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

# PLANNING FOR VERIFICATION, VALIDATION, AND ACCREDITATION OF MODELING AND SIMULATION APPLICATIONS

Osman Balci

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

William F. Ormsby
John T. Carr, III
Said D. Saadi

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

## ABSTRACT

A comprehensive and detailed verification, validation, and accreditation (VV&A) plan and its proper execution are crucially important for the successful accreditation of a modeling and simulation (M&S) application. We provide guidance in developing and executing such a plan throughout the entire M&S application development life cycle.

## 1 INTRODUCTION

The U.S. Department of Defense (DoD) is the largest sponsor and user of Modeling and Simulation (M&S) applications in the world. DoD uses many different types of M&S applications, consisting of a combination of software, hardware, and humanware, under diverse objectives including acquisition, analysis, and training. The DoD Instruction 5000.61 states that "it is the DoD policy that: … models and simulations used to support major DoD decision-making organizations and processes … shall be accredited for that use by the DoD component sponsoring the application" (DoDI 1996).

The DoD component sponsoring the M&S application (MSA) usually hires an *independent* organization to assess the MSA acceptability and to independently make an accreditation recommendation. For new MSA developments, the independent organization is called the Verification, Validation, and Accreditation (VV&A) agent who participates in all verification and validation (V&V) activities throughout the entire MSA development life cycle (Balci *et al.* 2000).

The VV&A agent is responsible for creating a VV&A plan (VVAP) at the beginning of the MSA development life cycle. The IEEE Standard 1059 states that "the purpose of planning and plan documentation is to employ the V&V resources efficiently, to monitor and control the V&V process, and to allow the identification of each participant's role and responsibility" (IEEE 1993, p. 5).

The remainder of this paper is organized as follows. We provide guidance in organizing a VVAP in Section 2. Section 3 describes the acceptability assessment phases throughout the MSA development life cycle. Section 4 explains the generation of an accreditation report. Section 5 discusses the accreditation recommendation. Concluding remarks are given in Section 6.

## 2 ORGANIZATION OF THE VV&A PLAN

A VVAP organization is presented in this section to be in compliance with the IEEE Standard 1059 (IEEE 1993) and to provide additional information specific to MSA VV&A.

*Cover or Title Page:* This page typically includes VVAP title, version number and date, the VV&A agent identification, and the MSA sponsor identification.

*Signature Lines or Page:* Signatures of the persons responsible for writing and approving the VVAP are required in many cases.

*Revision History:* This page typically includes the description, date, and version number for each revision.

*Executive Summary or Preface or Foreword:* This includes a concise description of the VVAP.

*Table of Contents:* This should be provided with hyperlinks in the electronic version of the document.

*List of Figures, Tables, or Illustrations:* These items are listed separately. They should be provided with hyperlinks in the electronic version of the document.

*List of Acronyms or Glossary:* This section defines all project-specific terms and acronyms.

*References:* This section lists documents cited in the body of the VVAP, binding compliance documents, and any related documents required to supplement or implement the VVAP.

1. *Introduction:* This section provides introductory information about the M&S project and the VVAP.

1.1 *Background:* This section concisely describes the program for which the MSA is being developed.

1.2 *Organizational Responsibilities:* This section describes the responsibilities of all organizations involved in the development of the MSA including the MSA sponsor, MSA developer, and the independent VV&A agent.

1.3 *Purpose of this Document:* This section describes the purpose and scope of the VVAP.

1.4 *Overview of this Document:* This section provides a summary of the VVAP section by section.

2. *MSA Overview:* This section describes the MSA and defines the domains of applicability, intended uses, and acceptability criteria for the MSA.

2.1 *MSA Description:* This section typically describes the MSA development approach and environment, MSA overall architecture, and MSA components.

2.2 *MSA Domains of Applicability:* This section defines the domains for which the MSA can be used. For example, an MSA for simulating the National Missile Defense (NMD) system design may have the following domains of applicability:

- NMD system performance assessment
- NMD ground and flight test prediction, planning, and design
- NMD system integration support
- NMD deployment readiness review (DRR) support
- Operational Test Agency (OTA) analysis

2.3 *MSA Intended Uses:* The intended uses must be well defined for the accreditation to be conducted (Balci and Ormsby 2000). For the above example MSA, the following may be the intended uses in the domain of applicability of NMD DRR support:

- Estimation of the Probability of Integrated System Effectiveness under a given scenario
- Estimation of the Probability of Kill Single Shot under a given scenario
- Estimation of the Health and Status Reporting Time under a given scenario
- Assessment of the Situational Awareness under a given scenario

2.4 *MSA Acceptability Criteria:* The DoD Instruction 5000.61 defines Acceptability Criteria as "a set of standards that a particular model, simulation, or federation of models and simulations must meet to be accredited for a specific purpose. (a.k.a. accreditation criteria)" (DoDI 1996). For the above example MSA, the following may be specified as the acceptability criteria:

- The MSA must be developed based on sufficiently credible requirements.
- The MSA must demonstrate sufficient
  a. conceptual model credibility,
  b. design credibility,
  c. implementation credibility,
  d. integration credibility,
  e. data credibility,
  f. configuration management quality,
  g. overall product quality, and
  h. application documentation quality.

  The sufficient credibility or sufficient quality is determined based on the established requirements and the intended uses within the identified domains of applicability.
- The MSA experimentations conducted for each intended use must be sufficiently credible.
- The MSA must be developed
  a. under high quality project management,
  b. by following standard software engineering practices, and
  c. at Level 3 of the Capability Maturity Model (CMM) as defined by the Carnegie Mellon University Software Engineering Institute (CMU SEI) (CMU SEI 1994).
- When used for critical simulation-based acquisition decision-making, the MSA must demonstrate sufficient credibility that minimizes the risk of making the wrong (possibly catastrophic) acquisition decision.
- The MSA must be
  a. delivered within agreed-upon budgetary constraints, and
  b. cost-effective for its use, maintenance, technical support, and training.

3. *VV&A Overview:* This section describes the VV&A agent, schedule, VV&A resources, VV&A responsibilities, and VV&A methodologies, techniques, and tools.

3.1 *VV&A Agent:* This section describes the independent VV&A agent.

3.2 *Master Schedule:* This section describes the scheduling of VV&A tasks throughout the project life cycle by providing milestones including completion dates and deliverables. It identifies the VV&A tasks' relationships within the overall project environment. The schedule typically describes how V&V results provide feedback to the development process to support project management functions (IEEE 1993).

3.3 *Resource Summary:* This section provides an overview of the resources required to perform the VV&A tasks, including personnel, facilities, tools, finances, and special procedural requirements such as security, access rights, or documentation control (IEEE 1993).

3.4 *Responsibilities:* This section describes the specific responsibilities of each organizational element

assigned to performing a specific VV&A task throughout the project life cycle (IEEE 1993).

*3.5 Methodologies, Techniques, and Tools:* This section identifies the methodologies, techniques, and software tools employed for the VV&A effort. The purpose and use of each should also be described (IEEE 1993). For example, we employ the accreditation methodology described in (Balci 2000), some of the V&V techniques presented in (Balci 1998), and the software tool Evaluation Environment™ (Orca 1999).

4. *Acceptability Assessment Phases:* This section describes the VV&A activities to be conducted throughout the MSA development life cycle. It makes up the core of the VVAP and is separately presented in Section 3.
5. *Accreditation Report:* This section is separately presented in Section 4.
6. *Accreditation Recommendation:* This section is separately presented in Section 5.

*Appendices:* These sections can be used to include additional supporting materials.

## 3 ACCEPTABILITY ASSESSMENT PHASES

Two of the strategic directions in VV&A research and practice advocated by Balci *et al.* (2000) are:

- VV&A should be expanded from accuracy-centered assessment to quality-centered assessment.
- VV&A should be expanded from product-centered assessment to (product/process/project)-centered assessment.

The IEEE Standard 1059 indicates that "Software verification and validation employs review, analysis, and testing techniques to determine whether a software system and its intermediate products comply with requirements. These requirements include both functional capabilities and quality attributes" (IEEE 1993, p. 4). The IEEE Standard 1059 includes quality assessment within the V&V activities by listing 19 quality attributes (also called quality characteristics) including efficiency, interoperability, maintainability, reliability, reusability, testability, and usability.

The VV&A activities, including the assessment of the quality characteristics, (i) involve the measurement and evaluation of hundreds of qualitative and quantitative elements, (ii) mandate subject matter expert (SME) evaluation, and (iii) require the integration of different evaluations. Planning and managing such measurements and evaluations require a unifying methodology and cannot be performed in an *ad hoc* manner. We use such a methodology described in (Balci 2000). The methodology consists of the following body of methods, rules, and postulates: (a) employment of SMEs, (b) construction of a hierarchy of indicators, (c) relative criticality weighting of indicators and SMEs using the analytic hierarchy process, (d) using a rule-based expert knowledge base, (e) assignment of crisp, fuzzy, and nominal scores for the indicators, (f) aggregation of the indicator scores, (g) graphical representation of the indicator scores and weights, (h) hypertext accreditation report, and (i) interpretation of the results.

An *indicator* is an indirect measure of a qualitative concept (e.g., M&S design quality) or a direct measure of a quantitative concept (e.g., utilization). Indicators can be specified as metrics, measures, indexes, factors, or assessment questions such as the ones presented in (Carr and Balci 2000).

Our overall strategy for creating indicators for MSA acceptability assessment dictates the assessment of the (a) quality of the product, (b) quality of the process used in creating the product, (c) quality of the MSA project management, and (d) quality of the MSA documentation that describes the product, process, and MSA developer's quality assurance of the product and process. The strategy is summarized in Table 1.

Table 1: Strategy for Creating Indicators for MSA Acceptability Assessment

| | | | |
|---|---|---|---|
| Credibility | Product Quality | Accuracy | Verity |
| | | | Validity |
| | | Quality Characteristic 2<br>Quality Characteristic 3<br>:<br>Quality Characteristic n | … |
| | Process Quality | Quality of Approach Used: | … |
| | Project Quality | Quality of Project Management: | … |
| | Documentation Quality | Product Documentation Quality | … |
| | | Process Documentation Quality | … |
| | | Quality Assurance Documentation Quality | … |

Table 1 shows that the first product quality characteristic "product accuracy" is assessed by evaluating product verity and validity. Product verity is evaluated by conducting product verification and product validity is evaluated by conducting product validation.

- *Product verification* deals with the transformational accuracy of that product and addresses the question of "Are we building the product *right*?"
- *Product validation* deals with the representational or behavioral accuracy of that product and

addresses the question of "Are we building the *right* product?"

We refer to product verification and product validation as simply V&V throughout the MSA development life cycle.

More than 100 V&V techniques are available (Balci 1998, Binder 2000). The V&V techniques should be selected and used under the guidelines provided by the V&V principles (Balci 1997).

Other product quality characteristics change from one M&S project to another and are determined based on the MSA requirements and intended uses.

Process quality is assessed depending on the process methodologies and techniques employed by the MSA developer. CMU SEI (1994) has developed the CMM as an application of the process management concepts of total quality management to software. CMM is now very commonly used in the software industry as a means of judging software development process maturity and quality.

Project quality is assessed by evaluating the MSA management characteristics including personnel, resources, planning, and control.

MSA acceptability assessment is conducted to formulate an accreditation recommendation. The top-level indicators for assessing MSA acceptability are given as:

- MSA Requirements Credibility
- MSA Application Credibility
- MSA Experimentations Credibility
- MSA Project Management Quality
- MSA Cost
- MSA Risk

Each is further decomposed into a hierarchy of indicators. The hierarchy that we have developed currently has more than 400 indicators maintained in the Evaluation Environment software tool (Orca 1999).

### 3.1 MSA Requirements Credibility

The VVAP should provide the scheduling and description of activities and tasks for assessing the following indicators so as to assess the credibility of MSA requirements.

- *MSA Requirements Quality* is the degree to which the requirements possess a desired set of characteristics such as:
  - *MSA Requirements Accuracy:* is the degree to which the requirements possess sufficient transformational and representational correctness.
  - *MSA Requirements Clarity:* is the degree to which the MSA requirements are unambiguous and understandable.
  - *MSA Requirements Completeness:* is the degree to which all parts of a requirement are specified with no missing information, i.e., each requirement is self-contained. For example, "radar search pulse rate must be 10" is an incomplete requirement because it is missing the "per second" part. The requirement "missile kill assessment delay must follow the Uniform probability distribution" is incomplete because it is missing the range parameter values. Also use of the placeholder "TBD" (to be determined or to be defined), "TBR" (To be resolved), "TBP" (To be provided), and use of the phrases such as "as a minimum", "as a maximum", or "not limited to" are indications of incomplete requirements specification.
  - *MSA Requirements Consistency:* is the degree to which (a) the requirements are specified using uniform notation, terminology, and symbology, and (b) any one requirement does not conflict with any other.
  - *MSA Requirements Feasibility:* is the degree of difficulty of (a) implementing a single requirement, and (b) simultaneously meeting conflicting requirements. Sometimes requirements conflict with each other. It may be possible to achieve a requirement by itself, but it may not be possible to achieve a number of them simultaneously.
  - *MSA Requirements Modifiability:* is the degree to which the requirements can easily be changed.
  - *MSA Requirements Stability:* is (a) the degree to which the requirements are changing while the MSA is under development, and (b) the possible effects of the changing requirements on the project schedule, cost, risk, quality, functionality, design, integration, and testing of the MSA.
  - *MSA Requirements Testability:* is the degree to which the requirements can easily be tested. A testable requirement is the one that is specified in such a way that pass/fail or assessment criteria can be derived from its specification. For example, the following requirement specification is not testable: "The probability of kill should be estimated based on the simulation output data." The following requirement specification is testable: "The probability of kill should be estimated by using a 95% confidence interval based on the simulation output data."
  - *MSA Requirements Traceability:* is the degree to which the requirements related to a

particular requirement can easily be found. Requirements should be specified in such a way that related requirements are cross-referenced. When it is necessary to change a requirement, it should be easy to identify those requirements to be affected by the change by using the cross-references.

- *MSA Requirements Engineering Process Quality:* This indicator is decomposed further into a hierarchy of indicators depending on the characteristics of the requirements engineering methodologies and techniques employed by the MSA developer.
- *MSA Requirements Documentation Quality:* This indicator is assessed in terms of the following indicators:
  a. MSA Requirements Specification Document Quality (product documentation quality)
  b. MSA Requirements Engineering Process Documentation Quality (process documentation quality)
  c. MSA Requirements Quality Assessment Report Quality (quality assurance documentation quality)

Document (or documentation or report) quality is the degree to which the document possesses a desired set of characteristics. The quality of a document is mostly determined by the quality of its content; however, other quality characteristics are also important. We have developed a hierarchy of more than 80 indicators to assess the document quality other than the quality of the document's content, which should be assessed in another VV&A activity. The top-level indicators of this hierarchy are given as follows:

- *Accessibility:* is the degree to which the document enables its users to easily locate its components in a non-sequential manner. The accessibility is usually provided by table of contents, list of figures, list of tables, chapters, sections, hypertext links, citations of references, glossary of terms, list of acronyms, references, index, and appendices. All of the above access or navigation aids can be made clickable in an electronic document (e.g., Microsoft Word, PDF, HTML) to enable the user to jump to the desired location at the click of the mouse button.
- *Accuracy:* is the degree to which the document possesses sufficient transformational and representational correctness.
- *Completeness:* is the degree to which (a) all components of the document are specified with no missing information, (b) the document covers all required aspects of the entity it is intended to document, and (c) the document contains all of the items required by the documentation standard in effect.
- *Consistency:* is the degree to which the document uses uniform page layout, notation, terminology, symbology, and format.
- *Clarity:* is the degree to which the document is unambiguous and understandable. *Document unambiguity* is the degree to which each statement, expression, or definition of the document can only be interpreted one way. *Document understandability* is the degree to which the meaning of each statement, expression, or definition of the document is easily comprehended by all of its readers.
- *Maintainability:* is the degree to which the document facilitates updates and changes.
- *Portability:* is a quality characteristic of electronic documents and refers to the degree to which the document can be easily transferred from one computer platform to another. Today, if a document is provided in Adobe's Portable Document Format (PDF) or in Hypertext Markup Language (HTML), the document is considered to be very highly portable. Portability of documents in other formats such as postscript, Microsoft Word, and Microsoft Rich Text Format (RTF) should be judged according to the format's availability to the intended user community. The portability quality characteristic may conflict with others such as readability, accessibility, and maintainability. For example, the most portable document format is plain text that has the worst readability characteristic. The PDF is very highly portable but it is not as maintainable as the Microsoft Word format. Therefore, the PDF is commonly used for document distribution while the document is maintained in its original format (e.g., Microsoft Word, LaTeX, WordPerfect).
- *Readability:* is the degree to which the document can be read easily.

We use the hierarchy of more than 80 indicators under the top-level indicators described above for the quality assessment of each document, documentation or report mentioned in this paper for the VVAP. This hierarchy can also be used for assessing the quality of a VVAP.

### 3.2 MSA Application Credibility

The VVAP should provide the scheduling and description of activities and tasks for assessing the following indicators so as to assess the credibility of MSA as an application:

- MSA Conceptual Model Credibility
- MSA Design Credibility
- MSA Implementation Credibility
- MSA Integration Credibility
- MSA Data Credibility
- MSA Configuration Management Quality

- MSA Overall Product Quality
- MSA Application Documentation Quality

### 3.2.1 MSA Conceptual Model Credibility

This indicator is assessed in terms of the following indicators, each of which is further decomposed into a hierarchy of indicators.

- MSA Conceptual Model Quality
- MSA Conceptual Model Development Process Quality
- MSA Conceptual Model Documentation Quality: is assessed in terms of the following indicators: (a) MSA Conceptual Model Specification Document Quality, (b) MSA Conceptual Model Development Process Documentation Quality, and (c) MSA Conceptual Model Quality Assessment Report Quality.

### 3.2.2 MSA Design Credibility

This indicator is assessed in terms of the following indicators, each of which is further decomposed into a hierarchy of indicators.

- MSA Design Quality
- MSA Design Process Quality
- MSA Design Documentation Quality: is assessed in terms of the following indicators: (a) MSA Design Specification Document Quality, (b) MSA Design Process Documentation Quality, and (c) MSA Design Quality Assessment Report Quality.

### 3.2.3 MSA Implementation Credibility

This indicator is assessed in terms of the following indicators, each of which is further decomposed into a hierarchy of indicators.

- MSA Implementation Quality
- MSA Implementation Process Quality
- MSA Implementation Documentation Quality: is assessed in terms of the following indicators: (a) MSA Implementation Specification Document Quality, (b) MSA Implementation Process Documentation Quality, and (c) MSA Implementation Quality Assessment Report Quality.

### 3.2.4 MSA Integration Credibility

This indicator is assessed in terms of the following indicators, each of which is further decomposed into a hierarchy of indicators.

- MSA Integration Quality
- MSA Integration Process Quality
- MSA Integration Documentation Quality: is assessed in terms of the following indicators: (a) MSA Integration Specification Document Quality, (b) MSA Integration Process Documentation Quality, and (c) MSA Integration Quality Assessment Report Quality.

### 3.2.5 MSA Data Credibility

The VVAP should provide the scheduling and description of activities and tasks for assessing the following indicators so as to assess the credibility of MSA data.

- *MSA data quality* is the degree to which the MSA data possess a desired set of characteristics such as:
  - *Data Accessibility:* is the degree to which data are available or easily and quickly retrievable. For example, the MSA user may be required to specify a set of input data that is classified and unavailable or very time consuming to obtain.
  - *Data Accuracy:* is the degree to which data possess sufficient transformational and representational correctness.
  - *Data Clarity:* is the degree to which data are unambiguous and understandable.
  - *Data Completeness:* is the degree to which all parts of the data are specified with no missing information, i.e., each data specification is self-contained. For example:
    - "Radar frequency is 150" is an incomplete data specification since no unit (i.e., Hertz) is given,
    - "Threat detection times are exponentially distributed" is an incomplete random data specification since no mean value is given, and
    - The trace data file may not contain all of the data required for the MSA to complete its run.

    Also use of the placeholder "TBD" (to be determined or to be defined), "TBR" (To be resolved), "TBP" (To be provided), and use of the phrases such as "as a minimum", "as a maximum", or "not limited to" are indications of incomplete data specification.
  - *Data Consistency:* is the degree to which (a) data are specified using consistent measurement unit, uniform notation, and uniform terminology, and (b) any one data value does not conflict with any other. Commonly, the simulation clock is assumed to use only one time unit (e.g., seconds) and all time values are expressed without a unit designation as numeric time values assumed to be in the time unit of the clock. All numeric values designating simulation time must be expressed consistently in the same time unit of the simulation clock. In general, once a measurement

unit (e.g., seconds, meters, pounds) is selected for a particular data set, that unit must be used consistently throughout the entire simulation model. Notation (e.g., $S^2$ representing sample variance) and terminology (e.g., "mean" as opposed to "scale") should be used uniformly.
  - *Data Currency:* is the degree to which the age of the data is appropriate for the use of the data in the MSA.
  - *Data Precision:* is the degree to which data possess sufficient number of significant digits in their numerical values.
  - *Data Relevance:* is the degree to which data are applicable for use in the MSA.
  - *Data Resolution:* is the degree to which data possess sufficient level of detail.
  - *Data Reputation:* is the degree to which data are trusted or highly regarded in terms of their source or origin.
  - *Data Traceability:* is the degree to which data are easily attributed to a source.
- MSA Data Collection Process Quality
- MSA Data Documentation Quality: is assessed in terms of the following indicators: (a) MSA Data Specification Document Quality, (b) MSA Data Collection Process Documentation Quality, and (c) MSA Data Quality Assessment Report Quality.

#### 3.2.6 MSA Configuration Management Quality

Configuration Management (CM) is the process that controls the changes made to the MSA and manages the different versions and releases of the evolving MSA. The CM quality is assessed with respect to the CM plan.

#### 3.2.7 MSA Overall Product Quality

The desirable quality characteristics of an MSA change from one M&S project to another and are determined based on the MSA requirements and intended uses. The following list enumerates quality characteristics that may be specified: accuracy, adaptability, interoperability, maintainability, performance, portability, reliability, reusability, security, testability, traceability, and usability.

#### 3.2.8 MSA Application Documentation Quality

This indicator is assessed in terms of the indicators assessing the quality of all documents, documentations, and reports created for the entire MSA. This is an overall assessment indicator to show how well the documentation is carried out throughout the MSA development life cycle.

### 3.3 MSA Experimentations Credibility

The VVAP should provide the scheduling and description of activities and tasks for assessing the following indicators so as to assess the credibility of MSA experimentations.

- MSA Experimentations Quality: This indicator is further decomposed into the following indicators:
  - MSA Experiment Accuracy
  - MSA Experimental Scenarios Quality
  - MSA Experimental Results Clarity
- MSA Experimentation Process Quality
- MSA Experimentations Documentation Quality

### 3.4 MSA Project Management Quality

The VVAP should provide the scheduling and description of activities and tasks for assessing the following indicators so as to assess the quality of MSA project management.

- MSA Software Quality Assurance Program Quality
- MSA Development Plan Quality
- MSA CM Plan Quality: Based on the IEEE Standards 828 and 1042 (IEEE 1998, IEEE 1987), we have developed a hierarchy of more than 60 indicators for assessing the quality of an MSA CM plan.
- MSA Project Personnel Quality
- MSA Project Resources Quality

### 3.5 MSA Cost

Acceptability of an MSA may be denied because of the high cost. Cost is certainly a major issue in the acquisition of some MSAs especially those used for training purposes. The following types of cost are commonly considered:

- Cost of Use: including the cost of (a) acquiring the MSA, (b) hardware required to run the MSA, and (c) personnel required to operate the MSA.
- Cost of Maintenance
- Cost of Technical Support
- Cost of Training

### 3.6 MSA Risk

The VVAP should provide the scheduling and description of activities and tasks for assessing the following indicators so as to assess the MSA risk.

- *Product Risk:* is the probability that the characteristics of the MSA will be unsatisfactory.
  - *Acceptance Risk:* is the probability that the MSA will not pass the acceptance test with respect to the acceptance criteria given in the requirements specification document.

- *Integration Risk:* is the probability that the MSA components will not be successfully integrated.
- *Performance Risk:* is the probability that the MSA will not be capable of performing properly when it is experimented with to obtain results required for a particular intended use.
- *Reliability Risk:* is the probability that the MSA will crash during experimentation to obtain results required for a particular intended use.
- *Reproducibility Risk:* is the probability that the MSA cannot be reproduced for distribution to others. For example, if the MSA relies on a piece of specialized hardware to run and the specialized hardware is no longer in production, then the MSA may not be replicated at a reasonable cost.
- *Supportability Risk:* is the probability that the MSA cannot be properly maintained after its delivery. For example, there may be no technical support, no support for correcting bugs, and no support for making improvements and upgrades.
- *Utility Risk:* is the probability that the MSA will be less useful than required by the MSA sponsor.

❑ *Resource Risk:* is the probability that the acquisition of the MSA will exceed the allocated resources such as budget and delivery date.

- *Cost Risk:* is the probability that the acquisition cost of the MSA will exceed the budgeted amount.
- *Schedule Risk:* is the probability that the MSA will not be delivered by the required deadline.

## 4 ACCREDITATION REPORT

The VVAP should provide a description of the accreditation report to be prepared and submitted to the MSA sponsor. For example, using the Evaluation Environment tool, we generate the accreditation report as a hypertext document in HTML and RTF (Orca 1999). The report includes

- VV&A project documentation
- Information about the SMEs employed
- Hierarchical list of indicators
- Alphabetical list of indicators
- Leaf indicators report
- Kiviat graphs of
  - Aggregate scores for an indicator's children
  - Weights for an indicator's children
  - SME scores for an indicator
  - SME weights for an indicator\

The report can instantly be published on the World Wide Web (web) and viewed by all people involved in the MSA project using a web browser. When publishing the report on the web, the identity of each SME can be hidden.

Figure 1 shows an example Kiviat graph displaying the interval scores for 47 child indicators. A legend names each indicator represented as a number in the graph. Figure 2 shows the Kiviat graph representation of the criticality weighting of 11 child indicators. The numbers in the graph represent the fractional weights that sum to 1.

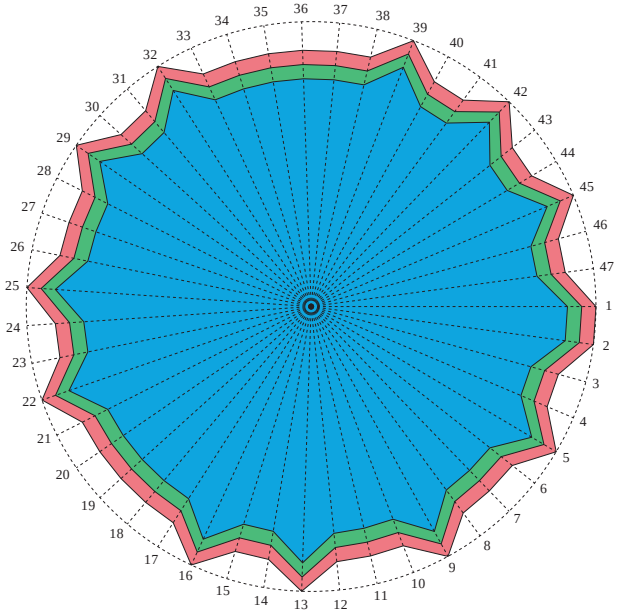


Figure 1: Example Kiviat Graph Showing the Interval Scores for 47 Child Indicators

All MSA acceptability assessment results are reported in hypertext format including graphical representations as shown in Figure 1 and Figure 2. The hypertext capability and the graphical representations facilitate the comprehension and interpretation of the complex assessment results. Based on the analysis of the assessment results, an accreditation recommendation is formulated.

## 5 ACCREDITATION RECOMMENDATION

Having a quality VVAP is a necessary, but not a sufficient condition. The quality of the execution of the VVAP is also very important. The VVAP quality and the execution quality jointly affect the confidence level at which an accreditation decision is formulated as depicted in
Figure **3**.

The impact of the quality on the confidence level is situation-dependent and changes from one M&S project to another as shown by the different curves with shape parameters $\alpha_i$, i=1,2,3,4 in
Figure **3**.

Comprehensiveness is one of the most important quality characteristics of a VVAP. By analogy, if an MSA corresponds to a "forest", its VVAP describes how the

"forest" will be evaluated. If a VVAP is structured to evaluate only, for example, 30% of the "trees" and their

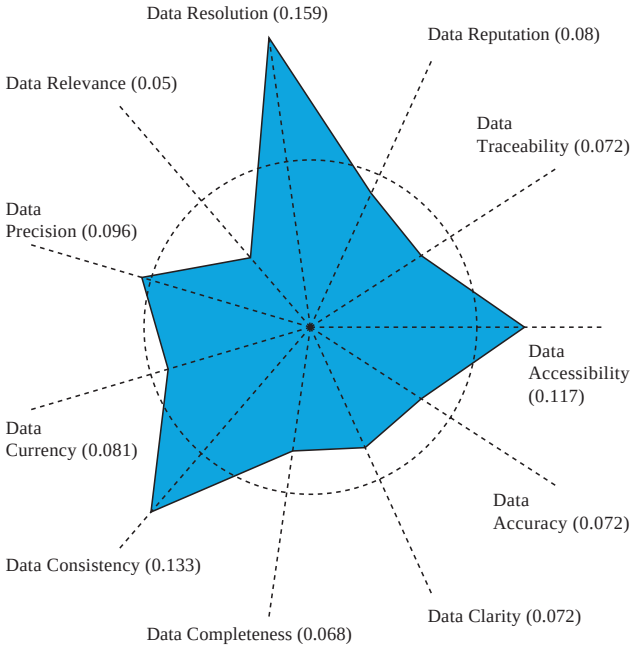


Figure 2: Example Kiviat Graph Showing the Criticality Weighting of 11 Child Indicators

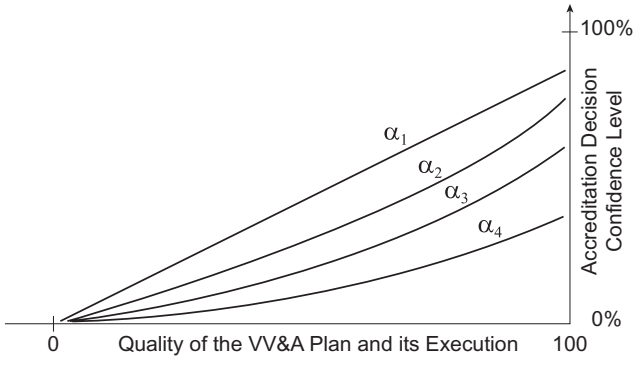


Figure 3: Quality versus Confidence Level

"branches", the accreditation decision can only be based on 30% of the "full picture" resulting in a low confidence level.

Sometimes, it is not possible to execute a VVAP completely due to factors such as schedule delays, loss of resources, changing requirements, and development refocus. In this case, the VVAP should show how much VV&A is conducted with respect to the comprehensive set of requirements specified in the VVAP. The amount of coverage should be taken into consideration in determining the confidence level at which the accreditation decision is recommended to the MSA sponsor by the VV&A agent.

Under no circumstance, should the VV&A agent develop and use an incomplete VVAP. If the VV&A agent develops and uses an incomplete VVAP and if the MSA fails the accreditation, the M&S developer may direct blame to the VV&A agent and attribute the failure to the incomplete VVAP.

In writing the accreditation recommendation letter, we should remember that accreditation decision is not a binary decision, where 1 implies "recommend" and 0 implies "not recommend." Accreditation should be recommended on a level of confidence ranging from 0% to 100%. However, it is usually very difficult to narrowly define the confidence level. Nevertheless, nominal characterizations such as "very high", "high", "average", "low", or "very low" can be used. For example, the following is a binary recommendation with no level of confidence specified: "We recommend that the M&S application … be accredited for the stated intended uses." The recommendation should be given by specifying a confidence level: "We recommend that the M&S application … be accredited for the stated intended uses with high confidence." Sometimes, the confidence level can also be expressed by specifying a number of caveats and the recommendation can be made as: "We recommend that the M&S application … be accredited for the stated intended uses by noting the caveats."

Two types of errors can be committed in making an accreditation recommendation:

- *Type I Error* is the error of rejecting the accreditation of an MSA when in fact it should have been accredited.
- *Type II Error* is the error of accrediting an MSA when in fact the accreditation should have been rejected.

The probability of committing Type I Error is called the *M&S Developer's Risk*. The probability of committing Type II Error is called *M&S User's Risk*. Committing Type I Error may result in further improvements of the MSA, which unnecessarily increases the development cost. On the other hand, the consequences of committing Type II Error can be catastrophic, e.g., resulting in the simulation-based acquisition of an unacceptable system. Therefore, the VVAP must be created and executed to significantly reduce the M&S User's Risk.

## 6 CONCLUDING REMARKS

One of the V&V principles dictates that V&V activities must be planned and documented throughout the entire MSA development life cycle (Balci 1997).

Planning for VV&A should begin early in the project so that the scope of the VV&A effort can be assessed. This assessment is necessary for assuring that the required VV&A resources are provided in the overall project planning. The initial VVAP may provide insights to the MSA sponsor concerning MSA development plans. (IEEE 1993)

Planning for VV&A should be closely coupled to the planning of the rest of the MSA project. The VVAP should

clarify how VV&A fits into the overall project life cycle and relates to all project entities (e.g., sponsor, developer, developer's subcontractors, developer's software quality assurance department, V&V agent, VV&A agent). Identification of some VV&A tasks may depend on how the project is organized. (IEEE 1993)

A VV&A effort is typically applied in parallel with the MSA development activities. Some VV&A tasks may be interleaved with the development process. A VV&A effort consists of management tasks (e.g., planning, organizing, and monitoring the VV&A effort) and technical tasks (e.g., analyzing, evaluating, reviewing, and testing the MSA development processes and products) to provide information about the engineering, quality, and status of MSA products throughout the life cycle. (IEEE 1993)

## ACKNOWLEDGMENT

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

## REFERENCES

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

Balci, O. 1998. Verification, validation, and testing. In *The Handbook of Simulation*, ed. J. Banks, 335-393. John Wiley & Sons, New York, NY.

Balci, O. 2000. A methodology for certification of modeling and simulation applications. NMD Project Technical Report, Orca Computer, Inc., Blacksburg, VA, Aug.

Balci, O., R.E. Nance, J.D. Arthur, and W.F. Ormsby. 2000. Expanding our horizons in verification, validation, and accreditation research and practice. NMD Project Technical Report, Orca Computer, Inc., Blacksburg, VA, June.

Balci, O. and W.F. Ormsby. 2000. Well-defined intended uses: an explicit requirement for accreditation of modeling and simulation applications. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J.A. Joines, R.R. Barton, P. Fishwick, and K. Kang. IEEE, Piscataway, NJ.

Binder, R.V. 2000. *Testing object-oriented systems: models, patterns, and tools*. Addison-Wesley, Reading, MA, 1191 pp.

Carr, J.T., III and O. Balci. 2000. Verification and validation of object-oriented artifacts throughout the simulation model development life cycle. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J.A. Joines, R.R. Barton, P. Fishwick, and K. Kang. IEEE, Piscataway, NJ.

CMU SEI 1994. *The capability maturity model: guidelines for improving the software process*. Addison-Wesley, Reading, MA.

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

IEEE 1987. *IEEE guide to software configuration management*. IEEE Standard 1042-1987, IEEE, Piscataway, NJ.

IEEE 1993. *IEEE guide for software verification and validation plans.* IEEE Standard 1059-1993. IEEE, Piscataway, NJ.

IEEE 1998. *IEEE standard for software configuration management plans*. IEEE Standard 828-1998, IEEE, Piscataway, NJ.

Orca 1999. *Evaluation environment user's guide*. Orca Computer, Inc., Blacksburg, VA.

## AUTHOR BIOGRAPHIES

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

**WILLIAM F. ORMSBY** is a Project/Program Manager for the Naval Surface Warfare Center Dahlgren Division (NSWCDD). He is currently the technical lead for the Ballistic Missile Defense Organization (BMDO) National Missile Defense (NMD) Independent Verification and Validation (IV&V) Program. In this capacity, Mr. Ormsby serves as Co-Chair of the NMD M&S VV&A Subgroup, which is chartered to ensure M&S VV&A processes are in place and executed to increase the confidence and reduce risk in the NMD system acquisition. He has produced software/hardware systems since 1982 when he received his M.S. degree in Engineering from Virginia Tech. He has

developed concepts, written requirements, designed, implemented and delivered products for space systems supporting U.S. Navy mission objectives. Prior to his NMD assignment he led the development of models and simulations used to verify tactical guidance software performance for the TRIDENT II missile system. His e-mail address is <OrmsbyWF@nswc.navy.mil>

**JOHN T. CARR, III** is a Scientist at the Naval Surface Warfare Center Dahlgren Division (NSWCDD). He received his M.S. degree from Virginia Tech in 1985. From 1973 to 1995, Mr. Carr worked on the research and development of space systems. His areas of expertise include space system simulation development, formulation of satellite orbit determination algorithms, satellite force model development, space system visualization, object-oriented software development, and M&S VV&A. Since 1996, Mr. Carr has worked on missile defense systems. Currently, he is the lead for M&S VV&A within the National Missile Defense VV&A Program at NSWCDD. His e-mail address is <CarrJT@nswc.navy.mil>.

**SAID D. SAADI** is the Deputy Program Manager for the Ballistic Missile Defense Organization (BMDO) National Missile Defense (NMD) Independent Verification and Validation (IV&V) program at the Naval Surface Warfare Center Dahlgren Division (NSWCDD). In this capacity, Mr. Saadi serves as the lead for the (a) NMD SSTB VV&A, (b) XBR and UEWR Sensor System representation assessment for the VV&A of the NMDSIM, SENTRY, and LIDS M&S applications, and (c) development of SAT and SFW. He joined NSWCDD in 1988 after receiving his B.S. degree in Electrical Engineering from West Virginia University Institute of Technology and has since completed graduate work in Radar Systems at Georgia Tech and Virginia Tech. Prior to joining the NMD IV&V and VV&A program, he served as the AEGIS SPY-1 Sensor System Subject Matter Expert. In that position, he led the development of representations of the AEGIS SPY-1 Radar in Track Initiation Probability/Probability of Detection (TRIPPOD) model, Extended Air Defense Simulation (EADSIM), and Extended Air Defense Test Bed (EADTB) for the Navy Area and Theater Wide studies. His e-mail address is <SaadiSD@nswc.navy.mil>.