Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

PROPOSED STANDARD PROCESSES FOR CERTIFICATION OF MODELING AND SIMULATION APPLICATIONS

Osman Balci

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

Said D. Saadi

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

ABSTRACT

Certification of modeling and simulation (M&S) applications is becoming more commonly practiced as M&S applications are used more and more for military training, analysis, complex system design evaluation, M&S-based acquisition, problem solving, and critical decision making. Certification is a very complex process, involves the measurement and evaluation of hundreds of qualitative and quantitative elements, mandates subject matter expert evaluation, and requires the integration of disparate measurements and evaluations. Planning, managing, and conducting such measurements and evaluations require structured standard processes and should not be performed in an *ad hoc* manner. The need for standard processes for unbiased, fair, cost effective, and consistent M&S certification is undeniable. This paper presents a scheme for practicing M&S accreditation and certification and proposes standard M&S certification processes.

1 INTRODUCTION

The U.S. Department of Defense (DoD) is the largest sponsor and user of M&S applications in the world. DoD uses many types of M&S applications (such as continuous, discrete-event, distributed, hardware-in-the-loop, software-in-the-loop, human-in-the-loop, Monte Carlo, parallel, and synthetic environments) for a variety of purposes including acquisition, analysis or training. DoD Instruction 5000.61 (DoDI 1996) 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."

Accreditation is defined in the DoD M&S community as "the official certification that a model, simulation, or federation of models and simulations is acceptable for use for a specific purpose" (DoDI 1996). On the other hand, the International Organization for Standardization (ISO) defines accreditation and certification as follows (Rae, Robert, and Hausen 1995):

- *Accreditation* is a "procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks."
- *Certification* is a "procedure by which a third party gives written assurance that a product, process or service conforms to specified characteristics."

The above ISO definitions conflict with the definitions commonly used by the DoD M&S community. To the best of our knowledge, all engineering disciplines, educational sector, and other areas use these terms as defined by ISO. We use the ISO terminology in this paper.

2 WHY STANDARD PROCESSES?

Certification is the independent award of a "Certificate", a "Seal of Approval" or a "Mark of Conformity" formally attesting that an M&S application fulfills specific quality criteria under a set of prescribed intended uses. The independent award is regarded by the M&S application sponsor as providing some form of guarantee of quality and credibility. Based on the guarantee, the sponsor decides to use the M&S results in making key decisions. The consequences of wrongly awarding a "Certificate", a "Seal of Approval" or a "Mark of Conformity" may be catastrophic. For example, based on M&S results hardware and tactical software maybe procured and fielded, which does meet the warfighter's needs or fails the mission.

Under the current practice today, a company is designated as an M&S certification agent without proper authorization that the company has the required processes and qualified personnel in place for successfully conducting certification. Certification is sometimes carried out in an *ad hoc* fashion with no discipline, methodology or framework. Such practice increases the probability that the award will be issued improperly and may cause the spon-

sor to make wrong decisions, acquire a defective system or incorrectly train military personnel.

Rae, Robert, and Hausen (1995) state that fairness, cost effectiveness, and reproducible results are prerequisites for a successful certification scheme. These prerequisites can only be met by having standard processes executed in an unbiased manner.

3 PROPOSED PRACTICE FOR ACCREDITATION AND CERTIFICATION

Similar to the manner accreditation and certification are carried out in engineering disciplines, educational sector, and other areas, we propose the comprehensive scheme shown in Figure 1.

We envision an accreditation authority at the national level. Example accreditation authorities include the United Kingdom Accreditation Service (http://www.ukas.com), Japan Accreditation Board for Conformity Assessment (http://www.jab.or.jp), and German Accreditation Council (http://www.dar.bam.de/indexe.html). We believe that the National Institute of Standards and Technology (NIST) can serve as the accreditation authority in the United States.

Under our proposed practice, those companies or organizations interested in serving as M&S certification agents apply to the accreditation authority. The accreditation authority examines the maturity of the applicant's standard certification processes and the qualifications of the key personnel who will execute the certification processes. Based on the examination results, the accreditation authority gives formal recognition that the applicant agent is competent to carry out the standard processes and provide certification which is unbiased, fair, cost effective, and consistent.

For example, many companies serve as ISO 9000 certification agents. These agents examine the processes of a company and certify that the company is compliant with the ISO 9000 standard. The ISO 9000 certification agents are accredited by an accreditation authority. A directory of ISO 9000 accreditation bodies is provided at (http://www.praxiom.com/accrediters.htm).

As the ISO definition indicates, certification must be conducted by a third party, where the first party refers to M&S application sponsor and the second party refers to M&S application developer. Of course, certification is meaningful when conducted in an independent manner. To achieve true independence, the IEEE Standard 1012 (IEEE 1998) requires technical, managerial, and financial independence as described below.

- *Technical Independence* implies that the certification agent determines, prioritizes, and schedules its own tasks and efforts.
- *Managerial Independence* implies that the certification agent reports to the M&S application sponsor independently of the developer organization.
- *Financial Independence* implies that the certification agent is allocated its own budget for the M&S certification and does not rely on the M&S development budget.

The accreditation authority establishes and publishes a set of criteria for accrediting M&S certification agents. The criteria should include:

- maturity of the applicant's standard certification processes,
- credentials of the key personnel who will execute the certification processes, and
- true independence of the agent.

The proposed practice is needed to provide the checks and balances required to minimize the M&S application sponsor's risks.

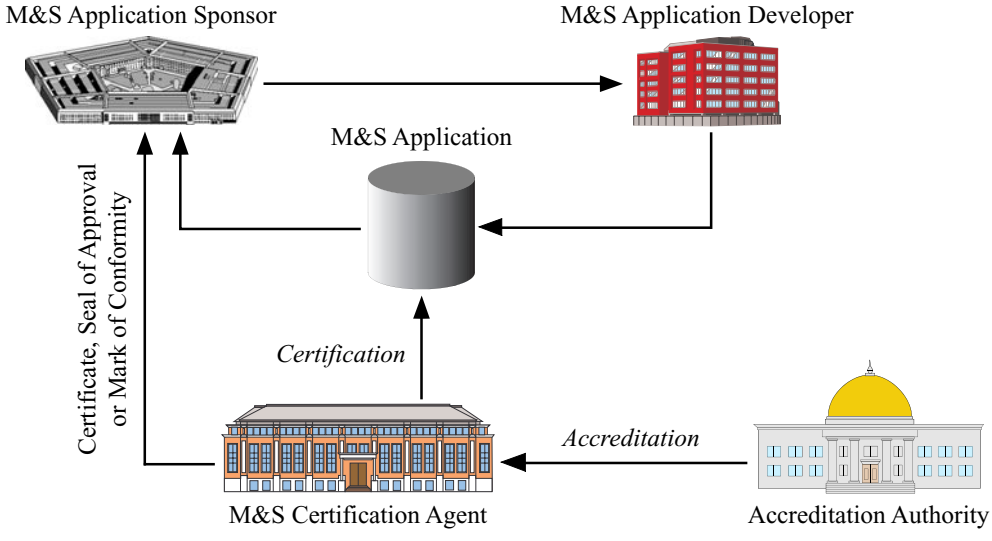

Figure 1: Proposed Practice for Accreditation and Certification

4 PROPOSED STANDARD CERTIFICATION PROCESSES

Certification may be conducted under the following three major scenarios:

1. Concurrent certification during the development life cycle of a new M&S application,
2. Certification of an already developed M&S application with some modifications, and
3. Certification of an already developed M&S application intended for use as is.

We present the proposed standard processes assuming that concurrent certification is carried out. Under concurrent certification, the M&S application sponsor hires a certification agent before the developer starts the development. This approach provides many advantages, including the following: (Balci 2001)

- The M&S application developer gets feedback for acceptability as the M&S development progresses throughout the life cycle.
- M&S errors and deficiencies are discovered early in the development life cycle resulting in significant cost savings.
- The complexity of M&S acceptability assessment is significantly reduced.
- The Probability of Type II Error (M&S Application User's Risk) is significantly reduced. Type II Error is the error of certifying an M&S application when in fact it should not have been certified.
- Communication between the independent certification agent and the M&S application developer helps improve the M&S quality.
- The M&S application developer is encouraged to implement an effective software quality assurance program.
- M&S product and resource risks are significantly reduced.

A standard evaluation process for certification of M&S applications

- is a disciplined process based on a sound and proven methodology,
- provides reproducible results,
- uses effective measurement and evaluation techniques for qualitative and quantitative elements,
- enables integration of disparate measurements and evaluations, and
- employs software tools to facilitate the measurement and evaluation, integration, and reporting.

Figure 2 shows the proposed standard evaluation processes for concurrent certification throughout the entire M&S development life cycle that we have developed. The life cycle consists of seven stages. A *stage* is defined to have an input work product (or artifact), an output work product (or artifact), and a process used to create the output product from the input product. A stage is conducted depending on the life cycle model employed. Many life cycle models exist including automation-based development model, exploratory development model, incremental development model, prototyping model, reuse-based development model, spiral model, and waterfall model.

We propose the seven standard evaluation processes depicted in Figure 2. Although an evaluation process is named after the output work product of the corresponding stage, the evaluation process must assess more than the output work product. A *standard evaluation process* is defined to measure and assess a particular life cycle stage's (a) output work *product*, (b) *process* used in creating the output work product, and (c) *project* characteristics (i.e., people, documentation, planning, quality assurance, capability maturity).

The three Ps (Product, Process, Project) of software engineering are commonly referred to for software measurement and certification. Voas (1999) presents a software quality certification triangle, which includes the three Ps as Product, Process, and Personnel, and advocates that certification can be approached from any one of these aspects, but a combination of all three will provide the best balance.

Product quality is the degree to which the product possesses a desired set of characteristics. 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* is substantiation of the transformational accuracy of the product and addresses the question of "Are we building the product right?"
- *Product validation* is substantiation of the representational or behavioral accuracy of the 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 development life cycle.

Process quality is the degree to which the process possesses a desired set of characteristics. The set of desired characteristics depends on the process methodologies and techniques employed by the M&S application developer.

Project quality is the degree to which the project possesses a desired set of characteristics. Project quality is assessed by evaluating a variety of characteristics including

configuration management, documentation quality, human resource management, personnel capability maturity, planning quality, and quality management.

Certification is a confidence building activity and can be best carried out if all three Ps are included. Concurrent certification, V&V, and quality assessment must be conducted in a manner integrated within the development life cycle. Therefore, we present the proposed standard processes for certification as integrated within the life cycle depicted in Figure 2.

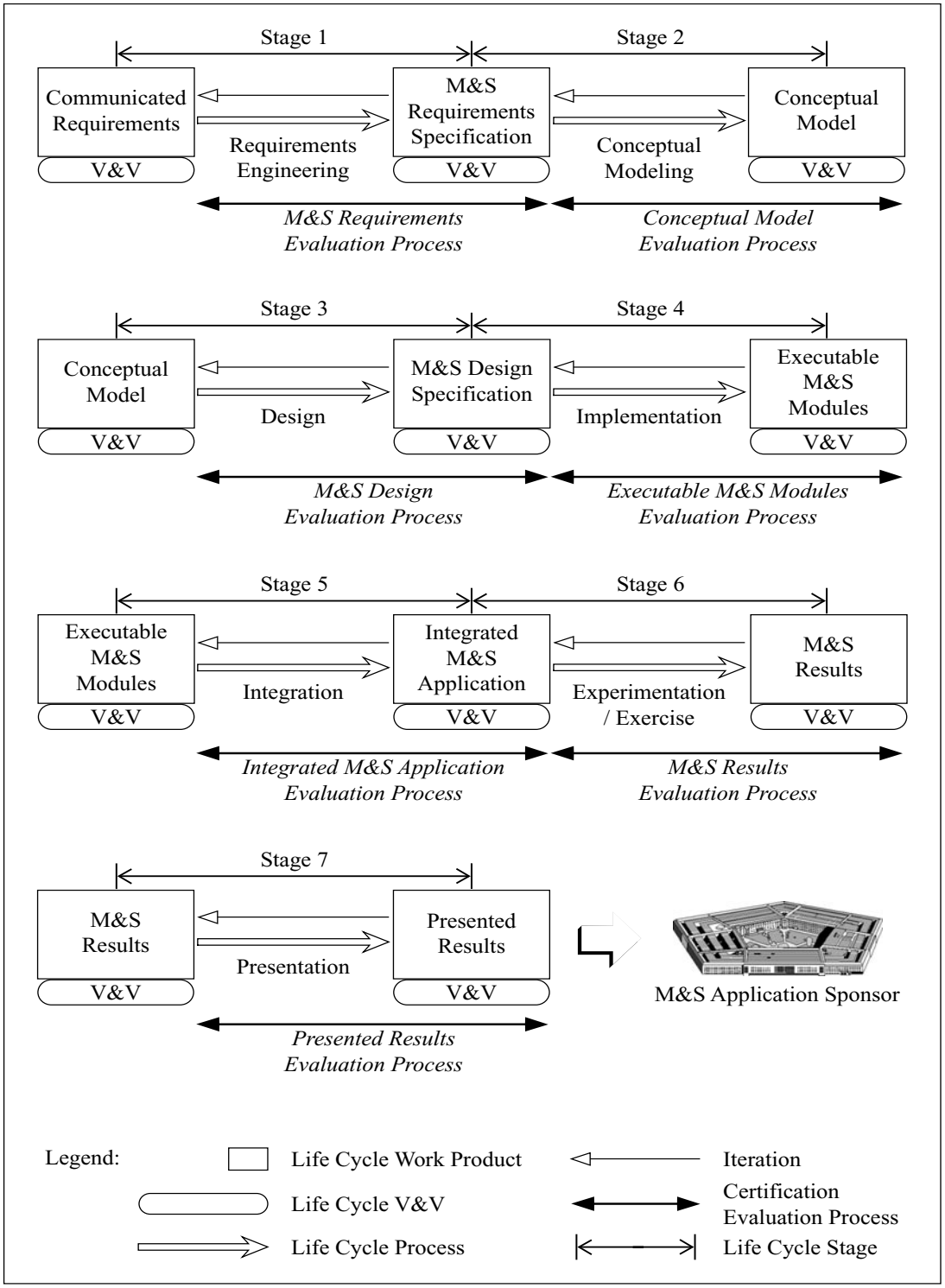

Figure 2: M&S Life Cycle and Proposed Standard Evaluation Processes for Certification

4.1 M&S Requirements Evaluation Process

This process evaluates the credibility of the M&S requirements created by conducting the requirements engineering process based on the communicated requirements. The process integrates the evaluations of: (a) M&S requirements quality, (b) requirements engineering process, and (c) M&S project characteristics related to life cycle stage 1. The following hierarchy of indicators can be used for assessing M&S requirements quality:

- Accuracy
  - Verity
  - Validity
- Clarity
  - Unambiguity
  - Understandability
- Completeness
- Consistency
- Feasibility
- Modifiability
- Stability
- Testability
- Traceability

4.2 Conceptual Model Evaluation Process

This process evaluates the credibility of the conceptual model created by conducting the conceptual modeling process based on the requirements specification document. The process integrates the evaluations of: (a) conceptual model quality, (b) conceptual modeling process, and (c) M&S project characteristics related to life cycle stage 2.

A *conceptual model* is a high-level architectural characterization of the M&S application based on the M&S requirements specification document. DMSO (2000) describes a conceptual model as consisting of (a) M&S context including functional description of the mission space, constraints, and bounds, (b) M&S concept including mission space representation and M&S space functionality, and (c) M&S elements including architectural relationships, assumptions, algorithms, and data.

4.3 M&S Design Evaluation Process

This process evaluates the credibility of the M&S design created by conducting the design process based on the conceptual model specification. The process integrates the evaluations of: (a) M&S design quality, (b) M&S design process, and (c) M&S project characteristics related to life cycle stage 3.

A reasonably large M&S application design is decomposed into modules to overcome the complexity of development and evaluation. If the M&S application is a federation of models, then each federated model is referred to as a module. If not, then each M&S component is referred to as a module.

Each M&S module design specification as well as integration and interoperability of the modules must be subjected to evaluation.

4.4 Executable M&S Modules Evaluation Process

This process evaluates the credibility of the executable M&S modules created by conducting the implementation process based on the M&S design specification. The process integrates the evaluations of: (a) executable M&S modules quality, (b) implementation process, and (c) M&S project characteristics related to life cycle stage 4.

An M&S module may be implemented by a team, group, or subcontractor. The implementation process creates executable modules, which can be evaluated by using dynamic testing techniques (Balci 1998).

4.5 Integrated M&S Application Evaluation Process

This process evaluates the credibility of the integrated M&S application created by conducting the integration process based on the executable M&S modules. The process integrates the evaluations of: (a) integrated M&S application quality, (b) integration process, and (c) M&S project characteristics related to life cycle stage 5.

One of the principles stated by Balci (1997) dictates that successfully testing each module does not imply overall M&S application credibility. Each module credibility is judged to be sufficient with some error that is acceptable with respect to the project objectives and M&S requirements. We may find each module to be sufficiently credible, but this does not imply that the overall M&S application is sufficiently credible. The allowable errors for the modules may accumulate to be unacceptable for the overall M&S application. Therefore, the integrated overall M&S application must be evaluated even if each module is found to be sufficiently credible.

4.6 M&S Results Evaluation Process

This process evaluates the credibility of the M&S results produced by conducting the experimentation / exercise process based on the integrated M&S application. The process integrates the evaluations of: (a) M&S results quality, (b) M&S experimentation / exercise process, and (c) M&S project characteristics related to life cycle stage 6.

If the M&S application is intended for training purposes, we exercise it typically in a distributed, interactive, and visual manner. If it is used for analysis purposes, we experiment with it to obtain the M&S results in a variety of forms including statistical averages, confidence intervals, graphs, charts, animations, and visualizations.

The design of experiments and statistical analysis of simulation output data are two major areas of discrete event M&S including techniques such as the following: (Banks, Carson, and Nelson 1996; Law and Kelton 2000)

- *Response-surface methodologies* can be used to find the optimal combination of parameter values which maximize or minimize the value of a response variable.
- *Factorial designs* can be employed to determine the effect of various input variables on an output variable.
- *Variance reduction techniques* can be implemented to obtain greater statistical accuracy for the same amount of simulation.
- *Ranking and selection techniques* can be utilized for comparing alternative systems.
- *Method of replication*, *method of batch means*, *regenerative method*, and others can be used for statistical analysis of simulation output data.

The experimentation process can be evaluated by using a hierarchy of indicators including the following:

- Are the algorithms used for random variate generation theoretically accurate?
- Are the random variate generation algorithms translated into executable code accurately?
- How well is the random number generator tested?
- Are appropriate statistical techniques implemented to design and analyze the simulation experiments?
- How well are the underlying assumptions satisfied?
- Is the problem of the initial transient (or the start-up problem) appropriately addressed?
- For comparison studies, are identical experimental conditions replicated correctly for each of the alternative operating policies compared?

4.7 Presented Results Evaluation Process

This process evaluates the credibility of the presented results produced by conducting the presentation process based on the produced M&S results. The process integrates the evaluations of: (a) presented results quality, (b) presentation process, and (c) M&S project characteristics related to life cycle stage 7.

The presentation process involves the

- *interpretation* of the M&S results,
- *documentation* of the M&S results, and
- *communication* of the M&S results to the decision makers.

Based on the presented M&S results, the decision makers formulate key decisions including acquisition of a military system, distributing scarce resources, selecting a business strategy, or training military personnel.

A *descriptive model* is a model that describes the behavior of a system without any value judgment on the "goodness" or "badness" of such behavior. All simulation models are descriptive models. Therefore, simulation results must be interpreted. For example, by experimenting with an M&S application, we can estimate the probability of kill as a 95% confidence interval $[0.89 \leq P_{kill} \leq 0.93]$. This M&S result must be interpreted by the analysts to determine if it is a "good" P_{kill} or a "bad" one.

The presentation process also involves the documentation of the M&S results. The documentation quality can be assessed by using a hierarchy of indicators including accessibility, accuracy, completeness, consistency, clarity (unambiguity and understandability), maintainability, portability, and readability.

The communication problem between technical and non-technical people should be recognized and the M&S results should be communicated to the decision makers in an understandable form without any technical jargon.

Due to the complexity of some M&S results, failing to properly interpret, document, and communicate the M&S results may lead to wrong decisions in spite of the fact that the M&S results are sufficiently credible.

5 CONCLUDING REMARKS

The V&V and certification activities must be tied to a well-structured M&S development life cycle. V&V is not a stage but a continuous activity carried out hand in hand with the M&S development throughout the entire life cycle. The use of a well-structured M&S development life cycle is critically important for effectively conducting the V&V and certification activities.

For new M&S application development, concurrent certification is recommended. Selected standard evaluation processes can be conducted to certify an existing M&S application with or without modifications.

Successful certification requires the certification agent to have full access to the M&S application with its associated documentation and data. However, the M&S developer has full control of the M&S application and might not fully cooperate in providing the required material and information to the certification agent. Sometimes, developers view certification as a performance appraisal activity, and they fear that their reputation and potential future funding are at stake if the certification agent identifies problems. Therefore, they sometimes show no desire to cooperate and behave in an adversarial manner against the independent certification agent personnel. The M&S application sponsor has a critical role in resolving this problem. (Balci et al. 2002b)

We have developed a methodology (Balci 2001) and a Web-based software system (Balci et al. 2002a) for certification of complex M&S applications. The proposed standard evaluation processes described herein can be executed under the guidance of the methodology by using the Web-based Evaluation Environment software system, which can be accessed at https://www.orcacomputer.com/ee.

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. New York, NY: John Wiley & Sons.

Balci, O. 2001. A methodology for certification of modeling and simulation applications. *ACM Transactions on Modeling and Computer Simulation* 11 (4), 352-377.

Balci, O., R. J. Adams, D. S. Myers, and R. E. Nance. 2002a. A collaborative evaluation environment for credibility assessment of modeling and simulation applications. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. Piscataway, NJ: IEEE.

Balci, O., R. E. Nance, J. D. Arthur, and W. F. Ormsby. 2002b. Expanding our horizons in verification, validation, and accreditation research and practice. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. Piscataway, NJ: IEEE.

Banks, J., J. S. Carson, and B. L. Nelson. 1996. *Discrete-Event System Simulation*, 2d ed. New York, NY: Prentice-Hall.

DMSO. 2000. *DoD Verification, Validation, and Accreditation Recommended Practices Guide*, millennium ed. Alexandria, VA: Defense Modeling and Simulation Office.

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

IEEE. 1998. *IEEE Standards for Software Verification and Validation.* IEEE Standard 1012. Washington, DC: IEEE Computer Society.

Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3d ed. New York, NY: McGraw-Hill.

Rae, A., P. Robert, and H.-L. Hausen. 1995. *Software Evaluation for Certification: Principles, Practice, and Legal Liability*. London, UK: McGraw-Hill.

Voas, J. M. 1999. Guest editor's introduction: certification: reducing the hidden costs of poor quality. *IEEE Software* 16 (4): 22-25.

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. He is the Editor-in-Chief of the international journal *Annals of Software Engineering*; 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 M&S International (SCS) and as a Director at Large for the SCS Board of Directors. Most of Dr. Balci's research has been funded by DoD since 1983. Since 1998, he has been providing technical services for the National Missile Defense program. His current areas of expertise center on Software Engineering; E-Systems Engineering (e.g., e-solutions, e-business, e-commerce); M&S; and VV&A, IV&V, and certification of M&S applications and software systems. His email and web addresses are <balci@vt.edu> and <http://manta.cs.vt.edu/balci>.

SAID D. SAADI is the Technical Director (TD) for the Missile Defense Agency (MDA) Ground-Based Midcourse Defense (GMD) Independent Verification and Validation (IV&V) program at the Naval Surface Warfare Center Dahlgren Division (NSWCDD). In this capacity, Mr. Saadi directs the efforts of the NSWCDD Task Leads in Battle Manager and System Software IV&V, as well as LSI Integrated Distributed Simulation (LIDS) Verification, Validation and Accreditation (VV&A). He also leads the GMD Sensor Simulation Test Bed (SSTB) VV&A, NSWCDD GMD Analysis efforts, and the MDA Model and Simulation Working Group (M&SWG) VV&A Subgroup for the GMD Joint Program Office (JPO). 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 as well as numerous other course work in the field of defense acquisition and program management. He is certified at Level III of the System Planning, Research, Development, and Engineering. Prior to joining the GMD IV&V program at NSWCDD, he served as the AEGIS SPY-1 Sensor System Subject Matter Expert. In that position, he designed, executed and analyzed Electronic Counter Measure tests for the SPY-1 Radar, and led the development of representations of the AEGIS SPY-1 radar model in the Extended Air Defense Simulation (EADSIM), and Extended Air Defense Test Bed (EADTB) for the Navy Area Wide, Navy Theater Wide, and US-Japan Bilateral studies. His e-mail address is <SaadiSD@nswc.navy.mil> or <Said.Saadi@mda.osd.mil>.