Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

MODELING AND SIMULATION (M&S) ISSUES IN OPERATIONAL TEST AND EVALUATION (OT&E)

Joseph Ramos III

Office of the Secretary of Defense
Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington, D.C. 20301-1700, U.S.A.

ABSTRACT

This paper addresses the current guidance and major issues facing the operational test and evaluation (OT&E) community in the use of modeling and simulation (M&S). With the increasing sophistication of M&S tools, to what extent can M&S be used in determining operational effectiveness and operational suitability? How does this evolving technology affect the way OT&E balances the legislative demands of Congress and the cost and schedule concerns of acquisition? These issues are among those facing DOT&E that may result in additional policy direction.

1.0 INTRODUCTION

M&S is often a primary tool of the development community throughout the acquisition process. Its use begins with the development of a Cost and Operational Effectiveness Analysis (COEA). Here M&S is used to compare alternatives against operational measures. The OT&E community is identified with encouraging the development test (DT) community to use realistic M&S to ensure readiness for operational tests, but with being averse to using M&S for operational testing or evaluation.

This paper addresses the issues of (1) how M&S can reduce cost and accelerate the schedule of OT&E, yet provide at least the same degree of confidence in assessing a system's operational effectiveness and suitability; and (2) how M&S can address items which cannot be tested due to constraints, such as safety.

This paper uses the definitions of M&S from DoDD 5000.59, *DoD Modeling and Simulation Management*:

> **Model**: A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.
>
> **Simulation**: A Method for implementing a model over time. Also, a technique for testing, analysis, or training in which the real-world systems are used, or where real-world and conceptual systems are reproduced by a model.

2.0 CURRENT GUIDANCE

A summary of the current guidance by Congress, DOT&E policy, and related DoD directives affecting the use of modeling and simulation in OT&E follows:

2.1 Public Law

When the position of the Director, OT&E (DOT&E) was created by an act of Congress in 1983, Congress intended that DoD systems be tested in operationally realistic field conditions. *Title 10 U.S. Code Section 139* (recently superseded Section 138) defines OT&E as:

> the **field test**, under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for determining the effectiveness and suitability of the weapons . . . for use in combat by typical military users."

Where operational effectiveness is defined in part as "the overall degree of mission accomplishment of a system" and operational suitability as "the degree to which a system can be placed satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, . . . " (and other "ilities").

Congress further directed DOT&E in *Title 10 U.S. Code Section 2399* which states in part:

> The Director shall analyze the results of the Operational Test and Evaluation . . . (and) . . . shall prepare a report stating the opinion of the Director as to

> (A) whether the test and evaluation performed were adequate; and (B) whether the results of such test and evaluation confirm that the items or components actually tested are effective and suitable for combat.

Therefore, by law the Director's evaluation must not (A) be based exclusively on computer modeling or simulation, since OT&E is defined earlier as " the field test, . . ." and in subsection (B) of this law, the Director's evaluation should confirm, not speculate, that the system as tested is effective and suitable for combat. However, opinion gives the Director great latitude.

2.2 DOT&E Policy

Does this then preclude DOT&E use and support of M&S in OT&E? No. DOT&E's *Policy for Application of Modeling and Simulation in support of OT&E*, January 24, 1989, states:

> Models and simulations are tools which can potentially augment and/or complement actual field tests and provide decision makers necessary information to assess the progress of a system toward fulfilling the operational needs.

DOT&E has always encouraged analysis using properly validated and accredited M&S, especially during early development phases. Also they should be used to assess those areas that, due to safety or test capability limitations, cannot be directly observed through testing. If hardware and software are not at equivalent stages of maturity, then hardware/software surrogates that emulate the least mature element may be necessary. The use of these methods, while not a substitute for testing, can reduce testing costs by supplementing actual test data, and can address areas which cannot be tested. DOT&E also encourages the use of computer models which can be useful tools to investigate variability of outcomes when random variation is an important factor.

2.3 DoD Directives

DoDD 5000.59, *DoD M&S Management*, empowers the Under Secretary of Defense for Acquisition and Technology (USD(A&T)) to "Strengthen the uses of M&S in the Department of Defense." It further establishes policies that:

> DoD Components shall establish verification, validation, and accreditation (VV&A) policies and procedures for M&S applications managed by the DoD Component.

and

> M&S applications used to support the major DoD decision making organizations and processes . . . shall be accredited for that use by the DoD Component for its own forces and capabilities.

2.4 Summary

M&S has a valid role to complement actual testing, but cannot replace it.

3.0 HIERARCHY OF M&S

A discussion of the issues surrounding the use of M&S can be put in perspective by first understanding the hierarchy or levels of M&S. Each of these levels are made up of either interactive or stand-alone M&S:

Interactive: Requires a human to actively participate, also called man-in-the-loop (MIL). Examples include flight simulators and distributed interactive simulation systems such as those using Simulation Network (SIMNET) to link manned and unmannned remote simulation nodes together for real-time interaction.

Stand-Alone: Non-interactive. Needs only input data to run, most M&S falls in to this category, and is useful for sensitivity analyses.

3.1 Component/Subsystem/System

This level is the most recognized form of M&S and includes such mature technology and institutionalized processes as those associated with aeronautical and space systems. All of us are aware of flight simulators that provide very realistic representation of the aircraft and associated systems. Crew members regularly train on flight simulators for proficiency to augment actual flying time. Further, most of these models and simulators have been verified, validated, accredited, and continually updated with actual flight information. As a result, OT&E use of M&S to augment the evaluation of the mission utility of these systems has been the subject of less debate. Another major example of this level of M&S is the emulation of "threat" systems that our military weapons are built to counter such as foreign air

defense systems. The same arguments as above apply. However, today's ease of acquiring the real thing makes the high cost of developing threat simulators less necessary.

On the other hand, advances have not been as dramatic in other areas where the processes are not institutionalized such as the Command, Control, Communication, Computer and Intelligence (C4I) systems and Automated Information Systems (AIS). M&S of C4I systems usually concentrate on the performance characteristics of the last "Cs" or the communications and computer system elements of C4I. These M&Ss typically perform well. This is not true for the first two "Cs." Since OT&E evaluates the mission utility of these tools, the mission utility of communications and computers to the function of command and control has not been easy to define. The usefulness of these systems is primarily human dependent. The use of artificial intelligence technology in M&S to emulate human decision making, perceptual processes, and other human factor issues have a long way to progress. As a result, its uses in OT&E are more controversial.

3.2 Single Engagement

This next level of sophistication addresses models and simulations associated with the system under development/test as it reacts in a "one-on-one" or "one-on-few" scenario. Here, the system under development/test is exposed from a performance standpoint against a threat environment of real and simulated threats. The Air Force 4-Star Modeling and Simulation Review in 1993 found that the availability of these resources is less than 10% of the component/subsystem/system level M&S resources discussed above. High fidelity MIL simulations exist in government and industry with sophisticated anechoic chambers, virtual reality systems, and other technology. Again, OT&E use of this level of M&S for evaluating operational effectiveness "fog-of-war" issues and operational suitability issues such as human factors, reliability, availability, maintainability, and the other "ilities" are usually not possible. So, VV&A of these resources is an issue.

3.3 Mission Level

This emulates several systems under development/test in a "one-on-many" and "few-on-few" scenario and considers more variables normally encountered in a mission. The Military Services have several M&S resources available in this category, such as the high fidelity joint Army-NASA Crew Station Research Development Facility (CSRDF) for light helicopters. Although these have the same limitations discussed previously, they are useful in early determination of mission utility or areas of interest for actual field tests.

3.4 Campaign

This highest level in the hierarchy of M&S emulates a major exercise or a "many-on-many" robust environment. These are also the least mature of the M&S resources. The Military Services have several useful computer models that can be used to determine sensitivities to test outcomes because of changes to certain variables. In aeronautical T&E, there are examples of integrating MIL simulations with actual flight tests from distributed locations linked by SIMNET. Operators interact with the simulator as blue, red, and grey forces in conjunction with real-time flight test data. Many assumptions and initial conditions are required, and variables such as skill levels, weather, etc. are often too complicated to consider. Also, the low level of simulator fidelity is a major shortfall. Short of conducting a war, validation of the M&S results is a difficult task. However, the results are extremely useful in augmenting actual field tests.

3.5 Summary

Use of M&S at all levels of hierarchy are useful in OT&E. The artificial world that allows manipulation of variables provides important information about what causes the observed in the real world. This results in early identification of critical issues and in a more informed OT&E plan. Other benefits include early operational feedback to system designers and should also reduce the scope of required field tests. However, as a model or simulation becomes more complex in order to address higher order levels of interaction, the confidence or credibility in the observations reduce due the decreasing fidelity in the design of the M&S. One approach to improving the credibility of higher order M&S in weapon systems operational effectiveness analysis is to develop high fidelity R&D MIL simulators that are integrated into a SIMNET environment. On the other hand, when it comes to M&S tools for use in operational suitability analysis, there is a severe shortfall in all levels of interaction.

4.0 ASSESSMENT

As already established above, there is little argument on

the usefulness of M&S in the acquisition process, to include OT&E. There are existing generic policies on when it should be used in OT&E (augment field testing and used throughout the acquisition life cycle); what should be used (M&S that has undergone VV&A); who manages the VV&A process to include configuration control (Heads of the DoD Components); who oversees OSD M&S acquisition policy (USD(A&T)), through DoD Executive Council for Modeling and Simulations (EXCIMS) and Defense Modeling and Simulation Office (DMSO). What then are the OT&E issues?

● **To what extent should DOT&E encourage the use of M&S in OT&E?** Early use of M&S tools can help formulate sound operational concepts and realistic tests. Use of contractor developed M&S would be useful if there are safeguards against conflict of interest. To a degree, an OT is actually a simulation since weapon systems are tested in a simulated combat environment. When OT is conducted during military exercises, these are as close to combat conditions as practical since the military trains the way they fight and fight the way they train. Even in these conditions, the OT is often piggybacked on a non-interference basis and often conducted in areas where instrumentation is limited. Evaluations are based on a single test or series of tests rather than incorporating developmental test data and M&S for a more comprehensive evaluation.

The Defense Science Board made the following recommendation in its December 1989 report:

> Do not employ simulations to prove or disprove things, but instead exploit their ability to isolate high sensitivity areas. Simulation has an important role in providing analyses, and as a method of focusing on system engineering issues early through operational tests.

● **What level of VV&A should DOT&E accept for a model or a simulation to be usable in OT&E?** Do the same levels of VV&A have to be met for the various hierarchies of M&S?

● **How does VV&A mature with the maturity of the system under procurement?** The use of developmental test results to validate performance assumptions in M&S will enhance their credibility. The Defense Science Board made the following recommendation in its December 1989 report regarding the use of models in the accreditation process:

> Confidence in models can be enhanced by employing them for excursion and sensitivity analyses, and focusing on critical issues by running tests and validating the results. It is not feasible or cost effective to set up a central office to accredit models.

● **What are the sources of M&S for OT&E?** Should they be derived from the COEA, modify or use existing models and simulators or develop new ones? Developing new simulators, for example, is expensive. And they may take as long to develop as the prototype system the simulator is being built to support. Further, if M&S is needed in support of T&E, it must be of better fidelity than the system under test, or results may be in question.

● **How does the cost of VV&A factor in?** For example, in the F-22 program, the Air Force Operational Test and Evaluation Center (AFOTEC) planned to contract for an independent VV&A of the F-22 model, but because of high costs, they chose to use existing models. How do these compromises affect T&E adequacy?

● **To what extent should COEA models, their assumptions, and limitations be considered in the subsequent use of M&S for OT&E?** COEAs justify the acquisition program based on a cost/benefit analysis of alternate capabilities. Mr. Louis Rodrigues, General Accounting Office (GAO), stated during congressional testimony on OT in March 22, 1994, "that if a program is 'sold' on meeting certain requirements, DoD's commitment to that program should be strictly limited until those requirements have been adequately demonstrated." Since OT&E is based on the Operational Requirements Document (ORD) which ideally evolves from the COEA, to what extent does OT&E consider the use of COEA M&S, assumptions and limitations in its planning and execution?

● **Should M&S results be part of the operational test readiness reviews and DT&E certification process?** If M&S was used throughout the T&E process in both DT and OT, there should be enough indicators of risk to predict readiness for OT&E.

● **Should there be differences in OT&E policy on the application of M&S between operational effectiveness and operational suitability?** Unlike M&S capabilities that emulate system performance which can be useful in operational effectiveness evaluations, today's M&S tools

have only demonstrated minimal contribution to suitability evaluations. Tools to better assess potential human factors, maintainability, and other "ilities" problems may be forthcoming with increased use of Virtual Reality and other artificial intelligence tools. Reliability Growth Models, for example, are extensively used as a management tool. But there is little or no evidence for the universal applicability of these models. Reliability numbers during OT are typically much worse than during DT. This is because soldiers are harder on equipment than engineers, the environment is tougher than the lab, repairs are less expert, and there are more failure modes during real operations. DOT&E evaluations are based on the "as tested" configuration.

4.1 Examples of Productive Use of M&S in OT&E

Joint Tactical Information Distribution System (JTIDS): The JTIDS program is an example of how M&S was used to augment Operational Tests of a C4I system. JTIDS is a digital communications system that will provide secure, jam resistant information for display in over 15 different multi-service tactical platforms. AFOTEC conducted initial operational T&E (IOT&E) of the JTIDS Class 2 terminal from August 1986 to April 1987. The performance of JTIDS-equipped F-15s was compared against non-JTIDS-equipped F-15s in Defensive Counter-Air missions.

M&S Augmenting T&E: The IOT&E was conducted in two segments: M&S MIL simulation using the McAir F-15 simulator in St. Louis MO with EW modeling using the TAC JAMIT digital computer model at Kirtland AFB to predict JTIDS performance in jamming environment, and flight testing at the Eglin AFB Air Combat Maneuver Information (ACMI) range. Both segments complemented each other, and each added elements which could not be accomplished by the other. Whenever possible, the maximum amount of replication and substantiation was accomplished between the two. For instance, the same F-15 pilots participated in each segment, the same scenarios were employed, and the same principal measures of effectiveness were applied.

The simulation provided sufficient data to complete a statistical comparison of baseline and modified F-15 mission effectiveness. It permitted repeatable test trials, reduced the number of baseline F-15 missions required during flight testing, and permitted the evaluation of conditions which were not feasible in flight tests. The model also provided insight to performance sensitivities. During flight tests, JTIDS and F-15 radars were subjected to DIA validated jamming threats.

Limiting Factors: There were limiting factors, such as several key threats and friendly elements that were not representative of future systems, thus requiring substitutions by less realistic surrogates; terminals and support equipment were not production representative; insufficient quantities to stress the system; and tactics followed peace-time rules of engagement.

T&E Results: The JTIDS class 2 terminal in the F-15 was considered operationally effective. The McAir simulation showed a statistically significant improvement in most measures of mission effectiveness and no degradation in the others. Flight test corroborated these results. Operational suitability however was unsatisfactory due to the terminal's high failure rate and incomplete technical orders, crypto variable problems, and the terminal software had not reached an adequate level of maturity. However, these results supported a decision for low rate interim procurement of JTIDS.

OT&E of Navy JTIDS integration in Navy platforms is currently on-going. Use of M&S to augment OT&E by the Navy is minimal. However, in the Multinational Information Distribution System (MIDS) program (smaller version of JTIDS), the Navy plans to use the McAir MIL F/A-18 simulator and a realistic ACMI range at the Navy Air Station Fallon, NV, to augment MIDS OT&E in the F/A-18.

4.2 Example Where M&S Is Of Little Benefit:

Single Channel Ground and Airborne Radio System (SINCGARS): The Army's second source of the SINCGARS is one example of what was to have been a straight forward alternate source procurement of a radio that was a form-fit-function "duplicate" radio already produced. However, the internal design was substantially different. Use of M&S in design or T&E was not considered necessary since performance was already known (an equivalent radio existed), and reliability was to be better since only more reliable parts were to be used. The second source radio was to have been transparently interchangeable with the first source radio. The second source contractor planned to take advantage of new technology and new manufacturing techniques to significantly reduce cost and have a better performing and more reliable radio. As a result, the second source acquisition strategy included an accelerated T&E program.

Unfortunately, the new design resulted in initial high failure rates and unexpected human factors problems when the radio was given to Army operators for initial field testing. One example was a change in antenna

mounting that proved to be disastrous. In an IOT&E using field conditions, the radio operators usually grabbed the radio by its antenna instead of the handle. This broke the antenna, making the radio useless. These and other "unpredictable" suitability design problems found in subsequent operational tests resulted in a two-year slip in the program. These radios eventually successfully passed OT with high marks and entered production. In retrospect, the problems found were not unusual for new designs and were equivalent to the time it took to field the first source radios.

A lesson learned is that just because there is low risk in the technology, and laboratory performance of equipment/systems may not show problems, there is no substitute for actual field operational tests. New systems will be placed in users' hands inevitably when the system is first fielded with the first military unit receiving the system. However, an OT&E under controlled conditions finds problems sooner and usually under conditions that are easier to help developers isolate the problem. This results in cost savings over the system life cycle.

5.0 CONCLUSION

M&S can provide a significant contribution to OT&E. It can augment OT&E field test results by compensating for real-world limitations such as safety factors or density of threats and targets. It can also provide tools for better assessments of test and evaluation plans, early identification of critical issues, and early participant system familiarization. In areas such as C4I systems where the effect of human interaction is key, and in strategic weapons (e.g. Ballistic Missile Defense) where safety is a concern and conducting large scale testing for statistical significance is impractical, M&S can be even more useful. However, more work in developing higher fidelity M&S tools in these areas are required. Overall, M&S can reduce (but not eliminate) the scope of actual field testing required for OT&E.

The key issues in employing M&S in OT&E are:

(1) Identifying specific cost effective applications with M&S limits for each application;
(2) Acquiring M&S, i.e., modify existing or develop new capabilities;
(3) VV&A process;
(4) Identifying assumptions, variables to control, other limitations, and specific values added to the OT&E process.

AUTHOR BIOGRAPHY

JOSEPH RAMOS III is a Staff Assistant for C3I Systems in the Office of the Director for Operational Test and Evaluation. He has 26 years experience in all phases of the acquisition and T&E process. This includes 13 years as program manager of many acquisition programs in support of T&E, and 13 years major systems test and evaluation experience directing and overseeing developmental and operational T&E of aeronautical fixed wing and rotary wing aircraft, C3I systems and Automated Information Systems. He has also been an adjunct professor from 1978-1988 for three universities lecturing graduate level Electronic Engineering and Business Management courses. He received a B.S. degree in Physics from University of San Francisco in 1968, a M.S.E.E. degree from University of Southern California in 1973, and a M.S. degree in Management from Stanford University in 1982 as a Stanford SLOAN Fellow.