Lower levels of benefit will lead to fewer V&V requirements.

ESTABLISHING THE LEVEL OF REQUIRED V&V

Recognizing that M&S outputs are used to make decisions that either produce benefits or reduce risks, the importance of credible information increases as the benefits or risks increase. Since V&V data are used to reduce uncertainty about the model outputs and improve credibility, the level of V&V is directly related to the level of benefits and risks (in the absence of any corroborating information). As a result, the benefits and risk assessment matrices can be considered as matrices of the required level of V&V for the M&S used in an application.

In using these matrices there are two factors that must be considered. The first is that there can be multiple beneficial impacts each affecting a different percentage of the population. The second is the availability of information to corroborate M&S results.

The first factor (multiple impacts with different population percentages) should be addressed by considering each impact and percentage combination as a separate benefit. The set that results in the greatest level of V&V is the controlling set. This same technique can be used if there are different benefits that might be realized from a decision. Again, each is analyzed separately and the highest level of V&V that is identified should govern the selection of V&V techniques.

The second factor (effects of corroborating information) should be addressed by reducing the level of credibility needed. It is suggested, as a rule of thumb, that the necessary level of credibility can be reduced by one level if corroborating information will be available and used. The amount of reduction, if any, is definitely dependent on the quality and quantity of corroborating information. In every case, the amount of reduction should be consciously determined and that determination reviewed as part of the accreditation decision.

ESTABLISHING V&V FOCUS

Knowing the level of V&V that is required for a particular application is part of the total problem of deciding the scope and depth of V&V. Another part is determining whether V&V is needed at the overall model level, or at the level of the individual functions within the model. This decision is also dictated by the type of application. V&V should be done at the functional level for applications where pieces of information about some particular aspect of the problem can significantly impact risks. For example, if a model is used to help select design alternatives for components in a system, V&V will probably be necessary at the functional level. For other applications (such as evaluating alternative force employment strategies) V&V at the model level will probably be sufficient. Table 7 provides a suggested guide for determining whether V&V is needed at the functional or model level. This table was also developed based on subjective judgment, and should be modified to accommodate individual applications.

Table 7: Guide for Focusing V&V

APPLICATIONS	IMPACT LEVELS			
	CATASTROPHIC	CRITICAL	MARGINAL	NEGLIGIBLE
ACQUISITION				
COEA Milestone 0	M	M	M	M
COEA - Other	M	M	M	M
System Design	M&F	F	F	F
Test Planning	M	M	M	M
Test Safety Assessment	M&F	M&F	F	F
T&E Extrapolation	M&F	M&F	M	M
ANALYSIS				
Strategy Evaluation	M&F	M	M	M
Tactics Development	M&F	M&F	M	M
Wargaming	M&F	M	M	M
TRAINING				
Individual or Unit	M&F	M&F	F	F
Mission Cmdrs	M	M	M	M
Campaign Cmdrs	M	M	M	M
Command Staffs	M&F	M&F	M&F	M&F

SELECTING V&V TECHNIQUES

The remaining step in determining the scope and depth of V&V is to select the V&V techniques that are best suited to provide the level of confidence needed for the application. Certain techniques provide greater confidence in the model results than other techniques. Also, techniques may vary from one application type to another, and also between M&S communities. Within the acquisition community, the menu of V&V techniques developed by the SMART Project is becoming more widely accepted as a norm for mature M&S. Table 8 provides a guide for selecting the appropriate set of techniques according to the necessary level of confidence and the focus on either the model or functional level. This table was developed based on experience with several acquisition programs, and with V&V on selected M&S used in aircraft survivability analyses. It is meant as a guide and should be modified to suit individual applications.

SUMMARY

The tables and methods outlined in this paper provide an M&S user with a guide to help in identifying the V&V techniques that should be used to develop sufficient information to justify accreditation for any application. The necessary precursor steps are the development of risk assessment and benefits matrices, both of which are tailored to the application. These matrices are key factors in selecting appropriate V&V techniques, and also in determining whether V&V should be done at the functional level, the model level, or both.

The major benefit of this process is that it follows a logical flow that clearly identifies points at which subjective judgments must be made. Furthermore, the judgments and underlying criteria are clearly spelled out, so that differences of opinion can be more readily discussed and resolved. Use of this process will aid the accreditation authority in identifying minimum V&V requirements. It will also help managers to understand the effect of not performing one or more of the required V&V tasks because of schedule or budget constraints.

Table 8: Guide for Selection of V&V Technique

	Model Level			Functional Level		
	High	Medium	Low	High	Medium	Low
Baseline Definition	X	X	X	X	X	X
Determine C/M Attributes	X	X	X	X	X	X
Assess M&S Documentation	X	X	X	X	X	X
Establish VV&A Status, Usage History	X	X	X	X	X	X
S/W Quality Ass'mt	X			X		
ID Assump & Limits	X	X	X	X	X	X
Produce Design Documentation	X	X		X	X	
Perform Logical Verification	X	X		X	X	
Detailed Code Verification	X			X		
Sensitivity Analysis (Model Level)	X	X		X	X	
Sensitivity Analysis (Function Level)				X	X	
Face Validation		X			X	
Results Validation (Model Level)	X			X		
Results Validation (Function Level)				X		

VERIFICATION AND/OR VALIDATION; WHICH ONE WHEN?

by Barry O'Neal

Once you've got a handle on what is important to your problem, you have to ask yourself, "When do I do verification, and when do I do validation?" While there are many immediate benefits from software verification activities, everybody wants to know if the model has been validated. But validation is expensive; it turns out that, in terms of cost effectiveness, verification gives you more bang for the buck than validation. But everybody wants validation. So what do you do? The following article discusses the tradeoffs between the two. (ed.)

Early in the course of the SMART Project, it became clear that the programmers who develop M&S, the analysts who use them, and the people who accredit them all had different ideas about what VV&A actually meant. A favorite (albeit facetious) definition was suggested by a model manager, who said, "If the code compiles and links, it's verified; and if it executes, produces output, and terminates normally, it's validated." Ideas about verification range from unit level testing to final acceptance software testing, to audits of code design by independent agents, to "making sure this version works like the last one." Notions of validation range from reflection of theory (especially in the absence of empirical data), to correlation at the trend level (as in economic models), to exact replication of each data point in a test result. Some of this breadth and variability of definitions was documented in an Accreditation Requirements Study [1], which summarized interviews with DoD decision makers, model developers, and persons employed by agencies and departments responsible for VV&A.

When the project was initiated, numerous papers identifying and describing VV&A activities and techniques used historically had been published by Pace [2], Waite [3], Sandhu [4], and others. Most of these activities and much of the methodology described by Kneppell [5] were incorporated into the V&V process ultimately developed by SMART, which placed them in a hierarchy (work breakdown structure) designed to support M&S accreditation. A few were not included due to their marginal contribution to credibility or to a conflict with emerging definitions. The Military Operations Research Society (MORS) proposed definitions for VV&A in 1994 [6], which SMART adopted as they were published in [7] and which are:

Verification - ...the process of determining that a model implementation accurately represents the developer's conceptual description and specifications;

Validation - ...the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

Note that verification is not the critique of an implementation, but rather an assessment of its correlation with intended design, yet independent V&V agents may presume the former due to the confusion over locally accepted definitions (see, for example Feldman [8]). Note also that validation is from the perspective of intended uses of the M&S and not from that of the developer or other requirement source.

Users that need to accredit a model are faced with questions about how much V&V is needed to satisfy credibility requirements, and how to acquire enough V&V information to convince a review board that the model is appropriate for its intended use. After working out functional requirements and the acceptance criteria for establishing suitability, the hopeful accreditor comes to where the rubber meets the available budget: to V, or not to V.... but which V first, and most important, do you really need both V's? This is where minor distinctions in their definitions become important.

Because the first V(verification) deals only with software, it will usually be easier and cheaper than the other V. This second V (validation) involves assessments of how well model outputs, or predictions, match reality, and these usually require acquisition, comparison, and interpretation of data collected via some form of testing. Therefore, it will always be more difficult, expensive, and risky than verification.

The essence of the accreditor's problem is in determining where verification will be sufficient and where validation will be essential. Using clearly specified credibility requirements and acceptance criteria, critical functions and appropriate tasks can be identified and resources can be proportioned among various V&V activities via a plan that will maximize results from the available budget. Cost estimates associated with accreditation support are provided in [9] and can be used to prioritize activities for targeted model functions. This approach will focus efforts on the adequacy of implementation (verification) for required functions while limiting validation efforts to critical functions or those for which test data may be available.

VERIFICATION

The best thing about verification is that given sufficient time, almost any programmer or analyst provided with the code and a reasonable amount of documentation can figure out what the software actually does and how it does it. Thus, it's implementation methods, concepts, and underlying design can be established and this information can be evaluated in light of other acceptable and/or reasonable designs. Such an evaluation results in a verification of program logic that can be documented and used to support an accreditation recommendation. Although more time-consuming, but also more effective, a detailed comparison (audit) of algorithms and data sources implemented in software with a conceptual specification or design reveals how well the code implements the intended (required) functionality. The only stipulation for performance of such audits is that they should be done by someone other than the person or group responsible for production of the software itself. Independence in V&V is a good requirement, but some latitude should be allowed due to organizational and contractual situations. Model developers should consider establishment of VV&A or quality assurance teams as suggested by Davis [10].

As with other aspects of V&V, activities related to verification can be applied to varying levels, or degrees, which will yield equally variable results. The amount of effort required for verification will be in direct proportion to the level of detail of the examination. Verification planning is done by spreading available resources over available time and volume of code. The level of verification required will be a function of the application, the accreditation requirements, the model intended for accreditation and its dependence upon correct design implementation.

The important thing to remember about verification is that in addition to being relatively easy to perform, it's also very effective in locating software problems, which is why many (of us) feel that it should be an integral part of software development efforts (see Hall [11]). Because such integration is rarely the case, verification is typically done on M&S that have been in use for some time. Such software is rarely accompanied by adequate design documentation needed to perform verification. This deficiency imposes additional costs on the V&V effort at hand. An inherent weakness in the verification process is the dependence upon accurate documentation of the M&S under examination. If it doesn't exist, someone must be paid to produce the necessary documents.

Conceptual Specifications and Software Designs

Software developers have usually attempted to produce some form of design documentation (even before the days of DoD and MIL standards), because most have realized its value in facilitating the ultimate implementation. Unfortunately, rates of development and requirements for change often outpace the ability of programmers to produce meaningful documentation. Thus, complex M&S abound and are used without sufficient descriptions of their underlying designs or the assumptions and limitations tied to those designs. Unfortunately, this type of documentation is exactly what is required for meaningful results from verification activities.

Software designs typically evolve during meetings, where problems with previously proposed designs are examined in an attempt to achieve the best compromise among functional requirements. Given the hardware and software development environments, modified designs are tested to maximize performance in application specific areas. Even when original versions of design documents exist, their final version counterparts are rarely found. This is as true today as it was when DoD STD-2167A was written. If an adequate conceptual model specification or design document is not available to support verification of the selected software, then a process like the one shown in Figure 1 will be necessary to produce it.

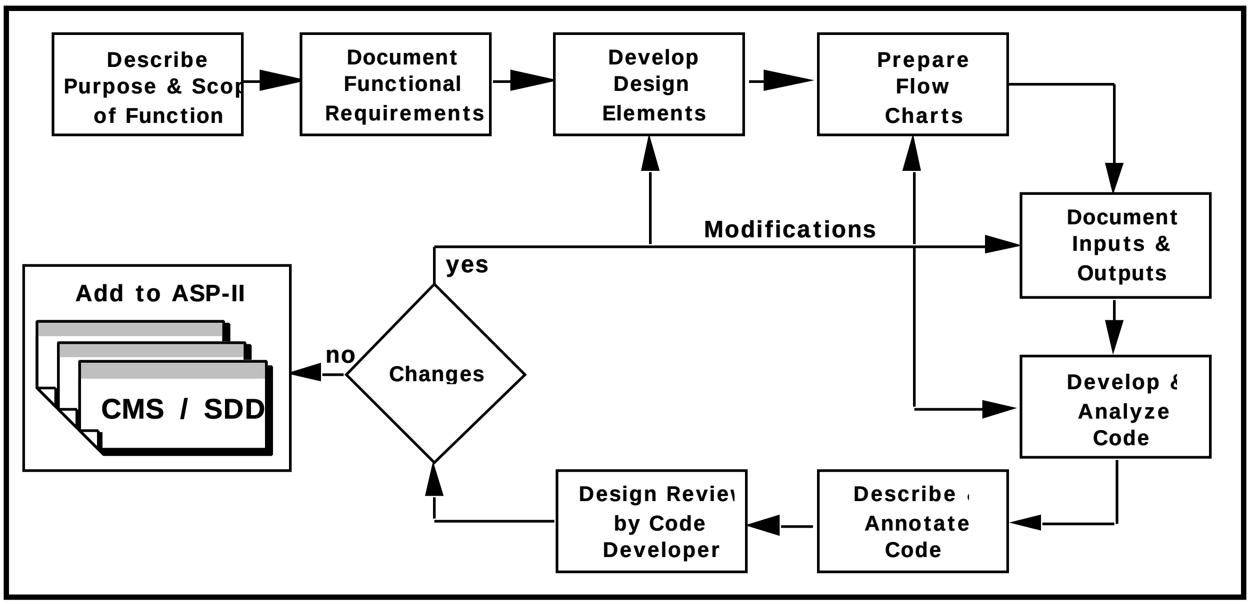


Figure 1. CMS Development Process.

Documentation we produced via the SMART process resulted in a conceptual model specification (CMS), or a reverse-engineered software design document (SDD) for many M&S functions. Evolutionary precursors to these were the post-development design document (PDDD), the verification source report (VSR), and the verification support document (VSD), all of which contained variations on the same types of design information. What the documentation is called is unimportant, but it must include enough information about design requirements, design elements, structures and sources to support verification.

Regardless of the name, format, or structure of the documentation, translation of functional requirements into a design is the most critical aspect of implementation, next to production of the correct code reflecting that design. Depending upon the level of effort and formalism applied to design and coding, an examination of either or both may be required to establish what the code is actually doing with respect to what is, or was required. At best, examination of the design by experts in the field can be used to lend credibility to the M&S or to perhaps question its novel implementation. These are the outputs of logical verification, which has been a large part of the traditional approach to establishing M&S credibility. A problem with such 'expert reviews' is that their own credibility may be questioned and that the ultimate question of implementation with respect to design requirements often remains unresolved. In order to perform truly independent (and certifiable) verification, a detailed examination of source documentation and actual code is required. A notional flowchart depicting the detailed verification process is provided in Figure 2.

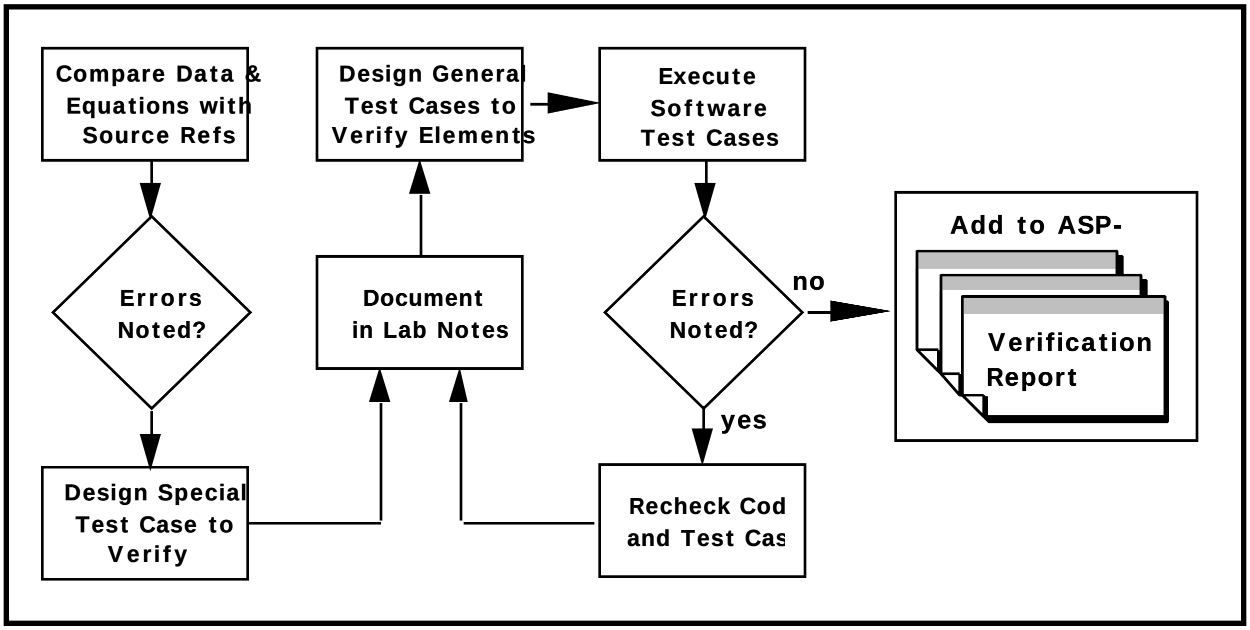


Figure 2. Verification Process.

Desk Checking

Most algorithms implemented in software reflect a theory or empirical finding that can be traced to a source in the applicable literature or unpublished report. Verifying these sources and the implementation of their theories and/or test results is the essence of independent verification. Although not always a strict requirement, the independent verifier should be a peer, or member of the community experienced with the functions being simulated. If their familiarity with current theories and/or methods is unquestionable, their endorsement of the sources and concurrence with design approaches will lend more credibility to the implementation.

On the other hand, designs that are not based upon accepted sources become difficult to document to the satisfaction of peer reviewers. This fact would suggest that desk checking of algorithm and data sources be conducted by peers of the programmers who are implementing them. If inventive developers have come up with novel ways to implement a design, they should document the implementation so that it becomes legitimate and can be verified. Desk checking can be performed by virtually anyone familiar with the intended application and programming language employed, and provides the most concrete, indisputable data on the implementation with respect to what was required.

Software testing

Verification of a particular implementation of theory or test results is accomplished via detailed testing of branches, paths, calculations, and interfaces. This is the most time-consuming (and expensive) aspect of verification, but it is also an activity that is routinely performed during the development process. Because results are rarely documented, the testing must be repeated, and formally documented during this phase of the V&V effort. This aspect of verification requires a working knowledge of the language used to implement the function in software as well as an ability to design and execute tests that will either substantiate or discount implementation correctness with respect to the intended design, given certain assumptions and limitations.

The problem with software testing is that it is tedious and the benefits often appear marginal when compared with other V&V techniques, especially validation. The good thing about verification is that once completed, the code can be certified as error free until it is changed. Code changes will necessitate repeated verification unless managed and controlled through an established tracking and updating process that can be used to focus efforts on areas where designs have been affected. However, very few examples of effective configuration management (CM) have been found among M&S widely used in the analytical community, which is unfortunate given the potential reduction in V&V costs over the life cycle that could be achieved. The combined result of the primary verification activities (desk checking and software testing) provides the most credibility that can be afforded short of comparisons with data from the real world. This statement assumes, of course, that the verification results are reported and recorded.

Reporting

A V&V effort that goes undocumented is one that is destined to be repeated. The purpose in documenting anything is to educate future users who may or may not wish to repeat a particular verification experience. Several hundred thousand lines of code should only need to be verified once and then updated at infrequent, semi-regular intervals.

An example of how a typical verification result can be documented is shown in Figure 3. The table identifies each design element, where the implementation is located in the code (module name and line numbers) and the desk check result, which is indicated by a check mark or an identification of a discrepancy discussed in the report. Similarly, test cases are identified by numbers correlated to their respective design element and results are shown as either satisfactory (check mark) or by a similar number that points to a deficiency described in the report. Detailed descriptions of software tests and results are also important when tracking down potentially serious errors.

Verification Matrix for Signature Fluctuations.

Design Element	Code Location	Desk Check Result	Test Case ID	Test Case Result
4-1 Geometry	GLINT 165-239	D1	4-1,2,3 4,5,8	4-1 4-2
4-2 Glint Half-Power Frequency	GLINT 245-249	√	4-6	√
4-3 Glint Error and Target Position	GLINT 264-346	D2	4-7,8,9,10	4-8 4-9
4-4 Exponential Draw	SINTL8 30-36 EXPDRW	√	4-11	√
4-5 Chi-square with Four Degrees of Freedom Draw	SINTL8 36-42 EXPDRW	√	4-12	√
4-6 Time Correlation	COREL8	√	4-13	√

Figure 3. Typical Summary of Verification Results.

As with software development, documentation of V&V is painful, but unlike development, results of undocumented V&V efforts are gone and forgotten in a short time. Undocumented M&S hang around for years, especially when we wish they wouldn't. To minimize this problem the SMART Project defined and adopted a standard format for reporting verification results. If this standard were widely adopted by M&S developers, model users could easily assess the adequacy of their implementations from these standardized V&V reports. But alas, verification reports (even non-standard ones), are as rare as the source documents needed to produce them.

VALIDATION

The best thing about validation is that positive results (i.e., demonstrated correlation between model outputs and the real world) are usually conclusive and rarely debated. The worst thing about validation is that such events are extremely rare because models are not designed to and physically cannot replicate all real world conditions. Furthermore, the notion that validation is a one-time certification of model credibility (akin to the Good Housekeeping Seal of Approval) is so pervasive that any real attempt to establish a baseline of validation results and maintain or expand them over time is often futile.

So why even attempt it? Because the results can be extremely important for critical applications where grave consequences could result from acceptance of erroneous model outputs. SMART was quite successful in validating significant portions of some important M&S, but the perception of what was validated relative to the total cost of the project leads one to the popular conclusion: validation just costs too much. This opinion seems to be the one shared by those who see no value in verification as well as those who contend that only total and complete validation of a model or simulation is sufficient for accreditation.

In truth, the SMART Project demonstrated that significant credibility could be gained through a variety of techniques related to traditional validation and that such increases or enhancements to credibility were affordable. Some of the activities associated with validation are performed in conjunction with normal software development efforts and could, if documented, contribute to model credibility. However, as with V&V, results of these are seldom recorded.

Sensitivity Analysis

Most software developers (programmers) perform some sort of sensitivity analysis on the code they write, but they are often focused on a few test cases that will demonstrate proper functionality. Time to exercise model functions over allowable ranges of performance is rarely available. This is most unfortunate because the results of a sensitivity analysis can contribute significantly more to model credibility than some other, more expensive activities associated with validation. The testing of a software module, or function (collection of modules) in a way that illustrates correct relative response is easy, cheap, and instructive. It's also what most people do when they use a model for the first time. Results of such analyses, although not a substitute for traditional validation, should be exploited and applied to application specific suitability questions.

Face Validation

A documented product that complements sensitivity analysis, the face validation report results from comparisons of M&S outputs with outcomes that are considered reasonable or appropriate by those experienced with the behavior of the systems/processes being simulated. This type of validation is frequently employed, primarily because of its low cost and lack of dependence on the acquisition of test data. It is most appropriate for those M&S functions that will have a low probability of detailed validation success due to their complexity, variability, or dependence upon data that would be difficult (if not impossible) to obtain. Model functions that deal with noisy processes or natural phenomena (like weather) or behavioral statistics are examples of good candidates for face validation.

Similarly, M&S performance criteria can often be derived from software requirements specifications (if they were documented), which can form the basis for expected behavior in the place of experts. Face validation at the functional level can also be established from documentation of unit level and integration testing, which are limited forms of sensitivity analysis. A typical result is shown in Figure 4, which illustrates the different manual tracking error histories produced for operators with three skill levels (i.e., low, medium, and high). The model in this case is based on two variables, minimum discernible error and reaction time, so the errors generated demonstrate a reasonable decrease in the magnitude and frequency of tracking error (in mils) from the low skill level to the medium and high one.

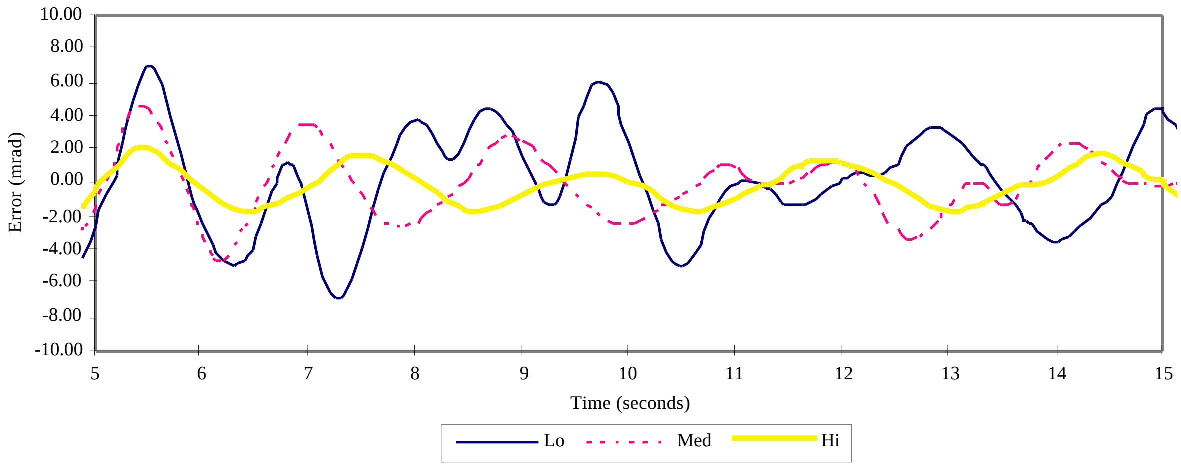


Figure 4. Face Validation Example.

Because of its cost-effectiveness, it's important not to overlook possible opportunities to take advantage of face validation. It's not a substitute for traditional validation with test data, but in many situations it's definitely the next best thing.

Data VV&C

Activities associated with verification and/or certification of the data that are used as inputs or calibration factors for models overlap to some degree the desk checking activities associated with verification. The determination by experts or knowledgeable individuals that data sources and values used are correct and appropriate from the perspective of intended uses of the software (valid) attributes a credibility beyond that of verifying their use in the code. This area of V&V activity has become increasingly important for simulations that are driven by large databases. Such M&S usually generate few predictions at the functional level, but integrate the complex interactions among many database entities in order to predict a higher level outcome. Thus the correct definition of those database entities, their capabilities and how they interact with other entities is critical to the result ultimately predicted. It has been suggested, perhaps rightly so, that certification of the databases used as inputs to aggregate-level simulations may diminish the need to validate internal functions, which tend to be more like accounting and statistical functions that lend themselves readily to face validation activities. Unfortunately, this doesn't mean that V&V of such M&S will be less expensive. Databases by nature are easily changed by users, so certification and control of their configurations pose significant problems for those with VV&C responsibilities. Configuration management (CM) of certified databases may have to be accomplished via network servers or CDROM media that preclude alteration by users in order to be viable.

Test Planning

If test data are required to support validation efforts, plans must be developed to specify data elements and their requirements for collection. The SMART Project was able to "piggyback" or cooperate with ongoing DoD test and evaluation (T&E) programs with considerable success. This was partly due to documentation of specific data requirements (i.e., accuracy, resolution, and quantity) and, in some cases, due to contributed instrumentation needed to collect specific values for assessment of M&S functions. The value of this interaction with the T&E community has perhaps been undermined by the completion of the project, but it served to illustrate and refine the type of documentation required to acquire data for validation purposes.

Clearly, more work is needed in this area to enhance cost-effective data collection, but it can be assumed that some portion of every test being conducted by the DoD can produce data that would be useful in validating some type of model or simulation. Thus, every such test should have data requirements and test plan documents designed to support data collection efforts for validation of those software products. The lack of real world test data has been a historical problem in the T&E community where substantial sums are spent collecting data, but no data archives are maintained to allow data reuse for V&V or other analytical purposes. This has also exacerbated the DoD-wide perception that validation is too expensive.

Data Acquisition and Reporting

Collection of data is not a validation activity per se, but if one needs to support a validation effort, they must assume some responsibility for its acquisition. Furthermore, participation by the validator in the acquisition process can be invaluable later in the reduction and/or transformation of the data for use in comparisons with M&S predictions. Knowledge of testing anomalies, especially bad ones, can also reduce later efforts to use data that won't compare favorably with anything. It is also essential in the preparation of reduced data packages and reports that can be used by M&S validators. This is an area where the SMART Project focused significant resources and the value of the documentation produced cannot be overstated. Unfortunately, efforts to document and archive test results will probably not be pursued by individual programs or test ranges in the current competitive environment for test range work and survival.

Assuming that your data acquisition effort was successful, pay careful attention to the form and nature of the data when preparing to make comparisons with model predictions. Many validation attempts have been thwarted by failure to simply rotate, transpose, or convert data fields or units to their proper representations in the simulation of interest. The roll angles shown by the plus (+) and circle (o) symbols in Figure 5 resulted from rotations of data downlinked via telemetry from two roll rate sensors in a test missile during flight. Note that all the test data lie between ± 90 degrees and that there are no data between ±90 and ±180 degrees. Agreement between the two channels is good between 1 and 6 seconds, but poor elsewhere. In addition to the sign/phase ambiguities that place values opposite where they should be, the data dropouts at 6-7 and 8-9 seconds also made hand editing of the data necessary to achieve the normal 'Integrated Roll' data shown by the solid line in the figure. The vertical lines connect roll angles at ±180 degrees.

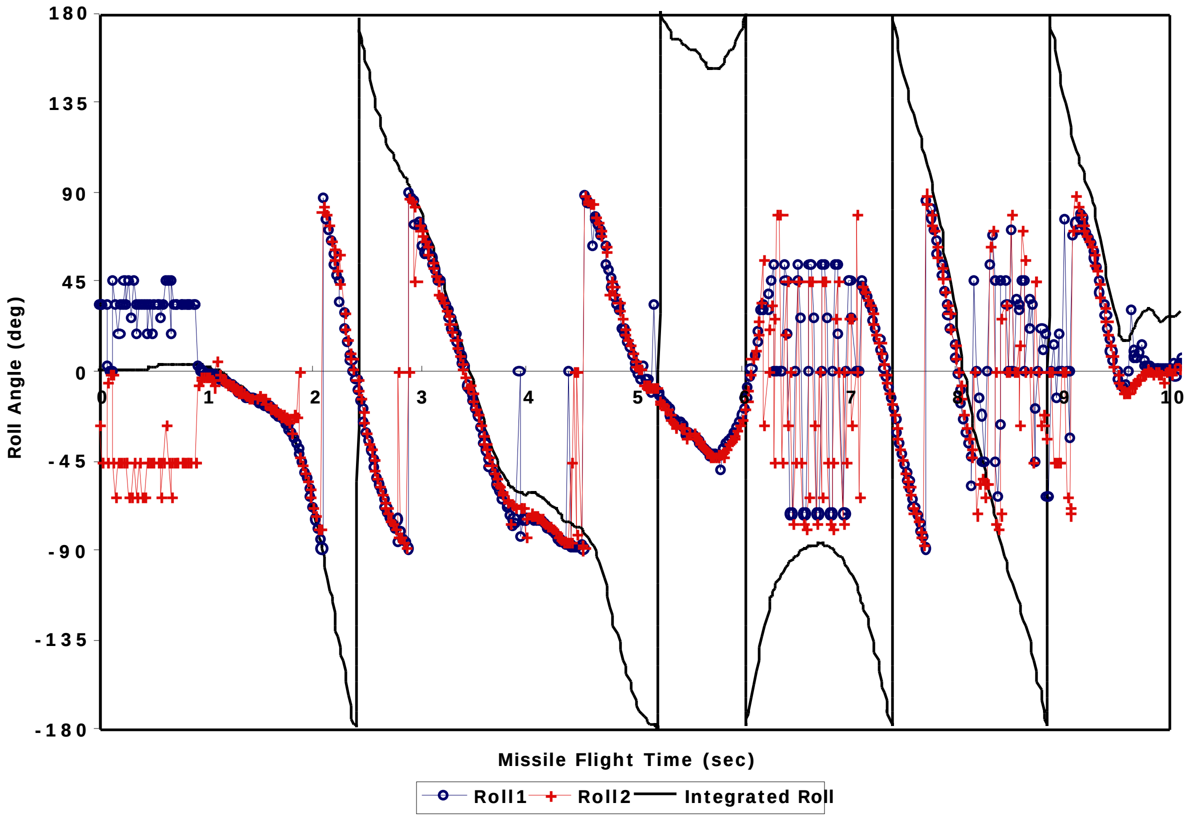


Figure 5. Data Manipulations Required to Make Comparisons with M&S.

During this 10 seconds of flight after launch, the missile completed one revolution at 3.5 seconds, then it rolled back about 115 degrees after reaching 205 degrees at 5.5 seconds, and then made another roll from just after 7 to just before 10 seconds. It took 2 years for SMART to get this data (courtesy of an Air Force test program), which highlights yet another problem with validation: it can take more time than you have available. Subsequent comparison with the flyout simulation, however, resulted in a degree of correlation and model credibility that was worth the wait.

Most test engineers would agree that the days when everything works and the data looks good are infrequent, but care must be exercised in the suggestion or discussion of problems with test data. The T&E community has been threatened by the possibility of M&S replacing the need for certain aspects of testing for some time. Even though this possibility is a remote one, those responsible for collecting the data are as sensitive about it as are the software developers who are crushed when model deficiencies are found and reported by independent V&V agents. The good news is that data quality and quantity are improving and this bodes well for the future of validation, even if we think we can't afford it today.

Comparisons with Predictions

In the process of correlating model outputs with test data, it may be necessary to calibrate functions or data values to those appropriate for the test conditions. In addition to test scenario geometry, timing, and the application of stressful conditions, assurance that internal data associated with frequency, wavelength, functional limits, and other system specific factors should be established before proceeding. Perhaps the most important factors that may need to be reflected in the simulation prior to execution are those artificial limitations required by the nature of the test range or facilities. This process of calibrating the model may need to be iterative as initial comparisons raise questions about why results from measured and predicted sources differ in unexpected ways.

In fact, one should be surprised if initial model predictions compare favorably with measurements because the probability is usually low that all factors affecting measurements were accounted for in the model. Typically, some unknown factors will affect the data in some unforeseen way. For example, tracking error comparisons like the one shown in Figure 6 suggested that range data values were an order of magnitude greater than model predicted errors.

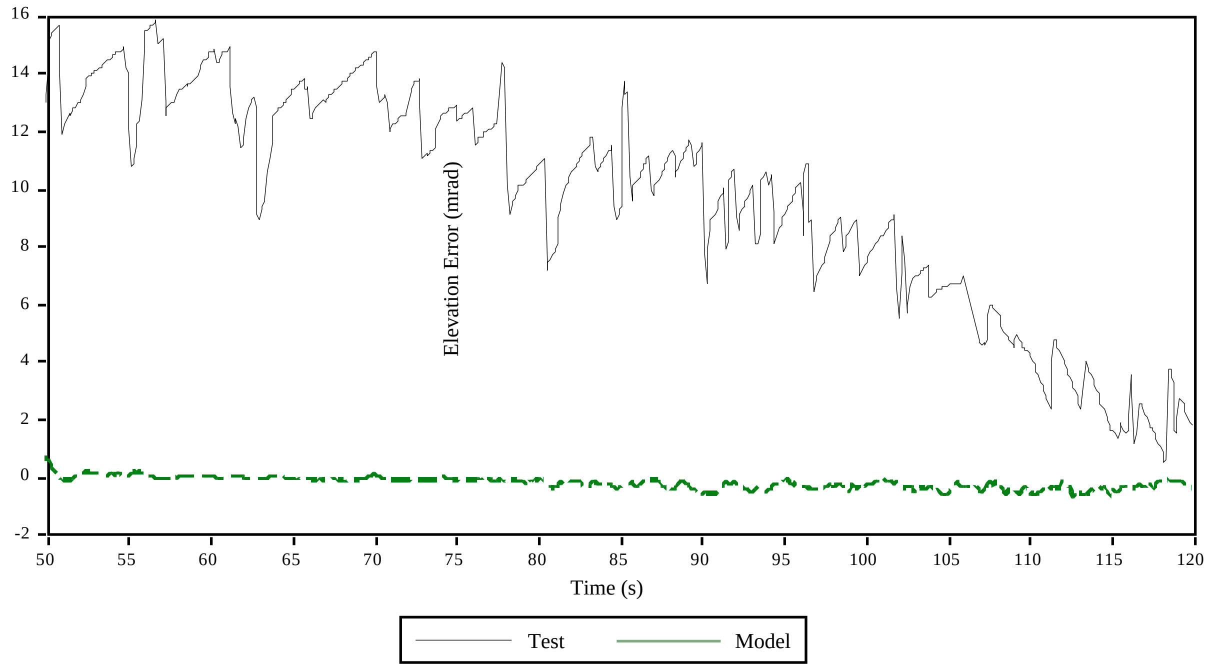


Figure 6. Initial Comparison of Predicted and Measured Tracking Errors.

After looking at the potential effect of such large tracking errors on endgame conditions, it was concluded that they couldn't be that large for an effective system and were divided by ten. The unexplained positive bias (i.e., no negative errors) in the test data was also removed and some semblance of correlation resulted as shown in Figure 7. Why were test data values ten times larger than predicted? Instrumentation gain is most likely. Biases or shifts in the data are often due to calibration timings used to post-process the data or can be due to those software programs used to rotate, scale, and shift the data to desired datum locations.

The important thing to remember is that unnoticed (undocumented) factors tied to test conditions, instrumentation, reduction algorithms, or calibration can affect the data adversely enough to nullify the validation effort. That is why reports that describe test conditions and limitations are so important to those tasked with comparisons. It is unfair to expect test data to be perfect and those involved with its acquisition have made great strides in the advancement of collection equipment and processes. Therefore, one charged with using it should attempt to understand and help in the improvement of data rather than be quick to blame it for failure to replicate model results. Remember that without indications of real world performance derived from test data, validation of M&S would not be possible.

In comparing test data and model results, various approaches can be taken depending upon the nature of the data, the model type, and the limitations imposed by the collection process. While visual comparisons and heuristic methods shed light on model credibility while also aiding in the understanding of phenomena not modeled, statistical methods are usually required to convince purists and project managers.

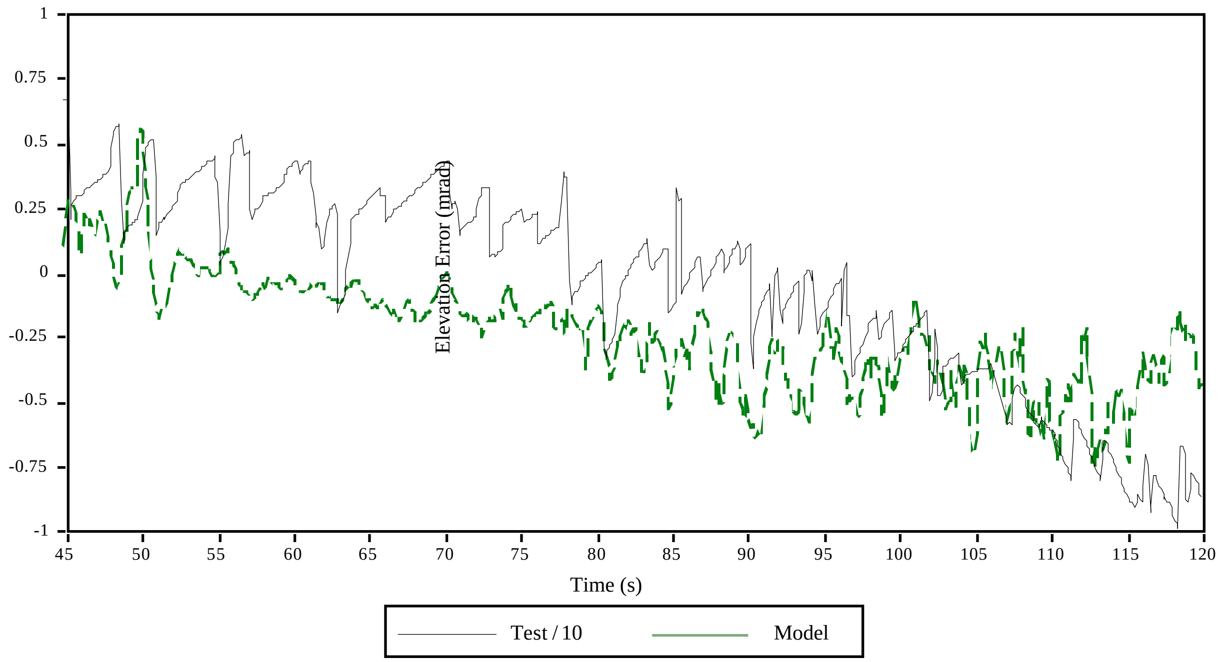


Figure 7. Subsequent Comparison of Predicted and Measured Tracking Errors.

SMART demonstrated and documented several methods in an appendix of their process description [12] that should be applicable to a wide array of validation efforts. The problem common to these methods is that confirmation of the degree of correlation with the data is rarely achieved, but a probability of similarity (or lack thereof) is produced for a given confidence level. While this is sufficient for some applications and some users, others remain suspicious of such conclusions and will only be convinced by perfect correlation. This points out another problem with validation: how close is close enough for the intended application? Accreditation proponents have difficulty defining a threshold or requirement for acceptable correlation, opting instead for the "looks good to me" or "correlation was good" conclusion to support their justification for model use.

It is important to realize that in many cases, a comparison of model predictions with test results constitutes a single data point on the model credibility report card. Thus the expense of checking all the possible boxes for a complex simulation often prohibits total and perhaps even adequate validation. This fact is further impacted by the rate of change experienced in most models. The practice of using M&S in a variety of applications causes them to be in a continuous state of development such that functions validated in the past need to be revisited as new model versions are produced. An effective CM program helps mitigate this situation by tracking software changes well enough to guide future V&V efforts. With standardized V&V reporting and effective CM programs, the M&S credibility report cards can continue to be filled out as additional V&V is done. This is true despite the fact that both model users and model developers are interested in adding more and more features to the model to increase its utility.

Reporting Results

As with verification, failure to document validation efforts will have the same effect as if they were never attempted, much less successful. A major benefit of the SMART effort was the development of standard formats for reporting V&V results so that credibility assessments for different types of M&S could be compared for purposes of accreditation. Although these formats will probably not be adopted by all agencies, the validator should make every attempt to document validation results as well as lessons learned for the benefit of future model users. Based on comments received concerning V&V documentation produced by the SMART Project, it is apparent that the value of well documented credibility assessments is increasing. However, in the minds of M&S users it may never approach the value of having the latest and greatest version of the code.

THE BOTTOM LINE

Many in the M&S community contend that verification, at least, should be a routine portion of M&S development and maintenance. This view appears very rational as the developer can easily provide design documentation for the software, as well as sources of algorithms and data used in specific modules. These can be used to support verification by others if the developer is not required or able (and some aren't) to do so. If the accreditor must produce this documentation, then the credibility afforded by verification may only be marginal. This was often the case in the SMART Project, which examined thousands of lines of 'legacy code' and found very few significant errors. Quite a few typographical errors, missing sources of data and algorithms, and lots of documentation bugs were found, however. In order to do this we also developed several large volumes of software design documentation and agreed on how to document the results. Then we tried to read them without falling asleep.

A validation effort is a luxury that few M&S users can afford, though many have attempted to do it justice in their own insightful way. Linked to reality by the T&E process, it promises hope for the programmers who have analyzed and attempted to simulate the problem and excitement for everyone else hanging around just to watch the smoke and fire. It must be approached cautiously, however, for the path to successful correlation of M&S predictions with test data crosses chasms wide enough to empty all but the deepest program pockets.

In summary, verification is just documentation of sources and software testing that every developer or sponsor agency should provide if they want users to have confidence that their models are well built and thoroughly wrung out. Validation is a landmark event that only occurs after a dedicated effort by people in both M&S and T&E camps. Consequently, it's also expensive, and this makes it risky for budget-constrained projects in which management must seek the most feasible (cheap) course of action. It should be attempted, however, whenever applicable test data becomes available from whatever source, and any program that has a T&E budget should be viewed as a target of opportunity for those M&S used by or associated with it.

The ultimate objective, accreditation, is difficult to achieve because of its dependence upon V&V results, which provide the basis for determinations of suitability to address specific problems. A problem common to most accreditation efforts is that they often start at square one without the benefit of prior experience with similar requirements or perhaps even the same suite of M&S. A well maintained and accessible repository of accreditation support information gathered over time would improve this situation and make VV&A less painful and more cost-effective than it is today. With on-line access available to any accreditation proponent, accumulation and distribution of historical V&V results and accreditation decisions should be accomplished to help achieve this worthy goal.

References

1. Laack, D. R., *Accreditation Requirements Study Report*, *Volumes I and II*, JTCG/AS (AIR 4.1.8), Naval Air Systems Command, Washington, DC (JTCG/AS-93-SM-020).

2. Pace, D. K., *Naval VV&A Process*, PHALANX, September 1993.

3. Waite, W. F., Nuzman, D., and Phillips, S., *Planning for VV&A: An Example*, Proceedings of the 1994 Summer Computer Simulation Conference (ISBN: 1-56555-029-3).

4. Sanhu, G. S., *Verification and validation techniques*, Proceedings of the Society for Computer Simulation 1983 Symposium, pp. 882-887.

5. Knepell, P. L. and Arangno, D. C., Simulation Validation: A Confidence Assessment Methodology, IEEE Computer Society Press, Los Angeles, 1993.

6. Sikora, J. J., and Williams, M. L., *SIMVAL '94*, PHALANX, December 1994.

7. Department of Defense, Department of Defense Dictionary of Military and Associated Terms, Washington, DC, DoD, 23 March 1994 (Joint Publication 1-02).

8. Feldman, P., *Verification of Two Functional Elements (Multipath and Moving Target Indication) in the ALARM92 Radar Model*, Illgen Simulation Technologies, Inc., Goleta, CA, January 1993.

9. Muessig, P. R., "Cost Vs. Credibility: How Much VV&A Is Enough?", *Proceedings of the 1995 Summer Computer Simulation Conference*, Oren, T. I. and Birta, L. G., Ed., July 1995, pp 166-175.

10. Davis, P. K., *Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations*, RAND National Defense Research Institute, 1993, (R-4249-ACQ).

11. Hall, D. H., "Integrating V&V Into M&S Development", SMART Project VV&A Lessons Learned, February 1997.

12. Muessig, P. R., Hall, D. H., et al, *V&V FROM A TO Z: A SMART Approach to VV&A for Acquisition M&S*, Joint Accreditation Support Activity (JASA), Naval Air Warfare Center, Weapons Division, China Lake, CA, January 1997.

THE FUNCTIONAL APPROACH TO V&V: WHY BOTHER?

by Barry O'Neal

Suppose you compare your model output with some test data, and it doesn't match very well. Is the model totally useless? How should the problem be diagnosed? This article talks about the benefits of dividing the model into parts that can be analyzed, understood and validated. This incremental process contributes to ultimate model credibility over time and focuses model improvement efforts where they are most needed.

Certainly the most common and recurring problems encountered during the course of the SMART Project were those associated with the many different preconceived notions people had about what the terms 'verification and validation' or 'VV&A' really meant. In the beginning, many in the modeling and simulation (M&S) community had their definitions exactly backward from what was proposed by the MORS [1] and adopted by the DoD [2], while others, who acknowledged these, campaigned for additional refinements or specialized interpretations (e.g., intelligence, policy) [3, 4]. Those responsible for M&S development also segregated 'software V&V' and the activities associated with unit level, integration and acceptance testing from what we were calling 'functional V&V', which produced assessments of how well M&S components were implemented and/or how well they predicted performance, or (sub)system capability.

An underlying premise of the V&V process tested by the project was that tackling large M&S in small pieces would be more efficient and cost-effective than the traditional approach [5], which encompassed the entire simulation (i.e., from end-to-end) [6, 7]. The idea behind functional decomposition stemmed from the fact that each of the M&S initially evaluated by SMART performed similar functions (e.g., target acquisition and tracking, weapon flyout and intercept, probability of kill calculations), but in different ways and (of course) via different codes. We believed that by performing V&V on the same function implemented in three different models, cost savings could be realized by reuse of the same test data. In addition, the best of the three implementations would be obvious from the results, which would recommend its replacement of the other two. Thus, model improvement and functional commonality would be cost-reducing by-products of V&V, in addition to the credibility afforded by the V&V alone. We later learned that incremental application of functional V&V, coordinated with M&S upgrades over time via a configuration management (CM) process, also reduced maintenance costs by eliminating the need to re-examine entire software packages each time a new version was released.

These concepts were not well received, however. In the beginning, nearly everyone we told about them came up with reasons why it just wouldn't work. From a classical software IV&V viewpoint, it was heresy. Even the independent verification agent (IVA) we hired to look at each functional element (FE) in the models was skeptical, but willing to take the money and give it a try. The test and evaluation (T&E) community was willing to cooperate with anyone who thought they could validate models with test data, perhaps because they had seen so many of these software geeks come and go, validating their conclusion that the V&V folks (who actually wanted to measure 'terrain roughness') were operating at or beyond the fringes of sanity. Thankfully, they were kind enough to give us some test data we could use for validation.

WHAT IT IS

Classical approaches to validation tend to evaluate model outcomes or predictions via calibration to the test conditions and execution from end-to-end. The problem with this process lies in its failure to address contributing factors or functions. Figure 1 illustrates what can happen when initial target and launch data are used to drive a model of a missile flyout to a target. The measured target and missile ground tracks from the test are shown by short and long solid lines (respectively) and the modeled one by a dashed line. Is the flyout simulation valid?

Even though tested and modeled missiles flew a very different trajectory, they both guided to the target and missed. Model-predicted miss distance was 45 meters and the one calculated from the test data was 75. This type of comparison is just too simplistic for the simulation under consideration and only raises questions about underlying causes of the obvious trajectory differences. In other words, it's the wrong way to approach the validation of this type of model.

There could also be test-related measurement problems, but more on that later. Given that our model-level comparison is not convincing, we would like to diagnose the underlying sources of error in the model. We focus our attention on contributing functions that can be compared with other test data, which in this case we are fortunate to have.

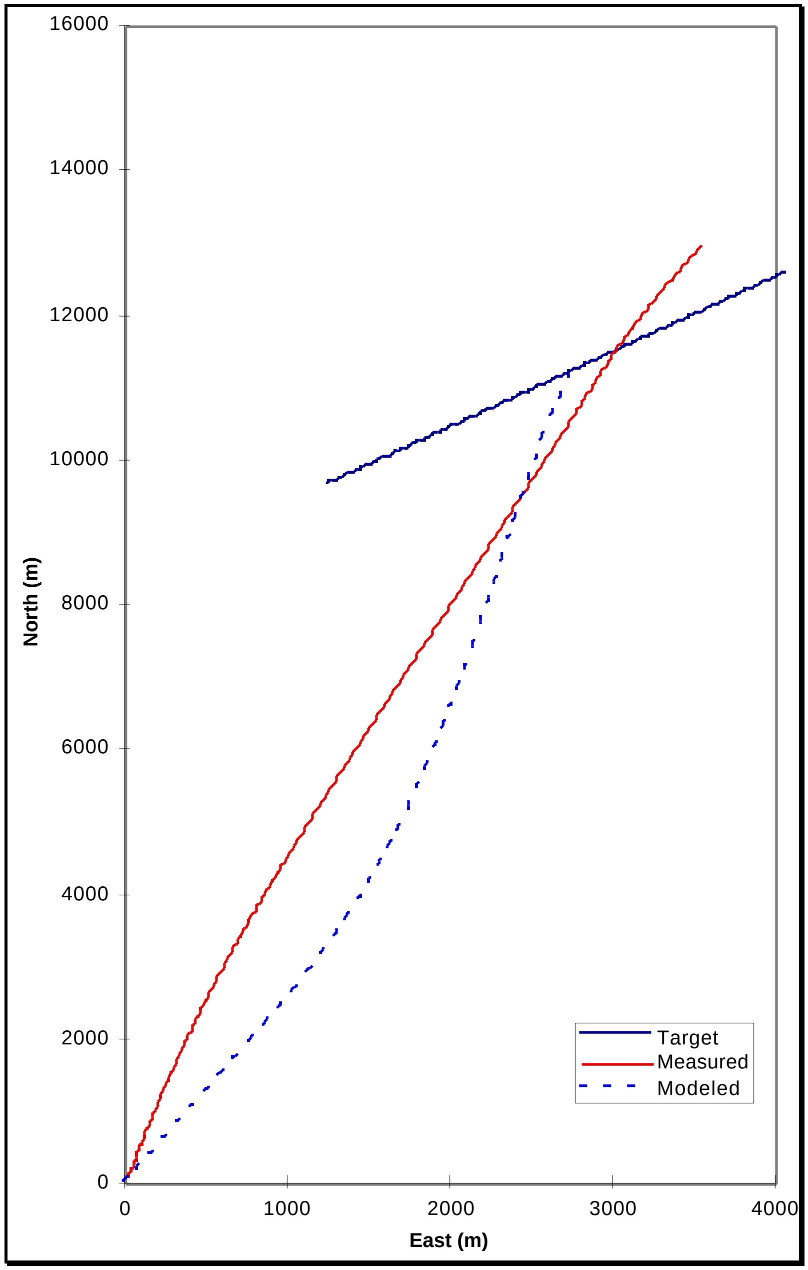


FIGURE 1. Typical Missile Model-to-Test Comparison.

Platform Missile

- 1.0 Attributes
  - 1.1 Configuration
  - 1.2 Movement
    - 1.2.1 Propulsion
      - 1.2.1.1 Boost
      - 1.2.1.2 Cruise
    - 1.2.2 Aero/Kinematics
      - 1.2.2.1 Force/Moment Generation
      - 1.2.2.2 Missile Movement
- 2.0 Sensors
  - 2.1 Seeker
    - 2.1.1 Transmitter
    - 2.1.2 Receiver
    - 2.1.3 Antenna
    - 2.1.4 Signal Processing
  - 2.2 Fuze
    - 2.2.1 Transmitter
    - 2.2.2 Receiver
    - 2.2.3 Antenna
    - 2.2.4 Signal Processing
- 3.0 Weapons
  - 3.1 Warhead
    - 3.1.1 Blast Pressure
    - 3.1.2 Frag Density Distribution
- 4.0 Comm Devices
  - 4.1 Uplink
  - 4.2 Downlink
- 5.0 CM/CCM
- 6.0 DME
  - 6.1 Guidance
    - 6.1.1 Proportional Navigation
    - 6.1.2 Command
  - 6.2 Autopilot
    - 6.2.1 Lateral
    - 6.2.2 Roll

FIGURE 2. FE Template for Missile Simulation.

The functional V&V process tested by SMART was based on the hierarchical (or sequential) decomposition of M&S into functional elements (FEs) that can be isolated, examined, and evaluated (validated). A sample decomposition for a missile flyout simulation [8] is shown in Figure 2. Each of the missile subsystems and their components are represented at a level of detail suitable for validation (i.e., their inputs and outputs can be measured). Similar templates have been constructed for aircraft [9] and ground vehicle [10] platforms and additional functional areas were defined for command, control, communications and intelligence functions simulated by mission level analysis models [11].

The functional design in this case is a system, or platform with attributes (e.g., weight, propulsion, configuration) and specific functions. These functions (e.g., guidance, control, countermeasure detection, fuzing) are simulated by subsystem, or component models and result in airframe movement to arrive at a predicted position and orientation. Figure 3 illustrates a notional design: guidance commands from the target tracker are fed to the autopilot, which activates the canards, fins, or wings, producing pitch, yaw, and roll moments on the airframe, which responds according to its thrust, lift, and drag coefficients. Missile body rates are fed back to the autopilot to stabilize and maintain control of the system throughout its flight. Autopilot response typically varies with flight speeds and/or regimes.

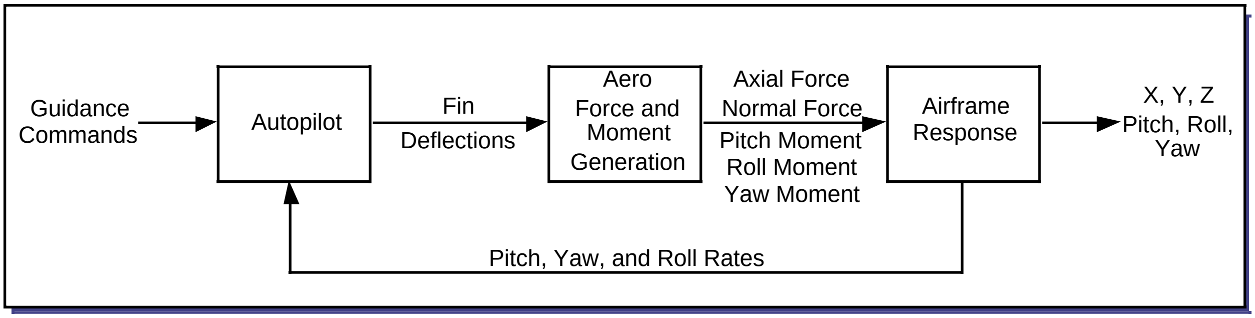


FIGURE 3. Missile Simulation Block Diagram.

The validator seeking to establish the credibility of the composite simulation must be prepared to examine each of the software modules, or objects in the chain that contribute to the ultimately predicted trajectory. This examination can be guided by the functional decomposition, the design and what data elements can be measured during testing. Consider the autopilot response shown in Figure 4, which exhibits significant correlation except for the negative peaks. Note the small positive bias in the test data during the first five seconds, which if removed, would improve correlation in several areas.

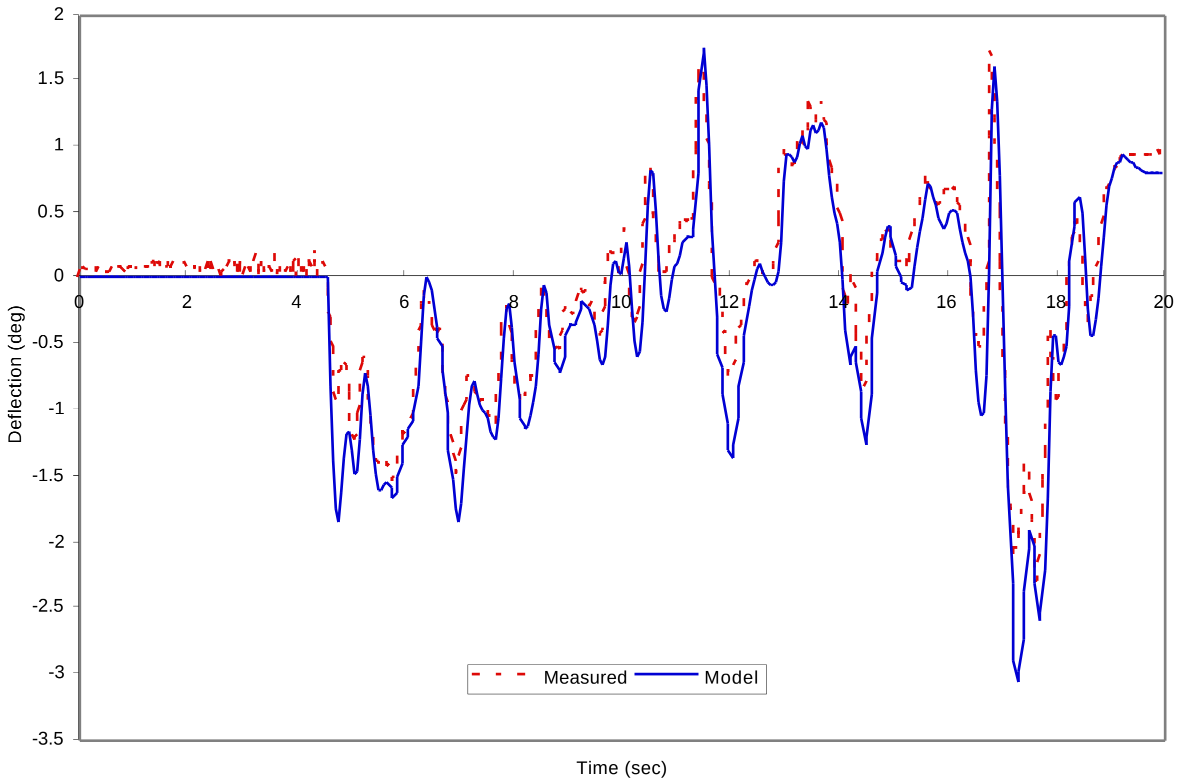


FIGURE 4. Measured and Modeled Autopilot Response.

The linear, or chain reaction design of this model is common among simulations of dynamic systems, and can be further complicated by similar processes that may be executed concurrently. The more modules needed to complete the simulation, the greater the magnitude of potential error propagation from end-to-end. In this example, autopilot errors contribute to lateral acceleration errors as shown in Figure 5, which also illustrates a typical test measurement problem between 2 and 5 seconds. During the boost phase, on-board accelerometers recorded an unbelievable, noisy response (perhaps due to vibration) that settled out when guidance begins. The model prediction is more believable that the data in this region.

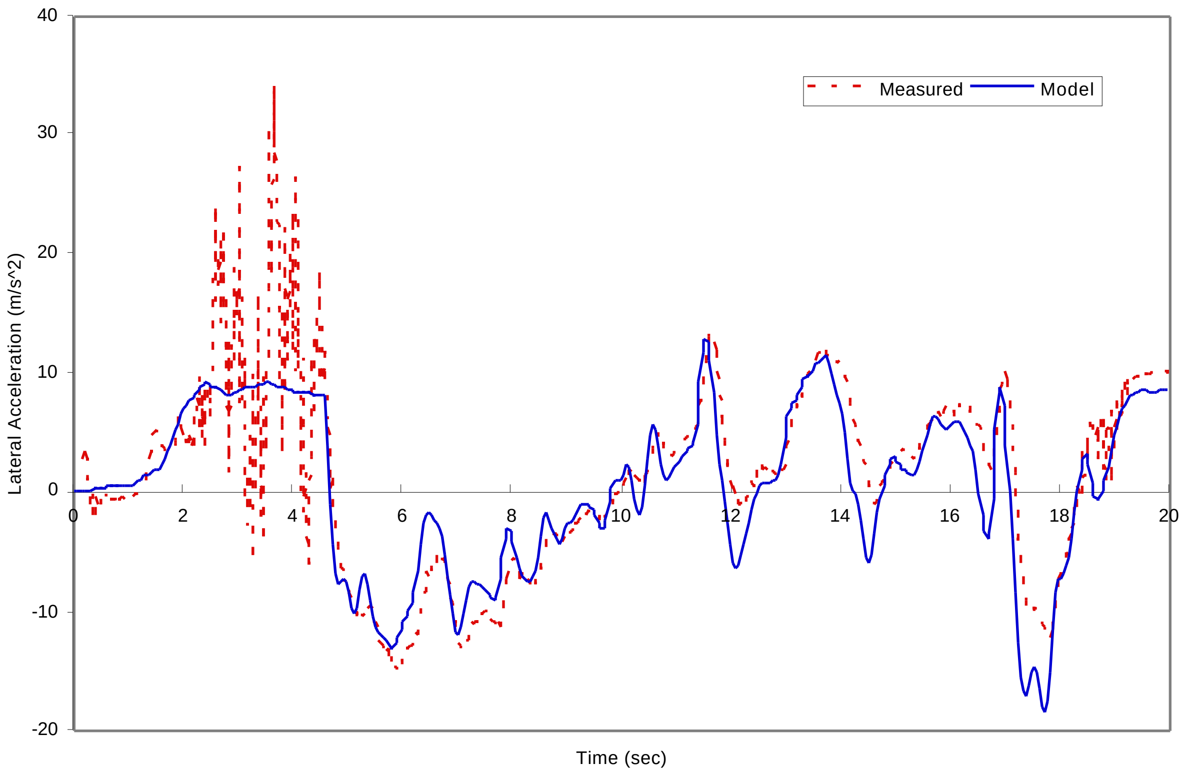


FIGURE 5. Measured and Modeled Lateral Accelerations.

Differences in airframe response to accelerations yield the pitch and yaw angle errors in Figure 6. The discrepancy during the first three seconds is an interesting one, apparently due to the establishment of flight. After leaving the launcher, the missile pitches over and then back up as airspeed increases and control surfaces begin to produce lift and/or stabilization. Relative heading (yaw angle) comparisons show a similar behavior that is clearly not represented in the model, but reflected in the ground track plot in Figure 1. At the end of boost, the test and modeled missile headings were more than 10 degrees apart. The initial pitch/yaw maneuver gave the test missile a greater lead angle relative to the target, which eliminated the need for a mid-course correction as predicted by the model. Should we expect the model to simulate this behavior? More data would be needed to characterize it well enough for simulation purposes.

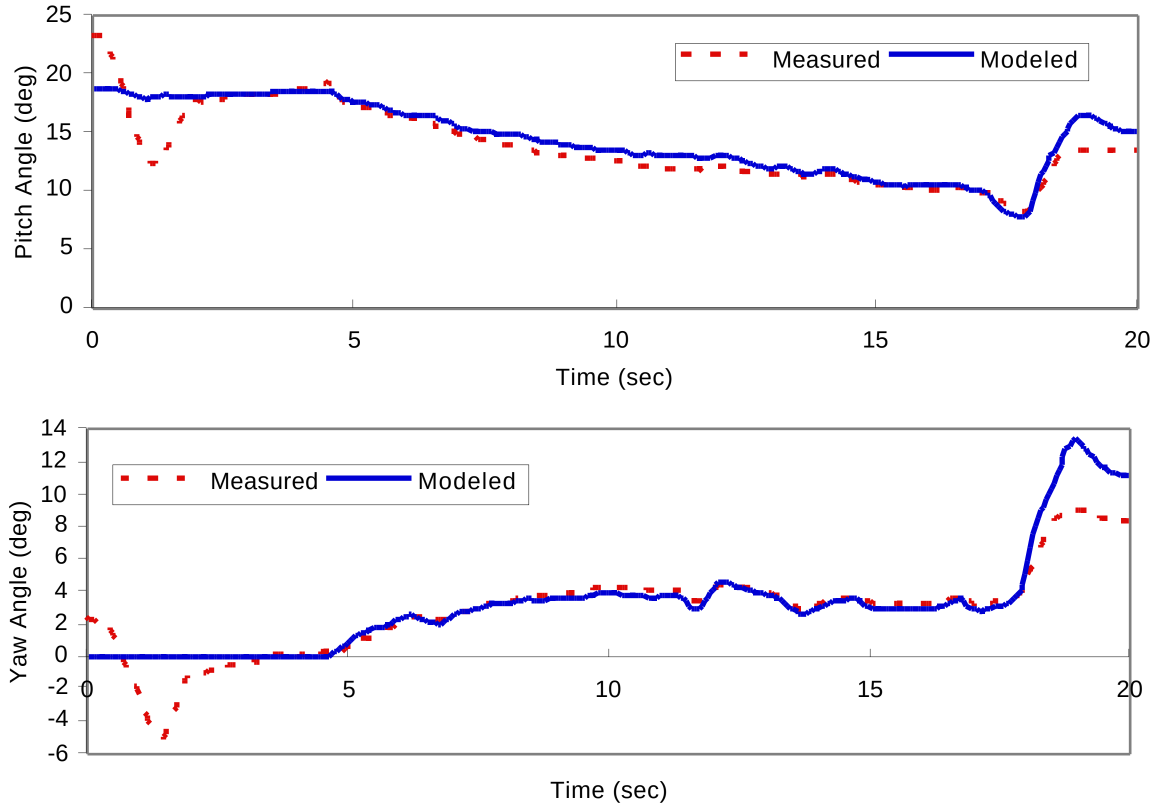


FIGURE 6. Measured and Modeled Airframe Pitch and Yaw Angles.

A more meaningful comparison for evaluation of the flyout model is shown in Figure 7, which illustrates the model-predicted altitude profile that results when the guidance from the measured one is injected. It is interesting that the model, which exhibited slightly higher pitch through ten seconds of the flight, produced a lower altitude trajectory than was measured. If the measurement is correct, this result points to differences in the thrust and/or drag attributes of the missile model, which could be explored and verified by a sensitivity analysis.

Given that each flyout model execution provides one of many possible trajectories and that the test data missile and target position errors are a function of range from the measuring device(s), how close must the model be to the measurement to be considered valid for its intended use? Here lies the current challenge for developers of these types of simulations: where and how much noise (variability) should be introduced to better represent the family of systems, given that there will be measurable round-to-round variation?

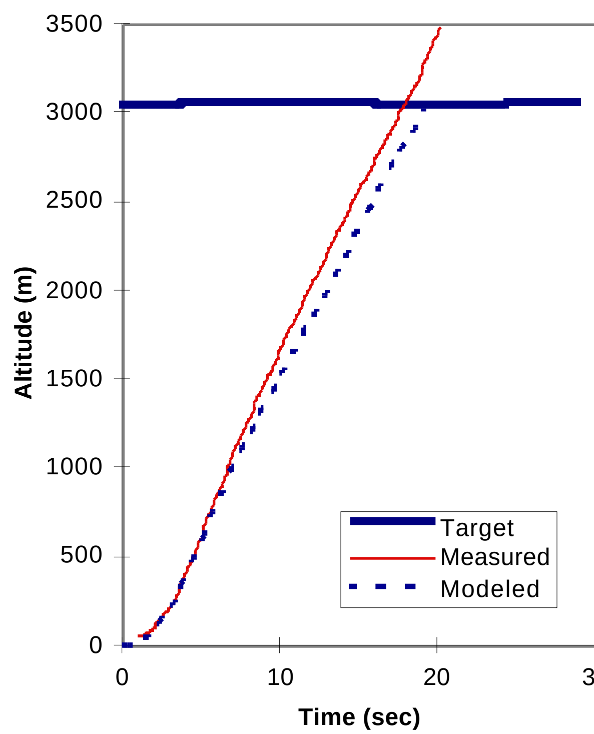


FIGURE 7. Measured and Modeled Response to Pitch Guidance.

Recalling our concern about thrust profile attributes above, Figure 8 illustrates measurements for six missiles of a different type (with a two-stage rocket motor). The solid thin line marks the best estimate used in the model prior to testing.

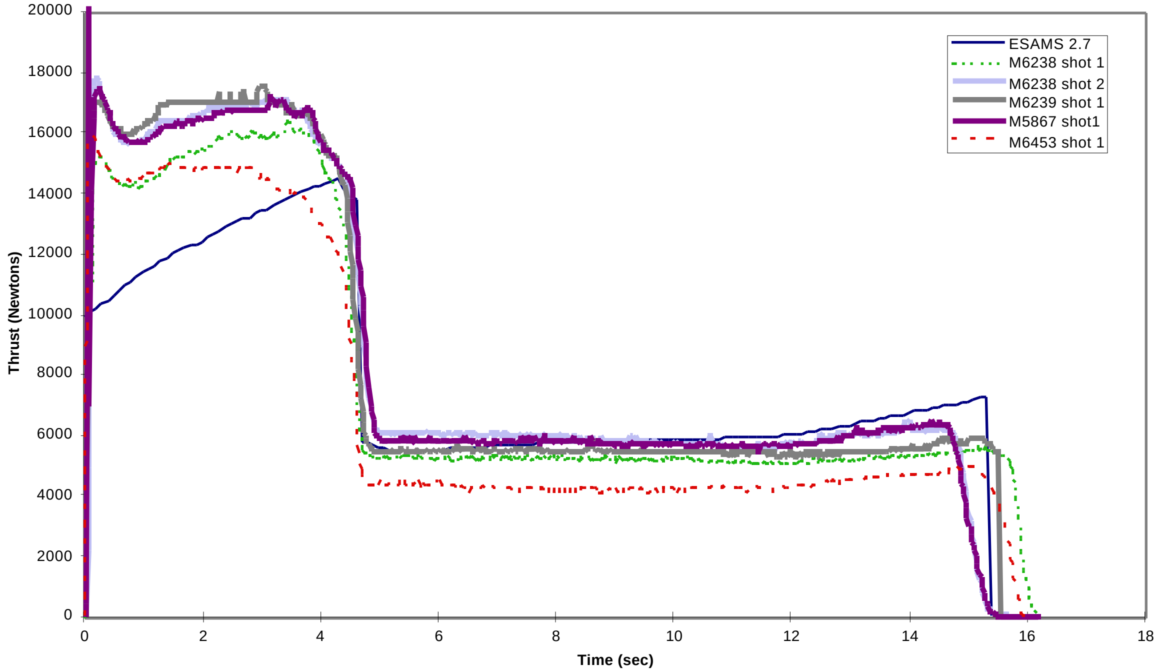


FIGURE 8. Measured and Modeled Thrust Profiles for a SAM.

In order to see how much of the difference in Figure 7 were due to thrust variation, we could bound the problem by making model runs using extreme values and then vary thrust about the mean until best time of flight (TOF) agreement was achieved. The resulting thrust profile would also improve correlation with the lateral acceleration (aero force and moment generation) and the resulting body angles (airframe response). Using an improved thrust model and launch angle bias of about 10 degrees, the measured trajectory in Figure 1 can be duplicated with the additional confidence that each of the subsystems are replicating real world response.

WHY DO IT?

Complex models and/or simulations (M&S) are difficult to verify and/or validate (V&V) for a number of reasons. They are typically composed of several (or many) lower level (or component) models integrated together via some type of software architecture so that errors produced by one are propagated to others, which may in turn amplify or reduce their impact upon ultimate results. In simulating the interactions among such models, the physical phenomena involved can also be complex and involve multiple sources of error due to assumptions about properties not modeled. Finally, M&S are usually built to perform specific functions in response to perceived needs, which may have been documented, but in many cases have not, so the intended scope of application and any associated limitations are often left to the user's imagination or experimentation. Today, we find many M&S in wide use throughout the DoD that have not been subjected to credibility assessments that would assure proper application to specific problems.

The idea behind V&V applied at the functional level was driven by the realization that total, integrated, or end-to-end, independent model validation was expensive, often impractical, and sometimes impossible. This is essentially due to the problem of compounding or compensating error, which precipitates a requirement to understand where and how results are being affected.

When the observations don't agree with the M&S built using the best information available, the cause of the discrepancy is immediately sought and results in an investigation of the inputs (components) that contribute to it. Our experience with combat prediction models and sensor simulations produced at least four reasons why functional V&V should be considered.

1. M&S are too complex to V&V from end-to-end.

The days when a DoD model or simulation could be verified, validated, and accredited (VV&A'd) once and for all are gone. There is some question as to whether such days ever actually existed. M&S commonly used throughout the DoD have grown from a few hundred or thousand lines of code in a few modules to millions of lines of code in thousands of modules. This growth has resulted in more flexible and capable M&S that perform or simulate many more functions and provide multiple execution and output options. The V&V requirements to address such M&S have also become increasingly complex and the ability to cover the entire software package cost prohibitive from a "classical" V&V perspective. That is, each module and line of code cannot be verified against its design, often because one does not exist, but also because of the time required to do so. By the time a typical application can be completely verified, a new version is available and ready for verification. Further, the investment required for validation of an application's many functions and capabilities can rival the one necessary for its development.

2. T&E efforts are focused on functional performance.

Data (observations and measurements) derived from testing activities comprise the best information to support validation efforts and these are most often focused on specific functional areas of the systems under evaluation. Testers in the DoD have long realized that a single experiment, or trial, cannot address more than a few objectives due the difficulties associated with controlling independent variables while collecting data on dependent ones. Thus, the results of tests can often be readily compared with M&S functions, or portions of the code that comprise them, but comparison with aggregate, more complex measures is not as easy or feasible with typical test range data products. Analysis is usually required to accomplish such objectives and attribute real worth to the test results as well as the M&S performance measures.

3. Mature M&S add functions and capability incrementally.

M&S developments are focused on specific functional areas due to user demand for capabilities to address specific operational and/or analytical issues. Given that M&S of some form exist for virtually every system operated or envisioned by the DoD, shortfalls or deficiencies in them have been identified and used to direct developments that will be incorporated into subsequent versions. Model versions are produced at quasi-annual rates in order to keep pace with funding, development, and testing of evolving and emerging DoD systems and each new version boasts new capabilities in terms of functions and/or outputs not previously available. If the previous versions of the model have been V&V'd and it can be shown that their functionality has been unaffected, then only new functions need to be V&V'd to maintain its accreditation status. If the model is rewritten for each version release, then previous V&V efforts are either repeated or more often abandoned.

Envision the plight of program managers, analysts, training ranges, research and development organizations, and operational units within the DoD that are faced with needs potentially suitable for the application of M&S. These groups typically canvas the vast array of M&S available in search of suitable programs and software packages that advertise capabilities to address their specific requirements for production of data and results. Even more important than advertised capabilities is the V&V information about how well a given model, simulation, or software package performs the desired functions. Furthermore, if candidate M&S have been accredited for similar purposes, or widely used to support similar types of usage, determination of suitability for the application at hand will be facilitated (if evidence of these facts exists).

However, this situation rarely occurs for a variety of reasons, but mostly because communities within the DoD responsible for M&S have concluded that V&V efforts required to support accreditation are cost-prohibitive. Data collected from V&V activities during SMART Project efforts suggest that this conclusion may not be valid [12]. It is the incremental application of the V&V process to individual M&S functions that yields more affordable credibility over time than that derived from repeated global V&V efforts.

Many activities can contribute to M&S credibility; from the documentation of an appropriate design, software integration test results, prior uses and accreditations of the model, or its inherent assumptions and limitations (as well as known errors), to certification of each data element and verification of each line of code Functional V&V provides added value through insight into how M&S are constructed and how well they can be expected to represent reality. It directly supports functional requirements for accreditation, is compatible with incremental software development and improvement efforts, and facilitates the diagnosis of problematic comparisons between M&S and test results. The process described and tested by the SMART Project relied heavily upon functional V&V, and demonstrated the potential to make V&V cost-effective and feasible for many M&S used in the DoD.

REFERENCES

1. Ritchie, Adelia E. (Ed.), *Simulation Validation Workshop Proceedings* (SIMVAL II). Military Operations Research Society, 2 April 1992.

2. Department of Defense, *Department of Defense Dictionary of Military and Associated Terms*, (Joint Publication 1-02), 23 March 1994.

3. Witt, Betsy Stone, *Validation, Verification, and Accreditation — A Perspective from the Intelligence Community*, PHALANX, June 1994.

4. Davis, Paul K., *Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations*, R-4249-ACQ, RAND, 1992.

5. Joint Technical Coordinating Group for Aircraft Survivability (JTCG/AS). *V&V From A to Z, A SMART Approach to VV&A for Acquisition M&S*, by Paul Muessig, et al, Joint Accreditation Support Activity (JASA), NAWCWPNS, China Lake, CA, March 1997. (UNCLASSIFIED.)

6. Pace, D. K., *Naval VV&A Process*, PHALANX, September 1993.

7. Balci, Osman, *Principles of Simulation Model Validation, Verification, and Testing*, TR-94-24, Virginia Polytechnic Institute and State University, June 1994.

8. Joint Technical Coordinating Group for Aircraft Survivability (JTCG/AS). *Accreditation Support Package for ESAMS*, by Baty, R.S., et al. BDM Engineering Services Corp., Albuquerque, NM, for the SMART Project, NAWCWPNS, China Lake, CA (JTCG/AS-92-SM-009, 010, 011, publication UNCLASSIFIED).

9. ______. *Accreditation Support Package for BRAWLER*, by Gary Eiserman, et al. Decision Science Applications, Inc., Arlington, VA, for the SMART Project, NAWCWPNS, China Lake, CA (JTCG/AS-96-SM-011, publication UNCLASSIFIED).

10. ______. *Accreditation Support Package for RADGUNS*, by Traci McCormick, et al. ASI Systems International, Ridgecrest, CA, for the SMART Project, NAWCWPNS, China Lake, CA (JTCG/AS-92-SM-012, 013, 014, publication UNCLASSIFIED).

11. ______. *Accreditation Support Package for EADSIM*, by Hal Cronkhite, et al. Teledyne-Brown Engineering, Inc., Huntsville, AL, for the SMART Project, NAWCWPNS, China Lake, CA (JTCG/AS-96-SM-012, publication UNCLASSIFIED).

12. Muessig, P. R., and D. R. Laack, *Cost Effective VV&A: Five Prerequisites*, Proceedings of the 1996 Summer Computer Simulation Conference, Ed. Ingalls, V. W., J. Cynamon, and A. V. Saylor, July 1996.

BUILDING A FUNCTIONAL AREA TEMPLATE

by Barry O'Neal

One of the basic tenets of the SMART approach to structuring verification and validation activities is that it is possible to develop a functional area template (FAT), representing the functions in a model, that is generic enough to also apply to other models built with overlapping functionality. While this has been proven to be true, at least for the models assessed by SMART (and later JASA), there is perhaps some value and insight to be gained from recounting the embattled history of the FAT and the variations on it that have been put forth by model developers and critics.

ENGAGEMENT LEVEL TEMPLATES

The first functional area template (FAT) SMART developed was for the ESAMS (Enhanced Surface-to-Air Missile Simulation) engagement analysis model and it is shown in Figure 1. The structure of this FAT was derived from a threat description document (TDD) that had recently been published by the Defense Intelligence Agency (DIA)'s Missile and Space Intelligence Center (MSIC) for a threat SAM system[1]. The RF SENSOR (upper) portion of this template was developed in concert with the developers of ALARM (Advanced Low Altitude Radar Model) and later also applied to the RADGUNS (Radar-Directed Gun System) package of antiaircraft artillery simulations.

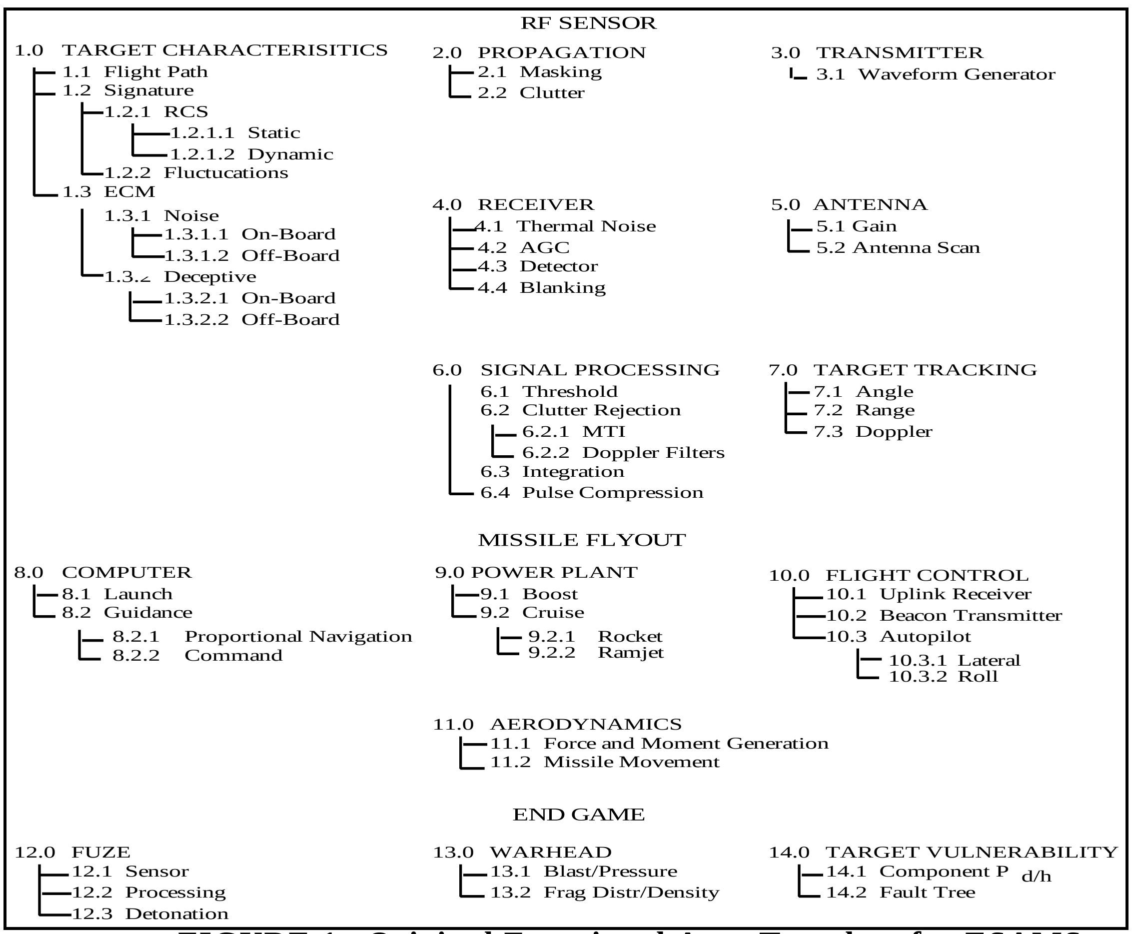


FIGURE 1. Original Functional Area Template for ESAMS.

There were several problems with this original template, not the least of which was the inclusion of the TARGET CHARACTERISTICS functional elements (FEs) in the RF SENSOR functional area. This was a problem because all of the documentation produced was organized by FE, which meant that the target Flight Path (1.1) was the first FE, which was always blank. Really more of an annoyance than a problem, but flight path data constituted input for each model and thus no need to perform V&V on it was envisioned. Another problem was that environmental FEs were also included under a PROPAGATION subarea portion of the RF SENSOR. This caused a problem because it was the first area of V&V investigation and all of the models treated those FEs independently of the sensor except for the consideration of transmitted frequency and power. So in an effort to construct a template with FEs common to three models, we ended up with only a few things in common (transmitter, receiver, and antenna) and a FAT that needed fixing.

MISSION LEVEL TEMPLATES

When the project sponsors asked SMART to examine the Brawler and EADSIM models, it became clear that this FAT simply would not do. These new models dealt with command, control, communications, and intelligence (C3I) issues as well as decision making elements (DMEs) [2,3] which we had not addressed beyond some operator functions (reacquisition and optical tracking) in RADGUNS. We decided to let the experts do the decomposition for us, so we briefed each of those model developers on the FAT concept and asked them to come up with a template for their own model. The plan was to merge them as best we could and then proceed with functional decomposition of the code. The Brawler folks came in first, and their template bore no resemblance whatsoever to the one we started with. They came up with four functional areas:

1. Analysis Environment, which included all of their pre- and post processing tools,
2. Simulation Architecture, which addressed startup, execution, and shutdown functions,
3. Operator Models, which covered pilots and AWACS or GCI controllers, and
4. Physical Models, which covered the various sensors, weapons, and countermeasures.

The template itself spanned twelve pages of Appendix A in their first document and some FEs were nested six levels deep (e.g., 1.3.2.2.1.1 Plume Radiation). A portion (one page) on the Aircraft platform is shown in Figure 2. The good news about this particular portion was that there was some similarity between their Propulsion, Aerodynamics, and Signatures template and the Flight Path and Signature sections of our original template. At least we got something right! The bad news was that it also bore no resemblance to the one provided by the EADSIM camp, so our idea about merging them together suddenly seemed like a stupid one.

The EADSIM developer started out with a Systems Functional Area (a portion of which is shown in Figure 3), which broke down into Elements, of which there were seven subtypes; Sensors, Weapons, Comm Devices, Aircraft, IR Signatures, RCS, and Rulesets. The Rulesets were subdivided further into 27 different types ranging from Fighter or Bomber to AWACS or Airfields, Intel Centers, SAM Launchers, Commanders, and so on. There were also separate templates for the Command & Control (C2) and Network functions, which we eventually decided was a good idea.

The six level hierarchy (e.g., 1.3.2.2.1.1) used by both developers to cover the level of detail in their software was problematic because it made the scope of FEs for purposes of V&V too narrow. Of course, these templates were so far apart in approach and level of detail that combining them in any way was hopeless, and adopting one over the other seemed unfair. So we set about building one from a perhpas more objective point of view that covered all aspects of offensive and defensive platforms participating in a mission.

IV PHYSICAL SYSTEMS

1. Aircraft

- 1.1 Propulsion
  - 1.1.1 Compute Engine State (seteng)
  - 1.1.2 Compute Resultant, Effective Thrust (effthr)
  - 1.1.3 Fuel Flow (fflo)
- 1.2 Aerodynamics/Kinematics
  - 1.2.1 Control (rates)
    - 1.2.1.1 Compute Initial Aircraft Flight State (flacti, flact0, acarea)
    - 1.2.1.2 Roll & Pitch Lag Filter Frequencies
    - 1.2.1.3 Max Roll/Pitch Rates (ratmx1, ratmx2)
    - 1.2.1.4 Commanded Roll/Pitch Rates (gwreq)
    - 1.2.1.5 Resultant, Filtered Roll/Pitch Rates (pnow)
    - 1.2.1.6 Resultant Yaw Rate
    - 1.2.1.7 Determination of Lift/AOA Needed for Pitch Response (aoanow)
      - 1.2.1.7.1 Lift Limit Mods to Pitch Rate (gldmx)
  - 1.2.2 Compute Final Aircraft Flight State (flact1)
    - 1.2.2.1 Angular Rates
    - 1.2.2.2 Orientation
    - 1.2.2.5 Position, Velocity, Acceleration
- 1.3 Signatures
  - 1.3.1 Static
    - 1.3.1.1 RF Cross Section
  - 1.3.2 Dynamic
    - 1.3.2.1 RF Cross Section
    - 1.3.2.2 Infrared
      - 1.3.2.2.1 Computation of Signal
        - 1.3.2.2.1.1 Plume Radiation
        - 1.3.2.2.1.2 Airframe Radiation
        - 1.3.2.2.1.3 Hot Part Radiation
        - 1.3.2.2.1.4 Reflected Solar Radiation
        - 1.3.2.2.1.5 Reflected Sky Radiation
        - 1.3.2.2.1.6 Reflected Earth Radiation
      - 1.3.2.2.2 Compute Background
      - 1.3.2.2.3 Compute Foreground
    - 1.3.2.3 Visual
      - 1.3.2.3.1 Threshold Contrast
      - 1.3.2.3.2 Probability of Detection Calculation

FIGURE 2. Aircraft Portion of FAT Proposed for Brawler.

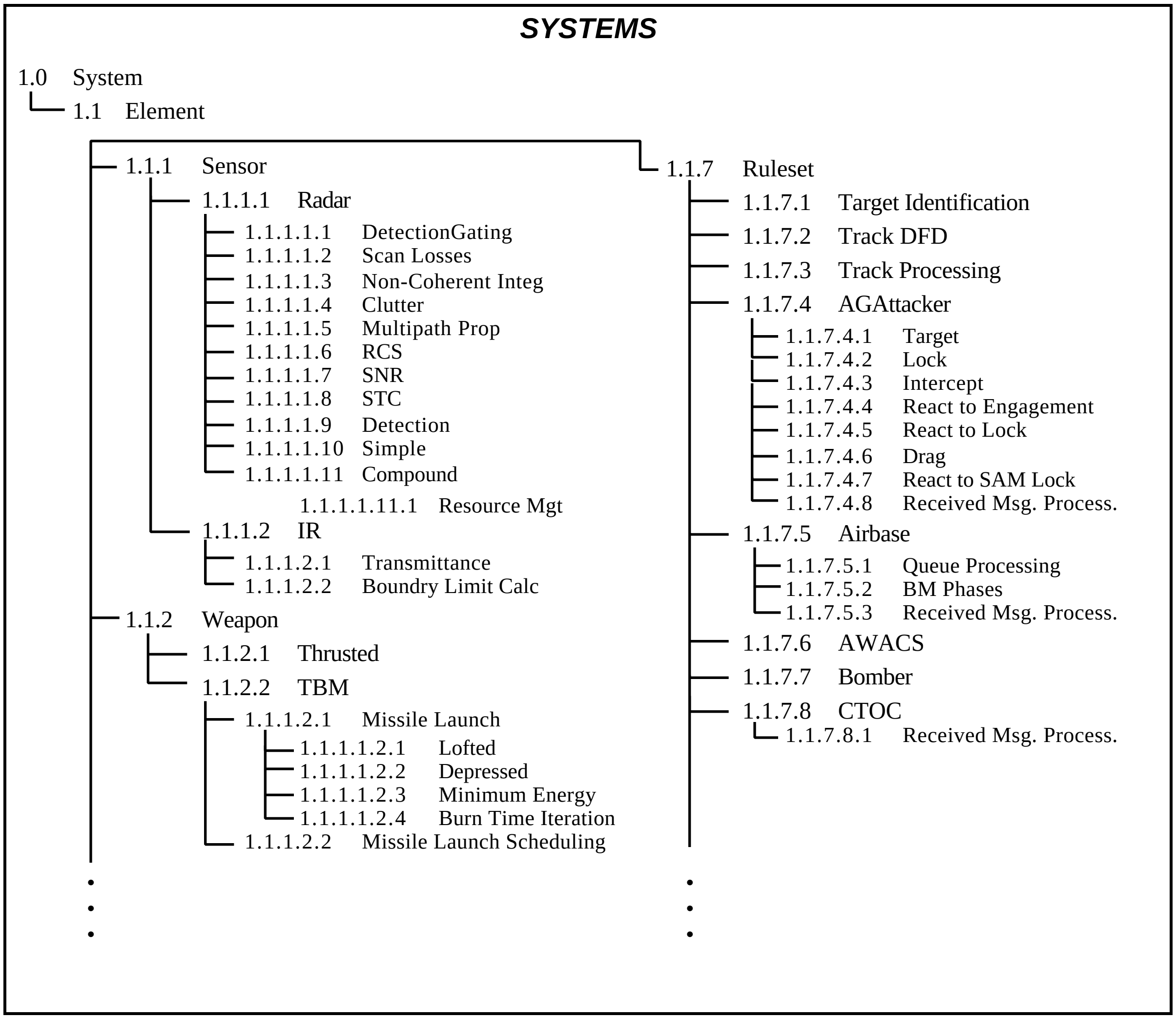


FIGURE 3. Systems Portion of FAT Proposed for EADSIM.

We sat down [4] and settled on the platform as the lowest common denominator, which we later realized was also true for the engagement level M&S we had been working with. Because both EADSIM and Brawler had some Command, Control, and Communications (C3) capabilities, we separated out this functional area, as well as one for the Environment, and focused on the platform (which could be an aircraft, a ship, a tank, a company or a platoon or a soldier), and defined the attributes and objects that could be related to any of them. The idea was to construct a generic template that the Brawler and EADSIM modeling experts could pick and choose appropriate elememts from, or adapt their simulations to.

In addition to attributes, which included configurations, mobility, signatures, and vulnerability, we identified five other functional (object) areas:

1. Sensors - EO, IR, RF, and/or UV;
2. Weapons - Guided, ballistic, laser, or microwave;
3. Communications Devices - Radios, computers, data links, and voices;
4. Countermeasures - Anti-weapons, chaff, flares, jamming, or mode changes; and
5. Decision Making Elements - Pilots, commanders, soldiers, fuzes, or fire control computers.

We wanted the template for EADSIM also to be applicable to the Suppressor model, which we were more familiar with, and to be generic enough for application to any mission level, or M-on-N analysis model (Brawler in this case). So we elected to use the same FAT structure for all of them and indicated functions that were not modeled by a shadow, or italicized typeface. The nice thing about the FAT we came up with was that it could be instantiated for each type of platform or object modeled in the software (like the target, the radar, and the missile), using the same form, but with only slightly different nomenclature.

A COMMON MISSION LEVEL FAT

After weeks of email back and forth with the model developers and compromises over wording and arrangement of functions, we settled on the platform template shown in Figure 4 for both EADSIM and Brawler. An interesting thing happened when we briefed this concept and result to attendees of the next SMART Quarterly Progress Review (QPR), however. Even though we had made peace with the EADSIM and Brawler developers, who both claimed that the template was workable, the ALARM and ESAMS developers accused us of making a model-specific template rather than a generic one. Even though they didn't admit it, I suspected that they were aware of some of the problems with the existing templates, but also didn't want to admit that our ideas represented an improvement over their original versions. Change is a painful process.

We tried to allow for expansion of the FAT to accommodate higher modeling resolution in any area and made minor changes to accommodate the EADSIM Ruleset feature. Once we got the developers to see our perspective, there were no further problems with the remainder of the model decomposition effort, the results of which fit our FAT quite nicely in most areas. In fact, we thought we could make the old engagement models fit it, too.

An important point from a V&V perspective is that none of these templates involved much code that performed housekeeping, utility, or input/output functions, which were not essential for V&V activities. Our experience has shown little benefit in performing V&V on those functions, because it is usually obvious whether they work or not and they tend to be the functions most often wrung out during the debugging and testing processes.

Applying our new FAT to the old engagement level models we had started with resulted in a template for the aircraft target and environment shown in Figure 5. This functional hierarchy of attributes and objects (subsystems) should be applicable to any model that attempts to simulate any of the elements shown (or indicated as not modeled, as in this ESAMS example). Note that the separate environment template is quite simple, but it covers all of the known propagation effects except amplification (the opposite of attenuation), which we were to soon learn about from another model developer.

When we tried to fit the RF Sensor FEs into this new structure, the SAM Platform FAT shown in Figure 6 resulted. Notice that several unmodeled functions become apparent (which is perhaps why the developers didn't like it), but these are inconsequential. In fact, the more FEs a model template has in shadow italics, the easier it will be to verify and validate, since the fewer functions it actually models the less data and effort will be required to assess their correctness and validity.

Instantiating the same template structure for the missile in flight, resulted in the sensor and fuze portions, more detail than had been shown by the previous template (see Figure 7), but all the other FEs fit nicely into their respective locations. Note that all of the elements in the original template are captured, but just in a different structure.

PLATFORM AIRCRAFT/SAM/TBM

- 1.0 Attributes
  - 1.1 Configuration
  - 1.2 Movement
    - 1.2.1 Propulsion
    - 1.2.2 Aero/kinematics
  - 1.3 Signatures (EO/IR/RF/UV)
    - 1.3.1 Static
    - 1.3.2 Dynamic
  - 1.4 Vulnerability
- 2.0 Sensors
  - 2.1 Acoustic
  - 2.2 Electro-optical
  - 2.3 Infrared
  - 2.4 Radio Frequency
  - 2.5 Other
- 3.0 Weapons
  - 3.1 Guided
  - 3.2 Unguided
- 4.0 Comm Devices
- 5.0 CM/CCM
  - 5.1 Noise
  - 5.2 Deception
- 6.0 Decision Making Elements
  - 6.1 Capabilities
  - 6.2 Knowledge Base
    - 6.2.1 Experience (Training/Proficiency)
    - 6.2.2 Doctrine/Tactics
    - 6.2.3 Mission Objectives
    - 6.2.4 Intel
    - 6.2.5 C^3
  - 6.3 Logic Processes
    - 6.3.1 Perception
    - 6.3.2 Assessment
    - 6.3.3 Decisions
    - 6.3.4 Actions

FIGURE 4. Platform FAT Used for EADSIM and Brawler.

Platform Aircraft TARGET

- 1.0 Attributes
  - 1.1 Configuration
  - 1.2 Movement - Programmed/Reactive
    - 1.2.1 Propulsion
    - 1.2.2 Aero/Kinematics
      - 3 DOF
      - Blue Max
  - 1.3 Signatures
    - 1.3.1 EO
    - 1.3.2 IR
    - 1.3.3 RF
      - Static
      - Dynamic
      - Fluctuations
  - 1.4 Vulnerability
- 2.0 Sensors
- 3.0 Weapons
- 4.0 Comm Devices
- 5.0 ECM
  - 5.1 Noise
    - 5.1.1 On-Board
    - 5.1.2 Off-Board
  - 5.2 Deceptive
    - 5.2.1 On-Board
    - 5.2.2 Off-Board
- 6.0 DME

ENVIRONMENT

- 1.0 Atmospheric
  - 1.1 Attenuation
  - 1.2 Refraction/Ducting
  - 1.3 Radiance/Transmittance
- 2.0 Topographic
  - 2.1 Clutter (Terrain/Sea/Culture)
  - 2.2 Multipath/Diffraction
  - 2.3 Masking

FIGURE 5. Proposed Target and Environment FAT for ESAMS.

OTHER TEMPLATE DEVELOPMENTS

While these efforts to develop and standardize functional breakdowns for M&S used to simulate combat missions and engagements were going on, an entirely separate group called the Electronic Digital Simulation Working Group (EDSWG) was developing similar breakdowns for threat models used to analyze electronic combat (EC) effectiveness. Comprised of experts from a host of DoD agencies and components, the EDSWG produced detailed functional templates for an array of current threats. An example of one of the EDSWG templates for a target tracking radar (TTR) is shown in Figure 8. The template starts with the Track function at the top, which is just a breakout of the Sensor function in our high level, generic template. A similar template could be constructed for target acquisition radar (TAR) functions that would branch from the same point (Sensor), but the processor might not have any Doppler, Range, or Angle functions, depending upon the system. The EDSWG also created Aircraft (Target) and Environment templates which were very similar to the ones we came up with, only in more detail.

Platform SAM System

- 1.0 *Attributes*
- 2.0 Sensors
  - 2.1 *Acoustic*
  - 2.2 *Electro-optical*
  - 2.3 IR
  - 2.4 RF
    - 2.4.1 Transmitter
      - 2.4.1.1 Waveform Generator
    - 2.4.2 Receiver
      - 2.4.2.1 Thermal Noise
      - 2.4.2.2 AGC
      - 2.4.2.3 Detector
      - 2.4.2.4 Blanking
    - 2.4.3 Antenna
      - 2.4.3.1 Gain
    - 2.4.4 Signal Processing
      - 2.4.4.1 Threshold
      - 2.4.4.2 Clutter Rejection
      - 2.4.4.3 Integration
      - 2.4.4.4 Pulse Compression
    - 2.4.5 Target Tracking
      - 2.4.5.1 Angle
      - 2.4.5.2 Range
      - 2.4.5.3 Doppler
- 3.0 Weapons
  - 3.1 Guided
    - 3.1.1 Surface-to-Air
  - 3.2 *Unguided*
- 4.0 Comm Devices
  - 4.1 Uplink
  - 4.2 Downlink
- 5.0 CM/CCM
- 6.0 DME
  - 6.1 *Operator*
  - 6.2 Launch Computer

FIGURE 6. Proposed SAM Platform FAT for ESAMS.

The focus of evaluations performed using the EDSWG templates was on simulation capability, but they could also have been used to plan and support V&V efforts. Functional requirement narratives were developed for each of the lowest level FEs so that a determination could be made as to whether the function as implemented in a model satisfied the requirement for EC analysis purposes. This type of activity is very close to verification, except that sources of data and algorithms and software testing were not performed. The focus of their effort was on a functional comparison of several models rather than V&V of any particular one.

V&V could, however, be accomplished using these very detailed templates to guide the process, but if each of the lowest level functions was treated separately, the results might be prohibitively expensive. For SMART, we attempted to roll up lower level functions as much as possible in an effort to make function level V&V more cost effective. It was not uncommon for subcontractors to spend the same amount of time and money verifying or validating a function with a hundred lines of code as they would for a function with a thousand lines of code. It's still not clear whether this apparent lack of economy of scale is due to the inherent nature of V&V activities or to the nature of contract procedures and the inability of contractors to work together cooperatively on V&V efforts.

Platform Missile

- 1.0 Attributes
  - 1.1 *Configuration*
  - 1.2 Movement
    - 1.2.1 Propulsion
      - 1.2.1.1 Boost
      - 1.2.1.2 Cruise
    - 1.2.2 Aero/Kinematics
      - 1.2.2.1 Force/Moment
      - 1.2.2.2 Missile Movement
- 2.0 Sensors
  - 2.1 Seeker
    - 2.1.1 Transmitter
    - 2.1.2 Receiver
    - 2.1.3 Antenna
    - 2.1.4 Signal Processing
  - 2.2 Fuze
    - 2.2.1 Transmitter
    - 2.2.2 Receiver
    - 2.2.3 Antenna
    - 2.2.4 Signal Processing
- 3.0 Weapons
  - 3.1 Warhead
    - 3.1.1 Blast Pressure
    - 3.1.2 Frag Density Distribution
- 4.0 Comm Devices
  - 4.1 Uplink
  - 4.2 Downlink
- 5.0 CM/CCM
- 6.0 DME
  - 6.1 Guidance
    - 6.1.1 Proportional
    - 6.1.2 Command
  - 6.2 Autopilot
    - 6.2.1 Lateral
    - 6.2.2 Roll

FIGURE 7. Proposed Missile Platform FAT for ESAMS.

When we began to support accreditation efforts for acquisition program customers, we became exposed to another simulation called SWEG (Simulated Warfare Environment Generator), which had the capability to execute in both constructive and virtual modes [5]. A descendent of the Air Force's Suppressor model, SWEG had been converted from FORTRAN to C++ and interfaced to human-in-the-loop and hardware-in-the-loop simulators by the Navy. When our customer decided to accredit the software for use in their analysis process, we went to the developers, showed them our functional templates, and asked them to produce a FAT for SWEG along the same lines. What we got, in addition to a explanation of the processes of semiosis, cognition, and existence, was the template shown in Figure 9.

The interesting thing about this decomposition at the time was that the objects usually associated with combat (e.g., sensors, weapons, countermeasures, etc.) were embedded deeply within the existence process, which was inextricably linked to the other processes by status and effects and perceptions and decisions. Along with the explanation of these processes, a diagram like the one in Figure 10 was supplied to illustrate how SWEG simulated reality and perceptions, as well as player and umpire functions.

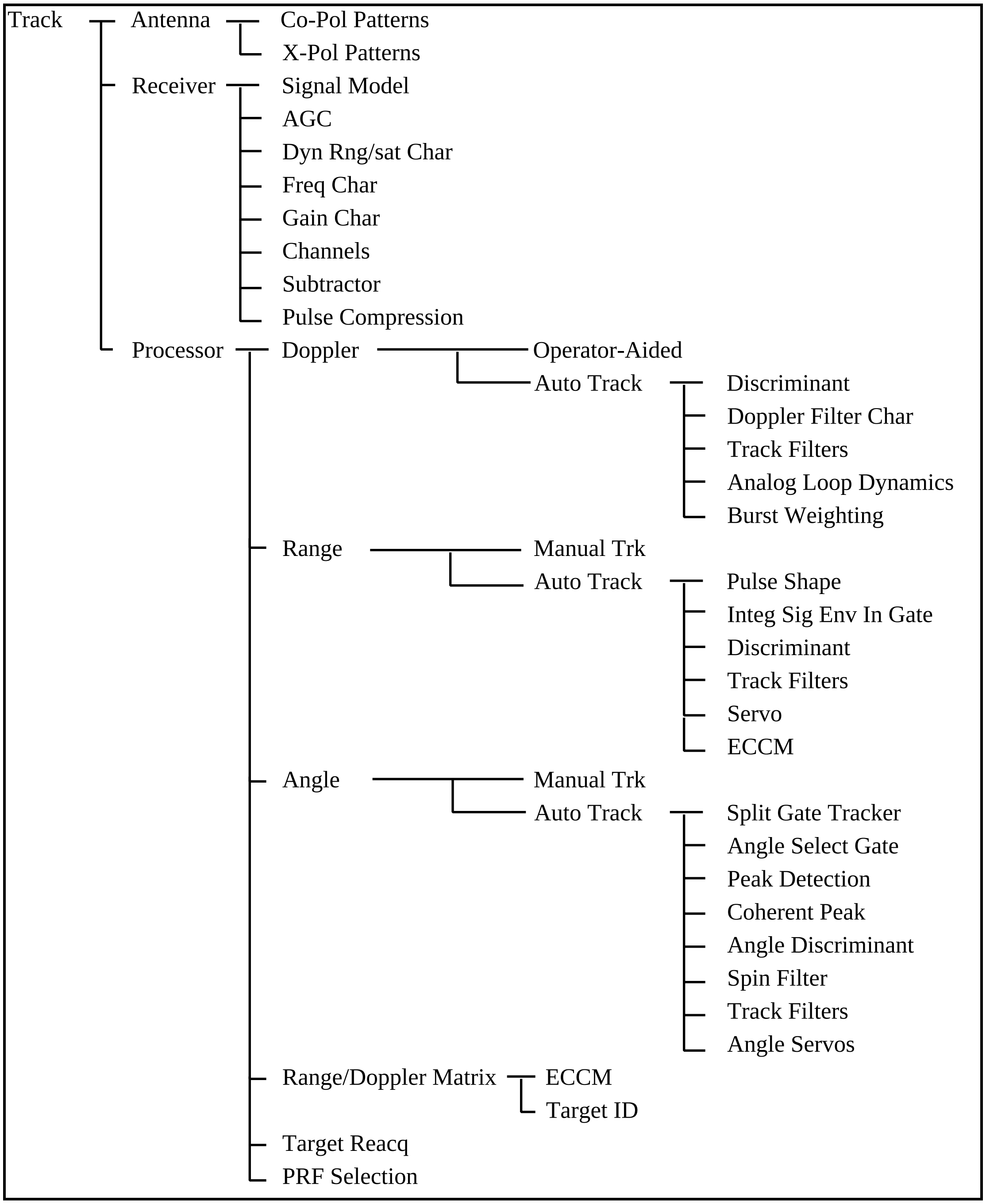


FIGURE 8. EDSWG TTR (Sensor) FAT.

If this was the way combat missions were simulated, it was back to the drawing board, and we'd have to rethink the work we had done on engagement level M&S. Surely these processes (semiosis, cognition, and existence) were happening at the engagement level, where perceptions by pilots and radar operators resulted in countermeasure actions and weapon launches, which resulted in energy exchanges between existing entities. After discussions with the authors, we concluded that the approach to functional decomposition taken by the SWEG developer was from the perspective of model development rather than that of the user; the latter was what we needed for

- 1.0 Semiosis
  - 1.1 Energy
    - 1.1.1 Emission
    - 1.1.2 Absorption
    - 1.1.3 Reflection
  - 1.2 Matter
  - 1.3 Movement
    - 1.3.1 Energy Interactions
    - 1.3.2 Matter Interactions
  - 1.4 External Interface
    - 1.4.1 Instructions
    - 1.4.2 Responses
- 2.0 Cognition
  - 2.1 Memory
    - 2.1.1 Pending
  - 2.2 Perceptions
  - 2.3 Tactics
  - 2.4 Orders
- 3.0 Existence
  - 3.1 Player
    - 3.1.1 Platform
      - 3.1.1.1 Element
        - 3.1.1.1.1 System
          - 3.1.1.1.1.1 Sensor
          - 3.1.1.1.1.2 Communications
          - 3.1.1.1.1.3 Weapon
          - 3.1.1.1.1.4 Disruptor
          - 3.1.1.1.1.5 Mover
          - 3.1.1.1.1.6 Thinker
          - 3.1.1.1.1.7 Expendable
            - 3.1.1.1.1.7.1 Ordnance
            - 3.1.1.1.1.7.2 Future Player
            - 3.1.1.1.1.7.3 Fuel
            - 3.1.1.1.1.7.4 Countermeasures
        - 3.1.1.1.2 Susceptibilities
      - 3.1.1.2 Shapes
      - 3.1.1.3 Kinematics
    - 3.1.2 Tactics
  - 3.2 Generic Functions
  - 3.3 Groups
- 4.0 Detail Changes
  - 4.1 Absorption
  - 4.2 Creation
- 5.0 Status Changes
  - 5.1 Sensor
  - 5.2 Communicator
  - 5.3 Disruptor
  - 5.4 Mover
  - 5.5 Thinker
  - 5.6 Shooter
- 6.0 Time Progression
- 7.0 Adiabatics

FIGURE 9. Proposed FAT for SWEG.

purposes of V&V. The developer agreed to try to fit their simulation (which seemed more like religion than software at this point) into our template, but warned us that most of the code associated with external semiosis functions would not be captured or represented by this format.

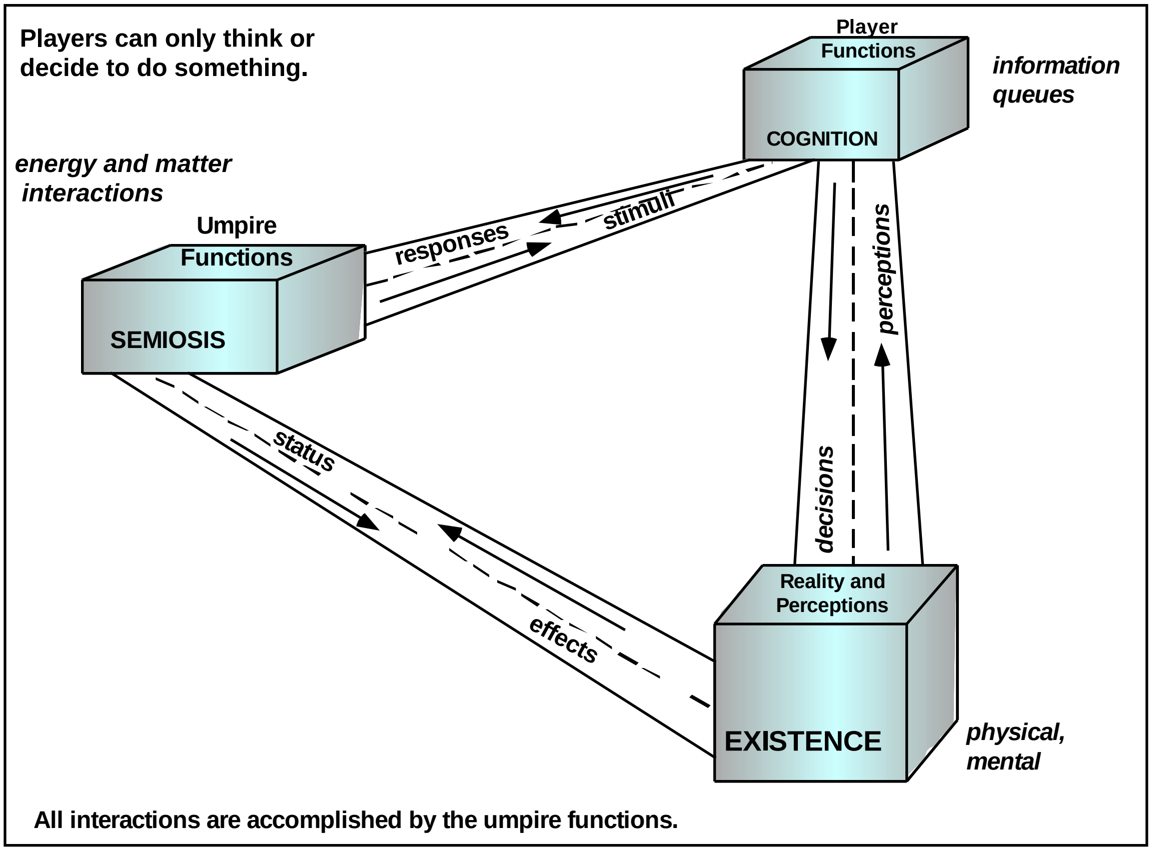


FIGURE 10. Processes Simulated by SWEG.

Fortunately for those users seeking V&V, models that simulate combat missions don't really do much that needs to be (or can be) validated. Being driven by large, complex databases that describe the platforms, systems, and command chain hierarchies for the scenario, they usually consist of an interpreter that analyzes problems with inputs, an event manager that schedules interactions among platforms or entities, and a database manager that keeps track of the outcomes of those interactions and accumulates their net effect. A much more important matter is the validity of the database representation of the opposing forces used to drive the simulation, which requires data verification, validation, and certification (VV&C). While there are functions within mission level models that can be validated with test data (e.g., command and control, target detections, weapon launches, timings, etc.), outcomes predicted by the multiple interactions among platforms and their support assets are naturally more difficult to validate owing to the expense associated with such testing [6].

SUMMARY

So what did we learn from our experience building functional area templates? It is possible to build a FAT which applies to a variety of models with overlapping functionality, both at the engineering level and at the mission level. However, it is not a simple task, due to the variety of approaches which model developers have taken in addressing similar problems. The most difficult part of this effort was in getting the model developers to agree that there may be another way to look at a problem than the one they used to develop their software. In each and every case the developer constructed a template that reflected their personal knowledge of the level of detail that their model treated as many functions as possible.

The key to making FAT development worthwhile is to focus on V&V requirements for model use, rather than model functional capibilities. The definition of functional elements must be made with the idea that verification and validation will be performed at that level. Too much detail in the FAT will mean that the V&V agent will spend too much time on a lot of little parts of the model, and the entire effort will be more expensive than necessary. Also, too much detail makes it difficult to show how functions in one model relate to those in another. Too little detail in the FAT will mean that the V&V agent will not have enough direction at each step in the process, since the FE's will be so large that it will be difficult to capitalize on what has been done for similar functions in other M&S.

References

1. Defense Intelligence Agency, *Threat Description Document for the Soviet SA-8 System*, Missile Space and Intelligence Center, Huntsville, (DIA/MSIC-92-00).

2. Decision-Science Applications, *The Brawler Air Combat Simulation Analyst Manual (Rev. 6.2) (Draft)*, Headquarters, U.S.Air Force Studies and Analysis Agencey, The Pentagon, Washington, DC, August 1995.

3. Teledyne Brown Engineering, Extended Air Defense Simulation (EADSIM) Manual Set Addendum, Version 4.01, U.S. Army Space and Strategic Defense Command, Hunstville, AL, November 1994.

4. Born, G.J., and Richardson, C.E., II, Personal communication during a meeting in Dayton, OH, July 1995.

5. Lattimore, Peter, et al, *SWEG 6.4 User's Guide, Volume I: Theory and Practice*, Bosque Technologies, Inc. Albuquerque, NM, 1996.

6. Born, G.J., *Validation Issues and Approach for Mission-Level Models,* Restricted Draft DRR-1257-OSD, RAND, Santa Monica, CA, November 1995.

3.3 VERIFICATION

How independent do Independent Verification and Validation activities need to be? Can the model developer be trusted? If so, with what, and how far? Do I really need to pay somebody else to come up to speed on a model they've never seen before, just to say I had somebody independent do it? For some illuminating answers to those questions, read on…

THE I IN IV&V: INDEPENDENCE OR INTELLIGENCE?

by Barry O'Neal

Independent Verification and Validation (IV&V) is usually deemed necessary due to the widely held belief that the person who developed the software cannot be expected to provide unbiased V&V information about it. This also holds true for opinions about related or similar software developed by others. There are several reasons for this, but probably the most prevalent is the pride of ownership or sense of responsibility that all software developers share when it comes to the fruits of their labors. In the past, much of the information used to support conclusions about V&V was subjective, or at least subject to some interpretation [1]. Today, as standards for the format and content of V&V information are emerging, these results are perhaps more objective, but implicit assumptions and interpretation of available V&V results still impact how well a user feels the software will meet his or her needs. Even though most agree that a model developer (MD) should be involved as a consultant when performing confidence or credibilty assessments and should be allowed to review the results [2], there is a general consensus that V&V results generated by a model developer cannot be trusted.

Tasking of IV&V agents (IVAs) to perform these functions has been the traditional solution, but there have been problems associated with this approach. First, IVAs cannot be expected to be experts, or even competent M&S evaluators over a very wide scope of applications. Aeronautical engineers capable of evaluating the performance of 6 degree of freedom (DOF) missile flyout simulations, for example, will find the intricacies of radar propagation, orbital mechanics, or human factor models difficult to evaluate. The basic principles and objectives will be within their grasp and they will be capable of evaluating them given sufficient time, but the problem facing the V&V project manager will be whether the time and resources required to bring an IVA up to speed on the M&S can be afforded or justified.

Second, even if the IVA is thoroughly familiar with the phenomenology or behavior associated with the model of interest, equal or greater familiarity with its approach to the problem or its history of application to problems is most often the property of the model developer (MD). This is a natural result of (typically) years of involvement with specifics of the problem area (domain) and proprietary knowledge of the approach used in developing the model as a tool to address those specifics.

Finally, given even intimate knowledge of the model's development history, its application to a variety of problems, and its approach to the methodology, the MD still has even more intimate knowledge of how the software is designed, structured, connected, or otherwise built. The IVA cannot afford to maintain this level of familiarity with any specific model without ongoing funding from the agency that already supports maintenance of the model, which also furthers the model developer's corporate knowledge of the code.

Thus, the use of IVAs generally incurs higher costs due to the learning curve associated with the M&S to be examined. The tradeoff between cost and value of V&V is shown in Figure 1, which was adapted from Sargent [3], and the heavy line represents the higher initial costs associated with IVA learning. Even in cases where the IVA is eminently qualified (like the aero engineers and the missile 6DOF above), there will be a period of time required for gaining enough familiarity with the particular model to be able to evaluate it fairly. This cost must be weighed against what the developer could provide over an equivalent amount of time, while giving consideration to the MD's known agenda of making the model look good in the light of the evaluation.

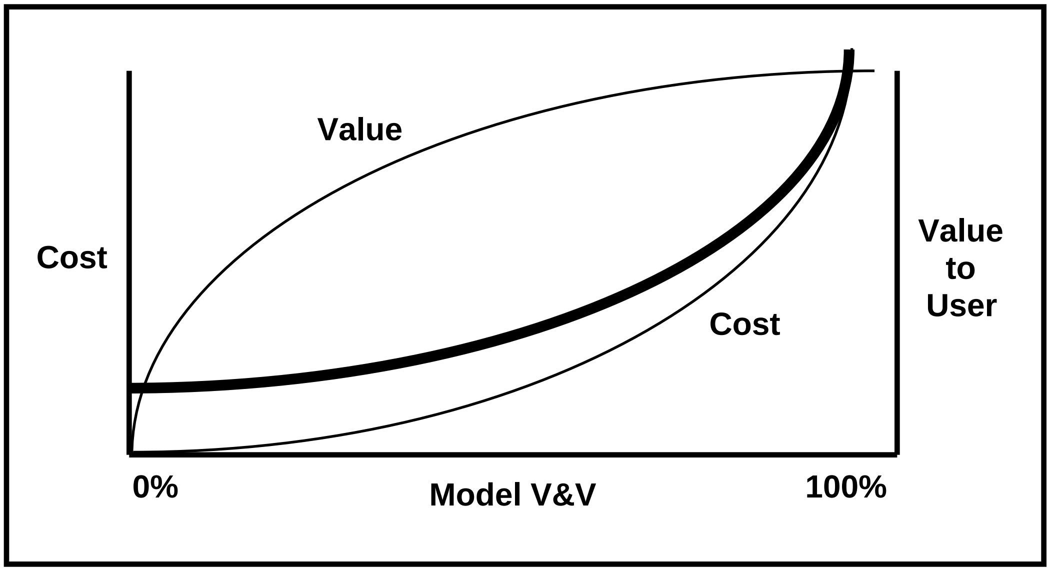


Figure 1. Additional Costs Required for IVA Learning

The fact remains that while everyone will believe what the IVA (who is assumed to have no personal agenda) says about the model, practically no one will believe what the MD reports unless it can be substantiated by other, independently derived supporting data. The other side of this tradeoff (from the V&V perspective) is that depending upon the environment in which the model is developed and maintained, the developer may be equally ignorant of the detailed requirements to be addressed and activities to be performed when supporting a V&V effort. This may come as a surprise to those who believe that V&V is somehow linked to the model development process currently employed throughout the DoD.

The challenge then for the manager charged with conducting a V&V project is how to make best use of the model developer's expertise without compromising the credibility of the V&V results. There is also the issue of how to make best use of IVAs to support the V&V process (with which they should be familiar), thereby lending credibility to the conclusions.

The SMART Project has, in the process of testing and documenting a VV&A process for M&S used to support analysis in the DoD, defined a number of V&V support and documentation tasks that the MD could be expected to perform. Furthermore, these incremental tasks do not require any subjective inputs and can either support other tasks assigned to IVAs, or contribute directly to the body of V&V information used to support M&S accreditation. In the remainder of this article, we describe those tasks, their outputs, and contributions to the V&V process.

V&V STATUS AND USAGE HISTORY

The single most important question asked by agencies attempting to accredit M&S for specific purposes is whether the M&S of interest have ever been accredited for similar purposes before. It is perhaps interesting (but sad) to note that given all the M&S developed and maintained by the DoD and used in support of the many programs in existence during any given fiscal year, documentation of M&S accreditations to justify their suitability for such purposes is virtually non-existent. Preservation of such documentation would probably do more to reduce the cost of VV&A than any other single activity that could be adopted by the DoD M&S community.

Compilation of this information by the IVA would require polling the user community via questionnaires and (usually) follow-up contacts by telephone or in person. The MD can usually identify important users and point to specific cases where usage and/or accreditation efforts might be of interest during the course of a one-hour meeting. Further, specific cases where the model performed well will be fresh in the minds of the MD (as will the failures, but these may be more difficult to wrest from their lips), whereas the IVA would have to research and acquire this information from other parties over a longer period of time.

SUMMARY OF ASSUMPTIONS, LIMITATIONS, AND ERRORS (SALE)

Another information element important in supporting accreditation decisions, the SALE can be used to identify the intended scope of M&S applications as well as limiting or simplifying assumptions that may preclude their use for prospective applications. Likewise, errors being reviewed or in the process of being corrected may also have direct impact upon specific application areas and thus may impact accreditations (adversely or positively, depending upon their status and severity).

Deduction of this information by the IVA would require careful and detailed examination of available documentation and source code, while the MD can usually describe basic assumptions and limitations in a few minutes. Documentation of known errors may require more effort if configuration management is lacking, but if software engineering principles are being applied to maintenance and development, accurate listings and descriptions of errors found during testing and reported by users will be readily available in files kept by the MD. This documentation of errors, their severity, and frequency can also contribute to another information element that is usually compiled by an IVA, the Software Quality Assessment.

SOFTWARE QUALITY ASSESSMENT (SQA)

This evaluation should be performed by an IVA, but the MD can contribute valuable data for some quality metrics. Defect density, or the number of errors found per line of executable code, is a measure that perhaps only the MD or Configuration Management (CM) agent, if one exists, can provide. Major software developers, most notably IBM Corporation, have analyzed and reported reductions in defect density that can be attributed to software engineering practices and have established goals for future development efforts. While defect densities of M&S examined by SMART have been considerably higher than these goals, tracking their trends over time provides an indication of software maturity and the effectiveness of CM and development discipline applied to maintenance over the software life cycle.

There are other quality factors, like complexity and modularity metrics for which the MD may also provide data, but most aspects of the SQA are aimed at determining the risk of verification and/or modification should they be necessary in order to accomplish accreditation. Subjective factors like readability, maintainability, and usability should be evaluated by IVA teams that have experience with the languages employed as well as with the functions being modeled, but the MD should not be expected to perform such tasks for the same reasons identified above.

FUNCTIONAL DECOMPOSITION

If more than a top-level look at a particular model is required or planned, the MD should be tasked to decompose the model into specific functions so that V&V can be performed in cost-effective increments instead of via the old beginning-to-end approach. Such a decomposition into functional elements (FEs) serves to break up complex M&S into manageable pieces and further identifies their interrelationships, importance, and hierarchy within the context of the system(s) or component(s) being simulated by the model [4].

Even if an IVA were familiar with the model of interest, the intimate and proprietary knowledge of the MD as to recent developments and approaches to FE implementations makes the IVA a poor choice for this effort unless the MD is not available. FEs are usually organized using a functional area template (FAT) that illustrates the hierarchy of functions simulated or performed by the model. These functions may also represent software objects, or entities, that interact within the simulation or with other objects in other simulations via data structures or communications services.

In providing documentation, the MD should describe the purpose of each FE, its relationship to other FEs (and the model as a whole), software modules (or routines) that comprise the FE, and data structures associated with the FE. The FAT and these FE descriptions (FEDs) can serve to guide the planning and execution of V&V tasking as well as the software changes that will be required as a result of those activities. If the model or simulation was designed using a structured or object oriented approach, these descriptions may be easily found in existing user or analyst manual documents; however, experience has shown this to be unlikely even for models touted as being developed using modern software engineering methods. Nevertheless, a FAT supported by well written FEDs is important to any V&V effort and should be produced for any M&S targeted for V&V; this should be done by the MD. Review of these FAT and FEDs by an IVA for completeness and appropriate levels of detail to support planned V&V activities may be a worthwhile task as well.

CONCEPTUAL MODEL SPECIFICATION (CMS)

As defined by Sargent [3], Davis [5], and the Society for Computer Simulation [6] the CMS is, essentially, software design documentation, and it should be compiled for a top-level perspective of the model as well as for each FE. The top-level design document has been described in various development standards (e.g., MIL STD 498) and tailored versions can be suitable to support verification that will be performed at the FE level. Detailed design documentation for each FE should be comprised of FED portions applicable to design requirements and design elements, which should be clearly stated and defined. As with all design document descriptions, CMS sections should also include flowcharts and descriptions of data elements that are input to and output from the software modules that comprise the particular FE. All of this data and information will be used in performing verification of the function, which should be accomplished by the IVA and reviewed for comment by the MD.

The CMS should be produced by the MD before any code is written, but this practice is rarely followed in the DoD, where demonstration of capability remains the criterion upon which funding of software development is based. Consequently, the MD usually produces a semi-working prototype which, if successful, is later augmented with additional code to flesh out required capabilities. Specification of the conceptual model in documentation as described by Kneppell [2] is usually not accomplished for a variety of reasons, but mostly because those funding the development see no need for it until verification of implementation is questioned. Even then a variety of testing activities may be performed in lieu of actual verification (e.g., integration testing, final acceptance testing, etc.) in an attempt to bypass what has been viewed as a too expensive proposition for programs that need models instead of more paperwork.

When verification must be performed, however, CMS documents must be produced. Even though SMART had considerable success in using IVAs to produce those documents, MDs were able to do it cheaper. Besides, they should be doing it anyway to document the software they're developing.

The process of producing CMS documents with IVAs is usually more iterative and costly than using the MD. When SMART tasked IVAs to reverse engineer design requirements and design elements for existing M&S, their first request was for the respective MDs to provide documented sources of algorithms, data values, functions, and methods employed in each FE along with descriptions of FE relationships, data inputs and outputs, flowcharts, and annotated source code listings. These input documents from the MD came to be known as Verification Source Reports (VSRs), which were reformatted and regurgitated by the IVA along with design requirements and design elements that could be verified at a later time. The resulting Post-Development Design Documents were reviewed by the MD, edited, finalized, and eventually included in the Volume II ASP in the form of CMS sections for each FE.

Later in the project, when the MDs came to understand the intent and significance of design elements, they were tasked to produce them by adding software design requirements to the VSR, which then came to be known as VSDs (for Verification Support Documents) to distinguish them from the earlier versions without design elements. In short, and to clarify a perhaps confusing history, a VSD contained everything an IVA needed to verify an FE, including annotated source code, and a CMS contained everything in a VSD except annotated source code.

If some form of these documents already exists or if they are produced by the MD for a specific verification effort, they should be reviewed by an IVA to make sure enough information is provided to fully support the desk checking and software testing that they will be required to perform. But it's cheaper and faster for the model developer to provide this documentation, with an IVA as reviewer.

SENSITIVITY ANALYSIS

While any model user with proficiency can perform this function, decisions about what relationships within an FE should be explored (as well as how they should be tested) can be expedited by consultation with the MD. Naive users (and IVAs) can spend lots of time and resources making model runs that are not very illuminating, or that just show expected relationships. While the latter may be used to support face validation efforts or review by subject matter experts, the MD should at least be used to focus and define the scope of sensitivity analysis performed by others.

The MD should also be called upon to review the results and conclusions. Results of these efforts (performed by the MD or others) can be subjected to review and critique by IVAs, but if the parties involved are DoD contractors (or Laboratories), a mediator may be required to settle the disputes and differences of opinion that will undoubtedly arise. Sensitivity analysis has a way of bringing out the best and worst of model performance by shedding light on areas that may never have been explored before. Surprises, incidentally, are rarely welcomed by the MD unless results are positive, which is why developers are not often pleasantly surprised.

WHEN AND HOW TO USE THE MD AND THE IVA

The decision to use or not to use the MD when performing V&V should be based upon the phase of the investigation and the outside expertise available with respect to the subject M&S and related phenomena. This is not a recommendation for exclusion of IVAs, which are critical to the credibility of final V&V results.

It is not always necessary, however, for IVAs to be external to the MD organization. Davis [5] has described ways to establish a 'VV&A regime within an organization,' and experience gained through execution of the SMART Project has proven the feasibility and desirability of this conceptual approach. IVAs that are an integral element of the MD team have several advantages over those that are truly independent consultants. They benefit from better communication and rapport with the code designers and developers. Working in concert with the MD, the IVA can assist with the design and development of code that requires less testing and that can be V&V'd before its actual release to users. While this has yet to be seen in practice, barriers to integrating V&V with development are disappearing due to the increasing demand for quality from the user community. However, IVAs within MD teams must maintain their integrity, which can be difficult at times and can result in tasking management to settle differences of opinion.

External IVAs, or consultants, have been less successful due to their perception by MDs as IG (Inspector General) agents sent from above to find out what's wrong with their model. These IVAs can also suffer from the notion that their ability to perform for their customer is measured by the number of errors or problems they report, which has contributed to the documentation of more V&V failures than success stories. We have also noted, when working with the consulting agency-type IVA, that ideas about what kinds of activities constitute V&V vary surprisingly. Even when definitions of V&V are initially agreed upon, interpretations have resulted in verification reports that say little about whether a given implementation meets requirements, but rather question the whole approach and suggest better ways to implement similar (or sometimes completely different) functions and/or algorithms.

The challenge facing those tasked with performing V&V is how to accomplish as much as possible with available resources, which are currently somewhere between zero and 15% of the investment in code development. Levels of tasking required will depend upon available documentation (especially of prior V&V efforts) and the status of the current model version (especially if those previous V&V results have driven later model improvements). Further investigation of specific model functions and development of additional capabilities will require functional V&V and (perhaps) more involvement on the part of the MD. IVAs must be used to certify conclusions and support recommendations and should be used whenever their appearance would be advantageous to VV&A endgame goals.

IVAs within MD teams should be exploited because they have squeezed their way into roughly ten percent of the software development action and are earning a living simply by establishing and maintaining credibility and quality. Their participation in independent V&V efforts should serve to justify their existence as well as provide the best available information about the M&S in question. Teams such as these are rare, but increasing in number, and deemed essential by simulation professionals who see integration of V&V with development as the only answer to a growing problem area, namely the proliferation of poor quality software.

SUMMARY

It is the experience of the SMART team that model developers should be an integral part of any V&V effort, independent or otherwise. Further, justification for inclusion of the MD can be based solely upon the value of their expertise and their experience with the software under examination. How they are used (tasked) by the V&V project manager can be critical to the credibility of ultimate results as well as the costs required to achieve them. If the development team contains or has employed an IVA group or committee, it will be a valuable source of information for the effort and should be used to the maximum extent possible. If external IVAs are used, they should not be tasked with analysis of the software unless the MD is unavailable or inaccessible. The most important contribution they can make will be review of findings, improvement of conclusions, and approval of the final results. They may also be used to present results to audiences that would be sensitive to contributions (or marketing pitches) by the MD.

The ultimate result of wise use of the MD and IVAs will be the most cost effective solution of the V&V problem at hand. The following rules apply:

1. The model developer usually has intimate knowledge (i.e., intelligence) that the independent V&V agent would otherwise have to acquire.

2. The independent V&V agent has a perspective and experience with M&S evaluations that the model developer may lack.

3. A mediator can be critical to getting bottom line V&V results and documentation products out the door.

By adhering to a consistent VV&A process built upon appropriate analysis and documentation, the credibility of a model or simulation can be established and maintained over time. How this is done and who performs critical functions will determine the level of credibility ultimately achieved as well as the cost effectiveness of the process. If an independent V&V agent or team within the model development organization can be used successfully, the cost of credibility will not only be initally lower, but should decline to acceptable levels over time.

References

1. Gupta, Uma, Validating and Verifying Knowledge-Based Systems, IEEE Computer Society Press, Los Alamitos, CA, 1991.

2. Knepell, P. L. and Arangno, D. C., Simulation Validation: A Confidence Assessment Methodology, IEEE Computer Society Press, Los Angeles, 1993.

3. Sargent, R.G., *Validation of Simulation Models, Proceedings*, IEEE 1979 Winter Simulation Conference Proceedings, pp. 497-503.

4. Kheir, N.A. and Holmes, W.M., *On validating simulation models of missile systems*, Simulation, March 1982, pp. 117-128.

5. Davis, P. K., *Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations*, RAND National Defense Research Institute, 1993, (R-4249-ACQ).

6. Schlesinger, S., et al, *Terminology for Model Credibility*, Simulation, March 1979, pp. 103-104.

3.4 VALIDATION

While verification is what everybody seems to get when conducting V&V, validation is what everybody wants. It's fear of validation cost that drives a program either to not do any of it (and maybe hope nobody notices), or to do what's cheap and easy (even if it's not what they need to do). The only alternative seems to be to spend a kabillion dollars and usually not end up any better than you started. But what's really needed is a way to figure out what you really need to do, and share the cost with others who need it, too. The articles below describe how to do that.

DATA SPECIFICATION AND COLLECTION FOR MODEL VALIDATION

by Chester Richardson III, Karl Simecka, Willie Stewart, and Stewart Turner

Validation is expensive, no question about it! The biggest expense, of course, is collecting suitable test data to compare with model results. There are, however, some ways to make that data collection more cost-effective. This can best be done by carefully specifying what data are required before test planning begins, and by sharing the cost of test data collection with others who need the same (or similar) data. This article talks about how to do both of those. (ed.)

A big part of developing and demonstrating a robust V&V methodology was the collection of test data suitable for conducting a "credible" validation assessment. It became obvious very early in the data collection process that this task was much more complicated than we initially imagined. The data collection process is long and cumbersome, entangled with bureaucratic rules, and often frustrating. Sometimes the process ends up with no data or incomplete data because of changes in test program goals, funding limitations, political turf battles, hurricanes, or priority changes due to war. Nevertheless, high quality data can be and has been collected from all of the services in an effective manner to support model validation.

But we had to overcome a considerable number of obstacles in the process. There were a number of thorny data collection issues which needed to be addressed before the validation effort could be conducted with any confidence. These issues are collected into four main categories in the following paragraphs. Many of these lessons learned have to do with conducting test data collection efforts in conjunction with another test agent. There is a significant cost savings in "piggybacking" on an ongoing test activity, but the savings can be offset if you don't have a well established and documented working relationship with the test range and the primary test sponsor.

PRETEST ACTIVITY

There are a number of activities that have to be conducted before testing begins, if model validation requirements are going to be met. These are probably the most important elements of effective data collection: unless everyone agrees up front what is required, the data collected will not meet your needs.

Establishing Memoranda of Understanding/Agreement

The first obstacle we encountered was ready access to test data. Test ranges did not have release authority for the data that had been collected on the range, since the data were owned by the organization actually conducting the test. Many of the organizations that possessed test data were at first reluctant to release the data for fear of their data being used inappropriately.

The solution to this problem was to establish a Memorandum of Understanding or Agreement (MOU/MOA) with the test organization which outlined the responsibilities and restrictions on the use of data provided to SMART by the test agent. When done properly, the MOU ensured the full support of the government agency's management in the collection of data for validation purposes. This support, in turn, made it possible for SMART Program representatives to move freely in and around the test facilities and all of the associated locations such as project offices, photo lab, math lab, etc. The MOU, in effect, gave the SMART Program representative a license to be treated as "one of the team" and not as an "outsider".

A well-written MOU states the advantages to the government agency as well as to the M&S validation effort. By distributing copies of the MOU to key agency personnel and explaining the advantages for the government agency, it is much easier to obtain cooperation from the agency and its personnel. This cooperation is mandatory if you are to collect any useful data.

The MOU's provided access to test data, and to collection opportunities that were previously unavailable and generally unknown. They also opened the door to acquiring data specifically useful for model validation that ordinarily would not have been collected by the test organization, by allowing us to pay for additional instrumentation that was not in the original test plan.

Data Specification

We searched through test data archives and databases looking for data to support M&S validation, to see if we could get by with previous test results. In general, the test data available in those archives tended to be outdated, insufficiently documented, lacked the fidelity required to perform detailed assessments of M&S components, or all of the above. It was clear that we were not, as a general rule, going to obtain the data we needed from previous testing. The implication of this result was that the data obtainable from (or at least normally documented by) the normal range data instrumentation packages were not going to meet many of the needs of model validation.

This issue was addressed by developing a Data Specification for collecting data based on the requirements identified through the examination of the M&S we planned to assess. This data specification set forth fidelity requirements, accuracy requirements, and sampling rates that would best suit an assessment of the M&S components. This document, the Data Requirements Dictionary (DRD), was distributed to various ranges to aid in the collection activity.

Unfortunately, the DRD did not work out, because the document was based on M&S requirements and was not translated into test requirements, nor did it address test range capabilities and limitations. The test range personnel were used to looking at test plans, and they could not relate to a document that identified M&S requirements unless it was related to testing. In fact, they found many of the "requirements" listed as rather humerous, and there is still a painted "SMART Rock" at one test facility which represented their interpretation of the clutter data requirements listed in the DRD.

The requirements document was subsequently modified to correlate these M&S requirements to a test requirement, and to ensure that requested data elements were compatible with test range nomenclature and limitations. This document was demonstrated to be very useful in identifying and recommending instrumentation requirements for performing detailed Science and Technology Investigations (S&TI) of acquired weapons systems.

Data collection for model validation purposes was clearly a benefit to both parties, especially the test sponsor. As an example, the addition of telemetry packages to a series of missile firings was determined to be advantageous to provide in-flight data for model validation. Initially, the project personnel were not interested in either the added expense or the schedule delay. When the advantages to the service were pointed out and an offer was made to pay for two of the telemetry packages, the service agreed to delay the firings and even pay for an additional four telemetry packages.

Test Planning

In order to ensure that the identified data elements were being collected, a written test plan or test plan supplement needed to be generated. Often times any data scheduled to be collected over and above the test organization's data requirements came after their planning process, so it was difficult to append to the existing test plan. In some cases, the test plan was unavailable for security or other reasons, so it was necessary to provide a test plan for the data collection which would support SMART's objectives.

Providing this test plan also ensured that any equipments, additional instrumentation, and personnel required to meet SMART's objectives would be present at the test event. The test plan was subject to approval by the organization performing the test, and the test range where the test was to take place. This allowed these organizations the advantage of incorporating the requirements outined in the SMART test plan into their own test plan if it was deemed necessary. The review and approval also provided a guarantee to the test organizations that this activity was being performed on a noninterefering basis on the part of SMART.

CONCURRENT WITH TEST

Coordination With Test Personnel

The most important aspect of SMART's involvement with any particular test was the working relationship that we had with the local engineers and data gatherers. The more involved SMART's test engineers were with the various engineers (test, systems, instrumentation, maintenence, and design) on a test event, the better the quality of data that was being taken. If immediate changes were required in order to collect data more efficiently or accurately, with a good working relationship the changes were more apt to be performed before the test event rather than after, thus saving time and resources. It also provided an avenue to better understand the perspectives of the personnel involved in the test.

Cultivating a trusting, open, and friendly relationship is extremely important. An adversarial relationship could prevent you from obtaining the kind of quality data required for model validation. Test project personnel don't react well to outsiders insisting they change the way they conduct their tests. A good relationship requires frequent discussions of the planned tests and how to most efficiently obtain the data for both parties. This cannot be done in a single session but can be done most effectively as a result of many short discussions over time to make the project personnel feel comfortable with the working arrangement.

The importance of this coordination is underscored by the particular type of data being collected. On several occasions the standard data package that was provided by a was not usable for detailed model assessment. However, when a test engineer representing SMART was at the test, data other than the standard product could be collected and distributed which directly supported detailed validation assessments.

POST TEST ACTIVITY

Documenting the Test Event

After a test was completed it was essential to document all aspects of the test event. This included any smoothing, or processing of the collected data. This documentation effort made the difference between being able to use the data for validation purposes, and either having to limit the scope of the validation effort as a result of not understanding the test, or not having sufficient detail to characterize the test event. Usually, the test does not go exactly as planned (which in many cases is quite an understatement). Because of that, it is important to record every critical element of what actually happened.

One of the most important aspects of data collection is the maintenance of records from all aspects of the test as it progresses from early planning, through set-up, conduct of the test, and finally data reduction. Often, the official service project records are not complete or they are unavailable after the test. The only way to ensure that a complete data package can be assembled after the test is to keep your own records. The type of records required include, but are not limited to:

A. The objective of each test and special requirements which may make one test differ from the next.

B. A description of each test item.

C. An instrumentation channel list for each type of data recorded; this could be digital, analog, telemetry, aircraft data, missile data, photographic, weather balloon data, etc. Include all associated calibrations, units of each variable, and arrangement of variables on the recording medium.

D. A record of all special equipment such as telemetry packages bought or built with traceable documentation showing sources, tests, calibrations, and, problems with each item of equipment.

E. All site survey data showing relative and absolute position of all test assets used in the tests.

F. A record of the sequence of the tests and all serial numbers or identifiers of each test asset of interest and the mission number in which the asset is used.

G. A record of all weights, dimensions, and characteristics of the test article and related assets (such as temperature) which may be of interest to the model validation process.

H. A record of all test conditions such as weather, launch angles, malfunctions, etc., for both target and test asset data.

I. Copies of quick-look data charts, photographic data, mission log sheets, engineers and controller's logs, and all electronic data products.

J. A record of any special test procedures used which would be helpful for model validation using the data.

Data Reduction Activities

The most important aspect of the post test activity is the data reduction effort. Virtually all data that come from a range require some processing or reduction. Often it was jointly beneficial for SMART to offer assistance in preparation of the data reduction plan and in the actual data reduction process. In most cases, the service personnel do not need nor want all of the data reduced that may be required to meet model validation requirements. The data reduction process is technically demanding and it goes much better if someone is available to provide technical guidance to the data reduction technicians. Otherwise, sample rates, scale factors, plot ranges, variable names, formats, etc., will not be accomplished correctly. Experience has shown that the data can assume random output formats if the data reduction process is not monitored closely. Consequently, an engineer that represented SMART was involved in the data reduction. His main function was to document the reduction process and to ensure that the data were being processed to reflect the requirements of the validation agent. If that engineer was not present then he would have to process the data himself to ensure the usability of the data. A secondary function was to provide a characterization of the data for archiving purposes.

SMART often provided funding for the portion of the data reduction process that was added due to model validation. We established a fund or account that was used by the data reduction personnel so that the agency project officer would know how much was contributed by the model validation effort. This demonstrated our willingness to support the project and work cooperatively as a team member.

Test Report

The test report, or test report supplement, compiles all of the work performed, from the initial coordination of the test plan to the final data reduction effort. In general, without our contribution the test reports would not have supported use of the collected data for model validation. The report needs to include the details of the test, how the data were processed, characterizations of the systems employed, notes from the various operators, and descriptions of how the data were collected. Discrepencies in the data would be identified, as well as recommendations and cautions concerning the use of the data. This report ties the data collection effort to the assessment by providing the particulars of a test event that are required to use the data properly in an assessment of a model.

INTERNATIONAL DATA COLLECTION

During the SMART project, we collected data from testing in other countries. The collection of data in other countries is quite different than in the US service test agencies. A different philosophy prevails and the total time required is much longer than your wildest expectations. The major items to consider when collecting data on a foreign test program are:

A. Be prepared to provide your own support if special equipment is needed for recording data. In general, they do not have the level of sophistication in test data collection that we have come to expect in the US. In support of data collection on the SMART Program, the US team had to provide key required instrumentation elements. In addition, the US team had to transport to the site all of the special support equipment and technicians required to calibrate, record, and download the data from the aircraft.

B. Recognize that politics play an important role in the collection and transfer of data. The foreign government likely will insist on maximum financial payment and/or access rights to data or software programs of interest to their national effort. Nothing will occur until the MOU is in place and the effort is fully covered by a Data Exchange Agreement (DEA) between the nations.

C. Expect very long periods of delay for data reduction, analysis, and transmittal. Everything may be inordinately slow and discussions and meetings can have very little effect on progress because the other participants have their own agendas. It took four years to receive some of the data.

SUMMARY

The importance of a clear and concise set of data specifications cannot be emphasized enough; those specifications must be in the form of test plans, so that test range personnel can readily relate to them. However, data specification is not the most critical element in ensuring that the data being collected will be usable for validation. It is the working relationship between the test engineers and the model validation agent. That interface between the modeling community and the test community provides the necessary point of contact to ensure quality data from the perspective of the modelling community. Without that good working relationship, data being collected will be of limited use.

STATISTICAL VS ANALYTICAL SIGNIFICANCE: HOW MUCH V&V IS ENOUGH?

by David H. Hall

A big problem for acquisition programs (or anyone else, for that matter) who need to perform V&V , is to know when to stop doing it! How do I know when I've done enough to satisfy whoever I've got to satisfy that my model is good enough? The article below addresses that issue from the standpoint of aiming your validation activities at showing that the model does what you need it to do. Spend your validation dollars on your own specific requirements, and spend "not a penny more, not a penny less." V&V for its own sake is not worth doing. - editor

Validation is the element which generally receives the most interest and scrutiny in the VV&A process, but it is also the most costly and elusive due to the difficulty and expense of obtaining adequate test data. As a result, in many cases subject matter expert reviews, or qualitative assessments of the model's "acceptance in the community" have been the sole basis for accreditation decisions. Many acquisition programs suffer from "validation avoidance." The root of the problem is that validation is not always understood to be a process, rather than an absolute end in itself. The question: "Is your model validated?" leads inevitably either to never-ending validation, since there is no definition of what's good enough correlation with test data to decide you can quit doing it, or it ends with the user deciding that validation is too hard (or too expensive) so he avoids it altogether. Validation really only has meaning in the context of an application: the user wants to know if the model is demonstrated to be good enough for his purpose. This means that the user has to analyze his application, determine what "good enough" means for him, and only do enough validation to determine if the model meets his needs. In other words, the user must search for "analytical significance" in validation results.

STATISTICAL SIGNIFICANCE

A typical approach to detailed validation involves decomposing a model into functional elements. This allows for (a) identification of "testable" elements of the code, and (b) examination of how each of those elements contributes to the overall credibility of the model for "end-to-end" model comparison with test data. That is, if we compare the model's final output parameters with test data, and the results do not correlate well, we must have functional level results to sort out which part of the model is causing the problem (assuming that the problem is not in the test data). And, if the results do correlate well, we want to be sure that it was not due to a serendipitous cancellation of compensating errors within the model. (Which in fact occurred on occasion during model level testing during SMART).

The breakdown of the model into "testable" functional elements supports the definition of validation data requirements; that is, in order to validate a model, we must translate model validation needs into test plans and test plan supplements to obtain the required data. Suitable data may be obtained from ongoing testing, but it requires pre-test coordination for data instrumentation requirements that may be unique to model validation issues. Data obtained from testing in this manner is at a low enough level in the model to support statistical comparisons between model predictions and test results. Often this is not the case when considering the model as a whole, since many model outputs are not directly measurable at a test facility, or may only be inferred from the data that are collected.

Figure 1 illustrates sample results from a test, as a function of time, when compared with model predictions under the same conditions as the test. This represents the comparison of a single model functional element, such as radar tracking performance, with actual test data. Because of the relatively small variation in the test data compared to the difference between test and model prediction, the difference clearly is statistically significant. Therefore, from a statistical standpoint, we should reject the hypothesis that the test data and the model predictions represent the same distributions. But what does that mean? That is, does the statistically significant difference between the test data and the model predictions really make any difference to the user? Are the test and model result differences "Analytically Significant?"

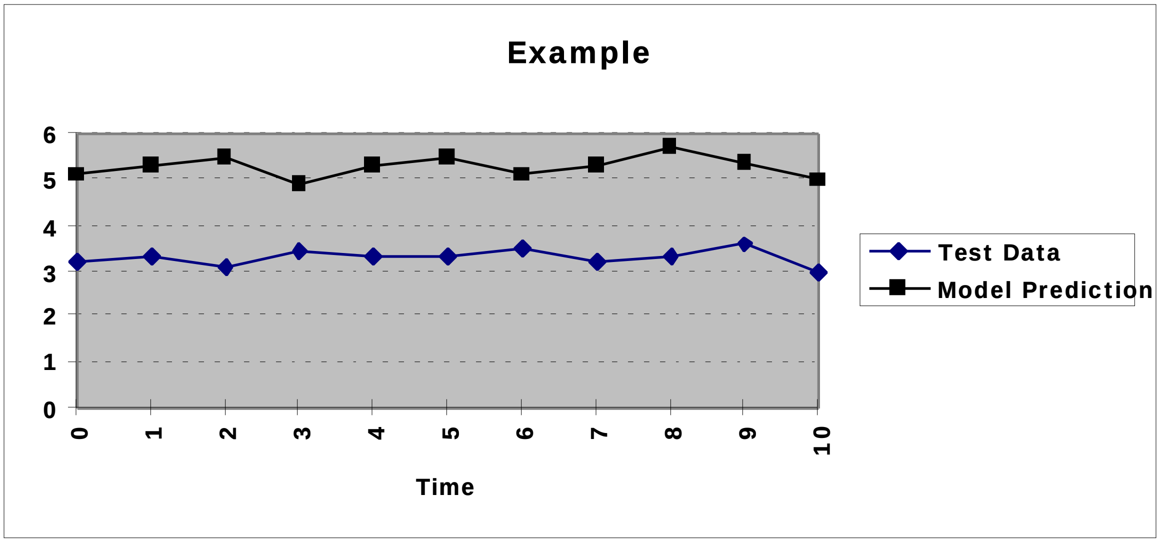


Figure 1. Sample Model Functional Element Correlation with Test Data

ANALYTICAL SIGNIFICANCE

The answer to that question depends on how the model predictions are to be used in the application of the model, and what questions the user is trying to answer. What are the user's "critical analytical issues (CAIs)"? This is really the meaning of analytical significance: is the difference between test data and model prediction significant in my context? Would that difference change my final answer?

In many cases, model or test results are used to feed higher level models which address user critical analytical issues; in other cases "testable" model functional elements are directly embedded into higher level models whose results address CAI's. Figures 2 and 3 represent the results of processing the functional element data represented in figure 1 through two different higher level models. In both cases, the model functional element predictions were used for one curve, and the actual test data were fed into the model for the other curve. Also shown on figures 2 and 3 are acceptance criteria for the applications using the models. In figure 2, for the model addressing CAI #1, the difference between the model functional element results and the test data results are "analytically significant," because the differences in model output using those data fall outside the acceptance criteria boundaries. In figure 3, however, use of the same data through another model yields results that are not significantly different, even though their differences were statistically significant. In that case, we would say that this functional element of the model is sufficiently accurate to support the requirements of CAI #2, or that the model supporting CAI #2 is not very sensitive to the functional element we tested.

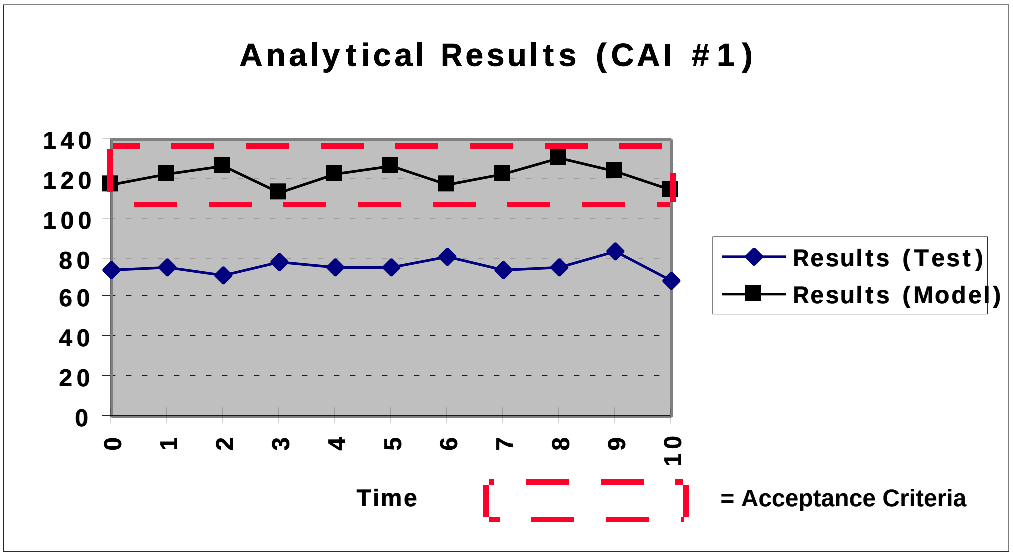


Figure 2. Model Level Results for CAI #1

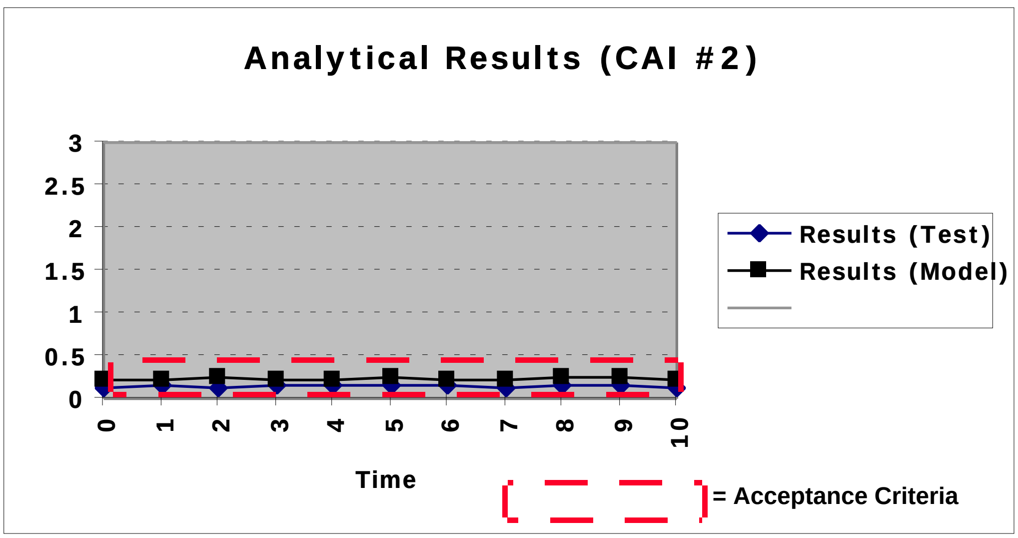


Figure 3. Model Level Results for CAI #2

Another more concrete example of this idea is shown in figure 4. That figure illustrates the miss-distance distribution of a surface-to-air missile against an aircraft target, as measured by two different measurement systems, for two different test series. The set of data labeled "MTS" clearly exhibits a Rayleigh distribution, with a fairly well defined "cutoff" limit at the extreme. The data labeled "GPS", on the other hand, presents a Poisson distribution, with outlying miss distances well beyond the cutoff of the MTS data. A statistical analysis of these data concludes that they do not come from the same distribution, even though they purport to represent the same missile system against the same target. (This may be due to a number of factors, including differences in the way that the two trials were set up, as well as differences in the measurement systems themselves).

But what does that mean? The fact that these two miss distance distributions are statistically different does not, by itself, mean anything other than that they are different. The question that needs to be answered is "does this difference mean anything to my use of the data?" The data in question were intended to be used to support estimates of probability of kill (PK) of the aircraft target by the missile system. Using the distribution of miss generated by the "MTS" system resulted in a computed PK of 0.74, with a confidence interval of plus or minus 0.02; the PK computed using the "GPS" system was 0.77, with a confidence interval of plus or minus 0.03. Considering that the program using these data would have been happy with PK estimates accurate to one decimal place, the difference between these PK's is insignificant. Thus, even though the two distributions are statistically different, for this particular application they are equivalent in an analytical significance sense. This result may seem surprising, but it illustrates the principal difference between statistical and analytical significance - a statistical difference may not make a difference to your use of the results of the model.

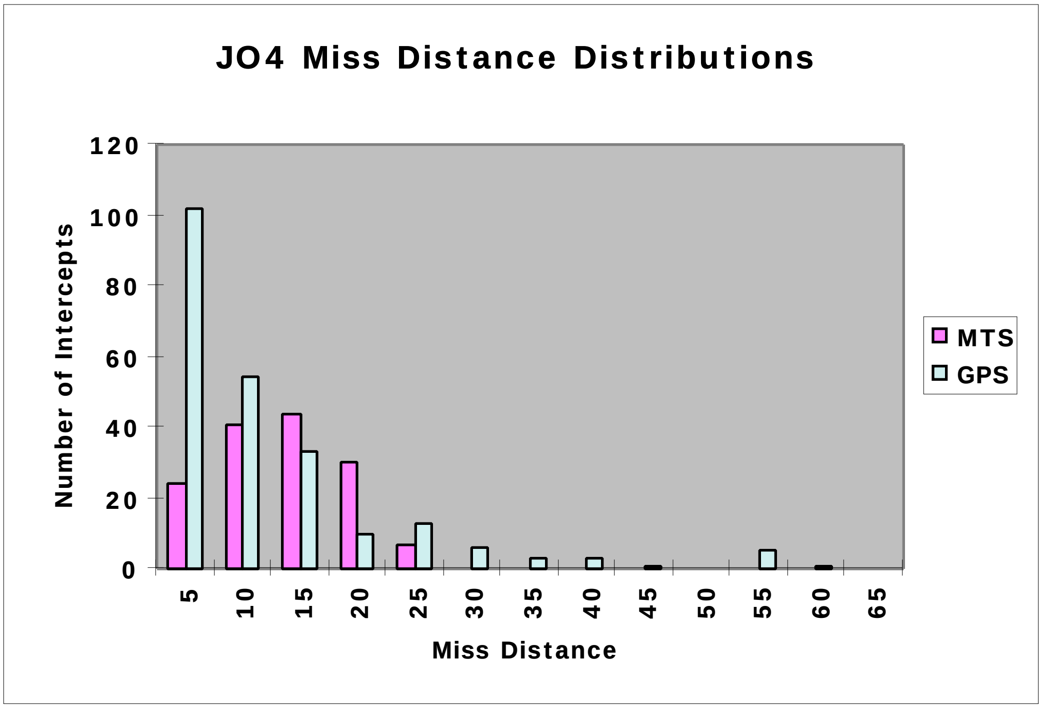


Figure 4. Miss Distance Distributions Resulting from Two Data Sources

But how do we determine the acceptance criteria boundaries shown in figures 2 and 3? The acceptance criterion in either case must come from fidelity requirements that are determined by an analysis that defines the "range of analytical significance" of the application. The key to making any V&V effort cost-effective is to tie the requirements for V&V activities to accreditation requirements. And those accreditation requirements come from the analysis of the application: What are the key study questions to be answered? What is the hierarchy of measures of effectiveness (MOE) that address those questions? How do those MOE's relate to model outputs? How do changes to those model outputs, and consequently to those MOE's, change the answers to my study questions? From the answers to those questions, particularly the latter, come M&S acceptance criteria for the particular application.

Once those criteria are established, they are compared with the model's capabilities to determine its suitability for the application at hand. The results of these comparisons are threefold:

(a) Data To Support M&S Accreditation (how the model fared compared to the acceptance criteria);

(b) A Risk Assessment (what areas of the model pose risks to the study questions?); and

(c) Requirements for Model Enhancements (what areas of the model should be "fixed" because of their high risk to the program?).

These are the results of comparing the model with test data for “analytical significance”: it ties the entire process to the real requirements for the model, in the context of an actual application. Tests of statistical significance are really suited more to an assessment of the behavior of the test data, rather than their significance to the problem. We can use statistical techniques to evaluate the suitability of the data for use in comparison with model predictions, but statistics themselves will never answer questions about whether the model is suitable for a particular application.

SUMMARY

The credibility of M&S used in the system acquisition process is crucial to informed use of those models in support of system design, T&E and training. The key to affordable validation activities is to focus efforts on the part of the model that’s important to your problem. Analytical significance is the key criterion for M&S comparison with test data, where the comparison is based on acceptance criteria that are determined by analyzing your M&S application. By taking this approach we can go beyond tests of statistical significance to tests of whether the model is good enough for what the user needs it to do. Validation for its own sake is never cost-effective, but validation conducted to support a specific problem focuses activities on key analytical issues and makes the process efficient and cost-effective. So ultimately, the answer to the question, “How much validation do I need to determine if my model is good enough?” is, “It depends on what you want the model to do!”

3.5 ACCREDITATION

Accreditation is a decision by somebody with some authority that a model is good enough for a particular application. Usually that somebody is the ultimate user of the model results. Accreditation is the reason for doing all of this work in the first place, and all of the V&V results should be focused on the goal of supporting an accreditation decision. It is the "summum bonum" of V&V (if we can be allowed an occasional Latin expression). The SMART project supported a number of real, live customers by showing them how to make cost-effective use of V&V data to support M&S accreditation; the articles below describe some of the lessons learned in the process.

THE SMART ROAD TO ACCREDITATION: LESSONS FROM THE FRONT

by Michelle Kilikauskas

For the last two years of SMART, project personnel provided model and simulation accreditation support to DoD acquisition programs (these activities are being continued under JASA). This article summarizes some of the keys to cost effective, meaningful accreditation activities that are derived from this experience working with missile, aircraft, and electronic warfare systems programs in various stages of the systems acquisition cycle. (ed.)

As the pressure to use modeling and simulation has increased, so also has the burden on program managers to defend the credibility of analyses based upon the results of those models. The staff of the SMART project and now the Joint Accreditation Support Activity have worked closely with systems acquisition programs endeavoring to comply with the DoD 5000 directives. This article summarizes the major lessons learned through this experience in designing and implementing cost effective VV&A programs to support M&S accreditation.

BASIC PRINCIPLES

First, keep in mind that accreditation is a decision. The responsibility of the accreditation support agent is to provide the accreditation authority with sufficient information to make an intelligent and defensible decision about whether or not a model or simulation is appropriate for their needs. The amount of information necessary to defend an accreditation decision (and therefore the amount of resources required to generate that information) is dependent upon several factors, including the importance of the decisions to which the model contributes, the ramifications of a wrong decision based upon model results which are not credible, and the dollar amount of the program. Because accreditation of M&S is but one of myriad concerns, program managers are motivated to spend only as much as is necessary to establish an acceptable level of confidence in model results. The question is, " How confident can I afford to be?" The accreditation support agent has a responsibility to help the decision authority to invest his limited accreditation support resources on those tasks which contribute the most to credibility for the least cost.

A corollary to the first principle is that accreditation is both a technical and a political decision. Insight into the proponents and detractors of the model and the influence they exert is crucial. As DoD increases the pressure to reduce the number of models in use, the selection of a particular model by high visibility acquisition programs will become even more politically charged. A well thought-out and well articulated rationale for selecting and accrediting a particular model over its competitors is critical for defending the decision to the Defense Acquisition Board (DAB), which approves transition from one phase of the acquisition process to another (from concept definition to engineering and manufacturing development, for example), and to Congress who ultimately controls the funding. From this point of view, accreditation support is a form of risk reduction and mitigation.

The second idea to keep in mind is that complying with directives such as the 5000 series is a secondary goal. The primary goal of accreditation is to understand the strengths and weaknesses of the model under consideration and their effects on the credibility of analysis based on model output. While a program manager's initial interest in VV&A may be driven by the need to comply with directives, the long term goal of the program is to make decisions which are well grounded and to produce a quality product. The accreditation support agent can be helpful in providing this perspective. Since VV&A must be addressed to check the box anyway, let's make it meaningful.

If insight into the credibility of M&S output is the goal of meaningful accreditation, a clear statement of the intended use of the M&S under consideration is the foundation. The better the intended use is defined, the more focused (and cost effective) the V&V and model assessment efforts can become. In the tailored VV&A process developed by SMART, the first step is to understand how the client envisions using the M&S. What analysis questions are being addressed? What measures of merit are will be used to quantify answers to those questions? What level of fidelity is required in the answers? What assumptions underlie the analysis?

Once those issues are clarified, M&S requirements can be developed based upon the intended use. The SMART process identifies three types of requirements: functional, fidelity, and operational. What functions must the model perform in order to address the questions of interest? To what level of detail must these functions be modeled, and what level of accuracy is required in model output to meet the decision maker's needs? In what kind of environment must the M&S run: computer platform and/or operating system, real time vs. non- real time, level of experience or training of model operators, compatibility with other analysis efforts, etc. The M&S requirements derived from the answers to these questions form a standard against which the model of interest is assessed for accreditation, and V&V efforts are focused on those features of the M&S and those portions of the code which are related to those criteria. The more the problem of interest can be defined, the less effort is wasted on examining features of the model which are not relevant.

LESSONS LEARNED

While clear definition of analysis goals and M&S implementation strategy is a key to cost effective accreditation, our observation has been that most acquisition programs begin with only a loose understanding of how analysis and M&S will be used. Transforming a general idea of the types of issues that need addressing and a vague notion of how M&S might play into a carefully constructed and well articulated plan is a long and painful process. The programs to which SMART/JASA have been the most help are ones in which the accreditation support personnel got involved with the analysis team early and stayed involved over time. This gives the accreditation support team the opportunity to listen carefully to the analysis team to make sure they understand the client's analysis goals (as well as the simplifying assumptions), and to keep up as the analysis plans naturally evolve. The greater the insight into the analysis effort, the more focused the accreditation assessment efforts.

Additionally, the accreditation support team's knowledge of the models and analysis areas of interest often prove useful in identifying potential risk areas early enough to allow time to mitigate or avoid the problem. Mitigation might involve making model improvements, gathering more credible input data, or using different tools in areas where the tool of choice is weak. The lesson here is to get involved early and stay involved.

Another basic lesson is that accreditation is an iterative process. As circumstances change over time, the accreditation issue must be revisited. Analysis questions and/or assumptions change from one phase of a program to another, model code and input data sets change on a periodic basis, and the importance of M&S in the resolution of analysis issues shifts from phase to phase. Careful documentation of previous accreditation assessments and disciplined configuration management of model code and input data sets on the part of both the model manager and the program office analysis staff minimizes the amount of effort required for a reassessment. Consistent involvement of the accreditation support personnel and the analysis team also minimizes the learning curve for each successive round of accreditation.

The accreditation support staff can also make a substantial contribution to the programs it supports by helping to decide whether and how to invest in model and data improvement. Because program managers are generally not modeling or analysis experts themselves, they are vulnerable to persuasive pleas for funding for model improvements which may be worthy, but are not high priority *given the analysis needs of the program*. The larger and more visible the program, the more appealing a target it is for model managers and model users interested in model improvements. The accreditation staff can be of help by consistently asking how a proposed model improvement effort is related to the analysis needs of that program in the near to mid term.

Many forces converge to cause program office personnel to have a short term focus. Program office personnel are generally military, and they rotate assignments every two to three years. They are evaluated on a yearly cycle, and they are barraged with issues which are pressing today. Accreditation support agents can play a valuable role by taking a longer term, more balanced view. Keeping one eye on the long term brings two benefits: first, it allows one to consider how to make the effort currently being expended of most use in the future, and it encourages one to look ahead to the impact of current modeling limitations on later phases of analysis, in time to implement model changes or select and develop proficiency with other tools. Limitations which may be of little importance to the questions of interest today may be “show stoppers” for questions in the next phase. The other eye needs to be firmly fixed on how the skills and knowledge base of the accreditation support personnel can be most helpful in the here and now. What of value can we do to meet today’s needs with the resources the client has available now?

Program managers generally do not have a modeling or simulation background, and accreditation is not generally on the top of their priority list. There are many benefits to having a central, independent activity like JASA which keeps current on policy statements and the way in which that policy is being implemented (both of which are in great flux today), has experience in conducting verification, validation and accreditation, has access to expertise in various modeling areas, has active contract vehicles available for performing VV&A related tasks, and maintains a repository of VV&A documentation. The availability of expert help increases the probability that programs will comply with VV&A policy and that their efforts will be meaningful.

Knowledge of what’s already been done (and documented!) reduces duplication of effort and maximizes the usefulness of the V&V work funded by individual programs or agencies to the larger analysis community. The availability of experienced personnel also reduces the cost to individual programs of designing and implementing a VV&A program. As noted earlier, the usefulness of an accreditation support agent to an acquisition program is also dependent upon the willingness to listen and learn about the particular needs of each client.

The quality and appropriateness of the model code itself is only half of the problem. Quality of model input data, its appropriateness to the model, and traceability to authoritative sources is the other half. Data quality and pedigree is a verification, validation, and certification (VV&C) issue. Regardless of the quality of the model code, bad input leads to bad output. Increasing emphasis within DoD on use of officially designated authoritative sources, in concert with the acknowledged shortfalls in some areas of interest create a great challenge for acquisition programs. Although official guidance on VV&C of data is currently less well defined than that on VV&A of model code, data quality and pedigree and the influence on the credibility of model output must be addressed as part of an overall accreditation effort.

Finally, some very practical observations. First, model usage and V&V history and a summary of assumptions, limitations, and errors are useful to a broad base of model users and they are two of the least expensive V&V activities. Usage and VV&A history can help build a case for provisional or preliminary accreditation based upon community acceptance, and the assumptions and limitations can help a potential user to quickly decide whether a model deserves consideration for a particular problem. Before SMART, these topics were generally not addressed in standard model documentation sets. Second, thick skin is a definite asset for individuals involved with accreditation support. The honest broker is often not everyone's favorite individual.

SUMMARY

The contribution of a competent and proactive accreditation support team can extend far beyond providing V&V underpinnings for meaningful accreditation statements. Experience by the SMART team suggests several keys for helping accreditation support personnel make the most of their efforts. Get involved early in the program and stay involved. Listen and learn. Keep one eye on the short term and the other on the big picture. Be proactive in contributing to other aspects of the program such as data and model improvement investment plans and configuration management efforts. Be flexible. Play the role of an honest broker even though this is often uncomfortable. Have a thick skin. Above all, help the customer become as confident in M&S results as he or she needs to be, within the confines of the time and money available.

<u>**ACCREDITATION SUPPORT PROCESS:**</u>

<u>**MORE LESSONS FROM THE FRONT**</u>

by Dennis Laack

Some techniques for providing accreditation support to customers worked better than others. And some customers were more receptive to being helped than others. The following article talks about some of the non-technical issues encountered in the process of showing up and saying, "Hi! We're from the government, and we're here to help!" (ed.)

During the last two years of SMART, the project provided accreditation support services to a number of programs that were faced with the requirement to accredit selected models for specific purposes. The nature of this support was to adapt and apply the SMART accreditation methodology to each particular program and, in most cases, to develop an accreditation plan for each program. This article summarizes some lessons learned during these efforts that deal more with organizational and personality issues.

INTERACTION LESSONS LEARNED

A number of important lessons were learned from these efforts in the realm of techniques for interaction, in addition to the area of accreditation procedures. The primary lesson relates to building trust and rapport between the support personnel (e.g. JASA) and the client organization so that the proposed concepts and recommendations would be considered on their merits.

In one instance, tasking came from an organization other than the program receiving VV&A support under these circumstances, it should be assumed that the supported organization may not recognize a need for any assistance. In fact, our support was viewed with some skepticism, and downright hostility at first. This was overcome through an earnest and open dialog, but it was a hurdle which had to be overcome before any meaningful technical discussion could take place.

This program already had a comprehensive effort underway to test the suitability of their software. One of the possible causes of the initial reaction to the SMART proposals for VV&A was confusion over the real meaning of different terms. After learning what testing was being done, the SMART personnel could easily relate their efforts to specific tasks or methods included in the SMART methodology. The lesson to be learned from this observation is that the interpretation of terms & concepts can cause confusion. Personnel engaged in providing accreditation support should learn and understand what efforts are already underway within a program and take pains to translate the accreditation support activities into terminology that is consistent with that already being used in a program.

Typically, the personnel in a client organization would have already developed an approach to meeting their objectives and would have confidence that their approach is valid. Therefore, initial dialogue must be focused on developing a common framework of objectives and approaches to achieving these objectives. This common framework should address the tasks that must be accomplished, alternative means of accomplishing them, and what support might help accomplish them most effectively and efficiently.

A related lesson to the one above dealing with co-opting terminology, is the need to use the on-going efforts as the basis for an accreditation support proposal. Program officials are much more likely to accept your recommendations for support if they clearly add to what is already being done in a logical fashion. If a totally new effort is proposed that does not take into account work already completed or in progress, there is less likelihood that the proposal will be accepted. The motto should be "Evolution vice Revolution".

Another lesson learned was that while it is important to have top level management support for the efforts of an accreditation support agent (ASA), it is even more important to have buy-in from the middle-level program management. In this particular case, we were called in to help develop a VV&A plan for the program by the program manager himself, who then assigned a deputy to oversee the effort. However, that deputy and the working level personnel under him had other ideas on how the job should be done. As a consequence, it was a considerable struggle to make any positive impact on the program's VV&A plan. It takes action from the top management to ensure that program personnel see the benefits of the support provided by an ASA, even if theirs is only an advisory role.

ADVISORY LESSONS LEARNED

In another effort, the initial concept envisioned the SMART personnel acting in an advisory role, facilitating and guiding program personnel in conducting the application analysis. After several months it became apparent that the role of the SMART representatives was not that of assistance and facilitation but one of actually performing the analysis. The lesson to be learned from this experience is that the specific responsibilities of the parties involved in a joint effort such as accreditation planning should be specifically agreed to and documented. A simple accreditation planning task agreement may be useful.

In laying the groundwork with the program to begin accreditation planning, briefings were used to outline the logical accreditation methodology and the need for a complete definition and analysis of the application. The purpose of these briefings was to motivate program personnel so they would devote sufficient effort to the definition and analysis tasks. These briefings did not achieve this objective. As a result, the SMART representatives were forced to conduct the analytical steps. The lesson to be learned is that some other means of motivation is needed or the process must be adapted to the realities of the program so the necessary tasks are accomplished.

Another problem area, not unique to this program, is that sponsors who plan on using M&S generally do not program adequate funding for M&S VV&A. It would be desirable to have a VV&A cost estimating tool that would help program personnel plan sufficient funds for VV&A. The SMART efforts have generated significant data on V&V costs. This data can aid sponsors in their budgeting efforts.

LESSONS LEARNED ABOUT ONGOING EFFORTS

In yet another program, their personnel had already undertaken an extensive V&V program for the models they were developing. Therefore, it was incumbent on the accreditation support personnel to utilize those V&V results to the greatest extent possible. Normally, under the SMART methodology, the requirements for the application would have been well defined and documented in an accreditation plan and these requirements would have served as the basis for model development. However, although the M&S requirements for the various program's applications had most probably been analyzed as a basis for their model development, the derivation of these requirements was not documented in a fashion that allowed a review of their logical foundations.

Although this approach of developing a model and performing V&V on it without formal documented justification for the M&S requirements is not the approach we recommend, it appears prevalent throughout the M&S community. This probably is due to accreditation and accreditation methodology not being well understood. Therefore, it is important to adapt the accreditation support concepts to the particular program needs vice trying to force the program to readjust their efforts to conform to the standard methodology.

In support of this program, the accreditation support approach was adapted to the program needs in three ways:

1. Normally, program personnel are expected to clearly define the program applications and to derive the acceptance criteria under the guidance of the support personnel. In this case, the accreditation support personnel used programmatic documentation to define and analyze each application and to ultimately derive the M&S requirements. Results of the analysis were checked with program personnel at each step to ensure completeness and correctness. This adaptation was implemented because of previous difficulty in getting program personnel to devote the time to these activities.

2. Typically, accreditation support personnel motivate program personnel to define and analyze each application through presentations on the accreditation concepts, pointing out the logic of the approach and the need for this type of activity. Sometimes this type of motivation is not effective. In this case, the accreditation support personnel recognized that the program personnel needed an Accreditation Plan prepared. The SMART personnel used that plan as a motivational tool by drafting the plan in stages, showing the plan outline at each stage to the program personnel, and identifying the pieces of information that were needed to make further progress on the plan. In this way the program personnel were motivated to provide the required information or the source material from which it could be derived. This motivational technique was at least marginally successful.

3. The standard approach to model validation which is part of the overall accreditation support methodology calls for model decomposition using a standard template. In this program, the models had been broken down into pieces for tracking V&V progress. The program personnel were told that this breakdown could serve as the first level for decomposition of the model into various functions. Again, the effectiveness of this adaptation was at least partially successful.

This program was the first time that accreditation requirements for several applications within one program were addressed as a unit. (The applications being addressed were: acquisition analysis, T&E, tactics development and employment support, and training). The technique employed to address multiple requirements was to develop the M&S requirements for each application separately and then to merge these requirements into an integrated list of requirements for the program, including all intended applications. This appeared to be a viable process for overall model requirements development.

SUMMARY

The primary lesson relates to building trust and rapport between the supporting organization and the program being supported. A common framework of objectives, approaches, and tasks should be defined.

Misinterpretation of terms & concepts can cause confusion. Accreditation support personnel should co-opt the program terminology for proposed accreditation support activities.

On-going V&V efforts should be used as the basis for any accreditation support proposal. “Evolution vice Revolution”.

Specific responsibilities of the parties involved in a joint effort such as accreditation planning should be mutually agreed to and documented.

Briefings on the logical basis for this accreditation support approach are not sufficient to motivate program people to perform the necessary application analysis. Some other means of motivation is needed or the process must be adapted to the realities of the program so that necessary tasks are accomplished.

Typically, program officials do not adequately budget for M&S VV&A. A cost estimating tool would be helpful. Personnel with VV&A experience should be involved in doing the financial planning for VV&A.

The accreditation support approach must be adapted to the needs of each particular program vice trying to force the program to readjust their efforts to conform to the standard methodology. Three examples of adaptation are:

1. Accreditation support personnel define and analyze an application to derive the M&S requirements instead of the program personnel.
2. The need for an Accreditation Plan was used to motivate program personnel to participate in application analysis. A draft plan was prepared with questions in certain sections to highlight the information needed from the application analysis.
3. The program's breakdown of a model for managing tasks was used as the basis for the model decomposition template.

If a model must be accredited for several different applications, all within one program, the M&S requirements for each application can be identified separately and then merged into an integrated set.

The explanation of the accreditation approach, especially the steps for defining the application and deriving the acceptance criteria are not easily understood. Briefings on this material must be simple and clear.

GETTING EXPERT RESULTS FROM EXPERT REVIEWS

by Dennis Laack

Since all of these VV&A activities can be expensive, one way that a lot of people get around some of the expense is to conduct "expert reviews". These reviews are conducted by a group of people who are expected to know something about the subject matter and/or the M&S in question. Unfortunately, all too often these reviews degenerate into a treatise on what the perfect model should be, and what should be added to your model to make it better than it is. But what you need is an assessment of whether you can use it, not for somebody to tell you that it isn't perfect. The following article contains some tips on how to set up an effective expert review, so that "better" does not become the enemy of "good enough". (ed.)

Expert reviews are commonly used in building model credibility because they provide a relatively low cost means of evaluating model capabilities. However, despite the widespread use of this technique, these exercises are often less productive than expected or desired. This article provides a number of things to consider when planning an expert review which should help ensure that the review will provide the desired results. The concepts described below were developed after observing and participating in a number of these reviews. In the discussion below, the term "model" represents either a model, a simulation, or both.

EXPERT REVIEWS

An expert review is an evaluation of a model or its outputs by a group of people who are trained and/or highly experienced in the phenomena, system, or evolution that is represented by the model. The purpose of such a review is to assess how well the model, its design, functions, or outputs compare with the experts' perceptions of the real world.

Expert reviews are not meant to evaluate how well the model conforms to software design guidelines, its computational efficiency, or any other measures of software quality. These issues are generally evaluated in design reviews that are conducted during the model development process. Thus, expert reviews are generally used in conjunction with model employment rather than as part of model development.

Expert reviews typically serve one of two purposes: they are used to evaluate either the degree of realism with which the model represents the real world (validation) or the suitability of a model for use in a particular application or problem (accreditation assessment).

Validation Reviews

Model validation through an expert review provides a comparison between the model and the experts' ideas and opinions of the real world. This comparison can be done in one of two ways. Experts can review the outputs of a model for a given scenario or set of input conditions and evaluate if these outputs are realistic. This is *face validation.* As an alternative, they can review how the various functions are implemented within the model and evaluate if these representations are realistic. This is a *design assessment.*

In either case, the quality of this comparison and, consequently, the confidence that one can place in the results of the review is directly dependent on the capabilities, knowledge, and experience of the experts that conduct the review. Thus, the value of an expert review varies with the degree of expertise and recognition that the team members possess.

When conducting a face validation, the estimates of real world outcomes against which the model outputs are compared are often only concepts in the minds of the reviewers. In some cases there may actually be real world data but it might not be suitable for detailed validation. Such data may be useful when combined with expert judgment. The outputs of a true face validation are a judgment as to whether the model outputs for conditions and scenarios similar to the intended application are realistic. If they are not realistic, the review report should include some identification of which outputs or which input conditions result in unrealistic model outputs.

The term *face validation* is mentioned in several publications and texts as a viable validation technique. It is identified as the fundamental validation method for many model users. In reality however, many users do not review model outputs when conducting an expert review. Instead, they delve into the design features and algorithms and assess whether the model design is realistic. Such reviews are actually a design assessment instead of a face validation. They involve explicit identification of assumptions and limitations associated with the algorithms used to represent the real world "thing". They are similar to design reviews conducted during development.

When conducted in support of a particular application or problem, the result of such a design review is a list of model shortcomings or inadequacies specifically related to the M&S requirements of the intended application. The review results directly support the accreditation assessment. In this type of review, as well as the face validation, the expert panel must make a judgment that the model or its outputs has adequate correlation with the real world for the intended purpose or application. Therefore, it is necessary that the expert panel fully understand the intended application so that these judgments can be made in context.

Expert reviews are one of the most common techniques used in validating M&S. There are two primary reasons for this phenomenon. The first is that an expert review provides an assessment of the total model at a relatively small cost. The costs for an expert review depend primarily on the extent of the planning that goes into the review and the number of personnel on the review team. There are other costs (e.g. conducting model runs, acquainting team members with the model, preparing the report) that are essentially the same from one model to the next. In almost all cases, expert reviews for validation purposes should be doable for less than 7 man-months of effort.

The second principal reason for using an expert review as a validation technique is that some models are unsuited for the classical, more detailed validation methods. Accurate data for validating models, such as those used for analyzing campaign strategies or battle field operations, cannot be collected easily. This is especially true if the data should be collected under controlled conditions, a requirement that normally exists in model validation problems. Thus, expert reviews are normally used for these types of models.

Accreditation Reviews

An accreditation review is an assessment of how well the particular model satisfies the requirements of the application for which it is intended. In theory, accreditation reviews differ from validation reviews in that a clear set of criteria are assumed to exist and to form the basis for judging model suitability. If explicit criteria do not exist, the judgment of suitability must be supported with logical rationale that explain why the model is adequate. Such detailed rationale are not generally included in the results of a validation assessment.

When these accreditation criteria cannot be explicitly stated before the assessment begins, it is often useful to have an expert panel conduct the accreditation assessment. Often the criteria to be used for accreditation are developed without a clear understanding of the problem and all its ramifications. In many cases these criteria are iterated as a greater understanding of the model, its capabilities, and the problems associated with its use are uncovered. Thus, an expert panel is frequently used to review all the material that supports accreditation, review the model functionality, and render a judgment as to whether the model satisfactorily fulfills all of the criteria for accreditation. Their rationale and logic in rendering such a judgment must be incorporated into the assessment results.

PITFALLS AND PROBLEMS

Although the expert review is a useful tool for evaluating models and their suitability for a particular application, there are several different phenomena that impede the review process and hinder delivery of a satisfactory product. Some of the more common pitfalls are identified in the following paragraphs.

Non-evaluative Validation Reports

One prevalent phenomenon in validation reviews is that the review does not consider the intended application. Either the application is ignored or not fully understood. As a result, the review evolves into an exercise that identifies the model's limitations in the way it represents the real world. Reports from these types of reviews present model strengths and shortcomings but do not provide an evaluation of how well model outputs match reality or whether the model's functionality is sufficiently realistic for the intended application. A listing of the model's strengths and weaknesses, although useful, is merely an input for a subsequent review that must determine if any of the weaknesses render the model unsuitable for the intended application. No model is perfect, so there are always areas where the model can be judged to be weak, but those areas may not make any difference to the problem at hand.

Straying from Objectives

A parallel phenomenon is "straying from review objectives". In many cases reviewers digress into discussions about model shortcomings instead of maintaining their focus on determining whether the model is "good enough". These digressions occur for several reasons. One, it is easier to identify and discuss model features that do not agree with reality than to discuss how such a deficiency might impact the problem. Second, the lack of emphasis on evaluating the model in terms of the intended use allows reviewers to stray into discussions about the model as compared to the system or operation it is representing. Third, the absence of a clear statement of intended use and a lack of understanding of the necessary M&S requirements for the intended application contribute to this digression into model design vice model suitability. In all these situations the review strays from the intended objectives and the final product typically does not address the mail.

Incomplete Reviews

One difficulty that is often experienced is an unfinished review. Typically, the review schedule is one of the first planning elements that is defined. Often the review schedule is set based on the time that was spent in other model reviews and not on a real estimate of the amount of study and discussion that might be expected. Consequently, when the review is actually in progress, the accomplishments do not meet the schedule.

This problem is typically addressed in one of two ways (assuming that long days and after hours discussions are already taking place). The first is to change the schedule, so that the formal report and the discussion needed to finally resolve issues is postponed; these are delayed to some time after the actual review period, after everyone has gone home. The second alternative is to accelerate the actual review of the model, its functions, or its outputs and curtail discussion of issues in the interest of finishing the task.

Either of these alternatives can lead to an incomplete product. The delayed report preparation eliminates any substantive discussion of issues that might arise in the course of finalizing and formalizing the logic of the analysis. For all practical purposes, this discussion does not take place, and the results presented may either lack depth or may reflect the opinions and views of only one or two reviewers. The accelerated review alternative also motivates review team members to pass over seemingly insignificant points that, under careful scrutiny, might actually be a model deficiency for the intended application. In either case, the results of the review are neither as complete as they need to be, nor as complete as they would have been if sufficient time had been scheduled for the review.

Indefinite Review Periods

In contrast to the problem of incomplete reviews due to a constrained schedule, another pitfall is an indefinite review schedule. This type of problem most frequently occurs when an expert team is established to review a model as it is being developed. Typically, review objectives are not clearly spelled out; regular meetings are held at which different model designs or modeling approaches are presented; and definitive results are not forthcoming unless some specific direction or tasking is given.

Because these types of new model review teams typically exist for a long period of time, membership changes over time. Also, the review goals either change or are forgotten. Consequently, a review report that presents an evaluation of how well the model meets specific requirements is often late or non-existent.

Delayed Reports

Another phenomenon that often characterizes expert reviews, both validation and accreditation reviews, is delayed reports. Reports of the results of expert reviews are frequently not prepared and distributed until well after the review is accomplished. This can be caused by inadequate time allotments in the accreditation plan. It is more likely, however, that this problem is caused either by inadequate planning for the documentation effort and/or the lack of dedicated resources to support the expert review.

In many cases the personnel planning the expert review do not identify the report preparation team; instead, they generally assume that it will be the full review team. Thus, team members do not block out time to prepare the report. Report preparation assignments are generally discussed at the end of the review and the actual writing is done partially during the review and finished on borrowed time afterwards.

Oftentimes, this lack of planning is exacerbated by the lack of any dedicated administrative support for integration of report pieces, smoothing, and editing. These activities are necessary since the report is generally a compilation of individual sections prepared by team members. This unfilled need for editing and finishing the document causes even further delays in distributing a final product.

TECHNIQUES FOR A “GOOD” EXPERT REVIEW

Although there are pitfalls that can cause an expert review to go astray, the key to running an effective review lies primarily in the planning and preparation. If proper preparations establish a sound review structure, a successful review is much more likely. The techniques for obtaining a successful review product are divided into three time-based phases: planning, execution, and follow-up.

Planning Phase

Proper planning is essential to achieving the desired expert review results. Good planning involves at least four critical steps: defining the application; selecting the team; establishing the review objectives, criteria, and grading schema; and establishing and disseminating the plans. The planning process must involve team members.

Application Definition

A comprehensive definition of the application is the most critical piece of information needed for an expert review. A clear statement of how the model is to be used and what it is intended to produce is necessary for the reviewers to make any judgments as to whether the model has adequate features and capabilities. The application definition must include a description of the overall problem to be solved and how M&S are to be used in reaching the solution. Specific questions that are to be answered with model outputs should be identified along with the criteria that will be used to make any judgments or decisions. If some particular scenarios are a part of the application, they should be identified along with any boundary conditions or limitations on the application. Typically, the application definition should be documented with an intended use statement or some other description of the application requirements

Selection of Team "Experts"

The next most critical element that contributes to a "good" expert review product is the selection of appropriate personnel as members of the expert review team. These personnel should include both subject matter specialists and analysts. Subject matter specialists should be knowledgeable of the intricacies and nuances related to the application and should have a general understanding of the overall nature of the application. The analysts should be fully familiar with the details of the application, especially the information needs and the criteria to be used to make any judgments. The analysts should also be familiar with the model that is being reviewed. As a whole, the team needs to have an in-depth understanding of both the application needs and modeling capabilities.

Team size should be limited to no more than 8-10 people to facilitate orderly discussions. This number should be sufficient for most applications and models that are being reviewed. In selecting team members, their projected availability should be considered. Team members must have sufficient time available to become familiar with the advance material and to fully participate in the review process. If the team members are to be involved in preparing the ultimate report, they must have time for that task also. Time availability is an important factor to investigate as part of the team screening process.

Agreement on Review Parameters (Criteria and Grading Schema)

The next important element of planning an expert review is development and specification of the evaluation criteria and grading scheme (if quantitative grades are to be assigned). The evaluation criteria are the metrics by which the team can determine if a model is good enough. These criteria generally depend on the application. Ideally, they should be quantitative. However, where quantitative criteria cannot be reasonably determined, qualitative criteria can be used if the team members are sufficiently experienced to judge a model in terms of such criteria.

If several models are to be ranked or if some quantitative score is to be assigned to a model, some type of grading system is necessary. To avoid lengthy discussions at the beginning of the actual review process, such a grading system should be definitized and agreed upon among the team members before the actual review begins. To be effective the grading system should be sufficiently detailed to highlight differences between two similar models. It should also include some weighting scheme so that items of importance to the application are given more significance than other, less critical items. Once the grading system has been agreed upon by team members, it should be communicated to the appropriate personnel outside the team for information or approval as necessary.

Establishing and Disseminating Review Plans

The actual review planning includes identifying team responsibilities, establishing schedules, selecting and distributing advance orientation material to team members, and, most importantly, defining the sponsor's expectations regarding the review products. Review planning also includes development of guidance to the model developer for any briefings or demonstrations that are needed to supplement team member understanding of the model. A review schedule should be promulgated and distributed to all participants.

In order to expedite the actual reviews, team members should receive advance information consisting of the final approved statement of application requirements (intended use statement) and model documentation in sufficient time so they can gain familiarity with the material before the scheduled review. They should also be notified of any support that is expected from them for preparation of the final report so they can adequately plan on providing such support. The review coordinator should get confirmation from each team member that they are available for the review period and can provide the expected support.

Execution Phase

Typically, most of the attention and discussion of expert reviews focuses on the execution phase. However, with proper planning, the execution phase will proceed smoothly. The execution phase begins with four steps that establish a common foundation among the team members. These four steps are:

a. A review of the application to ensure a common understanding among the participants

b. A review of the evaluation tools and criteria

c. A discussion and agreement on the intended report structure and substance (obviously not including evaluative judgments)

d. And finally an orientation to the model that is being reviewed

These four steps are followed by the actual model evaluation which should be the major part of the execution phase.

Understanding the Application

The actual review process should be based on a good understanding of the problem requirements and a notion of what capabilities or features are necessary to make a model adequate for this particular problem. (An expert review that focuses on model design features begs the question as to what is necessary and sufficient for the application and forces someone else to make a judgment of adequacy.) A good statement of the intended use is a basic requirement and such a statement should be reviewed and discussed at the start of the expert review so that all members have a common perception of what the model is to be used for. Based on this perception the team should develop an understanding of what capabilities and features a model should possess so that they can judge the model's suitability.

Evaluation Tools

Knowing how a model is to be used is the first step in judging suitability. The next step is identifying the criteria that will be used to quantitatively evaluate how well a model suits the application requirements. Along with these criteria, other review tools (i.e. review questionnaires and/or scoring systems) that will be used should be defined and discussed among the team members to ensure common understanding and to identify any potential deficiencies. The purpose of up-front discussions is to avoid changes part-way through the model review, which would necessitate a re-review of model information using the modified tools. Ideally, agreement on these review tools should have been achieved prior to the start of the review and these initial discussions only serve to refresh awareness and determine any minor issues over details.

Agreement on Report Structure

After the intended use and review criteria are discussed, the structure and desired information content of the final report should be discussed. This discussion provides a common framework for identifying and highlighting information that must be generated during the review to satisfy reporting requirements. Advance discussions of this nature help to save report preparation time during the follow-up phase. One important aspect of the final report is that it should contain both strengths and weaknesses of the model in the context of the intended application. Too often reports focus only on weaknesses but do not discuss strengths which are every bit as important to the intended user. Ultimately, a knowledge of what the report will contain helps reviewers to raise questions and focus discussions on those issues that must be included in the final report.

M&S Orientation

The next step is a prelude to the actual evaluation. It consists of model briefings or demonstrations that are meant to acquaint the reviewers with the model. These briefings and demonstrations should complement the advance information and should be directed at showing how the model satisfies the M&S requirements that are derived from its intended usage. If specific evaluation criteria have been developed, the model briefings should address comparisons between the model capabilities and these criteria. The M&S orientation should include a listing of previous applications that are similar to the intended one and a presentation on any model strengths or weaknesses discovered from those uses.

Any user feedback information, such as software trouble reports or suggestions for model improvements, that help to highlight model weaknesses are also useful items to include in the briefings about the model. Extraneous information such as model development history, conceptual overview, development rationale, etc. should be avoided unless the expert review is part of an accreditation assessment process.

M&S Evaluation

In all but design reviews, the review of the model itself should be done in terms of what is acceptable for the intended application instead of what is wrong with the model. Too often these reviews turn into design reviews with the team members pointing out how the model is deficient in the way it represents this or that function. The key to making the review a true validation or accreditation effort is to have a well stated and clearly understood statement of intended use and well defined criteria for determining acceptability.

The approach for discussing model suitability is flexible. The discussions can be organized by critical problem functions, model functionality, sequence of simulated activities, or any other structure that is suggested by the particular review. Maintaining a focus on suitability vice capability is the only important point to remember in organizing a review.

Follow-Up Phase

The follow-up phase is straightforward and contains only two steps: preparation and distribution of the final report, and evaluation of the process. The report preparation step, although seldom adequately planned, is always done. Oftentimes, report distribution lags completion of the review by several months because of the time that is required to prepare it. Better planning should reduce this time lag. Most of the time, the process evaluation is done haphazardly. When done, the evaluative comments are generally included in the final report.

Report Preparation

Preparation of the final report can be expedited through good planning. Report preparation is simplified if groupware and PCs are available to each member for preparation of individual report sections. During the discussion of the report format and structure, individual review team members should be assigned responsibility for preparation of individual report sections. Ideally the report structure, outline, and format guidelines should be developed during the planning phase. Preparation assignments can be by technical specialty, model functions, or report chapters. Clerical and administrative help should be provided and any necessary model information that will be included in the report (diagrams, etc.) should be made available by the model developer in electronic files that are compatible with the selected groupware.

These preparations, coupled with commitments from the review team members to support the report writing phase, will expedite report production. As much as possible, team members should be encouraged to prepare their portion of the report during the review itself. Time can be allocated for this on-site report preparation during the review planning phase. The only other aspect is an overall editorial integration of the individual parts. Again, this must be pre-planned and performed on a timely basis.

Process Evaluation

One aspect of the expert review process that is often overlooked is process evaluation. Such an evaluation is useful in improving the process from one review to the next. An after-action questionnaire should be distributed to the team members at or near the end of the review and comments on the process should be solicited from them. The evaluation questionnaire should be prepared by the organizer of the expert review process. Team consensus on the content of the questionnaire is not necessary but comments on its clarity should be obtained. All comments, on the process and the questionnaire, should be analyzed, consolidated, and included in the final report, either as a separate section or an appendix. Suggestions for process improvements should be developed from these comments and also included to assist organizers of future reviews.

SUMMARY

A good expert review can provide valuable information about how well a model or simulation fits a particular need provided the review team understands what is wanted, why it is needed, and any constraints or desires regarding how the information is presented. The secret to a good review is proper planning.

To help ensure that any expert review achieves the desired results, assistance and advice in planning the review is available from the Joint Accreditation Support Activity. Such assistance is recommended for those who have not previously organized and executed a successful review.

4. PUTTING IT ALL TOGETHER

How to make "V&V" support "A" in a cost-effective and continuing way is the subject of the next two articles. These articles deal with the costs of making M&S credible and how to facilitate sharing those costs across multiple users in the community.

4.1 MAKING IT COST EFFECTIVE

"How much V&V is enough?" or, perhaps more to the point, "How much do you have to spend to convince someone that your model is any good?". The answer is discussed in the following article in terms of our experience in providing accreditation support for weapons system programs within the Department of Defense (DoD). Programs must adhere to the dictum, "sumptus censum ne superet[3]", and so cannot afford to waste any resources on unnecessary V&V activities. It is by having available the results of everybody else's V&V efforts, and only doing the absolute minimum V&V that you need for your own application, that you can minimize the cost of V&V.

COST VS CREDIBILITY: HOW TO BALANCE THE TWO

by Dr. Paul Muessig

The recent focus on the requirement for M&S credibility[4] has come to be balanced by an equal concern for the cost of the V&V activities that contribute to it. The lack of a coherent scheme tying V&V products to M&S credibility requirements has hindered the identification of a minimum set of V&V activities that meet those requirements. The result has been a tendency either to overestimate V&V requirements (resulting in the frequently overheard lament that "V&V is too expensive"), or to dilute them (out of frustration) to the point of insignificance (resulting in accreditation by fiat).

Contributing to the inability to quantify the cost of V&V has been a lack of generically applicable cost data from efforts with broad experience in V&V tasking. V&V cost data are mostly anecdotal, and derived from efforts that may not have sampled widely enough from the V&V menu. Moreover, existing cost data are usually unrelatable to well defined V&V products that can be tied to specific M&S credibility requirements. With the maturation and stabilization of V&V techniques, however, has come an ability to relate V&V costs to M&S credibility in a more quantifiable way. The SMART Project has contributed to this quantification in three ways:

(1) From the myriad of possible V&V tasks and products, it has identified those which contribute most heavily to M&S accreditation decisions, and has developed significant practical expertise in each of them;

(2) From this core list of V&V tasks and products, it has developed an integrated V&V process (including elements of configuration management, or C/M) that generates the most frequently required V&V products at various levels of detail and cost, and;

(3) It has embedded both the V&V process and it's products within the context of an accreditation decision by developing (and applying) a procedure for defining objective requirements for M&S credibility based on intended M&S applications.

---

[3] "Live within your means" - sorry, we did it again…

[4] See, for example, DoD Instruction 5000.59, Air Force Instruction 16-1001 and Army Regulation 5-11.

SMART V&V PHASES

SMART divided V&V activities into three phases. The three "phases" of V&V activity are: model overview, functional characterization, and detailed V&V. By segmenting the process in this way, it allows the VV&A practitioner to implement a phased approach to planned V&V activities, where only those activities which contribute directly to a particular application need be executed. Also, by documenting the results of V&V activities in a standard format, subsequent users of the model can easily access previous V&V results, thereby minimizing duplication of effort.

V&V Phase Descriptions

Phase I V&V is geared toward characterizing a model. Typical questions addressed are: How is the model managed and supported? What has it been used for, by whom, and was it accredited for that use? What is its V&V history? How well is it documented? What is the quality of the software? What are the model's known assumptions, limitations and errors? The end result of Phase I V&V is confidence that the model so characterized and controlled will produce consistent results across a spectrum of users and applications, and that its predictions have been used by others with similar applications.

Phase II V&V is aimed at providing objective data to support a subjective determination of model "reasonableness". These data consist of: input data verification and validation and certification (VV&C); comparison of model outputs with intelligence data or best estimates; sensitivity analysis results; and a detailed conceptual model specification. Phase II information, in conjunction with Phase I model characterization, allows for evaluation of the functional suitability of the model for a particular application.

Phase III V&V is "classical" V&V: the kind of V&V everyone is afraid of. It typically consists of line-by-line verification of the code, including desk checking, software testing and comparison to design specifications, coupled with extensive comparisons of model predictions with all available sources of test data at both the detailed M&S functional level and the overall M&S output level. Because of its cost, Phase III V&V activities are not (or at least, should not be) performed on an entire model without reference to an application requirement. Rather, the scope of Phase III V&V should be tailored to each specific application, and only those M&S functions essential for use in that application should be subjected to this rigorous level of activity.

Cost Effective V&V

There are several aspects of this approach to V&V that have a potential to reduce the cost of M&S accreditation. First, low cost, low risk, high value V&V tasks are performed early in the process. These tasks are aimed at identifying M&S strengths and weaknesses in a short period of time, and permit an early assessment of M&S applicability to a wide variety of possible applications.

Second, placing high cost, high risk V&V tasks at the end of the process allows time for development of objective criteria by which to minimize the scope of the Phase III effort. This avoids artificially inflated V&V cost estimates (and worries) and mitigates against the tendency to "V&V for it's own sake." It also guarantees that any detailed V&V performed will have maximal applicability to the accreditation decision.

Third, the well-defined phases of V&V activity and the standard V&V reporting formats are essential to building cost effectiveness into accreditation decisions. V&V information developed in support of one accreditation benefits future users of the same model, who can build upon prior V&V information to reduce the scope of their own accreditation effort. Moreover, cost benefits for future users are guaranteed when each V&V effort is documented in a standard way, summarized in a central repository, and archived by a central accreditation support activity. Using this cycle, V&V becomes "market driven" by actual accreditation requirements, and no one sponsor has to pay for the V&V of an entire model.

COST ANALYSIS

But how much does all this really cost? To answer that question, we developed a Work Breakdown Structure (WBS) for each phase of V&V activity, and determined the level of effort (LOE) applied to each WBS element over the life of the project.[5] The result was a model-specific average LOE for each phase. Since these averages did not reflect completed products (none of the Phases were complete for any model at the time the data were collected), each LOE was normalized to the amount of progress actually achieved in each Phase. The normalized model-specific LOEs were then averaged to arrive at an estimated LOE for each phase that would apply to any model.[6]

To convert LOE figures into cost figures, we calculated an equivalent hourly rate based on a weighted average of contractor LOE and individual labor costs over the life of the project, including other direct costs such as travel and documentation. This weighted average labor rate turned out to be approximately $87.50 per hour, or $14K per man-month, assuming a 160 hour man-month. We then multiplied the average LOE for each phase (in man-months) by the average dollar cost per man-month to arrive at the average dollar cost of each phase of V&V.

The results of this cost analysis are shown in table 1. The table shows the average level of effort and dollar costs for each phase of V&V activity. The term "FE" used in the table refers to the "Functional Elements" that comprise the model. Identification of these FEs occurs in Phase I (Decompose Model). The first number in each WBS element identifies its phase, the second number identifies its primary activity (1=verification, 2=validation, 3=configuration management), and the third number identifies the task sequence within a primary activity. WBS tasks not listed in table 1 are those for which no reliable cost estimates were derivable because they vary with each application

Table 1 suggests some interesting characteristics about the incremental approach to V&V that should comfort those concerned about a perception of runaway V&V costs. For example, Phase I V&V tasks, whose products relate to the credibility of the model as a whole, have a fixed cost to complete in full (i.e., about $250K), while the V&V tasks in Phases II and III have a variable cost related to the number of specific M&S functions requiring credibility to support a particular application. Since the total cost of Phase II and III V&V for two FEs exceeds the cost of the entire Phase I V&V effort, it is clear that V&V costs are dominated by the number of FEs required to support a particular application. This suggests that significant attention must be paid to defining the smallest possible set of model functions required to support the model's use for a given application. It is in the rational and objective determination of this minimum set that the true path to cost effective V&V lies.

There are, however, other prerequisites for cost effective V&V. Consider, for example, two users of the same model who wish to accredit it for different applications. Assume further that both users have decided to complete Phase I, and that both require Phase II V&V for the same 5 FEs. If they both work independently, and have no knowledge of each other's efforts, they will both spend $462K according to the data in table 1, for a total of almost a million dollars of V&V. At the end of all that, each user will have completed the same set of V&V tasks. Several important questions arise: How did

---

[5] Reasonably complete cost and LOE data were available because SMART used Earned Value techniques to monitor project progress. These techniques require contractors to report actual progress against resource, cost and schedule estimates made at the beginning of each task. The WBS shown here is not the final version of the SMART WBS, however the costs are still representative of the experience over the life of SMART.

[6] The models V&V'd by SMART contained between 30,000 and 100,000 lines of code each. Larger, more complex M&S may require substantially higher levels of effort to account for the greater number of interactions and software and validation test cases that would need to be examined.

each user interpret the same V&V task? Based on the different possible interpretations of each V&V task, how different do the resultant V&V reports look? Based on how different the V&V reports might look, how can future users benefit from two reports that interpret V&V differently, that say different things about the same model, and that report those things in different ways? Could the two efforts have worked together to reduce the total V&V cost burden while still accomplishing their individual accreditation objectives? How can future users benefit from the work these two independent efforts have done? A few moments reflection on possible answers to these questions reveals some enlightening suggestions for making V&V more cost effective.

WBS #	TASK NAME*	LOE (MM)	COST ($K)
1.1.1	Assess Current Documentation	3	
1.1.2	Assess Software Quality	3	
1.1.3	ID Assumptions, Limitations, Errors	3	
1.2.1	Decompose Model	2	
1.2.2	Define Functional Templates	1	
1.3.1	Define C/M Baseline	1	
1.3.2	Evaluate Existing C/M Procedures	2	
1.3.3	Survey Model History	3	
	PHASE I TOTALS	**18**	**252**
2.1.2	Prepare Software Design Documents	2/FE	
2.2.2	Perform Sensitivity Analysis	1/FE	
	PHASE II TOTALS	**3/FE**	**42/FE**
3.1.1-3.1.4	Code Verification Tasks	2/FE	
3.2.1-3.2.7	Validation Analysis Tasks	6/FE	
	PHASE III TOTALS	**8/FE**	**112/FE**
	GRAND TOTALS	**18 + 11/FE**	**252 + 154/FE**

TABLE 1: TYPICAL V&V COSTS

Had each effort known of the other, a consolidated list of V&V tasks could have been drawn up from a coordinated review of M&S fidelity requirements for the two applications. In addition, a common understanding of V&V tasks could have been developed, as well as an understanding of how these tasks related to the previously defined fidelity requirements. Such a review could have led to a common V&V reporting format that would have served the needs of both applications without duplication of effort. By making this common V&V report available to other users of the same model, V&V requirements in support of other applications could be reduced substantially, by forming the nucleus of a body of evidence supporting the model's credibility, a body to which other users with other applications could have contributed, in turn.

Three prerequisites for cost effective accreditation are suggested by the preceding discussion. Simply put, they are: (1) a standardized V&V process; (2) a standardized V&V reporting format, and; (3) a structured approach to developing objective M&S fidelity requirements based on application requirements. Lest the faint of heart (or the hard of head) take umbrage at the words "standardized" and "structured," let us hasten to explain that we do not construe these words to mean "mandated" and "inflexible." By standardized we mean that V&V techniques applicable to different classes of M&S should be well defined, understood and documented by the M&S community served by these techniques. Likewise, the product resulting from each V&V technique should be specified in such a way as to facilitate an orderly accumulation of the evidence for M&S credibility. Finally, an objective means of narrowing the scope of required V&V activities based on application requirements is necessary, to avoid the temptation to "do it all" or "make do with nothing."

Having a history of prior V&V and accreditations is another essential factor in keeping V&V costs down. This suggests the need not only for a VV&A repository for M&S, it requires that V&V be conducted using techniques and definitions accepted by the various M&S communities, and that V&V be documented in a standardized format within each community. SMART developed a repository for VV&A information for the M&S which it has assessed, called the Accreditation Support Database, and it is available via the Joint Accreditation Support Activity homepage (http://www.nawcwpns.navy.mil/~jasa).

SUMMARY

As individual models are used again and again in support of different applications, substantive, objective evidence of their credibility grows. Although each user (or group of users) contributes only those aspects of credibility applicable to their current problem, over time an objective body of evidence grows (see Figure 3). SMART's standardized V&V process, coupled with its standard V&V reporting format (the three-volume ASP format), provides a convenient way to accumulate evidence of M&S credibility. It is the steady accumulation of such evidence that reduces the cost of subsequent accreditation for individual M&S, because accreditation builds on prior V&V, rather than on independent (mostly duplicative) efforts. Seen in this light, SMART's accreditation support approach provides the DoD community a pathway to lower cost accreditation.

SMART was charged with the task of transferring its V&V process and products, as well as its accreditation support experience, to the wider DoD community. We feel strongly that the incremental approach to V&V, dividing it not only into large phases, but into smaller model-level and function-level activities, helps ensure that precious V&V dollars buy meaningful V&V products.

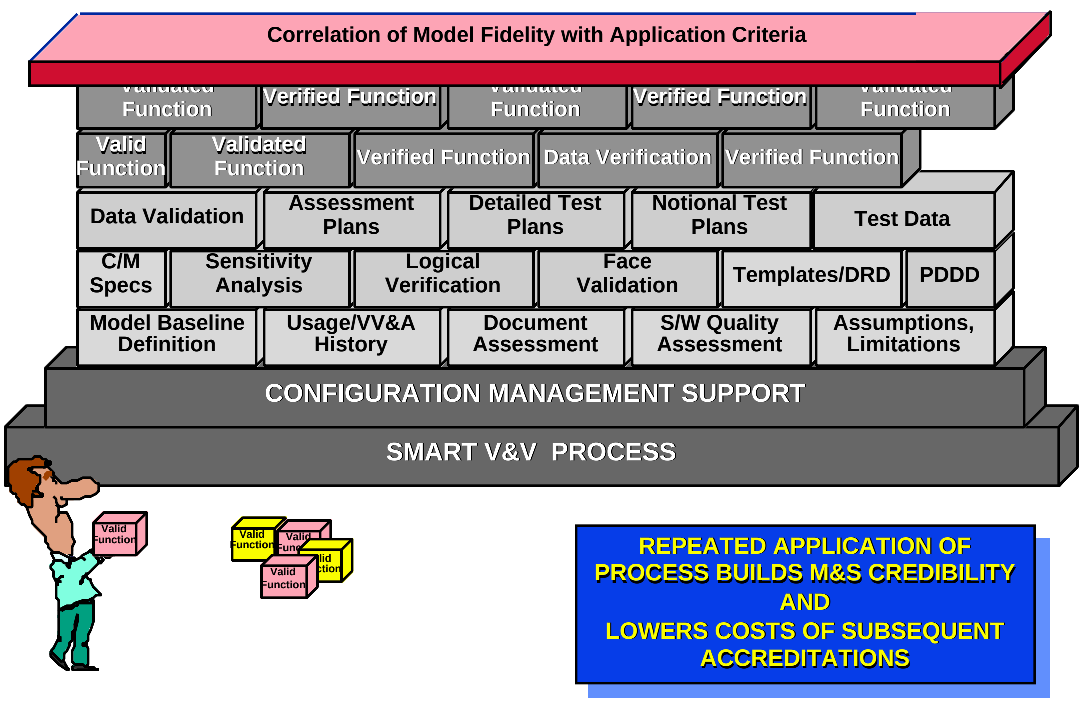


FIGURE 3: BUILDING M&S CREDIBILITY

4.2 AND KEEPING IT ALL TOGETHER

One of the biggest problems with VV&A activities is their "shelf life". Once a model has been improved or changed in any way, unless those changes are carefully recorded and evaluated, any V&V done previously on the model is suspect. This article describes how to make sure that V&V is an integral part of M&S management. It should be noted as well, that although the article speaks of "legacy" M&S, once a model or simulation is developed and in use in the community, it is by definition a "legacy" model. Thus the concepts in this article apply to new code developed in architectures such as J-MASS, as well as older, well-used FORTRAN code.

INTEGRATING V&V INTO DEVELOPMENT:
THE IMPORTANCE OF CONFIGURATION MANAGEMENT

by David H. Hall

This paper discusses some "lessons learned" during the SMART project about how verification and validation (V&V) activities for models and simulations (M&S) relate to the model development process. In particular, there are some valuable lessons learned about coordinated model development, not only within a particular model's life, but for a community like the survivability discipline, across models and simulations with similar requirements.

A common approach to configuration management, and to the inclusion of V&V in the development process, can greatly benefit models such as those distributed by the Survivability/Vulnerability Information Analysis Center (SURVIAC). Such an approach also facilitates the development and re-use of common algorithms used for various elements of those M&S.

TYPICAL MODEL MANAGEMENT, FOR NEW AND OLD

A typical situation for mature, or "legacy" M&S is that neither their development nor their management has been coordinated with any other related M&S activities. This is true of the models resident in the SURVIAC model repository, even though they represent the community accepted M&S used to support joint service survivability analyses. This is the case because model development is typically driven by who has funding to solve a particular problem, and not by coordinated community requirements. This results in conflicting database structures, and some duplication of effort across the community.

Also, insufficient focus and resources have been placed on V&V of models in general; model development is typically driven by a desire to add new features to the models, rather than by a need to correct identified errors or shortfalls. This is also true of new model developments that are currently ongoing outside of the SURVIAC model set. While lip service is being paid to V&V as a requirement, actual V&V in new model development is being deferred in favor of adding more features and enhancements (or sometimes simply in the interest of getting a final product out the door.)

Another issue is the management of these M&S, both for mature models, such as those in SURVIAC, and for new model developments. Mature model management is typically conducted by the individual model managers and developers, and their management styles are dictated by those individuals' understanding of the requirements for such a process, and by the funding available. Historically, the success of the varied approaches adopted for model management has been spotty at best.

And new model developments are not necessarily any better off: as with V&V, adding new model features typically takes precedence over management of the model itself. This uncoordinated model management and development leads to a waste of scarce resources, and it does not appear that new model developments are, in practice, going to fare much better.

The rest of this paper is focused on the M&S that are resident in and distributed by SURVIAC, but the concepts described here apply equally well to other M&S, especially those designed with "re-usability" and interoperability in mind, such as those to be put together under a J-MASS type architecture.

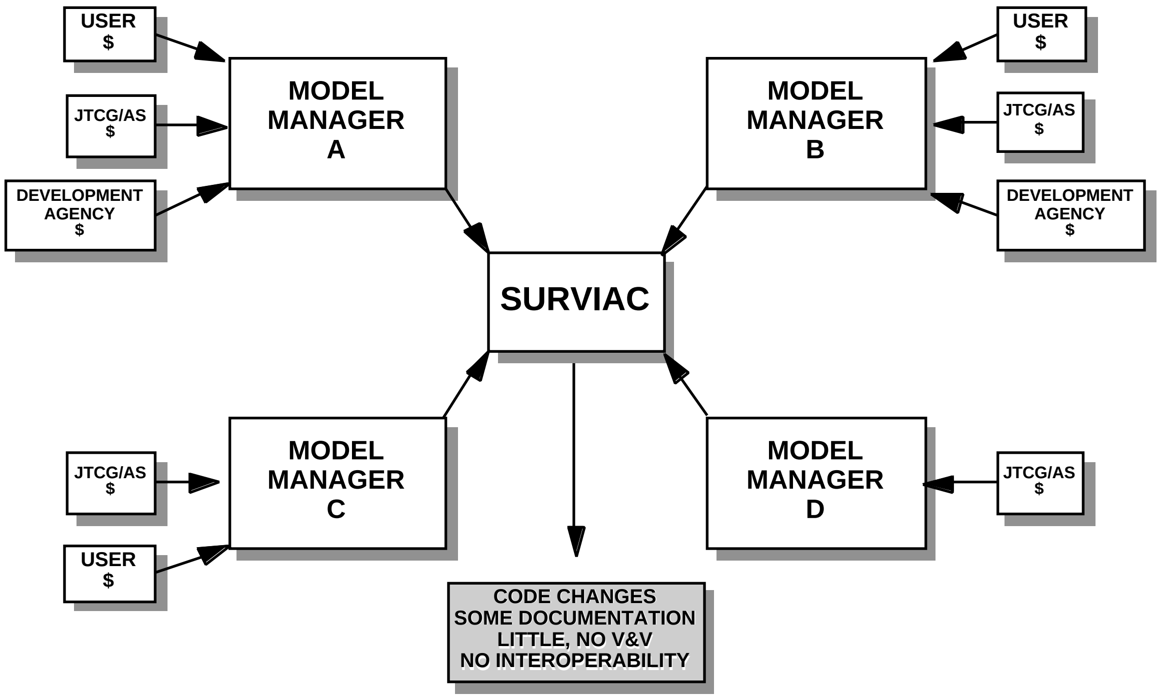


Figure 1. The Current State of Affairs

Figure 1 illustrates the situation as it has existed in the past for SURVIAC models: the various Service model managers have utilized what resources are available to them to develop model enhancements, documentation, and a little V&V. The results of these activities have been provided to users via SURVIAC, but SURVIAC has had little involvement in the development of these products. In addition, in the past there has been little effort at making the models interoperable, except for the recent introduction of the Digital Integrated Modeling Environment (DIME) into SURVIAC. As a consequence, it makes it difficult for a user to know whether he or she has the latest version of the model, in some cases what capability is in the version distributed by SURVIAC (if documentation was not forthcoming from the model manager), or how well the model works (if there is very little documented V&V). It also has made it difficult for users to operate these M&S in concert, since the work done on them has not been done in a coordinated manner.

However, this picture is changing. Partly as a result of recommended configuration management processes developed under the SMART project, and partly as a result of work accomplished under the JTCG/AS Methodology Subgroup, SURVIAC M&S are beginning to come under a coordinated model management approach, with a greater role (and associated resources) envisioned for SURVIAC to provide that coordination.

COORDINATED MODEL MANAGEMENT

The solution to the problem of model management is to coordinate model development, model management, and model V&V activities to ensure that users get well-managed, capable and credible M&S to support their needs. We can't sacrifice any of these elements for any of the others: a model with all the features the user says he wants, but without any V&V, provides no credibility; a model with full V&V but without the critical features that address the user's problem doesn't satisfy his requirements; and neither does a model with all the required features, in some version, but where it's not clear which version has which features or even which version the user has!

To coordinate model development, management and V&V requires a partnership between model users, managers, and in the case of JTCG/AS supported M&S, SURVIAC and the JTCG/AS itself. The goal is to execute a coordinated M&S development and management process with stable sources of funding to support it.

The first step to developing a coordinated model management process is to adopt consistent configuration management processes across SURVIAC models; this is the first step toward coordination of M&S development cycles. Configuration management of "legacy" codes varies considerably from model to model, and these must be brought into some synchronization for coordinated model management to be efficient.

The second step is to adopt common databases, input data structures and algorithms for common functions within the models, wherever appropriate. Historically, since these models have been developed independently, databases such as target characteristics, environmental functions, geometries, etc. are treated differently. By bringing these models into conformity in these areas as much as possible, we can not only reduce the required time and resources to generate input databases, but we can facilitate the transition to a common architecture when it becomes available. This is being facilitated by the results of SMART functional level validation across models; these results help identify the best algorithms to use for common model elements across M&S.

The final step to coordinating model management is to adopt common documentation standards and develop the documentation for each SURVIAC model supported by the JTCG/AS. Software documentation (for example, user manuals, analyst manuals, programmer manuals, conceptual model specifications) should be common among all the models. In addition, standard V&V documentation sets should be developed for all SURVIAC models; these should be in the Accreditation Support Package (ASP) format developed by the SMART project.

Effective Configuration Management

SURVIAC, under funding from the SMART project, conducted a configuration management study aimed at developing a recommended CM process for use by all SURVIAC model managers. The study was based on surveys of user preferences for CM practices, a detailed review of DoD standards and guidelines for CM, interviews with individual model managers, and the experience of the technical staff at SURVIAC itself. This study developed some basic assumptions and prerequisites for effective M&S configuration management based on a one-year CM cycle for updates to a baseline distributed version of the model. The study and process were documented[7], the process was demonstrated using the ESAMS model, and it is being implemented for all SURVIAC models starting with the 2.7 version of ESAMS.

Some basic assumptions and prerequisites must be understood and in place in order for any configuration management process to be effective. First and foremost, there must be acceptance in the community of the concept of a "baseline version" of the model, which represents the current "state-of-the-art" of the model at the time it is released. Once the current baseline release is out of beta testing, it becomes the new release, and it will not be revised or enhanced, other than bug-fixing. This ensures version consistency; otherwise a user has no idea what's in the version he or she has. Configuration management will be performed on distributed source code. The source code is distributed to allow the model manager and SURVIAC to take advantage of user developed enhancements; the CM process is

---

[7] *Configuration Management Requirements Study,* November 1995, Joint Technical Coordinating Group on Aircraft Survivability (JTCG/AS) JTCG/AS-95-M-005

set up to facilitate getting those enhancements into the next baseline version. In this way the users have maximum visibility into the code, while providing a mechanism to minimize version proliferation.

Enhancements, or the next version, will start from the current baseline, so that all users will know from whence springs the new version. Documentation will be concurrent with development: no documentation means there will be no new version delivered. Any independent enhancements (funded by users, and not coordinated by the model developer) will only be considered for inclusion in the new version if they come with the same documentation required of enhancements coordinated by the model developer.

V&V will also be concurrent with development; current V&V documentation will be released with the new version, which describes V&V conducted on the changes from the previous version.

The beta site process will be streamlined and coordinated. A beta-site test plan will be approved, with a few sites, each having specific test objectives for a specified portion of the model. This may require making funding available to the beta sites to facilitate the process.

Another prerequisite for effective CM is the existence of contract vehicles to support these activities. Active task orders must be in place with applicable statements of work to support model enhancements, corrections, documentation and V&V. Otherwise, there cannot be timely implementation of the process.

The definition of a "new version" must be understood: code that works, on media that can be read; a complete update description (What's new in this Version?); documentation updates (software manuals and ASP's); updated input databases and change descriptions; and sample test cases and results. All of these elements must be included in a "new version" of the model, or the user is not fully supported by the CM process. Too often users are provided with tapes they can't read, or no documentation, or no available databases, or no understanding of how this code differs from the last one they got (a month ago), or no sample cases to see if the code's even running right! We need to make using SURVIAC models as simple as opening a new commercial software package and installing it on a Macintosh (we're not going to mention PC users and *their* experiences).

Figure 2 illustrates the new stylized "CM Year". There is no real requirement for the CM cycle to be yearly, but that seems to be an acceptable compromise between version stability and user requirements for timely enhancements to the models. The first quarter of the year (release of the new baseline version begins the year) is devoted to defining what the requirements are for the next version. This is done in coordination with the users group, the configuration control board, the model manager, and in the case of JTCG/AS models, SURVIAC. From this phase comes prioritized and coordinated enhancements, a development plan, a beta test plan, and funding requirements.

The second phase includes actual development of the software according to the development plan, including integration and initial testing. Reviews are conducted by the users, the model manager, and the CCB. Documentation is developed concurrent with model development.

The last quarter of the year consists of beta testing and reporting, review of documentation, preparation of sample cases, and updates to input databases. All of these elements are prepared for dissemination upon approval by the CCB and the model manager. At the end of the CM year, the new baseline version is released, and the cycle starts again.

Figure 3 illustrates the desired result of this new CM process: coordinated development of not only each individual model, but M&S development coordinated with user requirements for other models in the SURVIAC suite. Only in this way can the JTCG/AS and SURVIAC provide the type of support needed by DoD decision makers who use the results of M&S.

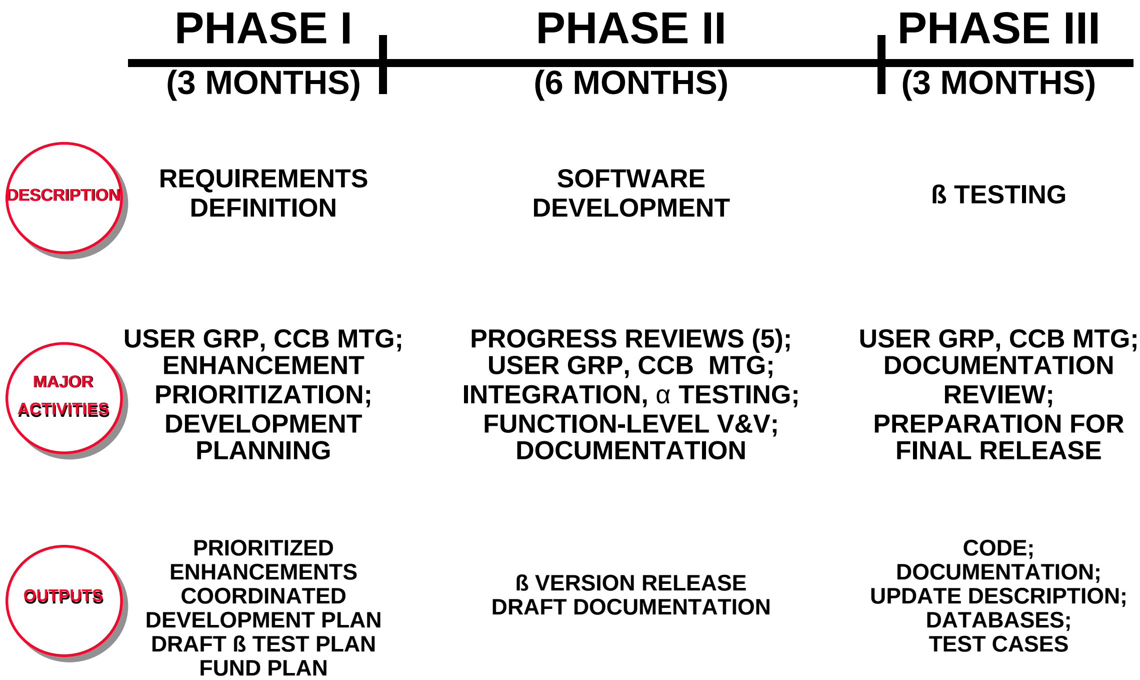


Figure 2 The Configuration Management Cycle

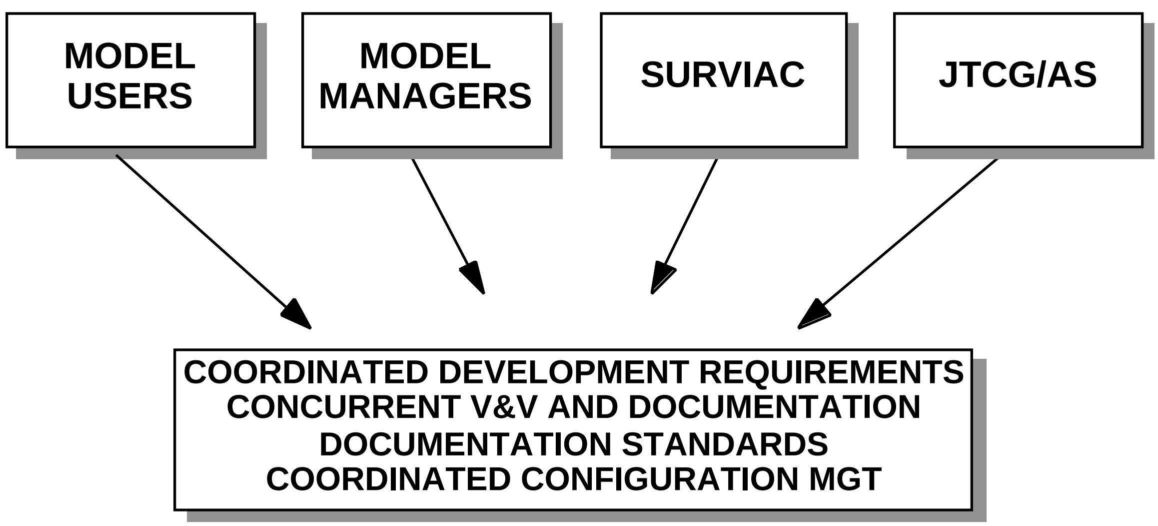


Figure 3 The Way Life Should Be

Roles of the Model Manager, Users and SURVIAC

There are distinct roles for each of the players in this process. The model manager must put together coordinated model development plans with inputs from JTCG/AS, SURVIAC, users and their parent service organization. He or she must develop “buy-in” from their parent organization, literally with funding to support this process. Model managers must adopt consistent configuration management by merging the key elements of this process with their current practices; minimizing model specific procedures and encouraging greater SURVIAC model involvement will ensure maximum benefits to the users. The model manager must make V&V and documentation a priority, concurrent with development. In the words of one model manager, “We must do less development better!”

The JTCG/AS must serve as focal point for coordinated model management activities. The JTCG/AS, through the Methodology Subgroup, can facilitate integration of model manager, user, and SURVIAC requirements and activities with respect to model development. JTCG/AS is also in a unique position to provide relatively stable funding to support coordinated model development plans and model manager initiatives through their annual budget and project prioritization process.

SURVIAC’s role must be expanded (and funded) to better support the configuration management process. SURVIAC is in a position to provide all required information to the users. SURVIAC can also support the CM process directly by collecting and disseminating information on model deficiencies, the status of model improvements, and minutes of user and CCB meetings. A good example of the support that SURVIAC can provide is the Electronic Bulletin Board maintained for the ESAMS model user community.

The model user’s role in this process is to “buy-into” the coordinated model management process. The users must make every attempt to minimize independent model developments and enhancements; model fixes and enhancements must be coordinated through the model manager and the CCB. If a user must enhance the model for his immediate requirement, those enhancements must be made available to the model manager as soon as possible, with applicable documentation, for consideration in the next baseline version. In order for this to happen, the users must participate fully in user group meetings to find out what else is going on with the model, and to help prioritize model fixes and improvements. SURVIAC can also make this feedback procedure a requirement for receipt of the model.

In order for this process to work efficiently, there must be an active user group for the model, that participates fully in the configuration management process. This is a “lesson learned” that applies equally well to new model developments. Any new model must cultivate an active and interested user group, who have been provided the source code to fully evaluate for their own application, in order to be an adequate resource.

BENEFITS

The principal benefit of this coordinated management process is better support to users. Orderly development is characterized by coordinated M&S requirements, guaranteed credibility (by really doing concurrent V&V), and adequate and up-to-date, complete and consistent software and V&V documentation.

Common input data structures are a critical step toward interoperability. Common algorithms developed for common functions in models are also a path to interoperability; the development of these common algorithms is made possible by the concurrent V&V of the models. The “best” algorithm can be selected from the set represented in the various models being evaluated because the V&V process evaluates how “good” each algorithm is. Thus V&V provides “directed M&S development.”

Coordination of all these efforts through the model managers and the JTCG/AS leverages funding from a variety of sources and saves the taxpayers money. As users come to the table with model requirements and funding to support the development of improvements, a coordinated plan of action will provide the highest benefit per dollar invested.

This type of coordinated model management and development, including concurrent V&V, is essential for new model developments as well. The lessons learned in applying sound model management principles to the SURVIAC model suite also should be applied to the development of any new M&S

which are intended to support multiple users through "reusability" and "interoperability." The relentless push for new capabilities cannot be allowed to displace the need for thoughtful and well planned model management: "more features" is not necessarily an improvement over "tried and true capability."

5. DOCUMENTING VV&A ACTIVITIES

So how do we document the results of all this? We should do it in a way that makes it easy for people to understand the results, to add any new results to the body of knowledge about the credibility of the model, and to share benefits (and cost). The best way to do this is to establish a standard reporting format for VV&A results.

THE BENEFITS OF STANDARDIZED VV&A REPORTING

by David H. Hall

Everyone in the system acquisition business wants well managed, community accepted, and well documented M&S which have a high degree of credibility. And they want someone else to pay for these features and benefits. The question of credibility, in particular, raises concerns among M&S users about the cost of establishing that credibility: how much is verification, validation and accreditation (VV&A) going to cost me? Hasn't somebody else already done this? Only by facilitating the sharing of V&V results can the community truly make model credibility affordable. And the best way to facilitate the sharing of those results is to put everyone's results in the same, standard format and update it every time new VV&A information comes in.

KEY CREDIBILITY METRICS

But what types of V&V data should be included in a standard report? As an aid to defining key credibility metrics for use in this process, SMART conducted an M&S Accreditation Requirements Study. This study included surveys of several dozen agencies and policy makers who are responsible for making accreditation decisions, in order to determine what processes they use to make those decisions, what information is required, and what key credibility issues are important to M&S users. Pertinent policies and procedures documented in DoD publications were also referenced. This study identified the key credibility metrics that are used by those agencies and listed in those policies, and organized those metrics into logical reporting elements for documenting the results of VV&A activities.

The key metrics identified by the Accreditation Requirements Study organize themselves into three general categories. The first category can be defined as "Model Overview," which includes elements relating to an overview of the suitability of the model for any application. The second category , called "Functional Characterization," includes credibility metrics that relate in detail to how the model is designed, and any reviews of that design that have been done by subject matter experts. The last category, identified as "Detailed V&V," includes those metrics which delve into the correlation between the model, its design, and the real world. These correspond to the three volumes into which the SMART V&V information was subsequently partitioned.

Model Overview:

Key metrics in this category answer the question: "Are the basic characteristics of the model well known and documented?" They consist of:

(1) <u>Model Configuration Management Baseline Definition</u>: What code and documentation set constitutes the "official" model baseline? How are the model, and changes to it, managed and supported? This metric tells a user if he can easily identify what version he has, and what's in it.

(2) <u>Summary of Assumptions, Limitations and Errors</u>: What assumptions, limitations and errors are known? What is the impact on model usage of each of those? By this metric the user can tell if any of these limitations of the model will affect his application.

(3) VV&A Status and Usage History: Who has used the model before, and for what? What is the model's V&V history and status? Who has accredited the model before and for what? A rich history of previous usage and VV&A activities lends confidence to the model's use for a new application.

(4) Documentation Assessment: How well is the model documented relative to accepted standards? This tells a user how easily he can count on becoming an informed user from available source documents.

(5) Software Quality Assessment: How "good" is the software relative to accepted standards? Software that is structured and easy to follow will lend far more credibility to results than will "spaghetti code."

The model overview elements provide enough information to a new user to determine at a relative glance whether this model is appropriate for further consideration for his application. It could be considered the "soup can label" for the model, listing the various ingredients in order of importance on the can, so that the user can put it back on the shelf unopened if he doesn't like what's in it.

Functional Characterization:

The metrics in this category answer the question: are the functional characteristics of the model defined, well designed, and reasonable? These metrics are:

(6) Functional Decomposition: What are the model's basic functional elements? This can tell a user if the model even addresses the requirements of his application.

(7) Conceptual Model Description: How are model functions integrated to produce model results? This addresses issues of model construction, and whether the model has the flexibility to address the user's particular problem.

(8) Detailed Software Specification: What are the design requirements for each of the model's functional elements? How are they coded? From this information the user can determine if the level of fidelity is appropriate for those functional elements that are important to his problem.

(9) Logical Verification: For what set of problems do the model's assumptions and limitations yield correct results? How do model assumptions, limitations, errors and approximations affect potential uses? Are the model's assumptions, limitations and approximations reasonable for specific applications? These are the results of assessments of the model by previous users that can be evaluated by a new user in the light of his application.

(10) Sensitivity Analysis: Are the model's sensitivities reasonable? Sensitivity analysis identifies function and model level impacts of variations in inputs; these analyses identify those functions which have the greatest impact on model results, and can in some cases substitute for detailed V&V activities. Sensitivity analysis also establishes accuracy requirements for validation data.

(11) Face Validation: Do model outputs agree with best estimates? These are the results of assessments of the model by previous users and subject matter experts.

The Functional Characterization Information Elements provide detailed information that allows a potential model user to evaluate the design of the model in comparison with the functional requirement of his application. They answer the question: Does this model have the functional fidelity that I need to address my problem? It might be equivalent to opening up that can of soup, and telling by the looks of it whether you want to heat it up and try it, or throw it down the garbage.

Detailed V&V:

The metrics in this category answer the questions: Is the code built in accordance with its design? How well do model inputs and outputs compare with the real world? The information elements in this category are:

(12) Data Verification, Validation and Certification (VV&C): Are input data well defined and consistently used? Do input data agree with best estimates or intelligence information? What is the impact on M&S use of any limitations discovered? Data VV&C tells a potential user if there are data issues which could impact his use of the model and his interpretation of its results.

(13) Code Verification: Does the code correctly implement the design? What is the impact on M&S use of any limitations discovered? Code verification metrics tell a user how well the software conforms to its design, what the configuration management process is (or is not) doing about any non-conforming code, and whether any of those non-conformities are important to his problem.

(14) Validation with Test Data: How well do model results compare with test data? What is the impact on M&S use of any limitations discovered? These results offer the best and final proof to the user whether the model is of sufficient accuracy for his purposes.

The information in this third and final category , "Detailed V&V," is the equivalent to heating up that can of soup, and tasting a little before you serve it. If it tastes OK, serve it to your company; if it doesn't, break out the peanut butter instead.

STANDARD DOCUMENTATION FORMATS

SMART has developed and implemented a standard reporting format for V&V results, organized around the critical metrics listed above. The information is collected in three volume "Accreditation Support Packages (ASP)"; the volumes provide, respectively, (1) Model Overview, (2) Functional Characterization, and (3) Detailed V&V metrics. It is especially important to note that one aspect of this standardized reporting format is that results from logical verification, sensitivity analysis, face validation, and detailed V&V include an identification of any assumptions or limitations that grow out of the findings. Also the implications or impacts of these limitations and assumptions on model usage are described. These elements facilitate comparisons between V&V results and M&S requirements for a given application. Without providing the reader an analysis of the implications of these V&V results, just having the V&V results themselves is of limited benefit. The prospective user needs to know what all this means to him (or her).

SPECIFIC BENEFITS

One of the perennial problems with M&S verification and validation has been inconsistent definitions of V&V by various people and organizations. While we now have accepted definitions of the terms verification, validation and accreditation in JCS Pub 1, the actual interpretation of those definitions is by no means standard. When you ask someone, "has your model been validated", whatever their answer is doesn't really tell you much unless you know what they mean by "validated." If they say "yes", some people mean they've done a detailed comparison with test data, some people mean that any intelligence data has been certified by the appropriate agency, and some people mean that a bunch of experts gave it a "thumbs up." If we can agree on a standard VV&A reporting format, when you ask that question, the answer can be "yes, and here are the results!"

The ASP format in essence defines what we mean by "verified and validated." You can look at the ASP's for the model and determine almost at a glance exactly what V&V has been done, and what has not. Thus the standard reporting formats provide a quick and easy understanding of the scope of the

V&V already accomplished. Similarly it provides a framework within which to specify additional V&V requirements.

A criticism of the SMART V&V approach has been that it costs too much to generate all the information elements that were identified as a result of the accreditation requirements study. No single user can afford all of the V&V tasks called for to produce a total V&V package with all the indicated types of information. As a consequence, the idea of a standard reporting format is thought to be not cost-effective. But this misses the point entirely! The point is that no single user can afford to do all of this, so we need to come up with a way to make sure that what each user does accomplish can be easily integrated with existing data in such a way as to be useful to subsequent model users.

The true benefit of a standard reporting format is felt more and more as the use of this documentation among a number of users increases over time. As more users require VV&A information about a particular model, and more users generate additional information specific to their own individual needs, the value of keeping this documentation current and available to the community becomes more and more apparent. If acquisition programs know where to look for this information, and what it's format will be, they can quickly and easily access it to minimize their need to conduct additional V&V. It allows users to concentrate on those areas of the model that are most critical to them. Standardized documentation also simplifies the problem of keeping the documentation current. With each user following the same documentation standards, there is no need to review and update past information in light of current findings.

Another problem which often occurs is locating, obtaining, reading, and interpreting existing V&V reports. Model users are likely to spend significant time in these activities, especially reading and interpreting past V&V reports in light of the current application. Standard reporting formats help reduce these problems in two ways. First, the standard formats facilitate identification of indices and pointers that can be used to index the reports along application lines. Second, standard formats allow the reader to scan the document and locate pertinent information for his particular application. Thus standard reporting formats reduce the time and effort involved in getting and using existing V&V data.

SUMMARY

The credibility of M&S used in the system acquisition process is crucial to informed use of those models in support of system design, T&E and training. A cost-effective credibility assessment process will have little value unless the results are documented in standard reporting formats. A list of 14 credibility metrics were developed through consultation with a wide variety of M&S users throughout DoD. A key to making the whole process cost-effective is to facilitate cost sharing of V&V activities across the whole spectrum of users of a model. The way to make this happen is to document and update all VV&A results in standard formats and provide them to the community.

6. WHERE TO GET MORE HELP

Perhaps one of the most important lessons that SMART learned over its five year life is that the acquisition community needs experienced help in doing VV&A. Recognizing that need, the JTCG/AS established an activity to provide that help to DoD customers, as described in the following article.

JOINT ACCREDITATION SUPPORT ACTIVITY (JASA) STANDS UP

by David H. Hall

The Joint Accreditation Support Activity (JASA) stood up in October 1996 to provide model and simulation (M&S) verification, validation and accreditation (VV&A) support services to DoD acquisition programs. These services include:

(1) Helping customers to identify M&S requirements and acceptance criteria for their particular applications.

(2) Comparing those requirements and acceptance criteria to what's known about existing M&S, including VV&A information.

(3) Developing cost-effective VV&A strategies and plans to meet customer needs.

(4) Applying an established, accepted and documented VV&A process to customers' M&S.

(5) Producing and maintaining standardized VV&A documentation for all M&S evaluated using the process.

(6) Developing risk assessments and risk mitigation strategies based on M&S shortfalls found with respect to customer requirements and acceptance criteria.

The JASA leverages the M&S VV&A support infrastructure and technical expertise developed under SMART. During that project, a cost-effective VV&A process was developed for engagement level M&S used in acquisition. The process has since been successfully expanded to cover engineering and mission level M&S. The process is in complete conformance with DoDD 5000.59, DoDI 5000.61 and the VV&A Recommended Practices Guide developed by the Defense Modeling and Simulation Office over the last two years.

A number of major weapons systems acquisition programs have been provided this type of M&S VV&A support. These programs have become the first customers of the JASA, continuing the M&S VV&A support activities conducted for them under the SMART Project. Acquisition programs are encouraged to investigate how JASA can assist them with their M&S accreditation needs.

Standardized V&V documentation, in the form of Accreditation Support Packages (ASP's) for the models evaluated so far can be obtained from the JASA Program Office technical point of contact mentioned below. These models include: the Enhanced Surface to Air Missile Simulation (ESAMS), the Advanced Low Altitude Radar Model (ALARM), the Radar Directed Gun System Simulation (RADGUNS), the Trajectory Analysis Program (TRAP), BRAWLER, and the Extended Air Defense Simulation (EADSIM). In addition to these models, JASA is developing ASP information for a number of other M&S, including the Joint Service Endgame Model (JSEM), the Computation of Vulnerable Area and Repair Time (COVART) model, the Joint Munitions Effectiveness Manual (JMEM) Small Computer Methods, and others. ASP's for SURVIAC models will also be available through SURVIAC, and will be distributed with other model documentation on request.

More information on the JASA, its customers and its support services can be obtained by visiting the JASA web site at

http://www.nawcwpns.navy.mil/~jasa/

The Washington, D.C. area point of contact is the JTCG/AS Central Office, (703) 325-0165; e-mail: rflores@msis.dmso.mil. Technical details of JASA support services can be obtained by contacting Dr. Paul Muessig at the Naval Air Warfare Center, China Lake , Voice: (619) 927-1271; fax: (619) 939-2062;

e-mail: paul_muessig@imdgw.chinalake.navy.mil