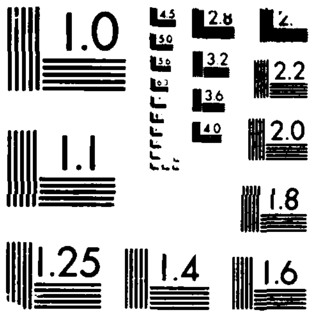
AD-A191 503 DOD SIMULATIONS: IMPROVED ASSESSMENT PROCEDURES WOULD INCREASE THE CREDIBILITY OF RESULTS(U) GENERAL ACCOUNTING OFFICE WASHINGTON DC PROGRAM EVALUATION AN.. 1/2

UNCLASSIFIED DEC 87 GAO/PEMD-88-3 F/G 5/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

DTIC FILE COPY

GAO

United States General Accounting Office

Report to the Chairman, Legislation and National Security Subcommittee, Committee on Government Operations, House of Representatives

AD-A191 503

DOD SIMULATIONS

Improved Assessment Procedures Would Increase the Credibility of Results

DTIC ELECTE FEB 2 2 1988 S D E

88 2 18 004

GAO United States
General Accounting Office
Washington, D.C. 20548

Program Evaluation and
Methodology Division

B-229237

December 29, 1987

The Honorable Jack Brooks
Chairman, Legislation and
National Security Subcommittee
Committee on Government Operations
House of Representatives

Dear Mr. Chairman:

This report presents our review of selected simulations the Department of Defense uses to provide information about the effectiveness of weapon systems and the efforts of the department to ensure the simulations' credibility. It is one of several GAO reviews on the quality of information being used for decisionmaking, and it focuses on the tools federal departments and agencies use to generate that information.

Unless you publicly announce the contents of this report earlier, we plan no further distribution of it until 30 days from the date of the report. At that time, copies of the report will be sent to the Committee on Government Operations of the House of Representatives and to the House and Senate committees on armed services and the budget. Copies will also be sent to the secretary of Defense, the secretaries of the Army, Air Force, and Navy, and others interested in the topic.

Sincerely,

Eleanor Chelimsky
Director

Executive Summary

Purpose

Multibillion-dollar acquisition decisions for major weapon systems should in principle be based on the results of testing weapons under conditions that replicate actual combat. However, subjecting complex and expensive weapon systems to the necessary number of such tests is sometimes impractical or impossible. One alternative is to use computer models to simulate performance, but simulation results must be as representative of real-world outcomes as possible. The need for representativeness generates the major objective GAO addressed in this report: to determine, using three case studies, that it is possible to assess the credibility of simulation-generated data. A second objective was to identify the steps the Department of Defense (DOD) has taken to foster the credibility of its simulations.

GAO posed three major questions: (1) What factors should be considered in a systematic attempt to assess the credibility of a simulation? (2) What are the results of assessing specific operational-effectiveness simulations of weapon systems with respect to these factors? (3) What efforts has the Department of Defense made to foster and reinforce simulation credibility?

Background

DOD uses developmental and operational tests and evaluations as part of a weapon-system's acquisition program to provide evidence that the weapon system performs as expected before proceeding through development phases to full-scale production. Field tests are important in determining the extent to which a weapon system satisfies operational requirements, but when such tests do not provide sufficient information, DOD often uses simulation models to generate supplemental data about a weapon's effectiveness. Although simulations are useful tools, they are always approximations to reality and, therefore, their credibility—the level of confidence that a decisionmaker should have in their results—is open to question.

GAO developed its own assessment framework and applied it to three operational effectiveness simulations developed for Army air defense system programs: the Carmonette and ADAGE computer simulations used in the division air defense gun (DIVAD) acquisition program and the COMO III computer simulation applied in the Stinger missile program.

Results in Brief

Using the framework in the accompanying table, GAO found that each simulation had strong points but found weaknesses and limitations that degraded their credibility severely enough to question their usefulness.

Area of concern	Factor
Theory, model design, and input data	1. Match between theoretical approach and real events being simulated
	2. Choice of measures of effectiveness
	3. Portrayal of weapon's immediate combat environment
	4. Representation of operational performance
	5. Depiction of critical aspects of broad- scale battle environment
	6. Appropriateness of mathematical and logical representation
	7. Selection of input data
The correspondence between the model and the real world	8. Verification effort
	9. Attention to statistical quality of results
	10. Sensitivity testing effort
	11. Validation effort
Management issues	12. Organizational support
	13. Documentation
	14. Full disclosure of results

One consistent weakness in all three simulations that potentially poses a major threat to credibility is the limited evidence of efforts to validate simulation results by comparing them with operational tests, historical data, or other models.

Guidance from the office of the secretary of Defense in the form of procedures would provide a structured way of assessing the simulations' credibility.

Principal Findings

GAO's Assessment Framework

GAO's assessment framework of 14 factors should be considered in attempts to evaluate a simulation's credibility. The number of factors could vary (other frameworks may contain fewer or more), but it is important that they cover the three major areas of concern: theory, model design, and input data; the correspondence between the model and the real world; and management, documentation, and reporting issues. Collecting and analyzing information about each factor should help identify a simulation's strengths and weaknesses and, therefore, its credibility. GAO's framework proved useful for the three case study simulations in this respect. (See pages 17-22.)

Assessment of Selected Simulations

GAO found that for all three simulations—the Carmonette, ADAGE, and COMO III—evidence of credibility was provided on only a few factors: measures of effectiveness, the representation of a weapon's engaging targets, sensitivity testing, and the disclosure of strengths and weaknesses of results. Even so, the simulations were still limited on these factors. (See pages 30, 34, 42, and 51.)

Generally, the principal weakness centered on the lack of validation of simulation results. Validation can be difficult, but it must be dealt with if simulation results are to be credible. (See pages 44-46.)

For most factors, the three simulations varied considerably. For example, the Carmonette simulation of the DIVAD was severely limited in its ability to portray a battle of area and duration appropriate for a division-oriented weapon. The simulations using the Carmonette and COMO treated attrition continuously throughout a battle with regard to mathematical and logical representation, whereas the ADAGE's approach only calculated attrition at the end of a battle period, a procedure that can introduce bias. The effort required to remove these limitations and some of those found in other areas might be considerable, but others could be corrected with relatively minor effort. (See pages 33, 36-37, and 39.)

DOD Guidance

The Department of the Army has been relatively active in fostering the development of organizations that can directly influence the credibility of simulation results. While DOD officials agree that credibility is important, and while there is some consensus about what should be done to achieve such credibility, DOD generally has not in fact established the credibility of its simulations systematically and uniformly. No guidance exists at the level of the office of the secretary of Defense that can be routinely used throughout DOD to review the credibility of military models. (See pages 54-56.)

Recommendations

GAO recommends that the secretary of the Department of Defense adopt or develop and implement guidance on producing, validating, documenting, managing, maintaining, using, and reporting simulations of weapon-system effectiveness. This guidance should include a way of routinely providing reviews of a simulation's credibility and, in this way, identifying problems that should be resolved. The secretary should explore requiring that a statement regarding validation efforts accompany simulation results.

GAO also recommends that the secretary of the Department of Defense direct the agencies responsible for managing the ADAGE, Carmonette, and COMO III models to explore the feasibility of correcting the limitations GAO has identified, especially the limitations in validation.

Agency Comments

In commenting on a draft of this report, DOD generally found the report to be technically correct and concurred with GAO's two recommendations. It has sent GAO's factors for assessing simulations to the services for review and evaluation.

DOD raised some concerns about the scope and focus of the report. One was about generalizing from three cases studies, asserting that GAO did, indeed, do this but without citing specific examples to support the assertion. GAO's purpose was to demonstrate from case studies that one can systematically collect and analyze information about a simulation that would permit one to assess its credibility. GAO did not intend to infer from these case studies anything with regard to the credibility of other simulations.

DOD also contends that applying GAO's framework gives only part of a simulation's picture and that people, input data, and a model's application are also important. GAO certainly agrees but points out that factors 1, 7, and 12, whose importance was defined in the draft report, do consider these. (See pages 62, 63, and 242.)

Other technical comments are found in DOD's letter and comments reprinted in appendix V.

Contents

Abbreviations

ADAGE	Air defense air-to-ground engagement
COMO	COmputer MOdel
DIVAD	Division air defense gun
DOD	U.S. Department of Defense
GAO	U.S. General Accounting Office
IFF	Identification friend or foe
NATO	North Atlantic Treaty Organization
OSD	Office of the Secretary of Defense
TRADOC	Training and Doctrine Command

Chapter 1

Introduction

Simulation is a two-phased process of constructing a model of an existing or a proposed system and conducting experiments with the model so as to understand the behavior of the system or evaluate strategies for its operation. A simulation is more than a static picture of the system; a simulation imitates the system's human and machine operation or behavior over time. In a military context, simulation can be a tool for analyzing the performance and operation of a weapon-system component (for example, the radar of a surface-to-air missile system), the total weapon system (for example, the complete surface-to-air missile system), or the total panoply of weapon and communication systems (for example, an air defense system).

The Department of Defense (DOD) uses development and operational testing and evaluation in weapon-system acquisition programs to provide the evidence that, among other things, a weapon system meets performance specifications and can perform as expected in realistic operating conditions. In principle, this evidence should be obtained empirically from developmental and operational tests for acquisition decisions. However, as weapon systems have become ever more complex and expensive and as attempts to expedite the acquisition process have increased, the willingness and sometimes the ability to subject them to extensive field testing to determine their effectiveness and suitability have diminished or become impractical. Acquiring the needed information efficiently during the acquisition process requires an appropriate use of available methods. Simulation can be used in conjunction with field experimentation and other analytical methods with the likely result that the benefit of the combination will exceed the benefits of the individual methods.

Evidence suggests that DOD uses simulation substantially in the developmental and operational test phases of the acquisition of weapons. However, questions arise about the credibility of simulation-generated data and DOD's practices for ensuring that simulations produce sound results. When simulations contribute information for multibillion-dollar weapon-system development and procurement decisions, it is important that they provide usable, high-quality information.

In this report, we describe our development of a method for reviewing simulations of the operational effectiveness of weapon systems. From information from assessment frameworks developed by other researchers, we developed a conceptual framework for systematically reviewing simulations and applied it to selected Army simulations used in the acquisition of air defense systems. We viewed our task as developing and testing a review framework to illustrate how it can provide insights into a simulation's strengths and weaknesses, especially in terms of identifying areas for improvements.

Simulation in Weapon-System Programs

Simulations can be and often are used throughout the life cycle of a weapon system. Simulations are used frequently in conjunction with other analytical methods and field experimentation, each approach contributing to the understanding of a weapon system's functioning. Contractors and the developing agencies during the concept exploration and early development phases of research, development, testing, and evaluation use simulations for such purposes as

- studying alternatives to a weapon system by conducting trade-off and parametric studies,
- defining a system's and subsystem's requirements, and
- determining a system's design.

During later stages of development, the test and evaluation agencies, the operational (or user) groups, the development agency, and the contractors use simulations for

- investigating a system's or subsystem's performance,
- identifying its problems and limitations,
- estimating operational effectiveness,
- determining logistic and support requirements, and
- determining tactics.

DOD has developed and uses a number of computer models that simulate weapon systems in combat. Models are complex computer programs for mimicking what happens in the real world when a weapon is used. Models used for operational effectiveness studies are ordinarily designed to simulate more than one type of weapon system. When simulations are needed for studies and analyses, DOD may choose existing models or develop new ones. The development and maintenance of major simulation models are usually the responsibility of specific organizational units within DOD.

Advantages and Disadvantages of Simulations

The overriding advantage of simulation is perhaps the opportunity to investigate questions and problems that could otherwise not be addressed and to investigate them systematically with numerous replications under controlled conditions. In a simulation, both the model of a system and the model of its environment can be altered in an organized manner. A model provides information on performance under assumed external conditions and permits the investigation of the system's response to changes in these conditions and to changes in the original characteristics of the system itself.

In addition, experiments can be performed on the model of a system that may not exist or that exists only in limited numbers or that operates in a physical environment that is not accessible. Simulations can provide information about a system's probable performance under conditions that cannot be tested because of costs, the lack of adequate equipment and realistic test environments, or safety and security restrictions. Simulation allows the exploration of more aspects of a system's performance more easily than is available from field experimentation with an actual system. Moreover, the development of a model and the simulation process do not consume or destroy a weapon system. After the possible consequences of using a weapon have been modeled, the results of simulations can be validated by field testing.

Simulation also has disadvantages. A model is an approximation, not the equivalent, of a real system. Inaccurate assumptions about a weapon or its environment may cause the results of a simulation to diverge from reality. Important variables or relationships may be omitted, and appropriate values for those that are included may be difficult to obtain. Data and resources for validating simulations may not be available. Statistical complexities may obscure the results. Simulations cannot be better than the analysts' understanding of the concepts, the hardware, and the relationships involved; unasked questions do not get answered in a weapon-system simulation. Conducting simulation experiments has its own set of problems. For example, different people and equipment are generally required for a simulation from those required in field-testing the actual system. And the simulation of a total system has its costs in terms of development time, staffing, and computer resources.

The Credibility of Simulation Results

Simulations can be valuable aids for decisionmaking, but there will always be some concern about drawing the wrong conclusions from them. Since simulations are abstractions or approximations of the real world, questions arise about their credibility. We define a simulation's "credibility" as the level of confidence in its results. To say that simulation results are credible implies evidence that the correspondence between the real world and the simulation is reasonably satisfactory for the intended use. Credibility is not an absolute condition but measured on a continuum.

While it is true that assessing credibility will always require some level of subjective judgment, it is also true that many parts of a simulation lend themselves to scientific and empirical tests and checks. Any framework for assessing simulations, including the one we developed, must therefore address the things that can be tested as well as those that must ultimately rely on informed but judgmental conclusions.

Objectives, Scope, and Methodology

In previous reports, we have addressed issues regarding simulation evaluation methodology and, more specifically, the modeling of weapon systems. A major focus and objective of this report was, using three case studies, to demonstrate that it is possible to systematically collect and analyze information about a simulation that would permit an assessment of the credibility of that simulation to be made. A second objective was to identify the steps DOD has taken to ensure the credibility of its simulations. To meet these objectives, we sought the answers to three evaluation questions:

1. What factors should be considered in a systematic attempt to assess the credibility of a simulation?

2. What are the results of an assessment of selected weapon-system operational-effectiveness simulations with respect to these factors?

3. What efforts has DOD made to foster and reinforce the credibility of its simulations?

The factors we identified in the first question provide a framework for collecting information about specific simulations. This framework allows for the identification of a simulation's strengths and weaknesses with respect to each factor. The strengths enhance the confidence a user might have in the simulation, and the weaknesses translate into threats to that confidence. Further, the weaknesses point to remedial efforts that could increase credibility.

The answer to the second question involved demonstrating that the framework can be applied as a guide for assessing three simulations of operational effectiveness and identifying areas where improvements would reduce threats to credibility. To answer the third question, we used information we collected while performing these case studies and additional data we collected during our review.

What Factors Should Be Considered?

To identify the factors that should be considered in a systematic attempt to assess the credibility of a simulation, we interviewed DOD officials, operations research analysts, other analysts, and test engineers, and we reviewed literature on the development and use of simulations. From this, we developed a framework of three major areas of concern and 14 factors, which we describe in chapter 2.

What Are the Results of Assessing Simulations With These Factors?

To answer the question on the results of assessing selected weapon-system operational-effectiveness simulations with respect to these factors, we applied our framework to three case studies. To select cases, we identified weapon-system programs that had used major simulations of operational effectiveness in support of acquisition decisions. We did this because we believe that the most useful process is to assess the credibility of a simulation in the context of its application in the study of particular issues. We also wanted, however, to examine general purpose models that had the ability to simulate several types of weapon systems.

We judgmentally selected two Army antiaircraft defense systems: the portable, shoulder-fired, infrared, surface-to-air Stinger missile and the division air defense gun (DIVAD, known also as the "Sgt. York"), a surface-to-air, radar-guided gun on a tracked vehicle. For these two weapon systems, we chose three simulations: for the Stinger missile, we chose the COMO III model, and for the DIVAD, we chose the Carmonette and air defense air-to-ground engagement (ADAGE) models. We describe these weapon systems and simulation models in chapter 3. (In appendix I, we also briefly describe how simulations were used in studies for the two weapon-system programs.)

We obtained general descriptions of the simulations and the use of their results in the acquisition process. We also reviewed documentation explaining how these simulations were developed and validated. We interviewed the analysts and test engineers who were involved in developing and using the simulations, asking for their perceptions as well as documentation pertinent to factors in our framework. We also interviewed several persons responsible for the maintenance of the simulations and for using the simulation results. We interviewed others who dealt with other aspects of the simulation development and experts in related subjects, such as operations research, combat environments, threat assessment, and field tests.

This provided us with information about the alternative theories, assumptions, data, and procedures that were used in developing, running, and reporting the simulations we reviewed. Using our framework to guide our analysis of these data, we identified strengths and weaknesses that could enhance or threaten the credibility of the simulations. Our summary findings for the three case studies are in chapters 4, 5, and 6, and additional detail on them is in appendixes II, III, and IV.

What Effort Has DOD Made Toward Credibility?

To address our third question—What effort has DOD made to foster and reinforce the credibility of its simulations?—we collected and reviewed information about DOD and Army regulations and policies relevant to simulation development, management, and assessment generally and to the simulations we reviewed specifically. We also interviewed DOD officials responsible for managing and performing simulations. Our findings, presented in chapter 7, provide information on DOD's mechanisms and procedures for gaining and maintaining the credibility of its simulations.

Our Study's Strengths and Limitations

We examined other assessment procedures and structures and based our framework on this body of work, but we found few examples of the application of other frameworks. We were able to use our framework with several Army simulations. Since one of our objectives was to demonstrate the feasibility of applying our framework, it was not necessary nor would it have been practical to review all or even a large number of the simulations used in major weapon-systems acquisition programs. The complex and technical nature of the simulations and our 14 factors called for a method suited to in-depth assessment. The case study method was the most plausible for illustrating the application of the framework. One limitation of this approach is, of course, that it prevents us from generalizing from our findings regarding the credibility of the simulations we selected to any other simulations.

The Structure of This Report

Our findings are presented in chapters 2 and 4-7. In chapter 2, we describe concepts others have used in assessing simulations and the framework we developed. In chapter 3, we describe the weapon systems and the simulations in our three case studies. This provides important background material for understanding our findings in the three subsequent chapters. In chapters 4-6, we address the three major areas of concern in our assessment framework. Table 1.1 shows this structure.

Table 1.1: The Structure of This Report

Question	Discussion
1. What factors should be considered in a systematic attempt to assess the *credibility of a simulation*?	Chapter 2
2. What are the results of an assessment of selected weapon-system operational-effectiveness simulations with respect to these factors?	
a. Background data on the 3 case studies	Chapter 3
b. The credibility of a model based on theory, model design, and input data	Chapter 4, appendix II
c. The credibility of a model based on correspondence between the model and the real world	Chapter 5, appendix III
d. The credibility of a model based on support structure, documentation, and reporting	Chapter 6, appendix IV
3. What efforts has DOD made to foster and reinforce the credibility of its simulations?	Chapter 7, appendix V

In chapter 4, we describe the importance of theory, model design, and input data as they contribute to credibility, and we discuss the applicable factors from our framework. We summarize examples from our analysis of the three case study simulations and include findings that illustrate their strengths and limitations. A more detailed discussion of these findings is in appendix II. We do the same in chapter 5 and appendix III, where the area of concern is the correspondence between a model and the real world, and in chapter 6 and appendix IV, where the area of concern is with a simulation's basic support structure, documentation, and reporting. In chapter 7, we examine the policies, regulations, and structures that DOD and the Army used to promote the credibility of the simulations with respect to their design, implementation, and management. Our findings are summarized in chapter 8, which also includes our recommendations to DOD. Appendix V contains comments from DOD about our draft report.

Factors in Assessing a Simulation's Credibility

Prior Research

Various procedures have been proposed to permit reasoned judgment concerning the credibility of simulation results. Several analysts have proposed structures for what are variously called "assessments," "evaluations," and "appraisals." While terminology and structure differ, a number of common themes appear. For example, S. I. Gass in 1983 proposed an assessment procedure that addresses 13 information items: (1) mathematical and logical description, (2) model documentation, (3) computer program documentation, (4) computer program consistency and accuracy, (5) overall computer program verification, (6) technical validity, (7) operational validity, (8) dynamic validity, (9) training, (10) dissemination, (11) usability, (12) program efficiency, and (13) overall model validation.[1] In 1979, we described 5 criteria necessary for evaluating models: (1) documentation, (2) validity, (3) computer model verification, (4) maintainability, and (5) usability.[2]

T. I. Oren in 1981 identified six components for systematically assessing the acceptability of a simulation study. They were (1) data, (2) model, (3) experimentation specification, (4) computer program, (5) methodology and technique, and (6) simulation results.[3] A framework is presented that allows an assessment of the concepts and criteria related to the acceptability of the components.

G. L. Harris's 3 items for gaining and maintaining credibility were (1) model qualification (focused on the simulated phenomenon's representation in theory and data), (2) computer model and program verification, and (3) general validation of the computer model.[4] Each item, in turn, was defined with a detailed procedural checklist.

Banks, Gerstein, and Searles developed a 7-step modeling structure that is both the framework for creating the model and the structure for performing the evaluation. The steps within the structure include (1) system feasibility, (2) requirements definition, (3) preliminary design, (4)

[1] S. I. Gass, "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," Operations Research, 31:4 (July-August 1983), 618.

[2] U.S. General Accounting Office, Guidelines for Model Evaluation: Exposure Draft, GAO/PAD-79-17 (Washington, D.C.: January 1979), p. 9.

[3] T. I. Oren, "Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference," Communications of the ACM, 24:4 (1981), 181.

[4] G. L. Harris, Computer Models, Laboratory Simulators, and Test Ranges: Meeting the Challenge of Estimating Tactical Force Effectiveness in the 1980's (Fort Leavenworth, Kansas: U.S. Army Command and General Staff College, 1979), p. vi.

detailed design, (5) coding, (6) testing, and (7) operations and maintenance.[5] A number of specific procedures and evaluation criteria are identified for each step.

Although the emphases may differ, the purpose of each assessment structure is to guide the analyst in determining a simulation's credibility. We used several structures in developing our framework. Since probably no framework can be exhaustive and also practical, we sought to highlight the most critical matters for determining the strengths and weaknesses of a simulation.

Our Framework

To assemble the factors necessary in any systematic attempt to assess credibility, we looked for factors that research and experience indicated should be linked to confidence. We found three major areas of concern and 14 factors.

Theory, Model Design, and Input Data

The first area of concern pertains to how a simulation model imitates a weapon and its environment. Matters of interest include the characterization of the weapon system and its operation in both its immediate environment and its larger combat arena, the mathematical representation of the real world, the indicators of the weapon's effectiveness, and the data for initiating the simulation and providing ongoing input. Briefly, the concern is with the theory that underlies the simulation, the design of the model, and the input data. These basic components in constructing a simulation determine the results and thereby seriously affect their credibility. We represent these concepts in the first 7 factors in table 2.1.

[5] J. Banks, D. M. Gerstein, and S. P. Searles, "The Verification and Validation of Simulation Models," School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 1986, pp. 5 and 28-118.

Table 2.1: A Framework for Assessing the Credibility of a Simulation

Area of concern	Factor
A. Theory, model design, and input data	1. Match between the theoretical approach of the simulation model and the questions posed
	2. Consideration of the weapon system's important operational measures of effectiveness
	3. Portrayal of the immediate environment in which the weapon will be used
	4. Representation of the weapon system's operational performance
	5. Depiction of the critical aspects of the broad-scale environment of the battle
	6. Appropriateness of the mathematical and logical representations of combat
	7. Selection of input data
B. The correspondence between the model and the real world	8. Evidence of a verification effort
	9. Evidence that the results are statistically representative
	10. Evidence of sensitivity testing
	11. Evidence of validation of results
C. The support structures, documentation, and reporting	12. Establishment of support structures to manage the simulation's design, data, and operating requirements
	13. Development of documentation to support the information needs of persons using the simulation or its results
	14. Disclosure of the simulation's strengths and weaknesses when the results are reported

Credibility as indicated by these 7 factors depends partly on how the simulation is intended to be used in decisionmaking. That is, it derives in part from the match between the simulation model and the purpose of the simulation. If critical features of the weapon system, its environment, and its operation in combat are not portrayed appropriately for the purpose of the simulation, the results may be inaccurate or irrelevant.

For example, if the ability of a missile's guidance system to function properly is an important concern to decisionmakers, then a model using a superficial characterization of guidance dynamics probably would not be suitable. But if the missile's guidance system is just a small part of much larger concerns about what happens in a multiweapon battle, it may be possible to model the guidance system in a very simple way without damaging the credibility of the results. Several of the first 7 factors focus attention on the match between the model and the proposed use of the simulation and its results.

Because all simulations depend heavily on judgment in selecting modeling techniques, identifying functional relationships, choosing scenarios, and selecting sources of input data in representing the real world, it is important that judgment be based on a knowledge of military operations, the physics of weaponry, the behavior of military personnel, logistics, and the results from tests of weapons and their use in combat. Incomplete knowledge and poor judgment may fundamentally distort the results, and evidence of such conditions will lessen the credibility of a simulation. The intent of several of the 7 factors is to manifest such evidence.

A Model and the Real World

The second area of concern is the correspondence between simulation outcomes and real-world outcomes, factors 8-11 in table 2.1. Of foremost concern in this context is the idea of "model validation," which refers to the process of determining the agreement between the real-world system being modeled and the model itself and, thus, determining whether the model is an accurate representation for a particular application.

Validation includes the application of tests to the simulation. Although no ultimate test or test sequence confers validity, a model can pass enough appropriate tests so that qualified researchers would say that it appears to be valid or that the results are credible. In the development and implementation of a simulation, attention must be given to the procedures (such as tests of face validity, or expert reviews of the model and its results) that will increase the correspondence between the results of the simulation and the results of operational testing, combat operations, and other simulations. For a number of reasons, such as limited resources and data, validity checks may be performed rarely or very weakly. Credibility is seriously threatened if little or no evidence demonstrates that results correspond closely to reality.

A related but narrower idea is that of "verification," which refers to the process for determining that a computer-based model performs as the program analysts intend, that the computer programming is correct and internally consistent.[6] The lack of evidence that programming errors have been sought and removed lessens the credibility of the model and its results, even when the theoretical formulation of the simulation is considered to be fundamentally correct.

Support Structures, Documentation, and Reporting

The third area of concern is the institutional process covering practices such as configuration management, oversight and review, and documentation and reporting, which help ensure that credible simulations are established and maintained. Factors 12-14 in table 2.1 deal with this area.

Simulation models that exist independently of the problems they can address are often revised in order to correct errors or omissions, reflect current information about systems or the environment, respond to specific modeling needs, and operate with revised computer languages and new equipment. An organization responsible for simulations should have an established process for changing the features of a model, such as modifying the input data, the computer programs, or its documentation and copies.

For simulation models that are used by many analysts over a long period of time, modifications not centrally approved or disseminated can result in users' not knowing what features are and are not included in a simulation. Such uncontrolled changes coupled with weak documentation can make it difficult for analysts and managers to understand how the results were derived. Furthermore, when the results are reported without sufficient detail about the simulation's capabilities and limitations, decisionmakers may risk using those results inappropriately. These threats to credibility undermine the user's ability to understand and use a simulation.

Summary

By addressing the 14 factors in our framework and by collecting and reviewing the information available for each of them, we believe one can identify the strengths and weaknesses that affect the credibility of a simulation. We did not attempt to weight the 14 factors for their relative

[6]These definitions are commonly used in the operations research and modeling communities and they are the ones most often found in DOD documents. A few scientists define verification as agreement with reality and validation as the investigation of internal consistency. The concept of simulation validity is sometimes used in the literature to refer to the totality of a review framework. As we use it, however, validation refers to the process of developing confidence in the simulation results by comparing the simulation output with data from other sources.

importance or formulate an overall rating that a weighting system would produce. We believe that if a simulation is sound, applying our framework to it will reveal its soundness and reassure decisionmakers about using the results; if it is not sound, the framework will indicate the weaknesses.

In sum, the credibility of simulation results has been defined in terms of how much confidence one has that a simulation closely reflects reality. We have argued that credibility is accumulated from three kinds of evidence: (1) a model and its input data have appropriately portrayed the important features of the weapon system being simulated and its environment, (2) the model produces results similar to results from the real world, and (3) the procedures followed in developing, maintaining, and using the model tend to minimize discrepancies between simulation results and real-world results.

Chapter 3

The Case Study Simulations

Early in our review, we believed it necessary to assess simulations within the context of their application, and we concluded that the best way to select candidate simulations for our case studies was to start with the weapon-system programs themselves. That is, by choosing a weapon system, reviewing its history, and talking with knowledgeable persons involved with it, we were led to the simulations that were used for it. We limited ourselves first to "major systems"—systems projected to cost at least $200 million for research, development, testing, and evaluation or $1 billion for producton. Then we imposed further conditions—a system's proximity to the full-scale production decision; the use of simulations in its research, development, testing, and evaluation; the existence of a body of empirical data; and its employment or control by low-level tactical units for which data were available. This led us to select the DIVAD and the Stinger as especially suitable weapon systems.

The Weapons

The air defense mission is to nullify or reduce the effectiveness of attack or surveillance by hostile aircraft or missiles after they are airborne, thereby supporting the fundamental Army function of conducting prompt and sustained land warfare operations. Protecting critical operational and strategic assets from enemy aircraft is a primary part of the mission; the attrition of enemy aircraft is secondary. Short-range air defense artillery units engage enemy close-air-support helicopters and fixed-wing aircraft, and when there is high-intensity conflict with enemy ground forces, engage ground targets in self-defense.

The DIVAD

The DIVAD was developed to replace the Vulcan air defense system, which was perceived as no longer able to defeat attack aircraft or armored assault helicopters. In addition to filling this void in the forward battle area, the DIVAD was to engage lightly armored vehicles, trucks, and personnel. The system was operated by a three-member crew.

The DIVAD's turret and other components, such as the prime power unit, were mounted on an M48A5 tank chassis, and, overall, the DIVAD closely resembled a tank. However, when its prominent radar antennae were extended, the system's height was 15 feet. The M1 tank's height, in comparison, is 8 feet. The DIVAD's major subsystems were the tank chassis; the turret, which contained most of the system's electronic equipment; and the radar, which was derived from the F-16 aircraft's radar. The radar was backed up by a fully integrated electro-optical sighting and ranging system consisting of a laser range finder and optical day sights.

Its primary armaments were twin 40-mm Bofors L70 guns that could be fired automatically or semiautomatically, either singly or in pairs. The ammunition for the system consisted of proximity-fused, point-detonating, and target-practice rounds. The system also had a 7.62-mm machine gun mounted on a pedestal next to the squad leader's hatch.

The request for proposals for engineering development for the DIVAD was issued in April 1977, and engineering development contracts were awarded to Ford Aerospace and Communications Corporation and General Dynamics Corporation in January 1978. After development and operational testing of the prototypes, Ford was awarded a fixed-price incentive contract to complete engineering development in May 1981. In May 1982, the DIVAD passed its program review, and the production of 50 systems was authorized. In May 1983, an additional 96 systems were authorized, and additional testing and evaluation followed. The DIVAD weapon-system program was cancelled in August 1985.

The Stinger

The Stinger is a passive, shoulder-fired, infrared-seeking, guided missile with an antiaircraft, air defense mission to fulfill Army, Marine Corps, and Air Force requirements. The 34.5-pound weapon system consists of a missile in a launch tube and a reusable gripstock containing the firing circuits and identification-friend-or-foe (IFF) electronics. Both the gunner and crew chief may acquire the target and fire the weapon, although the crew chief generally fires only when the gunner is engaged with another target. Acquiring a target includes an interrogation with the integral IFF system. If the target proves hostile, the missile is launched to intercept and destroy it. After the missile has been launched, the crew member is free to engage another target, take cover, or move to another location.

The Stinger's mission is to provide air defense support in forward battle areas and to high-priority resources throughout the divisional areas of operation. The Stinger's concept definition began in 1968 in response to combat deficiencies in the Redeye. The system's design was completed by December 1972. In April 1978, full-scale production began, and initial operational capability was achieved in February 1981. In June 1977, however, the Army had begun the engineering development of an improved version, known as the Stinger-POST, whose full-scale production began in July 1985. Another improved version, with a reprogrammable microprocessor, began development in September 1984.

The Stinger is used throughout the battle area. In the rear, it is used as a point air defense weapon for high-value resources, and in the forward area it is used against high-speed, low-level, ground-attack aircraft and helicopters. Additional capabilities are being designed so that it can be used at night, as an air-to-air missile for helicopter use, and in a new lightweight air defense system. In 1984, the inventory requirement for the Stinger was more than 60,000 missiles for the Army, Marine Corps, and Air Force.

The Simulations

Within the research, development, testing, and evaluation programs for the DIVAD and Stinger weapon systems, we found a number of simulations used to answer a variety of questions pertaining to the systems' concepts, engineering design and performance, costs, and operational effectiveness. The air defense air-to-ground engagement simulation (ADAGE—consisting of two "submodels," called "Incursion" and "Campaign") and the Carmonette simulation, both used in the DIVAD's acquisition program, and the COMO III air defense combat simulation, used for the Stinger's program analyses, were concerned with operational effectiveness; we focused on this because it is of interest to decisionmakers. These three simulations varied in a number of key features, including the type of simulation model, the treatment of uncertainty, size and duration of battle, attrition calculations, the coverage of air-to-ground interaction and ground battle, the coverage of resupply, and computer running time. These features are summarized in table 3.1.

Table 3.1: The Key Features of the ADAGE, Carmonette, and COMO III Simulation Models

Feature	ADAGE Incursion	ADAGE Campaign	Carmonette	COMO III
Model type	Functional	Functional	Combined arms	Functional
Treatment of uncertainty	Monte Carlo	Expected value, or deterministic	Monte Carlo	Monte Carlo
Size of battle	Division	Division	Battalion	All levels up to theater
Length of battle	Not applicable	Several days	Short, intense firefights, about 25 minutes	Short battles up to 2 hours
Attrition calculation	One-on-one models of each air defense weapon type against each target type	Probabilities developed in Incursion	Monte Carlo models of specific events using one-on-one data	Monte Carlo models of specific events using one-on-one data
Treatment of time	Sequenced by time	Calculated at end of mission	Sequenced by event	Sequenced by event
Air-to-ground interaction	None	Played	Played	Played
Ground battle	None	Played using data outside the model	Played	None
Resupply	Not applicable	Played	None	None
Computer time	Short	Short	Long	Long

Most of the features are self-explanatory or are covered in detail in later chapters and appendixes in this report. A few are described here. Functional models study a particular military function, such as air defense, whereas combined-arms models evaluate alternative combinations of combat forces, such as alternative combinations of armor, infantry, artillery, and air support for a given level of battle.

In the treatment of uncertainty by Monte Carlo modeling, important real-world parameters are described by means of probability distributions. A very large number of random inputs is sampled from those distributions and the simulation result itself is expressed as a distribution. In contrast, in the expected-value (or deterministic) approach, mathematical expectations, generally the mean of a distribution, summarize the random variables that describe real-world conditions. Such a model is deterministic because the result it produces is certain to follow from the initial conditions.

ADAGE

The ADAGE model is a functional simulation used to study the relative effectiveness of combinations of air defense weapons in a division. The Incursion submodel uses the Monte Carlo methodology to model the attrition of a single-threat aircraft from a single ground-based weapon. The Campaign submodel then uses these engagement attrition data from the Incursion submodel to calculate expected value results for a specific scenario of many weapons and targets.

The ADAGE Incursion simulates detection, threat reaction, the masking of the threat aircraft, reloading, and weapon-to-target interactions. The ADAGE Campaign simulates small raids by enemy aircraft attacking division ground targets over a span of several days. In the Campaign submodel, the number of air defense weapons and other ground weapons destroyed is based on an expected value derived from the number of attacking aircraft, the type of ordnance, and the type of target. Measures of effectiveness include the number of threat aircraft destroyed, the number of air defense and other ground weapons remaining and the number destroyed, the amount of air defense ammunition used, and the number of friendly aircraft remaining.

The ADAGE was developed by the U.S. Army Materiel Systems Analysis Activity specifically to study the DIVAD. It was used first for the division air defense cost-and-operational-effectiveness analysis conducted in

1977.[1] It was also used for the 1984 update of this analysis and earlier in 1979, for the short-range, air defense, portable force structure analysis and in 1985 for the DIVAD comparative analysis. The ADAGE has been used for other air defense studies as well.

Carmonette

Designed about 30 years ago, the Carmonette is a combined-arms combat model that simulates small-unit, ground combat involving the actions of individual soldiers and weapons. Analysts design small-unit engagements to examine specific questions such as, "In a battalion assault, what are the trade-offs between armor, infantry, and artillery?" The Carmonette includes all combined arms: infantry, mounted or dismounted; artillery, including air defense artillery, and mortars; and armored vehicles and helicopters. Even though the Carmonette was designed to simulate weapon-to-weapon duels, its proper use is for larger engagements of combined-arms actions in which weapon-to-weapon data are used as input. The focus of the Carmonette is the battle, not individual weapon systems. The Carmonette assumes an intense 25-minute battalion task force battle.

The Carmonette has been used extensively to model ground warfare. The U.S. Army Training and Doctrine Command has characterized it as an operational-effectiveness model in which the various systems on the battlefield are related in a way that allows for an investigation of their synergism. In addition to its ground warfare applications, the Carmonette was used in the 1984 and 1985 analyses of the DIVAD and in advanced-attack helicopter and antihelicopter studies.

COMO III

The COMO III, used primarily for studies of tactical air defense effectiveness, is a Monte Carlo, functional simulation in which particular submodels are combined to simulate a specific air defense environment. Weapon-system submodels include specific ground-based air defense and threat aircraft, and other submodels simulate functions such as communications and jamming.

The scale of battle can range from individual battles to a division to the theater. Time, in the range of 2 hours, generally represents a period

[1]This type of analysis is a comparative evaluation of alternative systems, their contribution to the force, and their costs in personnel and funds. Its purpose is to assist in the selection of a preferred course of action to meet a stated Army need. It is conducted prior to each acquisition milestone decision for major systems and other systems designated by the Army. Among its many subanalyses, the analysis of effectiveness is usually the most controversial.

short enough that logistic support is not an issue. It is a standard Army model for tactical air defense artillery effectiveness studies.

COMO III was developed in 1966 in the Netherlands by the technical center of the Supreme Headquarters of the Allied Powers in Europe as an advance over an earlier model. It has been used to investigate broad air defense concepts, the effectiveness of particular weapon systems, naval task force air defense, and the air defense structure of the Warsaw Pact nations, among others. It was used to evaluate the Stinger in conjunction with other air defense weapons and to determine the Stinger's support requirements. (The COMO III simulation report we examined was entitled the "Stinger Battery Coolant Unit Usage Study.")

Credibility Based on Theory, Model Design, and Input Data

In this chapter, we focus on the first area of concern in our framework, the simulation model and its underlying theory, model design, and input data and the 7 factors we identified for it in chapter 2 (see table 4.1). Information about how the ADAGE, Carmonette, and COMO III models were used in effectiveness analyses of the DIVAD and the Stinger may be found in appendix I, and a more detailed discussion of the findings in this chapter appears in appendix II.

Table 4.1: The Seven Factors for Theory, Design, and Data[a]

Area of concern	Factor
Theory, model design, and input data	1. Match between the theoretical approach of the simulation model and the questions posed
	2. Consideration of the weapon system's important operational measures of effectiveness
	3 Portrayal of the immediate environment in which the weapon will be used
	4. Representation of the weapon system's operational performance
	5. Depiction of the critical aspects of the broad-scale environment of the battle
	6. Appropriateness of the mathematical and logical representations of combat
	7. Selection of input data

[a]The two remaining areas of concern and 7 other factors are in table 2.1

The Match Between the Theoretical Approach and the Questions Posed

A simulation quite credible in the abstract may not meet the specific needs of its user, depending on the model's theoretical approach. The purpose may have been to create an engineering model to determine the optimal design of a weapon relative to its technical requirements, a functional model to aid in selecting the most effective weapon system from alternative systems performing the same functional element of combat, or a combined-arms model to compare alternative combinations of complementary weapon systems (for example, air defense weapons, infantry, helicopters, and tanks). Table 4.2 summarizes our case study assessment of this factor.

Table 4.2: The Match Between Theory and Questions

Weapon	Model	Strength	Limitation
DIVAD	ADAGE	Functional model designed for DIVAD and other air defense studies; useful for comparing alternative air defense systems	Expected-value approach; probabilities developed in the first submodel; incomplete consideration of the random factors of modern warfare in second submodel
	Carmonette	Combined-arms Monte Carlo model for broad questions of warfare; treats the random factors of warfare probabilistically	Emphasizes ground battle; not well-suited for studying the effectiveness of competing air defense systems; air defense a recent add-on, especially for fixed-wing aircraft; not focused on individual weapon systems
Stinger	COMO III	Functional Monte Carlo model for air defense issues; useful for comparing alternative air defense systems	Absence of ground battle modeling suggests that simulation of air defense in the more forward areas may be missing an important element of realism

A functional air defense model was a reasonable choice for studying the DIVAD's performance in comparison with other air defense alternatives. The ADAGE model emphasizes ground-based air defense weapons and otherwise generally focuses on how changes in air defense capability can change outcomes in ground and air-to-air battles.

The Carmonette was designed to answer broad trade-off questions beyond issues of air defense. As a combined-arms model, it is generally not as well suited to answering the questions about air defense alternatives that were posed about the DIVAD. The model attempts to portray an overall ground battle with limited air war features but is not focused on individual weapon systems.

The COMO III is similar to the ADAGE in that it is a functional model designed specifically to study air defense issues. In general, the COMO III model is properly matched to the questions asked about the Stinger. It was based on a standard scenario generated by the U.S. Army Air Defense Artillery School.

Operational Measures of Effectiveness

If the measures of effectiveness a simulation addresses are not related to the weapon system's mission, conclusions about the system's performance in combat may not be credible, even if the simulation is sound in other respects. The first mission of air defense systems is to protect critical resources from enemy aircraft; the second is to destroy enemy aircraft. Therefore, we looked for the coverage of measures of effectiveness reflecting these missions. Table 4.3 summarizes what we found.

Table 4.3: Operational Measures of Effectiveness

Weapon	Model	Strength	Limitation
DIVAD	ADAGE	Emphasizes the protection of critical assets as well as giving attrition factors	No coverage of effects of aircraft mission aborts; the effect of ground losses to enemy air attacks, an important factor in measuring operational effectiveness, appear excessive
	Carmonette	Reports mission aborts and helicopter remaskings caused by air defense artillery and radar warning	Emphasizes attrition factors with little coverage of protection of critical assets
Stinger	COMO III	Presents wide range of measures	No modeling of ground battles limits capacity to measure protection of critical assets; concentrates on attrition factors

Both the ADAGE and Carmonette simulations provide for the protection of critical resources to some degree; the former emphasizes it, whereas the latter emphasizes measures of aircraft attrition. Although the COMO III concentrates on measures of both attrition and weapon usage, it is more limited in its ability to use the preservation of resources as a principal measure of effectiveness, because ground war is not simulated. This threatens the credibility of the results of this simulation.

The Portrayal of the Immediate Environment

In looking at how adequately a simulation model portrays a weapon system in its immediate wartime environment, we focused on five attributes of a plausible battle scenario: the size of the battle, the duration of the battle, the nature and behavior of enemy targets, the deployment and movement of the weapon being evaluated, and the terrain over which the battle might take place. These attributes are summarized in table 4.4.

Table 4.4: Portrayal of the Immediate Environment

Weapon	Model	Attribute	Strength	Limitation
DIVAD	ADAGE	Battle size	A division model for a weapon with division-level responsibilities	
		Battle length	Covers up to 30 days, permitting the measurement of the cumulative effects of air defense	
		Target	Covers all potential targets, including helicopter and fixed-wing aircraft	Covers only nonjinking helicopters and aircraft with fixed flight paths
		Deployment and movement		Deployment of ground assets is static, movement is only indirectly modeled
		Terrain	A statistically general terrain that can be generalized to many areas	A statistically general terrain representing no "real" terrain
	Carmonette	Battle size		A battalion model for a weapon with division-level responsibilities
		Battle length		A 25-45-minute firefight that ignores the cumulative effects of air defense
		Target		Stresses helicopters, most studies did not include fixed-wing aircraft
		Deployment and movement	A fully dynamic model capturing the effects of movement of ground weapons	
		Terrain	A digitized, specific, "real" terrain	A digitized, specific terrain that cannot be generalized to other areas
Stinger	COMO III	Battle size	Covers all levels up to brigade, capturing the full range of air defense responsibilities	Portrayed a limited environment because a larger scenario would have been too intensive a use of computer resources
		Battle length		Covers short battles up to several hours, ignoring the cumulative effects of air defense
		Target	Covers the engagement of helicopters and fixed-wing aircraft	
		Deployment and movement		Static deployment of ground assets movement is only indirectly modeled
		Terrain	A digitized, specific, "real" terrain	A digitized, specific terrain that cannot be generalized to other areas

The evidence indicates that the ADAGE and COMO III can simulate a weapon system's immediate environment across these attributes with some limitations. Both are strong in characterizing the size of battle and the full range of targets. The ADAGE simulates longer battles but is limited by its uniform and static deployment of weapons. The COMO III portrays a shorter battle with the Stinger weapons; they are deployed realistically but do not move, a limitation for portable systems for which movement provides a form of individual defense at the cost of decreased operability. The COMO III and ADAGE use different approaches to portray terrain. The COMO III simulates specific terrain; the ADAGE uses a statistical portrayal. Neither is obviously superior to the other.

The Carmonette is more limited in its ability to portray the immediate environment than the ADAGE and COMO III. The battalion size, which is small, and the short duration of the battle are inappropriate for the DIVAD weapon, and the lack of fixed-wing aircraft targets for most of the analyses we examined resulted in an incomplete set of targets. These limitations were partially offset by the Carmonette's realistic portrayal of deployment, movement, and terrain but nevertheless threatened its credibility.

Operational Performance

We assessed the simulations across several attributes of a battle with respect to the weapon systems' operational performance, covering both detection and engagement. Four attributes pertained to the simulation of target detection: visual detection; factors that might lessen battlefield visibility; command, control, and communication, including the problem of distinguishing between friend and foe; and, for the DIVAD, radar detection.

Both the ADAGE and COMO III simulations are limited in the way they depict the detection of enemy targets. For example, the ADAGE only indirectly addressed the confusing elements of combat—battlefield obscurants; command, control, and communication; and IFF. The ADAGE also used indirect means to portray radar detection. The COMO indirectly includes battlefield obscurants and omits IFF. Our review of the Carmonette simulation, however, indicates its ability to address these more directly, although the features of the Carmonette that permit the simulation of IFF and command, control, and communication were not used in the DIVAD simulation. Our results are summarized in table 4.5.

Table 4.5: Portrayal of Key Detection Characteristics

Weapon	Model	Attribute	Strength	Limitation
DIVAD	ADAGE	Visual detection	A visual detection submodel used in early studies covered weapon's full range	Determined only in the first submodel; to achieve full range detection, later studies using a night vision and electro-optical laboratory model had to use forward-looking infrared capabilities that were not part of the DIVAD
		Battlefield obscurants		Only indirect play through probability of weapon participating in air battle. no night play
		IFF and command, control, and communication		Only indirect play through a visual detection submodel
		Radar detection	Covers gun's full range	Only indirect play through input data adjustments. aircraft do not react to radar warnings
	Carmonette	Visual detection	Fully dynamic but with range limits; later studies using the visual detection submodel for detecting fixed-wing aircraft covered DIVAD's range limits	Used forward-looking infrared in a night vision and electro-optical laboratory model to detect helicopters
		Battlefield obscurants	Covers night and most obscurants	
		IFF and command, control, and communication		Model capabilities not used
		Radar detection	Well detailed, early weaknesses overcome	
Stinger	COMO III	Visual detection		Limited range, using look-up tables and the same search procedures for fixed-wing aircraft and helicopters
		Battlefield obscurants		Only indirect coverage, using degraded detection probabilities
		IFF and command, control, and communication		Not modeled
		Radar detection		Not applicable to Stinger

Three of the attributes we examined pertained to a weapon's engagement of a target after detecting it. The first of these was the characteristics of the weapon system such as technical capability and operating modes. The second pertained to if and how an enemy target is actually engaged, called "engagement procedures." For example, a model might or might not include the engagement of an enemy aircraft flying past the air defense weapon en route to another target. And, finally, we looked at whether and how the models handle raids by multiple aircraft. See table 4.6.

Table 4.6: Portrayal of Key Engagement Characteristics

Weapon	Model	Attribute	Strength	Limitation
DIVAD	ADAGE	Weapon characteristics	Coverage of technical capabilities and targets	
		Engagement rules and procedures	Description of weapon and how it engages different types of aircraft; coverage of engagement of aircraft flying by or attacking defended targets	No play of duels
		Multiaircraft raids	Includes raids	Excludes spatial and temporal saturation effects
	Carmonette	Weapon characteristics	Corrected for erroneous early descriptions	
		Engagement rules and procedures	Prioritizes targets	Ignores aircraft flying past defended targets
		Multiaircraft raids	Permits selection from several targets	
Stinger	COMO III	Weapon characteristics	Uses separate weapon programs adaptable to studying weapon modifications, good description of weapon characteristics	
		Engagement rules and procedures	Allows player to select from alternative procedures; different firing doctrines can be specified	
		Multiaircraft raids	Saturation can be demonstrated; good vehicle for demonstrating Stinger operations in conjunction with other air defense weapons	

The evidence indicates that all three models portray engagement characteristics in considerable detail; COMO III has perhaps the best coverage. The ADAGE simulation was clearly limited in its treatment of multiaircraft raids, which did not adequately account for how the raids could saturate the defense; and the Carmonette model tended to ignore aircraft passing through the battle area. The relative strengths of these models in simulating the engagement aspects of a battle contributed to their credibility.

The Broad-Scale Battle Environment

When seeking to determine the effectiveness of a weapon system, attention is focused on the particular weapon, but other features of a battle must also be taken into account. Air defense usually does not operate in isolation, and other aspects of an ongoing battle may affect the operation of weapons such as the DIVAD and Stinger. In assessing these air defense simulations, we tried to take account of the bigger picture by looking at three battle attributes that we labeled the air war, the ground

Table 4.7: Portrayal of Broad-Scale Battle

Weapon	Model	Attribute	Strength	Limitation
DIVAD	ADAGE	Air war	Notes damage from fixed-wing aircraft; plays air-to-air war	Treats saturation attacks inadequately
		Ground war		Uses attrition rates generated only outside the model
		Interaction	Shows the relationship of air and ground wars	Plays air and ground wars not interactively but through expected values
	Carmonette	Air war		Fixed-wing aircraft not modeled in early studies and modeled only indirectly in later studies
		Ground war	Fully developed ground battle	
		Interaction	Fully dynamic interaction for helicopters	Uses a model similar to the ADAGE for multiaircraft raids by fixed-wing aircraft
Stinger	COMO III	Air war	Detailed model of air war	Excludes fratricide from air defense artillery
		Ground war		No ground war
		Interaction		No interaction except for ground damage inflicted by aircraft, no ground-war damage to air defense

war, and the interaction between the two. Evidence of the three simulations' capabilities is summarized in table 4.7.

Our assessment indicates that the Carmonette has considerable ability in broad-scale battle, probably more than either the ADAGE or COMO III, largely because of its fully developed simulation of the ground battle. However, its simulation of the air battle limits its usefulness for air defense analyses. The COMO III's lack of a portrayal of the ground war is a serious limitation for studying the full range of air defense activities. The ADAGE included all three aspects of combat but the realism of its portrayal was limited.

Mathematical and Logical Representations of Combat

Having looked at the extent to which various aspects of a battle are credibly accounted for in the overall design of the simulations, we looked at their mathematical and logical representations. We noted only minor problems for the Monte Carlo models, and overall the mathematical and logical features of the Carmonette and COMO III contributed to the credibility of their results (see table 4.8).

Table 4.8: Mathematical and Logical Representations

Weapon	Model	Strength	Limitation
DIVAD	ADAGE		Uses expected value in many-on-many engagements; poorly understood parameter determines the probability of various air defense weapons participating in battle; survivability is based on attrition rates applicable to weapon classes
	Carmonette	Simulates specific dynamic interactions between individual air defense weapons and helicopters	Early problem of squaring of kill probabilities; generation of only one set of random numbers; fixed-wing model uses basically the same approach as the ADAGE in multiaircraft raids; problems external to the model in issues of experimental design and adequate number of model runs
Stinger	COMO III	Simulates specific dynamic interactions between individual air defense weapons and targets	Same as the Carmonette with regard to experimental design and number of model runs

The events of a battle may be computed and expressed as expected values or they may be computed less efficiently, but more realistically, by the Monte Carlo technique. The two procedures may not produce the same results. Each method may provide information not available from the other. Our main concern with the ADAGE simulation was that its use of "kill probabilities" based on the interaction of a single weapon and a single aircraft neglects the complexities of multiple aircraft attacks and could lead to substantial distortions of what happens in the real world.

The Selection of Input Data

The results of a simulation are dictated in large part by the data that an analyst enters into the computer: missile firing rates, target damage probabilities, information about the terrain, and so on. If the input data are basically inappropriate or problems arise from tailoring the data before they are used in the model, the credibility of the results is likely to be diminished. In our assessment, we attempted to determine the data shortcomings in the case study simulations.

The Carmonette and COMO III appeared to have relatively appropriate data. In the earlier analyses, the ADAGE and Carmonette modelers differed in the selection of input data and models for the visual detection of approaching aircraft. In the later compromise, the data did not properly describe the DIVAD's detection capabilities. The ADAGE simulation had the most serious input data limitations, because some of its data were outdated and some key values (such as air damage to ground targets) produced results too large to be accepted by knowledgeable military officials. Table 4.9 shows that all three models had some limitations with regard to this factor.

Table 4.9: The Selection of Input Data

Weapon	Model	Attribute	Strength	Limitation
DIVAD	ADAGE	Data source	Uses data from a variety of recognized sources	
		Data quality	Visual detection data cover full range of gun for helicopters	Visual detection and terrain data are old; night vision and electro-optical laboratory data are inadequate for the DIVAD's ability to detect aircraft to the full range of the gun
		Data tailoring		Description of weapons in Incursion submodel is an integral part of the model and not addressed through a data base
	Carmonette	Data source	Uses data from a variety of recognized sources, some different from the ADAGE's sources	
		Data quality	Uses a visual detection submodel from the ADAGE for fixed-wing aircraft, early problems using Soviet ZSU-23 to model the DIVAD were overcome	Uses night vision and electro-optical laboratory data inaccurate for the DIVAD's visual detection of helicopters
		Data tailoring		Data tailored extensively to meet model requirements could affect results
Stinger	COMO III	Data source	Uses data from a variety of recognized sources, some different from the ADAGE and Carmonette sources	
		Data quality	Engineering data are reasonably reliable	Human-factors data are not as reliable as engineering data
		Data tailoring	Straightforward for engineering data	Data about the Stinger team's reactions may have been subject to greater adjustment or interpretation than engineering parameters

Some of the Carmonette's early data problems, such as an incorrect description of the DIVAD gun, were corrected, but the problems with disputed visual-detection data remained, and disputes concerning these data required the ADAGE modelers to change their detection data. The Carmonette and COMO III simulations require extensive tailoring of data in order to make the data usable in the models, opening the possibility that the results may depend as much on the judgment of the staff as on the operations the model simulated.

Summary

All three simulations had considerable capability with regard to portraying weapons engaging targets and simulating important aspects of measures of effectiveness. In almost all instances, however, the simulations we studied had some limitations. We believe that the effort required to remove some of the limitations we found might be relatively minor, but for others, much more work would be required. In a few instances, fixing the model might not be the appropriate response; using a different model might be more appropriate. For example, our assessment indicates that the Carmonette, as a combined-arms battalion-level model, was generally not as well suited to answering the original questions posed about the DIVAD as an air defense alternative, so that modifying the model is probably not a reasonable solution to the limitation.

Chapter 5

Credibility Based on Correspondence to the Real World

In this chapter, we focus on factors 8-11 in our framework; or the procedures with which the analysts demonstrate that a model is a good representation of reality and that the results are acceptable surrogates for results that might be collected in the operation of a weapon system. In table 5.1, the factors are repeated from table 2.1.

Table 5.1: The Four Factors for Correspondence to the Real World[a]

Area of concern	Factor	
The correspondence between the model and the real world	8	Evidence of a verification effort
	9	Evidence that the results are statistically representative
	10	Evidence of sensitivity testing
	11	Evidence of validation of results

[a]The two remaining areas of concern and 10 other factors are in table 2.1

While analysts can never provide absolute guarantees about the credibility of a model or its accuracy, they should be able to provide information so that the required decisions can be made with some degree of confidence. They can produce evidence that (1) the computer program operates as the simulation model's designers intended, (2) the output of the simulation represents the model's average output over many runs, (3) the results take into account sensitive parameters and alternative scenarios, and (4) a model's results bear sufficient resemblance to real-world results or results from other models or methods. In reviewing the simulations, we paid some attention to the use of COMO III with weapon systems other than the Stinger, because the information contributed to the credibility of the COMO modeling system. (A more detailed discussion of our findings is in appendix III.)

Verification

The process of verification, or determining that the computer programmer has translated a model into correct computer code, may be performed as part of the programming and checkout phases of a simulation's development. These phases are often not documented; that is, they may be performed, but the history of the performance is usually not recorded. Consequently, it is often difficult to find written evidence of verification.

In our case studies, no documentary evidence of verification was available for either the ADAGE or the Carmonette. DOD personnel involved with the ADAGE informed us that some checks of the computer code had been made and problems had been found and corrected. The Carmonette analysts reported that some peer review had been performed. We were unable to document any verification of the COMO III Stinger model or the variant that was developed for the Stinger's battery coolant unit analysis. (See table 5.2.)

Table 5.2: Evidence of Verification

Weapon	Model	Strength	Limitation
DIVAD	ADAGE	Analysts commented that computer code was checked and errors were corrected	No formal efforts documented
	Carmonette	Extensive peer review was reported	No formal efforts documented
Stinger	COMO III	U.S. Army Missile Command staff verify and validate contractors' simulations as a standard procedure	No specific verification efforts identified

The lack we found of documented evidence of verification presents a clear threat to the credibility of the three simulations. The recollections of some analysts have some value, but written documentation would be preferable.

Statistical Representation

Credibility rises as a model's users become assured that its statistically averaged results do not vary widely when the model is exercised several times. It is important to know whether the results of one or a few runs reasonably represent the values that would be developed if a simulation were operated an indefinite number of times.

The Incursion submodel of the ADAGE, using the Monte Carlo modeling technique, uses multiple runs to determine one-on-one kill probabilities that are then used in the Campaign submodel. Analysts who worked with ADAGE informed us that each Incursion scenario had been run 500 times and that the resultant mean was within 1 or 2 percent of the true mean at the 98-percent confidence level. This is substantial support for the simulation's credibility.

Each run of the Carmonette, however, required a substantially larger commitment of computer resources. Therefore, the analysts used a limited number of replications, generally 10 for a scenario. Replications of the scenarios brought many of the aggregated results to within 10 percent of the true mean at the 85-percent confidence level. Similar levels of confidence were not achieved for individual weapon systems, so that questions remain as to whether the Carmonette's battalion-level results can be extrapolated to the division. The Carmonette's analysts doubted that they would be able to improve the confidence with a reasonable number of additional replications.

The COMO III simulation of the Stinger made only one run for each scenario, and the report of the analysis did not address the statistical representativeness of the results. Thus, we do not know whether the differing results from scenario to scenario came from differences in the scenarios or random variation inherent in the model. The extent to which statistical representativeness supports or threatens credibility is quite mixed across the simulations, as can be seen in table 5.3.

Table 5.3: Evidence of Statistical Representation

Weapon	Model	Strength	Limitation
DIVAD	ADAGE	Probability of kill developed with multiple replications; statistical procedures developed kill probabilities within 2 percent of true mean at 98-percent confidence level	
	Carmonette	Multiple runs on many scenarios provided confidence in results; many results were within 10 percent of true mean at 85-percent confidence level	Either model variability or insufficient replications prevented development of confidence levels for some results
Stinger	COMO III		No evidence of testing for measures of the mean and variance of results prior to experimenting with alternative scenarios; simulation appeared to move directly to scenario and some parameter testing; information on confidence in results was not developed because there was only one run per scenario

The large number of replications and the quality of results in the ADAGE simulation enhance its credibility. For the Carmonette, the analysts addressed statistical representativeness but with only limited success. Thus it has some credibility but not that of the ADAGE simulation. The COMO III simulation appears not to have addressed the need for developing statistically representative values. This constitutes a threat to the credibility of the simulation.

Sensitivity Testing

It is important to know how sensitive a simulation's results are to errors or fluctuations in the values of its input parameters. Some parameters, such as the detection range of a missile system, may be in considerable doubt; others, such as visibility, may simply be subject to wide variation. If a model is especially sensitive to a parameter, then the credibility of the results will be lessened if the estimate of the parameter is in error. Sensitivity testing helps determine whether there may be a problem.

A related issue is that the effectiveness of a weapon system may vary substantially as the combat scenario changes. For example, a surface-to-air missile system may be effective against attack aircraft but easily defeated if jamming is used. A scenario can be tested by running a simulation model under a wide variety of realistic battle conditions in order to obtain a broad view of a weapon's effectiveness. This may be viewed as testing the sensitivity of the simulation results to variations in scenarios. Table 5.4 summarizes the extent and manner in which the ADAGE, Carmonette, and COMO III were tested for sensitivity in our case studies.

Table 5.4: Evidence of Testing for Sensitivity to Parameters and Alternative Scenarios

Weapon	Model	Test	Strength	Limitation
DIVAD	ADAGE	Parameters	In detailed analysis of four major parameters, three were found to have a major effect on the weapon's effectiveness	
		Scenarios	Scenarios investigated weapons, environment, and alternative threats	
	Carmonette	Parameters	Investigated in scenario tests; some scenario changes were slight enough to be equivalent to parameter changes	
		Scenarios	Investigating many scenarios gave insights on relationships between visibility and the weapon's effectiveness	
Stinger	COMO III	Parameters	Visibility parameter tested	Additional runs needed
		Scenarios	Range of scenarios tested	Only one run per scenario

According to the ADAGE documentation, including the comparative analysis and cost and operational-effectiveness reports, the ADAGE modelers tested four parameters they believed could cause substantial error in conclusions about the DIVAD's effectiveness if the parameters were in error. They experimented with scenarios for variations in threat levels, environment, and the use of other air defense weapons, thus developing valuable information on the simulation's response.

Extensive experimentation with scenarios was also performed with the Carmonette. More than 50 different scenarios were examined in the simulations presented in the 1984 and 1985 reports. Many involved a major change, such as the addition or deletion of a type of weapon system, but some were relatively minor and might be better thought of as sensitivity analyses of specific parameters. There was no formal, separate parameter testing for the Carmonette, although there is evidence that such testing was performed on earlier versions of the model that did not include the DIVAD component. Tests of alternative scenarios provided important insights on the effectiveness of both the DIVAD and total battalion defense with regard to visibility, mode of operation, and current versus mature DIVAD capabilities.

The report documenting the COMO III simulation analysis indicated that sensitivity testing was performed for visibility. The analysis addressed 11 scenarios that considered a broad range of air defense, threat, and visibility conditions.

Sensitivity testing can contribute directly to an understanding of a model's behavior and to its credibility, and it did so for all three we examined. The ADAGE analysts used both parameter testing and experimentation with alternative scenarios to examine simulation results. The credibility of both the Carmonette and the COMO also benefited from the use of parameter tests and alternative scenarios.

Validation of Results

Validation, in a narrow sense, is the comparison of simulation results to results from other methods, such as operational testing and evaluation or historical experience, or from models for estimating a weapon's performance that are believed to be substantially credible. The limited evidence from our case studies suggests that validation is not planned for or conducted routinely but is more likely to be performed when a disparity is found in the results of similar models or between the model and real system data. Analysts or others in DOD may then request a resolution or an explanation. Our conclusions about validation efforts for the simulations we studied are summarized in table 5.5.

Table 5.5: Evidence of Validation

Weapon	Model	Test	Strength	Limitation
DIVAD	ADAGE	Other models	Two major comparisons attempted with the Carmonette: early effort was thought to give good correspondence, but comparison after changes was unsuccessful	No validation prior to Carmonette comparison
		Operations		No operational tests identified
	Carmonette	Other models	Same as ADAGE	No validation prior to ADAGE comparison
		Operations	The model was validated, but not with the DIVAD, against a tank warfare field experiment	
Stinger	COMO III	Other models	The model, but not with Stinger, was compared with an Air Force model, with a satisfactory resolution of initial differences	
		Operations		No operational tests identified

We found that no formal validation efforts using real-world, DIVAD data were performed on the ADAGE or Carmonette. This is not to suggest, however, that there was no attempt at validation. The Army regarded the use of the Carmonette to model the DIVAD as itself a validation effort for the ADAGE. It was made when questions arose about the results of the DIVAD's effectiveness as shown by the ADAGE. Its results differed substantially from those of the Carmonette and other air defense models. However, further analyses that adjusted the models for consistency in inputs (for example, the same number of air-to-ground munitions) and scenarios (for example, the same size battle) made the ADAGE results reasonably comparable to those of the other models. Later changes in the Carmonette model, however, led to differences in the adjusted results with a cause that could not be pinpointed.

We did not find evidence of validation specifically for the Stinger simulation. We did, however, find evidence of an effort to validate the COMO III model by comparing its results to those from an Air Force model called SORTIE. The reasonable agreement of results when simulating similar conditions suggests that model-to-model validation can marginally strengthen credibility, especially when comparisons with real-world data are lacking.

Efforts to validate the ADAGE and Carmonette with respect to the DIVAD were limited to comparing the two models to each other and, to a limited extent, to other models. The lack of validation success with the model-to-model comparison threatens the credibility of the models. With no direct validation, the COMO III situation was similarly weak. Yet the comparison with the SORTIE suggests that validation should be attempted and that even comparison between dissimilar models may improve a model's credibility.

Summary

Some of the efforts of the simulation analysts to show that the models we examined closely represent reality were very limited. Some validation was not even attempted. In general, the efforts to validate simulation results by direct comparison to data on weapon effectiveness derived by other means were weak, and it would require substantial work to increase their credibility. Credibility would also have been helped by better documentation of the verification of the computer program and by establishing that the simulation results were statistically representative. Probably the strongest contribution to credibility came from efforts to test the parameters of models and to run the models with alternative scenarios.

Chapter 6

Credibility Based on Support Structures, Documentation, and Reporting

Many simulation models have a long lifetime. They are created and modified, become more complicated, and are sometimes used in several versions. Because of this, simulation models, like all other complex software, must be supported by an organization that documents its operation and ensures that decisionmakers understand both the strengths and limitations of the model. We believe that this will not create credibility where the underlying theory, computer representation, or validation procedures are weak, but it will help prospective users judge the applicability of a simulation to their needs and will add further credibility if the simulation is relatively strong. Table 6.1 shows from our complete framework the relevant factors that we address in this chapter.

Table 6.1: The Three Factors for Support Structures, Documentation, and Reporting[a]

Area of concern	Factor
The support structures, documentation, and reporting	12 Establishment of support structures to manage the simulation's design, data, and operating requirements
	13 Development of documentation to support the information needs of persons using the simulation or its results
	14 Disclosure of the simulation's strengths and weaknesses when the results are reported

[a]The two remaining areas of concern and 11 other factors are in table 2.1

Support Structures for Design, Data, and Operations

Looking at Army actions relating to the ADAGE, Carmonette, and COMO III, we looked for evidence that support structures had been established for controlling the three models and evidence that any resultant organizations were functioning as intended. We found that each model had been assigned to a formal entity for management: the ADAGE to the U.S. Army Materiel Systems Analysis Activity, the Carmonette to the U.S. Army Training and Doctrine Command (TRADOC) Systems Analysis Activity, and the COMO III to the U.S. Army Missile Command. In addition, the Army designated the deputy chief of staff for doctrine responsible for ensuring that doctrine, future concepts, and threats are properly portrayed in the models.

Illustrating one type of support, TRADOC plays a role in both managing and using simulation models. Its regulation entitled "Management: TRADOC Models" (regulation 5-4, August 20, 1982) provides guidance on managing the models under its control. TRADOC designates one agency responsible for each model—for the development of software and for the management of the data base and changes in a model's configuration. Although others may use the model and may even make changes for their own needs, the alterations are controlled in that the nature of the model must not be changed, the changes must be coordinated with the responsible agency, and the changed model must not be shared with a third agency.

Several other groups play roles in controlling the models. For example, an interagency group was established in 1980 to exert some control over the COMO III's configuration and documentation and the development of new models. In 1986, a COMO model resources group was formally convened, again with the aim of providing some control over the model.

In an effort to maintain oversight and review at a different level, TRADOC establishes study advisory groups to monitor the progress of individual studies using models under TRADOC's control. For example, in two DIVAD studies, a 1984 cost and operational-effectiveness update and a 1985 comparative analysis, study advisory groups played active roles regarding the use of the ADAGE and Carmonette.

Another kind of control is exerted by weapon-system program offices, which sometimes establish working groups to oversee engineering simulations. For example, the Stinger program office appointed working groups to define the validation requirements for models and to review and approve validation data.

We looked beyond the mere establishment of a support structure to see if the organizations we identified were actively managing the simulation models and the associated studies of weapon systems. Some organizations have had a long-term relationship with a particular simulation—as the COMO model management board has had with COMO III—and others have had a brief but intense relationship, such as the study advisory groups that have the authority to advise on the use of a specific simulation model, the input data, or the scenarios in an analysis. We believe the long-term relationship is more likely to lead to a substantive effect on the credibility of simulation results. Our review of the support structures is summarized in table 6.2.

Table 6.2: Support Structures for Design, Data, and Operations

Weapon	Model	Strength	Limitation
DIVAD	ADAGE	U.S. Army Materiel Systems Analysis Activity is responsible for management, study advisory groups oversee and review specific studies	U.S. Army Air Defense Artillery School has been considered the appropriate manager for air defense functional models such as the ADAGE a study advisory group is organized for a specific study and does not focus on long-term configuration control of models
	Carmonette	U.S. Army Training and Doctrine Command is responsible for management; study advisory groups oversee and review specific studies	A study advisory group is organized for a specific study and does not focus on long-term configuration control of models
Stinger	COMO III	U.S. Army Missile Command is responsible for management, COMO model management board represents users from various agencies and meets periodically to guide development, configuration, and documentation; a COMO model resources group was also established to facilitate greater coordination among users	U.S. Army Air Defense Artillery School has been considered the appropriate manager for functional models such as the COMO

The Army seems to have been at least partially successful in maintaining simulation models and controlling their development and use. It assigned formal responsibilities for control for each of the case study models and involved several groups within the Army that have an interest in the development of specific models. The present structure for managing COMO III recognizes the different interests of those various groups and their viewpoints toward simulation.

Documentation for Users

Well-documented simulation models inspire confidence that the models will be used correctly to address the types of issues for which they were designed. Conversely, if documentation is incomplete, and especially if a model has been evolving for a long time, we are concerned that a model may not be simulating the events and conditions the analysts think it is. We looked for evidence of clear and complete documentation. What we found is summarized in table 6.3.

Table 6.3: Documentation for Users

Weapon	Model	Attribute	Strength	Limitation
DIVAD	ADAGE	Completeness	Original documentation is complete	Recent changes are not yet documented
		Adequacy	No major problems reported; the developer and user communicate frequently	
	Carmonette	Completeness	An executive summary and list of input variables are available	Detailed documentation is not available
		Adequacy		Lack of documentation was reported as a problem in understanding the results
Stinger	COMO III	Completeness	Comprehensive and detailed programmer-user manual is available; comparably complete documentation is available for other models and for the overall system	
		Adequacy	No problems reported or identified	Basic knowledge of COMO is required to use the manual

We found the ADAGE relatively well documented, at least through September 1978. However, the cost and operational-effectiveness update study for the DIVAD required substantial changes to the ADAGE that were not accounted for in the documentation.

The Carmonette is documented relatively poorly, which became evident during the cost and operational-effectiveness update study, when analysts at the U.S. Army Air Defense Artillery School tried to reconcile disparities in the results produced by the ADAGE and Carmonette. The analysts expressed doubt about being able to reach a reasonable understanding of the Carmonette without better documentation. The chairman of the study advisory group charged with overseeing the update also expressed concern about the lack of documentation.

The COMO series of models has extensive documentation. Documentation was produced in the late 1960's and early 1970's at the technical center of the Supreme Headquarters of the Allied Powers in Europe, where the COMO was developed. Since then, much of the documentation has been produced by or for the Army Missile Command as part of the process of developing and validating individual weapon-system models and improving the COMO's program structure.

We found the main documentation for the COMO III simulation of the Stinger comprehensive and detailed. Although validation documents were not available for the Stinger, they had been produced for corresponding COMO simulations of the Patriot and Hawk missiles.

In sum, the COMO III, and to a lesser extent the ADAGE, has documentation that tends to strengthen the user's confidence in the credibility of the simulation. The considerable lack of documentation for the Carmonette detracts from the confidence that a user might have in its credibility.

Reports of Strength and Weakness

In examining reports from the simulation studies, we wanted to determine the extent to which the simulations' strengths and weaknesses were discussed. We believe that the candid and complete discussion of a model is associated with a positive contribution to credibility.

The reports we examined included the following. For the ADAGE, we reviewed the report on the DIVAD's 1977 cost and operational-effectiveness analysis and the draft reports for its 1984 update and the 1985 comparative analysis. For the Carmonette, we reviewed the 1984 update on the cost and operational-effectiveness analysis and the 1985 comparative analysis. For the COMO III, we reviewed the Stinger battery-coolant-unit usage report, a validation report for the Patriot missile studies, and the documentation for the Stinger model. Our observations are summarized in table 6.4.

Table 6.4: Disclosure of Results

Weapon	Model	Strength	Limitation
DIVAD	ADAGE	Explicitly stated objectives, strengths, and weaknesses of the simulation analyses; the 1977 cost and operational-effectiveness analysis report was especially comprehensive	The 1984 draft update report and the 1985 draft comparative analysis report contained less description of underlying assumptions; the later report included fewer division-level analyses
	Carmonette	Included major modeling limitations	Contained cursory description of theoretical bases for analyses; did not address how limitations affected results; variability of results was not addressed; some recommendations not supported by analyses
Stinger	COMO III	Included details about the model and its limitations; report on validation of Patriot models is highly detailed reporting of strengths and limitations	Omitted description of some methodological and modeling weaknesses

The ADAGE reports contained explicit statements of the study's objectives and the strengths and limitations of the simulation. The 1977 report provided the rationale for studying air defense in a division context and identified the major measures of effectiveness. It explained the logic of the simulation, the relationship between the Incursion and Campaign submodels, and the manner in which air-to-air and ground battle results are integrated. Although the implications of the analysis of ground battle damage were not fully discussed, it was, on the whole, an adequate treatment of the simulation's strengths and limitations.

The 1984 update, which was issued only in a draft version, also clearly specified the purpose of the simulation. It did not cover the background information as intensively as the 1977 report, but it did address changes to the ADAGE model after 1977, and it contained a section reconciling the ADAGE results with the results produced by the Carmonette and other TRADOC models. The analysis of alternative air defense structures stated the assumptions and limitations clearly. Thus, except for not repeating the underlying assumptions, this report also contributed to the credibility of the simulation.

The results from the 1985 comparative analysis (also issued in draft only) tended to concentrate on outcomes pertaining to the protection of forward combat units and gave less attention to the division context. A more balanced presentation would have been more appropriate. Several limitations of the simulation were discussed and an attempt was made to identify and reconcile inconsistencies in the results of the ADAGE and Carmonette.

The Carmonette's 1984 update report appeared to make recommendations that were not well supported by the simulation's results, and little or no attention was given to the theoretical basis of the analyses. While some of the model's limitations were discussed, the authors did not address how they might have affected the results. There was substantial variance in the results of the runs, yet they were accepted without discussion of the effects of their variance or instability. The 1985 comparative analysis clearly stated the purpose of analysis and some of the major assumptions and limitations of the model. But many of the important areas not discussed in the 1984 update were still not completely addressed, and the analysis was again based on a small number of replications and unstable results. A summary statement about the report identified several major limitations of the simulation that, in our opinion, cast substantial doubt on the ability of the Carmonette to study air defense alternatives, although the statement itself did not draw such a broad conclusion.

The Stinger battery-coolant-unit usage study clearly developed the rationale for the scenarios and identified the limitations of both the computer and the model. The COMO III model was described with a level of detail that would allow an analyst to examine the operation of the Stinger submodel in substantial detail. However, one limitation of the report was the implicit assumption that the submodel for another air defense weapon being simulated within the COMO III was sufficiently credible and accurate that the overall results would not be biased. Given the size and complexity of the COMO modeling system, however, it may not be reasonable to expect that an analysis of a particular weapon-system model can also address the credibility of other COMO submodels in detail. A second limitation was the lack of comment regarding the fact that only one replication for each scenario was produced and, thus, the unresolved issue of statistical representativeness in the results.

The COMO III modeling system functions with submodels that represent specific types of weapon systems. The reporting on the strengths and limitations of some of these submodels was complete and useful. For example, the report on the validation of the high-resolution Patriot missile submodel with three other surface-to-air submodels within COMO III was a thorough comparative analysis in which the results of each model were developed and compared for a wide range of scenarios. The report compared results such as detection time, launch time, and point of intercept rather than just presenting aggregated measures of aircraft kills. Recommendations were made for improvements to the models that would bring the results to greater uniformity. The strengths and limitations of each model were discussed, giving attention to the structural and logical differences in design that often accounted for differences in the results.

Summary

In examining evidence about support structures, documentation, and the reporting of simulation results, we found that the Army has established functioning support structures for simulation activities. We believe that although these structures have limitations, they contribute to the credibility of the simulation results. The quality of the documentation of models and results is mixed. The simulations of the ADAGE and COMO were made at least moderately more credible by detailed documentation. Inadequate documentation for the Carmonette led to questions about its credibility. Reporting practices could be improved, but the explicit treatment of strengths and weaknesses did contribute to the credibility of all three simulations.

DOD's Efforts Toward Credibility

Our third question—What effort has DOD made to foster and reinforce the credibility of its simulations—led us to look for formal guidance applicable to the three simulations we reviewed and to DOD's simulation activities in general. Formal guidance for controlling the quality of simulation activities, as for many other activities, might cover (1) initiation, (2) development, (3) assessment or evaluation, (4) documentation, (5) use, and (6) maintenance or upkeep. We believe that the guidance would not only designate the persons who are responsible for simulation activities and establish management requirements but also describe policies and procedures for these activities.

We asked two questions about formal guidance for establishing and maintaining credible simulations:

- To what extent has the office of the secretary of the Department of Defense developed regulations or other general guidance that addresses the development and assessment of simulations, even if it is not about specific models or simulations?
- To what extent has the Army or its organizations provided regulations or guidance on development and assessment for organizations that produce simulations?

Although our search led us to look for relevant guidance throughout DOD, we did not comprehensively review all related guidance, such as guidance in information resources management, automated data processing, studies and analysis, and testing and evaluation. We also limited our focus to the guidance found in our review of the three Army air defense simulations; Air Force and Navy guidance, therefore, is not included.

Guidance From the Office of the Secretary

We found no formal guidance specifically for simulations from the level of the secretary of the department. However, we did find related regulations from the secretary's office that could be applied to computer simulations. The more important ones are summarized below.

The need for information and the use of analysis to support weapon-system acquisition decisions is stated in DOD directives 5000.1 and 5000.2. These direct that some form of system-effectiveness analysis, in conjunction with analyses of costs and other factors, be performed to support milestone decisions. Directive 5000.3, on testing and evaluation, states that

"The use of properly validated analysis, modeling, and simulation is strongly encouraged, especially during early development phases to assess those areas which, because of safety or testing capability limitations, cannot be directly observed through testing."

While these directives encourage the use of simulations and other analyses, they do not give guidance on prerequisites for sound simulations, how to develop them, or how to assure their credibility.

Regulations on automated data processing and the management of information resources may be partly applicable, because simulations are run on computers. However, directives on these topics focus mostly on input-output processing and file structure. They do not always include other topics important to computer simulations, such as the construction of models, the treatment of assumptions and limitations, and the verification and validation of models. Guidance on automated data processing typically focuses more on the processing of input data than on creating data as part of the process. While DOD's directives and standards in this area may be useful, they are inadequate to guide the development and maintenance of computer simulations.

One example of guidance related to simulations is that dealing with the quality of computer software. The issue of software quality is not new to computer programming, and since the 1970's a great many professional papers have been published on various aspects of software quality and reliability. The concept of "quality" is somewhat elusive and includes a number of factors such as reliability, portability, usability, and maintainability.

One of DOD's major concerns with software quality began with the software used in weapon systems or "mission-critical computer systems." For example, the 1978 Weapon System Software Development addressed a number of issues related to quality.[1] Directive 5000.3, issued in 1979 and updated in 1986, also includes guidance for testing and evaluating the software as well as hardware components of defense systems. In 1983, a report to the office of the secretary about software testing and evaluation recommended modifications that would strengthen directive 5000.3 with respect to mission-critical applications.[2]

[1] U.S. Department of Defense, Weapon System Software Development, MIL-STD-1679 (Navy) (Washington, D.C.: 1978.)

[2] R. A. DeMillo and R. J. Martin, OSD/DDT&E Software Test and Evaluation Project, vol. 1, Final Report and Recommendations (Atlanta: Georgia Institute of Technology, 1983), pp. 1-2.

There are indications that DOD's interest in the evaluation of software is being extended to software systems in general. When Weapon System Software Development was revised in 1982, the draft title was changed to "Software Development" and its stated purpose was to establish "uniform requirements for the development of software for the Department of Defense," expanding the standard to a much broader class of software. The 1985 revision, issued as DOD-STD-2167, is entitled "Defense System Software Development." Another indication of this broadening interest is the April 1985 draft entitled "Software Quality Evaluation," which

> "establishes requirements for software quality evaluation . . . to be performed during the development and support of software in Mission-Critical Computer Systems (MCCS). This standard may also be applied to the evaluation of software in non-MCCS."[3]

Although this interest in the quality of software began with weapon systems, it may be generalized to all computer systems. However, among the military personnel involved with simulations, we did not find substantial interest in or recognition of the importance of a systematic approach for addressing software quality. Arguments that can be raised against designing, programming, and testing software to satisfy established engineering standards of quality include that it will take more time, at least early in the process; it will be more costly; and it is not mandatory for applications not mission-critical. These arguments may be appropriate for some simulations that are small and have a short-term or limited purpose. But the results of simulations that have a longer term, develop a community of users, and are intensive consumers of computer and personnel resources may influence major decisions in acquisition, allocation of forces, or operations. The cost of designing and testing the quality of software for these simulations becomes a necessary part of their development.

Army Regulations and Practices

The Army has issued regulations that address the management of models in the context of its models improvement program and in the management of studies and analyses that include modeling. The Army has made an effort to develop a hierarchical modeling system that reflects the guidance of the Army's models committee; it was spelled out on August 15, 1983, in regulation 5-11, the most detailed Army statement

[3]U.S. Department of Defense, "Software Quality Evaluation," draft MIL-STD-2168, Washington, D.C., April 1985, p. 1; the emphasis is ours

regarding modeling policy and practice among the documents that we reviewed. Its guidance is specific to the models in the hierarchy that the Army will include in the major modeling efforts it expects its many organizations will use over the next several years.

The purpose of the models improvement program is to develop, document, and implement a hierachical family of combat models that could be used to evaluate combat capabilities and determine resource requirements through an integrated system of models of theater, corps, division, combined arms, and support task force operations. The program's management is specifically directed to ensure that appropriate technical procedures are used in software development and application, assign responsibility for the control of the model's configurations, and identify and assign the data management responsibilities.

TRADOC provides specific guidance on managing models and on using and reporting on simulations that are part of studies. TRADOC's August 20, 1982, regulation 5-4, entitled "Management, TRADOC Models," sets forth the manner in which its models are managed to ensure that high-quality, responsive models are available for combat development and training. TRADOC's March 29, 1985, regulation 11-8, "Management: Army Programs—Studies Under AR 5-5" and the accompanying pamphlet, "Army Programs: Studies and Analyses Handbook," issued on July 19, 1985, provide guidance on planning and conducting studies as defined in the Army's "Management: Army Studies and Analyses" (AR 5-5).[4] The "handbook" discusses studies from inception to completion in considerable detail to help officers perform timely and high-quality studies. It includes a detailed description of the strengths and limitations of models, analytical tools, and guidance on reporting.

As we mentioned in chapter 6, TRADOC's regulation 5-4 assigns management-control responsibilities to various groups but does not set out procedures for maintaining models. That is, it does not describe how to systematically and routinely evaluate, coordinate, approve, or disapprove models or how to implement approved changes. Although it does not establish requirements for establishing and maintaining the basic configuration, it does include an outline of key attributes to be covered when describing models that are in TRADOC's inventory.

[4]In the October 15, 1981, regulation AR 5-5, "Management: Army Studies and Analyses," the Army took a broader view, prescribing policies, responsibilities, and procedures for improving the quality of its studies and analyses. In addressing a much broader area, this regulation contains no detailed guidance on modeling approaches.

The Army has established various groups to address the technical and management aspects of the studies and the modeling process. For example, TRADOC analysts participating in the 1984 workshop on consistency in TRADOC's studies addressed process, modeling, doctrine, scenario, and "enemy and friendly data." They noted problems that remained in areas already covered by their guidance and made many recommendations for improving the quality of TRADOC's simulations. One was the recommendation that the configuration of models be controlled, because the thorough validation and verification of a model that are not followed by a "benchmark run" and reasonably tight configuration control allow an unacceptable risk of inconsistency. They noted further that agencies studying models change them without audit and without documentation. They suggested that although configuration control is expensive, it might be placed in a body meeting periodically or as needed or might consist of the requirement that a change be provided to its proponents, the Combined Arms Center, TRADOC, and the like for review prior to its implementation.

The workshop reported that the effectiveness of the study advisory group that is the principal oversight and review body ensuring quality and consistency in the models when they are used in TRADOC's studies is often hampered, because it is not ultimately responsible for the quality of the simulations used in a study. The study advisory group is encumbered by the large number of members and observers who attend it and the lack of depth in its reviews. In addition, the logistics of setting up a large group, preparing for it, and attending it consume valuable time, especially for the agency conducting the study. The workshop suggested two options. First, active "working groups" of senior analysts should meet periodically throughout a study at critical junctures, not merely at convenient milestones, and conduct critical reviews in depth, analyze problems, implement solutions with some autonomy, and report to the study advisory groups. This would not only ensure more thorough review but would also permit more timely corrective action and redirection. Second, smaller executive groups of senior officials who could make immediate decisions would contribute to more productive dialogue and save time, personnel, and resources.

A further manifestation of the Army's intention to guide and manage its modeling activity are the two groups we mention in chapter 6 and appendix IV that were constituted at different times to oversee the development of the COMO modeling system. These groups drew their members from the many commands and organizations that have an interest in the development of the COMO models.

Summary

Overall, the Army appears to be concerned about the quality of its models and its responsibility to provide guidance for those who manage them. Over the years, various management and procedural improvements have been discussed and, at times, initiated in the form of both regulations providing guidance to developers of models and committees taking an active interest in the ongoing development of specific models and modeling efforts in general. We note, however, that the guidance generally concentrates on management aspects and does not provide substantive technical detail, especially concerning the systematic and routine evaluation of models.

At the level of the secretary's office, we found little guidance with direct relevance to simulations, although some DOD directives and regulations on related topics include information pertinent to them.

In one area, the interest in the quality of computer software was initially oriented to systems critical to military missions but has gradually broadened to encompass computer systems in general, reflecting developments taking place in the computer software field. We believe that stronger links between software development and computer modeling may facilitate more rapid integration of software advances into the programming of computer models. The adoption of practices for assessing and improving the credibility of simulations might be encouraged if management gives greater attention to such technical aspects of modeling as software quality, statistical analysis, and validation.

Summary and Recommendations

DOD used the simulations we examined to obtain information about the effectiveness of weapon systems for decisions about acquisition. These and other simulations were also used to evaluate improvements or changes in the systems, force levels, and operating doctrine. Because the credibility of the results of simulations used for major decisions is important, we posed three broad questions about credibility.

The Factors in a Systematic Assessment

We identified 14 factors that are useful in assessing the credibility of a simulation as applied in a particular study. The 14 factors fall into three broad areas of concern: (1) theory, model design, and input data, (2) the correspondence between simulation outcomes and real-world outcomes, and (3) the institutional process of configuration management, oversight, and review and documentation and reporting practices. Severe limitations in any one of these areas would lead to doubts about the credibility of a simulation but for different reasons. Problems with theory, design, or input data would pose questions about the basic integrity of the simulation's internal structure. Little or no evidence on the correspondence of outcomes would leave insufficient proof of the extent to which the simulation represents reality. The absence of efforts with respect to the institutional process would cast doubt that appropriate practices had been used to ensure quality in the first two areas, the continuing integrity of the model, and disclosure of its critical limitations.

Our framework appears to be appropriate for reviewing the credibility of simulations of operational effectiveness, which usually involve many weapons against many targets. We did not attempt to apply it to other types of simulations. For engineering simulations, which often involve one weapon against one target, and war-game simulations, which often involve confrontations between large forces, individual factors in the framework may have to be modified; the three major areas of concern should apply as they are.

We believe our framework provides a structured and useful way to review the credibility of the results of simulations of operational effectiveness. The 14 factors can guide data collection and analysis to help in understanding both the strengths of the simulations that would enhance confidence in using the results and limitations of them that threaten confidence and point to the need for remedial efforts.

The Results of Reviewing Operational-Effectiveness Simulations

Nonexistent or weak evidence of validation efforts (factor 11) posed a major threat to credibility in all three case study simulations. Validating a simulation's results by comparing them to real-world results is a difficult problem in weaponry. It cannot be solved easily but would be helped by more efforts first to identify appropriate data sources and methods for validation comparisons and then to use them.

According to our review, credibility was consistently supported by only a few of the factors in our framework for the three simulations. All three simulations were fairly strong, with some limitations, at including important measures of effectiveness (factor 2), modeling weapon-to-target engagement (part of factor 4), and testing the parameters of models and running the models with alternative scenarios (factor 10). The reports on all three simulations were relatively complete in discussing the strengths and weaknesses of the analyses (factor 14).

Despite these strengths, the limitations of other factors reduced credibility and thereby the usefulness of the simulations. Therefore, we believe it would be imprudent to use the results directly in major acquisition decisions without correcting the weaknesses. We believe that even with these limitations, the results can be used in an exploratory way to identify possible problems in the weapon systems. With greater caution, they might also be used for extending evidence on weapon-system performance to cover many more conditions than would be possible in field tests. A simulation's results may be quite valuable for these purposes within the constraints imposed by its limitations.

DOD's Efforts

The office of the secretary of the Department of Defense has issued no formal guidance specifically for the management of simulations or how to conduct them and assess their credibility. Although several directives and at least one military standard have some bearing on simulations, we found no documented evidence that the secretary's office has sought to develop and implement appropriate quality controls that could be expected to directly improve the credibility of simulations.

The Army has been more active in fostering the development of organizations and guidance that can directly influence the credibility of simulations' results. Several Army organizations—parts of the command structure as well as less formal working groups—have roles in overseeing and upgrading simulations. The Army has also issued several regulations and a handbook that emphasize specific aspects of configuration management and reporting results.

We conclude that the Army's efforts are noteworthy in both intent and performance but that additional actions, especially more guidance on the technical aspects of simulations and requirements for validation, would improve simulations and thereby enhance their credibility.

Recommendations

We support the efforts DOD has made to develop and sustain credible simulations. We recommend that to reinforce these efforts and to ensure that such practices are followed, the secretary of the Department of Defense develop and implement guidance on producing, validating, documenting, managing, maintaining, using, and reporting weapon-system effectiveness simulations. The guidance should include a provision for routine reviews of a simulation's credibility and, in this way, the identification of problems that should be resolved. The secretary should also explore the possibility of requiring that a statement regarding validation accompany the report of a simulation's results.

We recommend that to make the ADAGE, Carmonette, and COMO III models more useful in future applications, the agency responsible for managing each simulation explore the feasibility of remedying the limitations we identified, especially in the area of validation.

DOD's Comments and Our Response

DOD commented on a draft of this report; our response appears in appendix V. DOD attributed 21 findings to the report, concurring fully with 19 and concurring partially with 2. DOD concurred with the two recommendations presented in the report.

DOD's comprehensive and detailed review indicates clearly that simulation is an area of importance to DOD, one in which it agrees that improvements can and should be made.

The letter transmitting DOD's response raises concerns about generalizing from three case studies and asserts that the report does indeed do this without, however, citing specific examples to support this assertion. From our perspective, we made every effort to avoid inappropriate generalization, and we believe we were successful. A major focus of our study was to demonstrate that one can systematically collect and analyze information about a simulation that would permit one to assess the credibility of that simulation. Using operational-effectiveness simulations, our three case studies show the feasibility of an approach for simulations of that kind. We do not infer from these case studies anything with regard to the credibility of other simulations. Our recommendations are based on both our review of DOD's effort to foster and reinforce the credibility of simulations and our case study analyses.

In its letter, DOD highlighted one of the two "findings" to which it gave only partial concurrence—namely, that applying our framework to assess credibility gives only part of the picture because quality depends also on the persons involved, the input data choices, and the way the model is applied. We certainly agree that these are important contributions to an assessment of a simulation's credibility, but we do not agree that our framework excludes these factors. In fact, the application of models is considered under factor 1 of our framework, input data is the focus of factor 7, and persons involved is included under factor 12. In the report, we have tried to indicate the importance of these and other elements.

The other finding to which DOD gave only partial concurrence was our concern about the use of the expected-value method for representing the mathematical relationships in the engagement of multiple air defense weapons against multiplane attacks in the ADAGE Campaign submodel. By pointing out several limitations, we did not intend to imply that the expected-value approach is intrinsically bad. The concerns we reported were raised either by DOD personnel themselves or by experienced models practitioners. Moreover, we tempered our criticisms in this area with other statements in the report pointing out that the theoretical approach of the ADAGE was appropriate for addressing decisions concerning competing air defense weapons even though it was an expected-value model.

A Description of Simulation Models

In this appendix, we define terms commonly associated with simulation models and explain the simulations used in the weapon-system acquisition programs for the DIVAD and the Stinger.

Definition of Terms

Simulation is the overall process in which a system is modeled and the model is experimented with. In this report, "model" refers to the representation of an object, a system, an activity, or a situation by something other than itself. It might be a logical, mathematical, or physical representation or a combination of these. A model represents the system, its elements (or variables), and the relationships between the elements that govern their interaction.

The types of simulations or models of combat the military services use to support decisions are often described or categorized in several ways:

- in terms of the numbers of friendly versus enemy units or systems engaged in combat events, from one-on-one to one-on-few, many-on-many, or theater-level interactions;
- in terms of the organizational levels of the units engaged, from battalion to corps or division to theater;
- in terms of the degree of detail in depicting combat events, whether high-resolution simulations that depict smaller units in fine detail or low-resolution or large-scale simulations that depict larger units in highly aggregated variables.

Simulations are also categorized by the techniques they employ. A computer simulation is a model of a weapon's behavior in combat that is run entirely on a computer. A hardware-in-the-loop simulation substitutes one or more actual components of weaponry for a portion of the model, the remainder of the model being handled by computer. A man-in-the-loop simulation places a human being—a radar operator or pilot, for example—into direct interaction with the computer or hardware-in-the-loop simulation.

Simulation models may be further classified as stochastic or deterministic. A stochastic simulation model (described by some authors as a Monte Carlo or probabilistic model) has one or more random variables as inputs. Since random inputs lead to random outputs, they can be considered only statistical estimates of the true characteristics of the model. Simulation models that contain no random variables are deterministic. For a given set of input data, deterministic simulation models provide a unique set of outputs.

In the context of simulations and models, hierarchy refers to a vertical sequencing relationship in which the outputs of one model provide inputs to a more aggregated model. However, a sequence of models or simulations in weapons acquisition may refer to the order in which modeling and simulation are performed. Generally, the order is from computer simulations of subsystems up to the full system in its operational environment to hardware-in-the-loop simulations to man-in-the-loop simulations.

The Use of Simulations for Two Weapon Systems

Simulations were used extensively in the development of the DIVAD and the Stinger weapon systems. The program offices for both noted that as their budgets became tighter and the systems more costly, they made greater use of simulations to augment data from physical tests. The one-on-one, item-engineering models, with or without hardware-in-the-loop, were used to assess technical performance. Force-on-force simulations were used to assess operational effectiveness.

The DIVAD

Prior to the Army's 1976 decision to develop a new air defense gun to replace the VULCAN air defense gun, the Army Materiel Systems Analysis Activity had constructed and validated antiaircraft gun models. In 1971, during the gun air defense effectiveness study, a simulation model for the VULCAN was built and validated with field-test data. Later, other air defense gun simulation models were built and validated, using data from the gun low-altitude air defense test. These models—the Fire Unit Effectiveness model for the VULCAN and the Modern Gun Effectiveness Model for the DIVAD—were the basis for all the Army Materiel Systems Analysis Activity one-on-one air defense gun studies during the mid-1970's. The models were modified to simulate other air defense gun systems and validated with field data.

The two contractors that were selected to build the prototype DIVAD gun systems (Ford Aerospace and Communication Corporation and General Dynamics Corporation) were asked to develop computer simulations concurrently and to base them on the Modern Gun Effectiveness Model to represent their respective systems. In 1980, the Army validated these models with data from the field tests of the prototypes.

Since 1977, several studies and analyses have used force-on-force simulations to investigate the need for and contributions of the DIVAD gun. In 1977, the Army reported on the cost and operational-effectiveness analysis of the division air defense gun. The report examined whether the procurement of a DIVAD gun, as one component of future air defense weaponry, was the most cost-effective solution for air defense missions. The ADAGE simulation was created to perform this analysis, and a generic 35-mm gun was modeled. The recommendation was to proceed with the development of the DIVAD gun and to place 36 DIVAD guns per division in the field.

The division air defense gun cost and operational-effectiveness analysis update, completed in June 1984, addressed concerns regarding operational and developmental test results and new threat projections. The Army Air Defense Artillery School was instructed to use the Carmonette in this analysis, which was specifically designed to address the effectiveness of the performance of the gun as indicated both by test data (an "as tested" version) and by expected production characteristics (a "mature" version). The study, conducted by the Army TRADOC Systems Analysis Activity, concluded that force effectiveness increased when the DIVAD was added to the forces, even with performance shortfalls shown by testing and significant increases in the projected threat. The Army Air Defense Artillery School also conducted additional analyses using the ADAGE.

The DIVAD force structure analysis, an offshoot of the update, supported the recommendation of 36 DIVAD guns in the 1977 analysis. Decisions supported by these analyses led to the exercising of options I and II of the contract with Ford Aerospace and Communication. A decision on option III was deferred until the fall of 1985 to allow testing for operational effectiveness, suitability, and limited production. To support the review process for option III and assist the secretary of Defense in deciding whether to continue with the production of the DIVAD gun, a comparative analysis was directed by the Department of the Army. The analysis, which used the ADAGE and Carmonette models, examined the ability of the DIVAD to perform its designated mission within its postulated initial operational capability on the battlefield. It also examined the ability of alternative weapon systems to perform the same mission.

The ADAGE helped determine the effectiveness of air defense systems in terms of resources saved in a division. In a parallel effort, the model was also used to determine the operational effectiveness of the DIVAD for different levels of its performance parameters, and the results determined the levels of degradation at which the DIVAD would become less effective than the alternative systems under consideration. The results of the effectiveness analysis were used to compare the operational effectiveness of the DIVAD's alternatives. The Carmonette model examined the alternatives in the context of an intense battle with a battalion task force.

The Stinger

At the engineering level, digital, analog-digital, and hardware-in-the-loop simulations have played a major role in the Stinger's development and product improvement. At least three such simulation capabilities have been developed. General Dynamics, the contractor, verified a simulation with various types of flight and nonflight tests. When the output of the simulations was confirmed, the results could be used as a design tool. The Army used a similar simulation at U.S. Army Missile Command to validate the contractor's performance data and to investigate improvement alternatives. A third simulation was developed at the Office of Missile Electronic Warfare to evaluate electronic countermeasure and counter-countermeasure performance and to assess vulnerability.

The Stinger's operational combat effectiveness was assessed with the Tactical Air Defense Computer Operational Simulation in a cost and operational-effectiveness analysis reported in 1977. Several alternative, portable air defense systems were evaluated under identical situations, including a comparison of the relative effectiveness of the Stinger and the Redeye in various environments.

Another study focusing on operational employment issues used the COMO III to investigate the Stinger's battery-coolant-unit use rates in a wartime environment. This was the study we reviewed, because it was reasonably well documented, the model on which it was based was well documented, and the programmers, analysts, and managers were still available for interviews and questions.

Supporting Material for Chapter 4

The Theoretical Approach

A model's basic theoretical approach for evaluating the effectiveness of a weapon system may be engineering, to determine the optimal design of the weapon systems; functional, to aid in selecting the most effective weapon system from alternative systems performing the same function (for example air defense); or combined arms, to compare alternative uses of competing weapon systems (for example, air defense weapons versus helicopters versus tanks).

The Carmonette is a combined-arms model designed to answer broad trade-off questions about armor, infantry, artillery, and the like. It focuses on the total ground battle, not individual weapon systems; air defense considerations have only recently been added to the model. The ADAGE, in contrast, is basically an expected-value model, designed to study the effectiveness of combinations of ground-based weapons in providing air defense to a division. The COMO III was likewise designed to study the various factors involved in providing air defense but, like the Carmonette, it is a Monte Carlo model and it operates at high resolution. (We have summarized the three models' theoretical approaches in table 4.2.)

As functional models, the ADAGE and COMO III emphasize the adequacy of air defense; the other aspects of war, where they are included, are concentrated on how changes in air defense capability can change battle outcomes. However, the emphasis of both models is air defense, not the total battle. Even critics of the ADAGE agree to its usefulness in making decisions between air defense systems. The ADAGE and COMO III are also systems-analysis models in that they are designed to provide information to decisionmakers concerning various alternatives for providing air defense and are not useful for considering trade-offs between air defense and other wartime functions.

Why should the differing theoretical approaches of the Carmonette, ADAGE, and COMO make any difference? With the emphasis of the ADAGE and COMO on air defense, only a less-detailed portrayal of the remainder of the war may be sufficient to judge the trade-off between competing air defense systems. The Carmonette's emphasis on combined arms in the total battle means that some elements a 'e often omitted or aggregated in simulations of air defense in a manner such that important information can sometimes be lost.

In our op. ion, the basic approaches of the ADAGE and COMO are more appropriate for studying air defense trade-offs than a combined-arms model like the Carmonette, which has to be modified to accommodate air defense.

Operational Measures of Effectiveness

The protection of operational and strategic assets from enemy aircraft is the primary mission of U.S. air defense forces; the attrition of enemy aircraft is secondary. Although the Carmonette may be able to produce information on protection, its emphasis in the DIVAD analyses was on attrition. It stressed the comparison of the loss of enemy forces to the loss of friendly forces in the form of various exchange ratios. (We have summarized the Carmonette and the ADAGE and COMO III in table 4.3)

In its analyses, the Carmonette produced "killer-victim scoreboards," or matrixes comparing kills of all types of enemy aircraft by all types of friendly air defense weapons and kills of all types of ground targets by enemy aircraft. The figures from the matrixes were used for comparisons of the effectiveness of weapons. The principal force-effectiveness measures reported in the Carmonette were the loss-exchange ratio (or the total enemy losses divided by total friendly losses) and the fractional exchange ratio (the percentage of enemy losses divided by the percentage of friendly losses). Systems ratios permitted the comparison of losses of friendly weapons to losses of one target or all targets against which the weapon was used (for example, the DIVAD against the HIND helicopter or the DIVAD against all target aircraft).

The emphasis in all these comparisons was attrition. No differentiation was made between the relative worth of assets lost. Other measures of effectiveness reported in the Carmonette analyses were the number of helicopter remaskings caused by radar warning and the number of mission aborts caused by damage to enemy aircraft from ground fire. These measures were not covered in the ADAGE.

Although the ADAGE can produce statistics that can be converted into the same type of attrition statistics that the Carmonette does, the effectiveness measure emphasized in the ADAGE cost and operational-effectiveness analysis was the protection of assets. Friendly assets were assigned a value called "military worth," the assets having a military value to the enemy as well as to friendly forces. Military worth to the enemy was used in enemy air-raid allocations; the principal measure of effectiveness was the military worth of friendly forces remaining after enemy raids. The analysis also reported the worth of individual classes of targets remaining and showed how military worth declined over several days of fighting and how much of the loss of friendly military worth was attributable to the ground war only and to ground and air wars combined.

The proportion of loss attributable to enemy fixed-wing aircraft was a major source of concern to the study advisory group and the critics of the ADAGE. Ground damage attributable to enemy aircraft was so great that credibility was questioned in comparison to the Carmonette and other models. These concerns and the possibility that the ADAGE may overstate damage by fixed-wing aircraft means that this aspect of the ADAGE modeling may need refining. Nevertheless, from the theoretical perspective, it seems able to report measures of effectiveness that are appropriate to air defense.

Since the COMO III does not model interactions between ground forces, it is limited in its ability to use preservation as a principal measure of effectiveness. While analysis in the COMO III may concentrate on measures of attrition, its flexibility allows a wide range of measures of effectiveness. One example is its use in the analysis of Stinger battery-coolant-unit usage, where the output measure was the number of units needed to fire each missile.

A chronological description of the critical events of a COMO III simulation is available in summary form. The measures of effectiveness are the analyst's choice. They are based on the raw material of the simulation history, which includes detection attempts, detected targets, completed reloads, the availability of a system, missile intercepts, threat attrition, the amount of munitions used, and kill ranges, among other things. This information is available by fire unit, platoon, battery, battalion, or scenario, and it is further processed into report outputs summarizing the activity at a site and the effectiveness of threats and air defense.

Each simulation addressed measures of effectiveness in operational terms, the ADAGE better than the Carmonette or COMO III, since it produced measures related to protection in addition to the attrition of enemy aircraft and war-exchange ratios. The Carmonette might have produced this type of information, but it did not address this facet of the air defense mission. The COMO did not address this measure since it did not cover the ground war at all. However, the COMO III was able to produce a measure of effectiveness especially designed for the study of the battery coolant unit.

The Portrayal of the Weapon System's Immediate Environment

Once a model's theoretical approach is understood, one can assess how well it treats the critical aspects of a weapon system's behavior in tactical combat. How the model formulates them determines the critical variables to be considered and how the variables relate to one another in describing not only the behavior of the weapon system but also the overall war environment in which the weapon system is to be used. We believe it is important in the evaluation of the model's portrayal of the various characteristics of a weapon system to consider both the weapon's tactical environment and how it operates in combat. The tactical environment involves such features as the size and duration of battle, the potential target set of a weapon system, the deployment and movement of the system, and the terrain in which it is to operate. (We have summarized the issues of environment in table 4.4.)

Level of Battle

Since the DIVAD was to be a divisional rather than a battalion or some other air defense weapon, the ADAGE model, developed specifically to address the DIVAD gun, treats the weapon as a division weapon. The Carmonette, however, addresses sections of the battlefield only up to the battalion level and, thus, could preclude the weapon from engaging some targets it was designed to kill or suppress. Moreover, not all the Carmonette analyses included the effects of all battalion DIVAD guns because of the small block of terrain being modeled. Critics of the Carmonette as a tool for analyzing the DIVAD assert that air defense is a division responsibility and that some aspects of the surface-to-air battle are overlooked, because the focus is limited to a battalion battle.

Unlike either the ADAGE or Carmonette, the COMO III can be played at any level, one-on-one, battalion, division, or even theater conflicts. For the analysis of the Stinger's battery coolant unit, the analysts selected a front-to-rear brigade slice, a representation of an area they believed encompassed a sufficiently large number of air defense units and threat aircraft and helicopters to provide a realistic exercise. The activities of 99 Stinger units and more than 300 threat aircraft were represented in the analysis.

The fact that the Carmonette focuses on an intense, 25-minute battalion battle, as opposed to the ADAGE's small raids by enemy aircraft against targets in the division over several days, is also of some concern. A conflict simulated with the ADAGE can last up to 30 days, and logistics are included. The Carmonette battle covers less than 10 percent of the territory of an ADAGE battle and includes 4 DIVAD guns, while the ADAGE uses 36. The Carmonette emphasizes the effects of aircraft only in the main battle area, whereas the ADAGE also portrays the effects of aircraft against combat support units to the rear of the division. Moreover, when it comes to measuring the potential damage attributable to enemy aircraft, a 25-minute firefight cannot be directly compared to a battle of several days. In effect, the ADAGE purports to model the results of several Carmonette battles and measures the cumulative effect of enemy air attacks on the ability of friendly forces to wage war.

The analyst chooses the level of play—battalion, brigade, division, or higher—for the COMO III but the model is limited in its ability to play battles of extended length, since it does not model logistics. The study of the battery coolant unit, whose purpose was to determine the number of units each Stinger required in wartime, worked with the initial supply position and did not address resupply. COMO documents indicate that a typical simulation represents about 2 hours of real time. The complexity of the Stinger scenario and environment was limited in order to reduce the resources required for computer runs.

Targets

Another significant difference between the ADAGE and Carmonette in the treatment of the DIVAD was the weapon's potential set of targets. The ADAGE modeled nonjinking helicopters and fixed-wing aircraft with fixed flight paths as potential threats and dealt with the damage from fixed-wing attacks in the rear as well as forward areas of the division. Fixed-wing aircraft were not included in most of the analyses using the Carmonette. The TRADOC studies advisory group recognized the omission as a serious deficiency but did not demand changes to the Carmonette model.[1]

Even when the Carmonette finally addressed fixed-wing aircraft, it did so by using information produced by another model that addressed surface-to-air gun attacks in essentially the same manner as the ADAGE. The Carmonette was modified after the last DIVAD study to include a fixed-wing component, but no analyses of the DIVAD were made with it because the DIVAD program was cancelled.

[1]TRADOC's study advisory groups monitor the progress of its studies and review and provide advice on the planning, performance, and reporting of specific studies to both the agencies conducting them and the agencies directing that they be done. Group members represent interested organizations that know aspects of a particular study but are not directly involved in it. They meet three or more times at critical points during a study, and subgroups review the more technical matters, such as analyses, costs, scenarios, doctrine, and threats. The minutes of a study advisory group meeting can become directives.

The results of using an ADAGE-type modeling approach in conjunction with the Carmonette led to the conclusion that fixed-wing aircraft were not a significant threat to assets of combat ground units in the forward part of the main battle area, a conclusion that contradicted a conclusion from the ADAGE model alone. The difference came, to a large degree, from the Carmonette's focus at the battalion level, where fixed-wing aircraft may not be significant, compared to the ADAGE's focus at the division level, where the damage from fixed-wing aircraft is a more important consideration. It is not clear that including a fixed-wing component would overcome the difficulties resulting from the Carmonette's more limited concentration.

In the COMO III, the Stinger could attack helicopters and fixed-wing aircraft. The study of the battery coolant unit included both air threats. Most COMO III modeling, however, has concentrated on the threat from fixed-wing aircraft.

Weapon Deployment and Movement

Another important aspect of modeling the use of a weapon is how a model portrays the weapon's deployment and movement on the battlefield. The analyst determines the tactics, deployment, and decision rules that are to become input for the Carmonette. The reports on the Carmonette's simulation of the DIVAD indicate that the analysts studied the effectiveness of the alternative deployment of weapons. While there was some concern about the appropriate portrayal of the DIVAD's deployment in the Carmonette analyses, the concerns were about the analysts' input rather than the fundamental theory of weapons deployment.

The Carmonette has a submodel that uses mobility factors as inputs to treat movement on the battlefield. The Carmonette allows weapons to move in response to firing, permits well-defined movement patterns, and allows intermediate stops in them. At one time, the Carmonette would not allow the DIVAD to fire on the move, but this problem was corrected in the analyses. Movement rates in the Carmonette were affected by the environment: the mode of movement, terrain slopes, and ground conditions such as the presence of paved roads, dirt roads, no roads, and so on. On the whole, the Carmonette's treatment of weapon deployment and movement was suitable for the DIVAD.

In contrast, the ADAGE assumes a static deployment. It deploys weapons in rectangles or zones of terrain. A division's dimensions are input for the ADAGE model, and for purposes of computing aircraft attrition, it partitions a division into zones parallel to the forward edge of the battle area. Air defense weapons within one zone are assumed to be uniformly distributed. The ADAGE gives some indirect recognition to deployment, since the one-on-one Incursion places air defense weapons randomly relative to aircraft flight paths in several replications that determine one-on-one attrition factors. These factors are used in the many-on-many simulation in the Campaign submodel, in which air defense weapons are assumed to be uniformly distributed within each zone of the battlefield modeled. Thus, the ADAGE results are, in effect, the average of several randomly generated weapon deployments.

Not only does the ADAGE not directly portray how weapons are deployed; it also does not portray the movement of the DIVAD. The Incursion does not portray the movement of air defense units. It is possible that movement is portrayed indirectly in the Campaign, since it applies a probability-of-participation factor to ground-to-air attrition rates in determining final attrition rates. The movement of air defense units may be partially portrayed by adjusting these factors to represent the "nonavailable" time caused by the movement of the weapon. On the whole, however, the ADAGE's treatment of weapon deployment and movement has to be considered less adequate than the Carmonette's.

Like the Carmonette, the COMO III deploys the Stinger according to the analyst's specifications, but like the ADAGE, it does not specifically model the movement of defensive weapons, except aircraft. Rather, it addresses movement through the lessening of the probability of participation. The COMO III allows individual Stinger units to become operational or nonoperational at specific times, a capability that may be used to roughly simulate movement. The individual Stinger teams, however, are given specific locations by the analyst. The Army's field manual on the Stinger's team operations emphasizes that frequent movement as far as several hundred meters contributes to survival. Moving after each firing, unless there is another aircraft to be engaged, could affect the time that the team is actually in operation. Neither the greater likelihood of survival nor a decrease in operations because of the team's movement appears to be directly included in the COMO III model.

Terrain

How a simulation models terrain is important for air defense weapons like the DIVAD and Stinger, because helicopters, one of their primary targets, can use terrain to mask their intentions until moments before they fire. The ADAGE uses a statistical terrain; the Carmonette and COMO III use a digitized map of a geographic area. Problems associated with these approaches are worth commenting on. For example, the ADAGE's statistical terrain was based on empirical data of an extensive study of World War II tank battles that may not represent the line-of-site considerations appropriate for air defense in the 1980's. The model's terrain does not depend on the scenario, which can be viewed either as a strength, because the results can be generalized, or as a weakness, because the results do not seem real.

Terrain in the ADAGE was specified by a distribution of unmask-remask ranges that depended on aircraft altitude, type of terrain, and the height of the weapon site relative to the mean terrain. The terrain parameter specified only whether terrain was rough, rolling, or open, and intervisibility (the ability to see between two points) is calculated with a statistical model, given that parameter. For specific aircraft altitudes, weapon heights, and flight paths, the mean unmask range was determined, and random draws determined the probability of unmask and remask for each replication of the Incursion. Interruptions in intervisibility were not considered, and the aircraft was detectable from the first unmask until remask. It should be noted, however, that the ADAGE plays terrain only in the Incursion model, where it is used in developing the probability of kill; it is not explicitly incorporated in the Campaign, and it is not considered in the ground war.

In contrast, for the DIVAD study, the Carmonette modeled a specific area near Hunfeld, Germany, with terrain data from the Defense Mapping Agency and additional data on vegetation and traffic from a waterways experiment station. Although this provided a more realistic portrayal of terrain, the limitation to a single area was viewed as a deficiency, but no requirement for any other terrain was imposed. Whether other terrain would have changed the conclusions about the DIVAD is unknown.

The COMO III, like the Carmonette, uses digitized data that describe particular terrain areas in West Germany. Lines of visibility are determined for each Stinger unit and the aircraft that may become targets. That the COMO III appropriately considers visual masking is important, because many of the Stinger's targets are aircraft of relatively low altitude.

The ADAGE and COMO address the tactical environment reasonably well, whereas the Carmonette is weak in this area. Both the ADAGE and COMO simulate a battlefield of the size appropriate for air defense, and both simulate all the targets likely to be encountered in air defense. The ADAGE's coverage of the length of battle is the more appropriate for air defense, since its battle of many days best addresses the cumulative damage attributable to air attack. The ADAGE's portrayal of terrain allows generalizations more easily than that of the Carmonette or COMO III but it is less realistic. The Carmonette's strength regarding the environment is its ability to portray the movement of ground weapons, while the limited portrayals in the ADAGE and COMO are definitely weaknesses.

The Portrayal of the Weapon System's Operational Performance

A complete model of air defense weapons not only focuses on how a weapon engages and fires on enemy aircraft but also considers how that weapon works with other air defense weapons to maintain the ability of ground forces to resist an enemy invasion on land. A consideration of how a weapon system operates in combat involves such features as the detection of and engagement with its assigned targets. In air defense, detection can be either visual or by radar, either of which can be affected by battlefield obscurants or problems with command, control, and communications as they relate to identifying whether a potential aircraft target is a friend or foe. A consideration of engagement involves the physical characteristics of the air defense weapon system, the procedures of its engagment of attacking aircraft, and the application of those procedures when more than one aircraft is attacking.

For air defense weapons, an important aspect of modeling is how well computer models portray the way weapons detect and engage enemy aircraft. The important aspects of air defense include radar and visual detection, battlefield obscurants, battle management, IFF, and command, control, and communications as they relate to IFF. The important aspects of engagement include the characteristics of a weapon affected by the engagement procedure and the application of those procedures to multiple aircraft raids.

Detection of Enemy Aircraft

In table 4.5, we have summarized how each of the three simulations represented the critical aspects of the air defense mission related to the detection of enemy aircraft.

Visual Detection

Both the ADAGE and Carmonette modeled how the DIVAD gun detected enemy aircraft and included provisions for visual detection. Originally, the ADAGE used a separate visual detection model called VISPOE, developed by the U.S. Army Missile Command, and results from this model were used as input for the Incursion submodel of the ADAGE. The Carmonette used the visual detection model developed by the night vision and electro-optical laboratory. However, because differences between VISPOE and the laboratory's model could not be resolved for the cost and operational-effectiveness update and the comparative analysis, the ADAGE was modified to use data from the latter model for pop-up helicopters, whereas the Carmonette used data from the former for the detection of fixed-wing aircraft in the comparative analysis.

In the original ADAGE analyses, the VISPOE model incorporated gradual lessenings of expected visibility to the full range of the DIVAD gun by extrapolating limited Fort Knox field test data on helicopter detection ranges. In contrast, in the Carmonette analyses, a ground-to-ground detection model was modified to include helicopters; incorporated visual detection distances up to only 3 kilometers, considerably short of the DIVAD gun range; and treated this detection range as a "brick wall" beyond which no visual detection could occur. Because of this range shortfall and because the Carmonette analysts disagreed with the procedure of extrapolating VISPOE data, the Carmonette analyses of the DIVAD used the forward-looking infrared detection routine as a proxy for the visual detection of helicopters to the full range of the DIVAD gun. In addition, the basic probabilities of detection assumed that the ground observers in the night vision and electro-optical laboratory model had infinite time in which to detect targets, so the Carmonette modelers had to insert search-time limits in order to keep the model from accepting unrealistically long search times.

These two characteristics—the DIVAD's forward-looking infrared and search-time limits—were also incorporated into the ADAGE for the visual detection of helicopters. Since the DIVAD was not equipped with forward-looking infrared detection capability, its use as a primary visual detection model for helicopters resulted in a model that did not properly represent the operating characteristics of the gun. The Stinger model in the COMO allowed either the use of a simple probability of detection that would be the same for fixed-wing aircraft and helicopters or, alternatively, the use of tables showing the probability of detection as a function of the type of aircraft. Like the Carmonette, the Stinger model in the COMO appears to limit the visual detection search range and impose a "brick wall."

Battlefield Obscurants

Other aspects of visual detection important in the tactical environment include nighttime vision and smoke, dust, and glare. The ADAGE does not model night conditions while the Carmonette does. The developers of the ADAGE sought to include the direct effects of smoke, dust, and glare in their model but did not do so, apparently because of a lack of empirical data. For the cost and operational-effectiveness analysis update, the

ADAGE was given a provision to handle smoke the same way it handles bad weather—indirectly, by adjusting the input values of the probability of participation in the ground-to-air war.

Using a fully dynamic detection model, the Carmonette played the effects of smoke, dust, fog, rain, snow, and aerosols. The ADAGE permits the selection of weather conditions that determine the Incursion outputs that are used as Campaign inputs, but only visual-detection parameters are directly modeled in the Incursion. The COMO did not directly play the effects of smoke, dust, weather, or the time of day or night. These are included indirectly by allowing the analyst to input degraded probabilities and search ranges of visibility.

Command, Communications, and Control and IFF

Neither the ADAGE nor the Carmonette addresses command, control, and communications and IFF directly. While documents concerning Carmonette indicate some ability to play command, control, and communications, the Carmonette studies of the DIVAD specifically excluded their effects. The ADAGE gives some indirect consideration to command and control in its Incursion submodel, because these are considered in the visual detection model used in the ADAGE. Any command and control effects on the total battle are difficult to determine, however, since the Incursion produces only one-on-one attrition results, which are used as inputs to the Campaign battle model.

Command and control were not explicitly played in the Campaign. The ADAGE gave only indirect consideration to IFF in the Incursion by including it as one of several factors in establishing a DIVAD crew reaction time in engaging detected aircraft. Whether this provision for IFF in the Incursion has any effect on the battle in the Campaign is difficult to determine since the IFF effects on reaction time are not differentiated from any of the other effects. Moreover, the ADAGE plays friendly air in the Campaign, but the model structure does not permit the engagement of friendly air by friendly air defense forces, thus omitting a consideration of the potential failure to properly identify friendly aircraft. The Stinger weapon does not require a modeling of command, control, and communications, since Stinger teams are free to engage other targets or move. The Stinger model in the COMO does not model IFF since it does not allow friendly aircraft to become potential targets.

Radar Detection

Both the ADAGE and Carmonette provide for the detection of enemy aircraft by radar. Early versions of the Carmonette did not correctly portray radar-detection capabilities, but changes produced a model that is probably superior to the ADAGE in this regard. In the early stages of considering the use of the Carmonette to model the DIVAD, objections were raised because the Carmonette did not correctly play the primary mode of the DIVAD's operation—a combination of radar and optics—nor did it include the effect of electronic countermeasures in counteracting the DIVAD radar. In addition, the Carmonette originally was not able to model the full detection capabilities of the DIVAD radar. The Carmonette was modified to handle all these problems for DIVAD analyses.

The ADAGE does not model radar detection directly; instead, it includes radar effects, covering the gun's full range, in the input data. The ADAGE matches the flight path of approaching aircraft against radar boundary "footprints"—input data—to determine whether an aircraft can be detected and, if so, when. The effects of electronic countermeasures are included in determining the "footprints." This approach to modeling radar was used to produce a quick-running model. Not only does the ADAGE not play radar detection directly; it also does not portray how aircraft respond to radar warning. It assumes that a flight path does not change when an aircraft is likely to maneuver. Overall, therefore, the Carmonette appears to model radar detection better than the ADAGE does. Radar detection is not applicable to the Stinger.

In summary, none of the models provides complete coverage of the detection aspects of air defense. Visual detection is generally limited in range. Command and control and IFF are either not covered at all or covered only indirectly. The Carmonette covers radar detection and battlefield obscurants reasonably well, but tne ADAGE and COMO address them only indirectly, if at all.

Engagement of Enemy Aircraft

Once computer models indicate that air defense weapons have detected enemy aircraft, they must then model how those weapons proceed to engage and destroy enemy aircraft. All the models encompass this engagement-and-firing process, each having strengths and weaknesses in its approach. In table 4.6, we have summarized these strengths and weaknesses.

Weapon Characteristics

The ADAGE was developed specifically to study the proposed DIVAD gun, but the Carmonette originally based its modeling of the DIVAD on the capabilities of the Soviet ZSU-23-4, an antiaircraft gun. This version of the Carmonette was used in the antihelicopter study that first raised serious questions about the effectiveness of the DIVAD. Disclaimers in this study's report stated that no conclusions regarding the DIVAD should be made because of inappropriate modeling of aspects of the DIVAD. Consequently, corrections to the Carmonette were necessary and a revised model was used for the 1984 cost and operational-effectiveness update. Since the characteristics of the DIVAD gun have been similarly modeled in both the ADAGE and Carmonette, we believe that any further differences probably result from how the gun was modeled for use in combat.

The COMO Stinger model was based on the physical characteristics of the Stinger weapon system and its operational procedures. The physical characteristics can be altered if the intention is to evaluate prospective enhancements. Programming changes would generally be required to make changes in operation; however, one feature is that firing doctrine, which must be responsive to existing conditions, is selected in the data input phase.

Engagement Procedures

Both the ADAGE and the Carmonette model engagement procedures. In the ADAGE, all short-range air defense weapons could engage aircraft in the "fly-by" mode—that is, aircraft fly past the air defense weapon enroute to another target—or the vicinity-of-target mode—that is, aircraft maneuver during ordnance delivery on a target defended by the weapon. However, the ADAGE directly models one-on-one engagements only in its Incursion component, the results of which are used as input data in the Campaign many-on-many expected-value model.

The ADAGE many-on-many approach does not properly account for the spatial or temporal saturation of many enemy aircraft attacking at the same time. Other aspects of the ADAGE's failure to model many-on-many engagements directly are (1) the ADAGE does not permit guns to switch targets; (2) the ADAGE does not allow the number of aircraft to change during segments of a raid; (3) the ADAGE does not handle the effect of mission aborts properly; and (4) the ADAGE assumes perfect coordination between air defense units in seeking and engaging the same target.

Even in one-on-one modeling, there are problems with the ADAGE's portrayal of weapon-aircraft engagement. Since the Incursion did not model duels, the DIVAD could engage and kill threat aircraft but the threat aircraft could not directly engage the DIVAD. The DIVAD's attrition as a target class was played in the Campaign submodel and the destroyed guns were removed only at the end of a raid. Thus, the DIVAD could remain operational to inflict damage when it might otherwise have been destroyed. This approach is similar to the attrition of enemy aircraft and is a problem inherent in the expected-value approach. In addition, the ADAGE definition of the DIVAD target class permitted target overkill, which resulted in the destruction of fewer numbers of the DIVAD than in the Carmonette for the same number of threat missiles fired at it. Furthermore, aircraft in the ADAGE fly a constant heading and altitude and do not react to ground fire or radar warning.

The Carmonette plays rules of engagement but, unlike the ADAGE, concentrates on vicinity-of-target engagements. In the TRADOC study advisory group discussions about adding fixed-wing aircraft to the Carmonette, reviewers justified the exclusion of these aircraft by asserting that including a fly-by mode serves no useful purpose, since all it does is give the DIVAD more targets to shoot at without any effect on the ground battle at the battalion level. Omitting the fly-by mode appears to ignore the DIVAD's division-level responsibilities. While the Carmonette allows different engagement doctrines, air defense weapons generally commit to engage only after their particular targets have been recognized.

The Carmonette provides for selection from among several targets. It gives priority to the nearest target and then prioritizes targets according to type and speed, starting with hovering helicopters and going on to moving helicopters and fixed-wing aircraft. Some concern was expressed about this order. The Carmonette simulates the DIVAD's ability to continuously track multiple targets, retaining a track file for future engagements and continuously updating it with prioritized targets. The model did not play fire distribution command and control, so the DIVAD, which moved in pairs, could fire from the two guns on the same target.

The Carmonette models helicopters, including their reaction to radar warning and gunfire, but in 1984, it did not model fixed-wing threats. For the 1985 comparative analysis, the U.S. Army Material Systems Analysis Activity provided fixed-wing aircraft data relevant for the DIVAD in the form of tables generated by a gun-effectiveness model similar to the ADAGE. More recently, a fixed-wing aircraft submodel has been added that allows preset flight paths with varying heading and altitude but does not alter the flight path in response to radar warning and gunfire.

Since the Carmonette is an event-sequenced Monte Carlo model, it models each engagement between an air defense weapon and an aircraft as it occurs. Attrition occurs after an engagement between an aircraft and an air defense weapon rather than at fixed points in time, as in deterministic models like the ADAGE. It should be remembered, however, that the Carmonette models only a battalion-level rather than a division-level battle, like the ADAGE, and it models only 4 DIVAD guns, compared to the ADAGE's 36.

The COMO III provides extensive detail of how weapon systems engage their targets. Like the Carmonette, it includes the coverage of multiaircraft raids. Like the ADAGE, the COMO permits the engagement of all targets, in contrast to the Carmonette, which ignores aircraft flying through the battle area to and from deeper battle zones. The effect of a saturation level of aircraft attacking an area defended by Stinger teams can be demonstrated. The separate and overall effects of Stinger and the air defense weapon types can also be shown.

To what extent, then, did the three simulations we reviewed appropriately characterize the critical aspects of air defense weapons? We looked at specific aspects of the modeling of air defense under three broad areas of coverage—weapons system environment, detection of enemy aircraft, and engagement with enemy aircraft. We found that all the models had significant weaknesses in at least one of these general areas. Only one—the COMO—completely modeled even one of the general areas of interest. The ADAGE was generally weak in its portrayal of the detection of enemy aircraft; the Carmonette was weak in its portrayal of the weapon-system environment. The COMO provided reasonably complete modeling of the engagement of air defense weapons with attacking aircraft.

The Broad-Scale Battle Environment

The description of a weapon's tactical environment should be complete enough to cover all the critical variables in the total war that might affect the behavior of the weapon. In the three models, we found differing approaches to various aspects of modern warfare and their interaction. In table 4.7, we have summarized the coverage in the three simulations.

The air defense tactical arena includes air war, ground war, and the interaction of the two. Air defense artillery provides support for tactical, operational, and strategic warfare. Its mission is to nullify or reduce the effectiveness of attack or surveillance by hostile aircraft or missiles after they are airborne, thereby supporting the Army's primary function of conducting prompt and sustained land warfare operations. Short-range air defense and artillery units engage enemy close-air-support helicopters and fixed-wing aircraft and engage ground targets in self-defense when conflict with enemy ground forces is intense. Therefore, simulations appropriate for studying the effects of air defense weapons should cover fixed-wing and helicopter targets as well as the general effects of the ground war.

Air War

Throughout much of the Carmonette's modeling effort with the DIVAD, it failed to model one of the gun's primary targets—fixed-wing aircraft. A 1983 Carmonette study that originally raised questions about the effectiveness of the DIVAD, the antihelicopter study, did not include fixed-wing aircraft as an attacker and a potential target. Despite this concern, the study advisory group did not require the Carmonette modelers to develop fixed-wing model coverage, acknowledging that they did not have sufficient time to meet deadlines. By the time of the 1985 comparative analysis, Carmonette analyses did cover enemy fixed-wing aircraft, but friendly fixed-wing counterair and IFF were not included. Previous concerns about the failure to address friendly close air support do not appear to have been addressed.

From the beginning, the ADAGE noted the importance of fixed-wing aircraft to the battle and included almost all aspects of fixed-wing air play, omitting only the effects of friendly close air support. Although the ADAGE recognized the need for IFF in determining gun reaction time, it did not play IFF directly in its portrayal of the air defense war. Rather, the air-to-air war was a separate component of the model and was played only for egressing enemy aircraft—that is, friendly aircraft could be killed by enemy aircraft only after the air-to-ground, ground-to-air, and ground-to-ground battles had been played. This procedure did not permit friendly aircraft to become a target for friendly air defense. Other aspects of the ADAGE that limited its portrayal of the air war included (1) sequential rather than simultaneous multiple enemy air raids, (2) inappropriate treatment of saturation attacks, (3) perfect intelligence in enemy air-raid planning, and (4) the uniform distribution of air defense weapons in a division defense zone.

Because the COMO was developed primarily for tactical air defense systems, it has always given particular attention to modeling ground-based air defense weapons versus aircraft, but it includes a detailed model of the air war. A simulation can be as simple as playing the Stinger weapon system against a single type of aircraft or as complicated as playing a fully formed defense at the divisional or theater level against diverse air attack scenarios that may include helicopters, various fixed-wing attack aircraft, and other supporting aircraft. All these abilities are external to the Stinger submodel. Although air defense appears to be modeled, the fratricide of friendly aircraft by ground-based air defense is not included.

Ground War

Since air defense weapons interact with the ground war, the complete modeling of air defense weapons should include coverage of the ground war to determine both the effects on the primary mission of air defense weapons and the survivability of the air defense weapons themselves. The Carmonette is an event-sequenced, fully computerized simulation of ground combat. All combined arms are included: infantry (mounted or dismounted), artillery and mortars, armored vehicles, and helicopters. The Carmonette can model movement, target acquisition, firing, damage assessment, and communications. Resupply and evacuation, however, are not covered.

The ADAGE does not model the ground war dynamically but, rather, plays ground battle attrition external to its Campaign submodel. Ground battle damage to ground targets, including air defense weapons, is input in the form of externally generated attrition rates. Ground-target attrition rates vary by target class and day of the war, while air defense weapon attrition rates vary by type of weapon, air defense zone, and day of the war. Loss of ground targets is determined by applying ground battle attrition rates, and ground losses are assessed prior to each enemy air raid each day. Ground war damage to air defense weapons is distributed equally among all weapons of the same type to maintain uniform density of air defense coverage.

Moreover, attrition rates are independent of air-to-ground damage. The study advisory group was concerned about the ground war attrition input data but could not decide upon the most appropriate scenario for generating input data. Compounding this problem was the group's determination that there was no known relationship between an ADAGE battle day and a battle day in the scenario being used to generate attrition data. Even though some advocates of the ADAGE believe that complete coverage of the ground war is not necessary to study the relative effectiveness of air defense weapons, the study advisory group directed that ground war attrition be a part of the ADAGE model. Although dissatisfaction with the ADAGE ground war attrition rates had been expressed, the acting director of the TRADOC studies and analysis directorate stated that there was nothing basically wrong with the ADAGE model and all that it required were reasonable inputs. It seems, then, that if the ground war scenario problems can be solved, the concerns may be dispelled about the ADAGE's portrayal of the ground war.

Unlike the Carmonette and ADAGE, the COMO does not simulate interactions between ground forces and, thus, does not measure ground battle damage to either air defense weapons or any other ground target. To the extent that air defense weapons should be threatened by ground attack, the realism of the COMO modeling approach is diminished. However, to the extent that the scenario avoids playing the forward edge of the battle area or establishes a scenario in which ground attack is not a factor—such as air base attack—then the absence of the portrayal of ground attack is not critical.

The Interaction of Air and Ground Wars

Models of air defense should allow the air and ground combat to interact in a reasonable manner. The Carmonette treats events dynamically, but the ADAGE allows no dynamic interaction between losses from ground fire and air attack. The ADAGE calculated all ground damage, whether caused by the ground war or air attacks, by applying attrition rates to ground assets. The assessment of losses was calculated between waves of air raids rather than during them, and damage depended on the type of target, among other things.

The Campaign submodel of the ADAGE uses externally generated attrition rates for the ground-to-ground war. It uses probability-of-destruction input from the munitions effectiveness subgroup of the Joint Technical Coordinating Group's survivability program to calculate air-to-ground attrition. The ground-to-ground and air-to-ground damage calculations are separate subroutines and do not interact. The portrayal of air-to-ground damage considers such factors as ground target class, number of targets in that class, total number of raids in an air wave attack, the assignment of those raids to targets, ordnance loadings, the probability of locating assigned targets, and probabilities of destruction that, combined with aircraft probability-of-survival factors, produce a parameter called the fraction by which targetable elements are to be reduced. This procedure produces average damage for all targets in a class rather than damage to specific targets.

The ADAGE documentation indicates that this procedure may lead to overestimating air-to-ground damage in certain cases. The ADAGE's approach to the air-to-ground war produced results that its critics call unexplainable, unconvincing, and disconnected from reality and that resulted in an attempt to require that they be made consistent with other TRADOC studies. This consistency was to be considered not equivalence but reasonable agreement with attrition results developed in other models. The study advisory group suggested that consistency might be obtained by having the ADAGE use an air threat similar to that used in the SCORES V scenario.

In this connection, even the the ADAGE's critics say that the model is useful for air defense weapon-system comparisons and that its per-raid attrition did not differ much from the Carmonette's per-raid attrition; it was the accumulation of attrition over multiple raids that caused problems. The results being questioned—especially damage by fixed-wing aircraft—could not be resolved by the Carmonette's results until the Carmonette played fixed-wing for the comparative analysis of 1985. Even then, battalion rather than division portrayal raised questions of the appropriateness of comparing the Carmonette's results to those of the ADAGE.

Proponents of the ADAGE assert that it is appropriate for comparing air defense weapon systems even though the air-to-ground damage results may be "too high." They state that accurate numbers are not necessary when comparing the relative effects of different systems. Even its critics agree that the ADAGE produced similar results—major damage by enemy air—no matter how many excursions were run. These attrition rates, which were considered excessive, cannot be overlooked, but the consistency of air damage to ground targets using different weapon-system combinations in the ADAGE cannot be overlooked either. Further examination of the aircraft damage to ground assets appears warranted.

The only way ground assets are damaged or destroyed in the COMO is by air attack. These assets in COMO modeling are often air defense weapons, although other ground-based assets may be included. Loss of ground targets, like all attrition in the COMO, is played probabilistically. The destruction of ground assets depends on successful attack by and survival of particular threat aircraft.

How well do the models we reviewed address the critical aspects of the combat arena in which the weapon system is to be used? All the models have weaknessess in the portrayal of at least one critical aspect of the air defense combat arena. The ADAGE and COMO give inadequate consideration to the effects of ground war activities on air defense weapons, and they do not completely portray the interaction of air and ground activities. The Carmonette's treatment of the air war is incomplete, since it continually failed to include fixed-wing aircraft effects and only recently addressed these aircraft, even indirectly. The strength of the ADAGE and COMO lies in the portrayal of the air war, while the Carmonette's strength is its good portrayal of ground activities.

Mathematical and Logical Representations

Another critical area of concern in modeling the operational effectiveness of weapon systems is how the theory and the phenomena are mathematically and logically represented. As we have summarized in table 4.8, three areas of concern about the ADAGE are the expected-value approach in the Campaign for modeling engagements of multiple air defense weapons against multiplane attacks, its use of the probability of participation of air defense weapons, and its apparent exaggeration of the DIVAD's survivability.

The ADAGE does not account for the spatial or temporal saturation of enemy aircraft—that is, many attacking at the same time. Rather, it uses an expected-value approach, in which the probability of aircraft survival in a many-on-many raid is based on crossproducts of simple exponential expansions of the basic one-on-one survival probabilities of individual air defense weapon systems. Some authorities believe this approach is severely flawed because its results are simple extrapolations of one-on-one free-encounter attrition factors and ignore the totality of a configured many-on-many encounter with its many potential interactions. These extrapolations suppress the stochastic or probabilistic effects of many-on-many encounters, because to treat them analytically in an expected-value approach is unmanageably complex. Even a small engagement of 10 weapons versus 10 aircraft requires more than 10 million analytical steps.

Therefore, it is not possible to relate the analytic equations to the specific parametric performance of a given weapon or to relate that performance to lower-level decisions and engagement rules. The analytical approach relies on the use of expected values to represent the behavior of random processes, and many of the possible variations thereby lost are adequate, of themselves, to materially alter the course of the battle and destroy the relationships and effects being investigated. A complete Monte Carlo approach to modeling is generally recommended.

Aggravating this basic unsoundness of the ADAGE is the process used to determine the number of air defense weapons to be used in the ingressegress portion of a many-on-many raid. The number of air defense weapons encountered by enemy aircraft is strongly influenced by another parameter—the probability of an air defense weapon participating in the defense against enemy aircraft. Determining the probability of an air defense weapon participating in the air battle starts with several assumptions: (1) the gunner has survived, (2) the system is operational, and (3) the gunner and the system are in the right place at the right time. Since the ADAGE does not play the ground war dynamically but assesses damage to air defense weapons through "bookkeeping" routines that account for damage at the end of a wave of aircraft raids, all weapons available at the beginning of a raid are presumed to be available throughout that specific wave. This could overstate the total number of weapons available within a wave.

Once these assumptions are accepted, however, the probability that any air defense weapon will participate in the air battle becomes a function of the weapon type, the zone in which the weapon is deployed, the type of attacking aircraft being engaged, and whether the raid is ingressing, attacking the target, or egressing. ("Zone" refers to the fact that the ADAGE partitions the division into four zones parallel to the area of the forward edge of battle.) Some of the factors depend on tactics and doctrine, the tactical situation, the commander's guidance, and the intensity of the ground battle. Specific considerations are the operational availability of the gun, the suppression that may have taken place, the movement of air defense weapons, smoke and dust conditions, and raid saturation. A systematic mixing of all these considerations results in a set of probabilities of participation for each type of air defense weapon against each type of aircraft.

These probabilities anticipate likely participation by the DIVAD except against ingressing and egressing targets 2.5 to 5 kilometers behind the forward edge of battle. The concept of probability of participation was not clearly understood in the simulation, and the first cost and operational-effectiveness analysis on the DIVAD, which was based on the ADAGE, indicated that the probabilities of participation might be optimistic. Although one of the reasons cited for using the Carmonette in the 1984 DIVAD update analyses was to shed light on this parameter, we were informed that this subject was not studied, and no relevant information was discussed in the update report.

Another area of concern relates to the ADAGE's definition of target sets, which led to an apparent exaggeration of the DIVAD's survivability. The ADAGE does not model direct attacks by aircraft on the DIVAD itself, since it does not model duels. Instead, the attrition of the weapon was played in the Campaign, which uses expected-value equations to calculate the probability of damage to ground targets by class from air attacks and assumes a random selection of targets within one target class. Similar procedures were used to assess damage to DIVAD weapons in the ground war.

This approach led to a problem in which the DIVAD was labeled the "immortal DIVAD." ADAGE results implied that it took 10 times the number of air-to-ground missiles indicated by the Carmonette to kill one DIVAD. Analysis by the study advisory group indicated that classifying the DIVAD in a target class by itself caused the ADAGE model to shoot all the helicopter missiles for the class at the one DIVAD; hence, the problem was one of target overkill rather than of the DIVAD's survivability. The correction of this problem—reclassifying the DIVAD into a tank-mechanized vehicle target set—was a source of discomfort to the study advisory group, because this implied a change in enemy helicopter firing priority.

The Carmonette also had problems with mathematical and logical representations. Even though its proponents asserted that the mathematics of the model was rather simple and straightforward, early attempts to model the DIVAD included at least one basic mathematical error. Early in the Carmonette's use, reviewers from the U.S. Army Air Defense Artillery School discovered that the Carmonette routines were incorrectly squaring a probability-of-kill parameter in its gun submodel. This would obviously distort the effectiveness results but was corrected for the Carmonette analyses of the DIVAD.

A logical consideration involving the Carmonette's application of Monte Carlo techniques that was of concern to the same reviewers was the procedure used to generate random numbers for various randomly occurring events in the model. The Carmonette generated random numbers only once, at the begining of the run; it used the same random numbers throughout the run. For example, the degree to which detection sensors would be degraded by enemy electronic countermeasures was selected randomly at the beginning and used throughout the entire run. It is reasonable to assume, even for the short battles that are modeled in the Carmonette, that the effects of electronic countermeasures would vary and that a better representation of them should have been modeled.

Finally, the Carmonette did a reasonably good job of modeling the dynamic interactions of multiple aircraft against multiple air defense weapons, but the probabilities of killing fixed-wing aircraft were determined with a procedure basically similar to that used in the ADAGE. The model primarily addressed one-on-one engagements in a few-on-few context and used the same approach discussed earlier to determine the kill probabilities applicable to a multiaircraft, multiweapon context. This opens the Carmonette to some of the same criticisms applicable to the ADAGE for fixed-wing aircraft.

In the COMO, weapons are unavailable for further use as soon as they are destroyed by aircraft attack. Similarly, the availability of weapons to engage target aircraft is limited to the actual capacity constraints of communications channels, launchers, radar, and so on. The "bookkeeping" capabilities of the COMO are constantly in use to determine the resources that are available and whether the operation of the system is possible. If threat aircraft did not come within range of a Stinger unit, the unit would not be engaged, regardless of how many threat aircraft were saturating an adjacent area. The COMO thus avoids the pitfalls of the expected-value approach. In return, it requires realistic scenarios, not scenarios that have been specifically developed to take advantage of the model's limitations.

How appropriate are the mathematical and logical representations used in the three models? The expected-value approach of the ADAGE is severely flawed in determining the effects of multiaircraft, multiweapon engagements. While the Carmonette's Monte Carlo approach alleviates some of these problems, its basic mathematical formulations of fixed-wing aircraft engagements are the same as those of the ADAGE. Moreover, both of these models have other, less serious mathematical and logical problems that threaten the credibility of the results. Only the COMO appears to be free of serious problems.

The Input Sources

We have noted the appropriateness of input factors throughout the discussion. In assessments of the credibility of simulations, data considerations are important, since even the best theoretical model produces noncredible results if it is based on faulty input data. Since the whole simulation can falter when input data are not clearly relevant, complete information about the data is necessary. In table 4.9, we have summarized the more critical aspects of input data for the three simulations.

Data Sources

All the models used data developed by recognized sources. The ADAGE documentation cited the tactical air division of the office of secretary of

Defense for planning and evaluation as its primary data source. Damage to ground targets by enemy aircraft—even though a source of criticism for producing unconvincing results—was based on the data and methodology from the joint munitions effectiveness manual of the U.S. Army Material Systems Analysis Activity, which also supplied weapon-system characteristics, as did the weapon-systems project managers and the Army Material Development and Readiness Command. Ground battle data came from the Combined Arms Combat Development Activity; visual detection data came from the U.S. Army Missile Command and the Night Vision and Electro Optics Laboratory in Fort Belvoir, Virginia. Some of the data came from the Army Air Defense Artillery School.

All these are typical data sources for DOD simulations. Even when the study advisory group expressed concern about input values from these sources, it had difficulty recommending more appropriate sources. With respect to visual detection, however, it should be noted that the laboratory's sources were used in the ADAGE to detect pop-up helicopters, principally because the Carmonette's modelers would not accept extrapolations of the VISPOE results from the missile command, even though they adopted its methodology for visual detection in their modeling of fixed-wing aircraft, since the laboratory's method applied only to helicopters.

The input data sources for the Carmonette included the Defense Mapping Agency for terrain data, the Atmospheric Science Laboratory for smoke and dust considerations, the Night Vision and Electro Optics Laboratory and the Army Missile Command for visual detection, and the Army Material Systems Analysis Activity for weapons characteristics and lethality data. The waterways experimentation station and the Tank and Automotive Command were the source of ground vehicle mobility information. These appear to have been appropriate data sources. The Carmonette depends on input from the users of the model for a description of the processes to be simulated, and since weapon systems, terrain, and time are explicitly modeled, there is no inherent limitation on what it can simulate.

The COMO uses some of the sources that the ADAGE and Carmonette use and some that are different. Like the ADAGE, it receives detection data from the Army Missile Command. Like the Carmonette, it uses terrain data from the Defense Mapping Agency. Like both the ADAGE and Carmonette, it uses lethality data provided by the Army Material Systems Analysis Activity. Scenario information comes from the TRADOC Systems Analysis Activity and the Army Air Defense Artillery School.

Additional scenario data from the Concepts Analysis Agency were used. Weapon-system characteristics were provided by the Army Missile Command for friendly weapons and the Intelligence Security Command for enemy weapons. The COMO's data sources appear to have been appropriate.

The simulations shared some data sources, especially for lethality data, but each model also had unique data sources. Since some of the sources did differ, there is always the possibility of differing qualities of data inputs across the models. One such area was visual detection, for which the Army has not yet resolved disputes concerning the data.

Data Quality

The appropriateness and structure of data for use in a particular simulation can be a source of concern. If the data are basically inappropriate or problems arise from structuring the data for use in a simulation, the simulation's results may not be well accepted.

The ADAGE produced results that were unconvincing to some potential users and some data items, such as target damage tables, were thought by the ADAGE modelers to yield overestimates of damage in some cases. The terrain data were recognized as old, and ground-war attrition rates concerned the study advisory group, which also considered target military-worth data—used in the ADAGE to measure the worth of unlike targets such as tanks and air defense weapons and, therefore, directly related to the ADAGE's measures of effectiveness—to be consistent with similar data used in other models. All these elements taken together probably led to the conclusion of one TRADOC official that there was nothing wrong with the ADAGE program—all it needed was reasonable inputs. In defense of the ADAGE, its proponents asserted that even though some data elements that related air damage to ground targets might be too high, they were alright for the ADAGE's purpose, which was to compare competing weapon systems. Correct relative values are sufficient for this, and correct absolute values are not necessary.

Visual detection ranges were a source of serious disagreement between the ADAGE and Carmonette modelers. The compromise, which was to use results from modeling the DIVAD with forward-looking infrared sensors for long-range searches, resulted in the use of inaccurate data to accommodate a correct theory (coverage of the gun's full range) and points out the need to establish data sources that will measure visual detection over the full range of a weapon without including weapon characteristics in input data that do not exist.

One aspect of the ADAGE data-handling requires special attention—how the ADAGE models weapon characteristics in its Incursion component. Because of the complexity and uniqueness of weapon-systems input data, the ADAGE modelers wrote the weapon-system characteristics into the model rather than addressing them through an external data base. This is contrary to requirements suggested by the Joint Forward Area Air Defense Test Force and was considered a weakness by the test force reviewers. In addition, since many of the data elements were classified, these reviewers were not able to review the documentation for the Incursion component that contained the computer program. While changes to the program could be made, the ADAGE modelers required new data of appropriate format for the model. Changes to the computer code of a model, even though supposedly limited to data elements, always carry the risk of uncontemplated changes to the program itself. This is a legitimate concern that nevertheless seems secondary compared to the problems with the basic data values themselves.

Overall, the ADAGE input values are a source of concern clouding its general acceptability. At the same time, however, the ADAGE's basic approach—comparing different weapon systems competing in the same functional area—should be carefully considered before unnecessarily stringent data requirements are imposed.

The Carmonette depends on the user's input for a description of the process to be simulated. Since weapon systems, terrain, and time are explicitly modeled, there is no inherent limitation on what the Carmonette can simulate. Its input structure allows considerable flexibility but also places the burden of obtaining realistic simulations on the analyst and requires extensive effort in data preparation. We have already discussed several significant data problems: the Carmonette's early use of the ZSU-23 data characteristics to model the DIVAD was corrected but the failure to properly represent the DIVAD's visual detection capabilities resulted in the use of incorrect data to model the DIVAD's full range.

Other problems relate to the Carmonette's input structure. Input data have to be tailored to meet the model's logic and to make the results plausible, yet tailoring opens the possibility that the end results will depend as much on the judgment of the analyst as on the manipulations in the model. Changes that seem insignificant can produce a widespread effect. One can speculate that the difficulty in tracing the reason for the divergence in ADAGE and Carmonette results (discussed in chapter 5) might be related to this tailoring. Tailoring data is time-consuming. A principal reason for not including fixed-wing aircraft in the Carmonette for the cost and effectiveness analysis update was insufficient time to do so and still meet study constraints.

Many of the Carmonette's early data problems were resolved but it shares one problem with the ADAGE that still needs resolution—how to handle visual detection. Moreover, the Carmonette's data-handling requirements regarding both time and tailoring can and did limit its usefulness.

The COMO uses program modules that describe the characteristics and operations of specific weapon systems at varying levels of detail, depending upon the intended application. To produce the Stinger battery-coolant-unit usage study, it was necessary to increase the detail over that of the standard Stinger model by making program changes to the initial Stinger model. While Stinger engineering data are straightforward and reasonably reliable, human-factors data for Stinger personnel reactions and functions (such as detection and engagement processes) are less well understood.

Like the Carmonette, the COMO requires some tailoring of the input data, which must be evaluated to determine the appropriate factors to include. Thus, as with the Carmonette the data-tailoring may be important to the model's results.

In summary, there were problems with obtaining appropriate input information for the ADAGE and Carmonette. Some of these problems were corrected and some were not. The Carmonette and COMO required data-tailoring, which raised the question about whether the results depended as much on the data-tailoring as on the models' manipulations of the data. All the models used recognized data sources, although not necessarily the same sources. A result of differing sources could be differing quality of data inputs. Data problems did occur, at least in the ADAGE and Carmonette, and some of these problems were related to challenges of the results. Data-structuring presented problems to both the ADAGE and Carmonette. The Carmonette's extensive structuring requirements prevented a timely inclusion of fixed-wing aircraft. More attention to a model's data requirements should improve its usefulness and credibility.

Summary

Our review of the ADAGE and Carmonette models of the DIVAD and the COMO III model of the Stinger led us to these conclusions:

- The Carmonette has sound theory for a combined-arms analysis, but its approach is not the most appropriate for decisions regarding competing air defense weapons. The ADAGE and COMO III were designed with such decisions in mind.
- All three models have specific strengths in dealing with the critical aspects of air defense weapons but all also have serious weaknesses.
- All three models are in some respect restricted and incomplete in their coverage of the combat arena.
- Of the three models, the ADAGE has the greatest number of basic mathematical and logical flaws that raise concerns about the credibility of its results.
- All three models address operational measures of effectiveness, but the ADAGE appears to relate its measures more closely to protection, the ultimate mission of air defense, while the other models stress loss-exchange ratios.
- The ADAGE and Carmonette simulations of the DIVAD both had problems with obtaining appropriate data, and these problems affected the credibility of the simulation results; the Carmonette and COMO III require extensive tailoring of the data, and the effects of this cannot be easily distinguished from manipulations of the models.

All the models we reviewed had advantages that made them applicable for answering certain issues and disadvantages that detracted from their usefulness. We recognize that it is practically impossible for a simulation to fully address all aspects of an issue. The question becomes, Is the simulation sufficiently applicable to address the critical aspects of the issue?

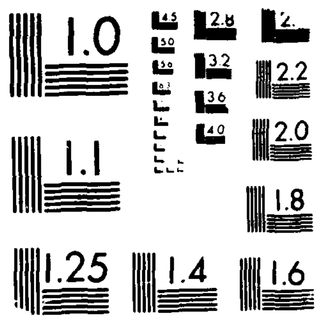
The basic theoretical approach of models is a key consideration. The ADAGE and COMO are functional models, designed to compare specific types of air defense weapons, whereas the Carmonette is a combined-arms model, focusing primarily on alternative strategies in ground war. From this perspective, the ADAGE and COMO are pernaps more appropriate than the Carmonette for their purpose—deciding between air defense weapons in a given scenario. Even critics of the ADAGE agree to its usefulness this purpose.

None of the models fully address the tactical environment of the weapon system studied; nevertheless, each has definite strengths. The ADAGE's strengths lay in its portrayal of the DIVAD gun in its intended environment and in its coverage of helicopter and fixed-wing targets and their ability to inflict serious damage. The Carmonette's strengths were its portrayal of the ground battle and its dynamic interactions. Like the

AD-A191 503 DOD SIMULATIONS: IMPROVED ASSESSMENT PROCEDURES WOULD INCREASE THE CREDIBILITY OF RESULTS(U) GENERAL ACCOUNTING OFFICE WASHINGTON DC PROGRAM EVALUATION AN.. 2/2

UNCLASSIFIED DEC 87 GAO/PEMD-88-3 F/G 5/1 NL

END
DATE
FILMED



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

ADAGE, the COMO portrays an appropriate air defense environment with its essentially unlimited battle size.

The Carmonette's weaknesses prevent the model from completely simulating air defense, since its scope is too small and, until recently, it failed to address fixed-wing aircraft, a principal target set. The COMO's failure to represent the movement of the air defense weapon causes the model to overlook portability, a principal characteristic of the Stinger, while its short timespan limits its usefulness for studying extended warfare. The ADAGE's approach to terrain detracts from the realism of its modeling while improving the ability to generalize from it. In summary, the ADAGE and COMO address ground-to-air activities reasonably well, while the Carmonette's strength lies in its treatment of ground activities.

Both the ADAGE and Carmonette modeled how the DIVAD gun detected enemy aircraft and included provisions for both radar and visual detection. They addressed visual detection differently because of differences in theory and input data. Neither model has yet appropriately modeled the DIVAD's visual detection characteristics, since disagreement over the visual detection components of the models has not yet been resolved, leaving unanswered questions as to whether any of the DIVAD studies have appropriately modeled the visual detection of enemy aircraft.

The COMO suffers from some of the same shortcomings as the ADAGE and Carmonette. Like the Carmonette, its coverage of detection throughout the full range of the weapon is questionable, and like the ADAGE, it lacks realistic coverage of battlefield obscurants. The Carmonette tends to give more complete coverage to radar phenomena than the ADAGE but only after significant model changes. Radar was not applicable to the Stinger in the COMO.

While the ADAGE and Carmonette address the same basic phenomena in modeling an engagement between the DIVAD and an approaching aircraft, differences could affect the acceptability of some of the results—some favoring the ADAGE and some the Carmonette. All things considered, the Carmonette probably models the engagement of enemy aircraft better than the ADAGE, since it models more phenomena directly and uses Monte Carlo throughout. Nevertheless, the Carmonette suffers from a more basic problem; it does not model the DIVAD in its intended environment.

We found differing emphases on the aspects of the air and ground wars that were modeled and how they interacted. All the models failed to address certain aspects of modern warfare and addressed other aspects inadequately, limiting the insights to be gained about the effectiveness of new weapons in battles of the future. Throughout much of the modeling effort on the DIVAD, the Carmonette gave inadequate coverage to fixed-wing aircraft. The ADAGE's expected-value treatment of the ground war raised concerns about its credibility.

The ADAGE covered nearly all aspects of the air war, including the damage to division ground assets by enemy fixed-wing aircraft. The Carmonette analysts did not include the effects of fixed-wing aircraft in the Carmonette, ignoring it completely in early studies and relying on data from other models in later studies.

The Carmonette was designed almost 30 years ago to simulate small-unit ground combat and addresses nearly all aspects of combined-arms ground warfare. The ADAGE does not play the ground war directly but relies, instead, on externally generated attrition rates. The ADAGE's approach to modeling the ground war attrition input data and its estimates of air-to-ground damage are principal areas of disagreement for its critics. The COMO does not play ground war at all.

The preservation of ground assets is the primary function of air defense. The ADAGE addressed this in its analyses, but the study advisory group appeared to be reluctant to consider requiring this measure in the Carmonette analyses. Consequently, the Carmonette results concentrate on various exchange ratios that are principally attrition oriented. Since the COMO did not play the ground war at all, its ability to address the protection of forward-area assets was limited. Therefore, of the three models we reviewed, only the ADAGE addressed air defense weapons in their primary roles. However, even the ADAGE failed to address one important aspect of air defense that was addressed by the Carmonette—the ability of air defense weapons to cause aircraft to abort their missions.

The ADAGE fails to address explicitly the time and spatial relationships of a many-on-many raid, relying rather on expected-value calculations. How much this theoretical and mathematical problem detracts from the results is difficult to determine because of the concurrent problems associated with the input data. While we found some fundamental errors in the theoretical approach to modeling air defense, many of the problems we noted appeared to deal with the appropriateness of data inputs. Sometimes the problems with the characterization of a phenomenon and its environment stemmed from using inaccurate data to achieve a correct theoretical approach.

All the models took approaches to data treatment that were unique in some respects and each had peculiarities worthy of note. While all the models obtained their data from recognized sources, how they used the data tended to differ. If the wide divergence in results can be explained and corrected for, then the ADAGE would appear to be able to give the most complete treatment of air defense weapons. Even adding fixed-wing elements to the Carmonette, it may remain less appropriate for air defense issues because of the level of battle portrayed. Questions about the ADAGE's portrayal of the ground war and the COMO's limited modeling of the ground war detract from their ability to measure air defense protection of ground assets in a combined-arms environment. The Carmonette appears capable in this area, but protection, air defense's primary mission, was never stressed as a measure of effectiveness in the Carmonette analyses.

Attempts continue to be made to solve the problems associated with aspects of the theory, model design, and input data in the Carmonette and ADAGE. We believe that as these efforts continue, both models may become more appropriate for analyses of the effectiveness of air defense weapon systems. As some of the problems are resolved, the results may become more comparable—that is, if the principal source of difference in results does not prove to be the size of the battle being modeled. The COMO, however, cannot be as comprehensive an analysis device until it too addresses the effects of ground war activity.

Supporting Material for Chapter 5

Verification

Determining that a computer program performs as the simulation analyst intended occurs during the development of the simulation. Verification efforts should also occur whenever substantial changes are made to the simulation. Even before verification begins, some of its components will have been defined by the selection of the computer simulation language. Programming conventions and policies such as structured programming further define the context for verification.

A number of techniques have been developed or adapted to assist in verification. Techniques include a "structured walk-through," a line-by-line code review performed by several members of the modeling team; program traces, listing the values of key data elements after each event during operation; computer runs made under an extremely simplified scenario; graphic displays of simulation output; and the intentional insertion of errors (or "seeding") prior to line-by-line review to develop estimates of remaining errors. (We have summarized evidence of verification for our three case studies in table 5.2)

We were informed that no formal verification effort had been conducted for the ADAGE but that some line-by-line checks of computer codes to develop an understanding of the model had uncovered some problems that were corrected. The Carmonette, originating in the 1950's, has undergone many changes since then and is still being changed. We found no evidence of verification efforts but were informed that the model has been subjected to extensive peer reviews. We were unable to document verification efforts related to either the standard Stinger model or the version that was developed for the battery-coolant-usage analysis. It was developed by a contractor, and the Army Missile Command informed us that the command performs verification and validation tests for model acceptance. We had no data on those tests.

Identifying verification efforts appears to be one of the more difficult issues of our framework. We did not identify documented verification efforts specifically related to the DIVAD or the Stinger. Our discussion and review of a number of simulations lead us to believe that, in general, there is no audit trail to identify verification efforts. Verification is an integral part of programming, but, like programming, it is often not documented.

Statistical Representation

Experts in simulation have noted that in many simulation studies, the greatest time and money are spent on design, development, and programming and that relatively little effort is given to analyzing a simulation's output data. Since Monte Carlo models produce results by sampling variables represented by probability distributions, sufficient numbers of replications and the appropriate statistical analysis of the simulation results are necessary to allow reasonable confidence that the simulation results are representative of the model's true values. The objective of the analysis is essentially to develop estimates of both the expected value of outcomes and their variance. In practice, it appears that the larger, longer-running simulations are less likely to be subjected to this analysis because of the major demands that they make on computer time. In fact, this behavior has received some theoretical support. As early as 1965, Brooks argued that only a few replications of a large battle model are needed to get good estimates of the gross results (emphasis ours), provided that the fate of a given weapon has strong influence on the fates of only a limited number of other weapons (Brooks, 1965). Many analysts, however, believe that multiple replications are especially needed when detailed results are examined. Some analysts have also given attention to developing statistical procedures that will reduce the required number of replications. (We have summarized the evidence of statistical representation in our three case studies in table 5.3.)

We were able to determine that substantial attention was given to identifying the true model mean for the Incursion submodel of the ADAGE. The only portion of the ADAGE model that is Monte Carlo is the Incursion submodel, which produces the one-on-one probabilities of kill that are subsequently used for each weapon system modeled in the Campaign submodel. The original cost and operational-effectiveness analysis of the DIVAD stated that the Incursion had undergone "a sufficiently large number of trials" before the probability of kill of an average engagement was calculated. Analysts involved with the ADAGE informed us that each Incursion scenario was replicated 500 times in producing probability-of-kill results. They noted that these replications yielded a 98-percent level of confidence that the Incursion results were within 1 to 2 percent of the true mean, although the specifics are not presented in their reports. The ADAGE analysts further stressed that this practical ability to generate a large sample size is an advantage that the ADAGE has over the Carmonette.

Because the Carmonette requires substantial computer time, only a limited number of replications are available to establish confidence that results have stabilized. In both the 1984 DIVAD update and the 1985 comparative analysis, the minimum number of replications required was 10. After 10, the analysts conducted statistical analyses to determine whether the results had stabilized. The criterion for determining stabilization was an 85-percent level of confidence that the results were within 10 percent of the true mean. The principal measures to which this criterion was applied were total enemy losses and total friendly losses. In the 1984 update, 21 of 26 scenarios tested met the confidence criterion within the minimum 10 replications. The largest number of replications needed was 17. In the comparative analysis, the analysts determined that all 29 scenarios tested met the criterion within the original 10 replications. For 2 scenarios in the update and for one in the comparative analysis, however, the Carmonette analysts accepted scenarios as stabilized that only approached, but did not meet, the 10-percent precision factor.

The Carmonette analysts elected to measure stabilization on total enemy and total friendly losses—rather than enemy aircraft killed by DIVAD and vice versa—because of the small number of guns and targets available in the battalion scenario. Nevertheless, it is noteworthy that these output measures were not nearly as stable as total losses, several scenarios showing standard deviations as large as or larger than the mean.

The Carmonette analysts justified the decision not to run additional replications to stabilize these variables by stating that since the mean values were so small, more replications would not necessarily produce a significant difference in the computed mean. Since standard deviations on these variables are often large, relative to the mean values, it would seem that wide variations in mean values could still occur. Unfortunately, the reports do not contain enough information to judge the volatility of potential variation in values or whether the values were beginning to converge at all on the acceptance criterion. Since the divisional tactical environment for the DIVAD includes the coverage of 36 guns, whereas the battalion-level Carmonette covers only 4 guns, there is some concern about the possible effects of using unstable results from a battalion-level model for projecting the operational effectiveness of the DIVAD in its tactical division environment.

The Stinger battery-coolant-unit usage study addressed the requirement for coolant units under varying conditions of visibility and types of threat and supporting air defense systems. All the computer runs generating the data were part of a total COMO simulation. An implicit assumption, however, that one run for each set of conditions was sufficient raises the issue of the number of replications required for large-scale simulations.

The battery-coolant-unit study included a total of 11 computer runs, one for each scenario. If multiple replications of at least one of the scenarios had been made, the analysts might have been better able to assess whether the values produced by one run were near the true mean. There was no indication in the report as to the variability of results—no calculation of model mean or variance. No reason was given for this omission. The analysts may have believed that a single run for each scenario was acceptable because of the large number of Stinger units operating within the simulation, but no arguments were advanced to suggest or support this rationale. The variability of output results may have been tested when Army personnel performed validation and verification testing on receiving the model from the contractor, but this was not documented in the report.

In the ADAGE and Carmonette cases, the evidence indicates that the analysts recognized the need to estimate some of the true model values. The credibility of the ADAGE simulation benefited from the multiple replications used to develop statistically representative values. In the Carmonette analysis, it is not clear that true model values were determined for enemy aircraft killed by the DIVAD and DIVAD guns killed by the enemy, although the attempt was made to determine them. In the Carmonette analysis and implicitly in the battery-coolant-unit study, there is an indication that the analysts tended to combine testing for underlying true model values with testing for changes in results stemming from parameter and scenario changes. This practice leads to a confusion of two important but distinct areas and, thus, to a decline in credibility. The Carmonette analysis did use multiple replications in its scenarios that enhanced its credibility in the development of statistically representative values, but there are still some concerns about the stability of some of its results. There was no evidence that the statistical representativeness of the COMO simulation was determined for either the model or the scenarios.

Sensitivity Testing

Sensitivity testing identifies how changes in a model's parameters affect the results in both direction and magnitude and provides feedback of the model's behavior to the analyst. When changes extend beyond the alteration of parameters, the process is recognized as the testing of alternative scenarios. Parameter testing is most likely to be explored early in a simulation's development, but scenario testing is generally performed in response to particular questions about the system's effectiveness under a range of threats and conditions. (We have summarized the evidence of testing for sensitivity to parameters and alternative scenarios in table 5.4.)

The ADAGE modelers dealt explicitly with sensitivity testing and testing for uncertainty, covering both in their published reports. In the 1985 comparative analysis, they conducted sensitivity analyses (or "parametric analysis") on four parameters: operational availability, reaction time, aim bias (that is, the offset of the center of the aim distribution from the target), and angular aim error (that is, the dispersion of aiming points around that center).[1] The report indicates that operational availability, reaction time, and angular aim error were critical parameters in determining the DIVAD's effectiveness. Moreover, the direction of changes in results was logically consistent with the direction of changes in parameter values.

The ADAGE modelers included a chapter on uncertainties in the original cost and operational-effectiveness analysis report. Their concerns about uncertainty in the modeling were

- operational employment concepts visualized for each weapon system,
- environments in which systems may be placed on battlefields of the future,
- threat levels and tactics to be encountered,
- system performance characteristics that directly affect effectiveness inputs.

The first element of uncertainty dealt with the use of weapons such as rifles and tanks in an air defense role, and analyses showed enough difference to conclude that ground weapons should be integrated into the air battle and air defense weapons into the ground battle. The environmental aspects dealt principally with the effect of uncertainties in visibility, and the results show an extreme sensitivity to visibility. The threat uncertainties dealt principally with expected enemy tactics and indicated significant increases in damage from heavily concentrated first-day enemy assaults. The effectiveness uncertainties were

[1]This dispersion is distinguished from ballistic dispersion. Angular aim error deals with the variability of a gunner's aiming ability. Ballistic dispersion is a function of the gun barrel and the projectile and is sometimes referred to as "round-to-round error."

addressed in two ways, first by holding the DIVAD's effectiveness constant and degrading the relative effectiveness of other air defense systems and, second, by allowing all air defense systems, including the DIVAD, to be equally degraded in effectiveness. The changes in effectiveness showed that the DIVAD held up well.

While there is evidence that the Carmonette modelers have conducted sensitivity testing on input parameters in the past, the published reports based on the Carmonette analyses of the DIVAD do not cover such testing. However, some of the scenarios varied so little that they were essentially the same as sensitivity testing. The Carmonette analysts addressed 26 different scenarios in the 1984 update and 25 in the 1985 comparative analysis. Many of these varied conditions too much to be called sensitivity tests (for example, the presence or absence of air defense, the presence or absence of certain types of weapon systems); others changed conditions only slightly and, therefore, are similar to sensitivity tests.

Although analyses in both Carmonette studies showed that changes in battlefield visibility had significant effects on the DIVAD versus enemy aircraft effectiveness, these effects were small and had only small effects on overall battlefield outcomes. The update analyzed the difference between 7-kilometer and 3-kilometer visibility ranges; the comparative analysis reported on the difference between 7-kilometer and 16-kilometer visibility ranges. Both studies also reported on the effects of changing the mode of operations for the DIVAD gun to show the effects of not using some of the DIVAD's radar capabilities to track helicopters. The update reported only on tests for 7-kilometer visibility days and showed that while the performance of the DIVAD itself is extremely sensitive to the mode of operation, overall combined performance changed only slightly. The 1985 comparative analysis reported the same pattern of results for 7-kilometer visibility days but showed little variability for 16-kilometer visibility days. Additional tests conducted in the update included the effects of modeling capabilities demonstrated in test firings versus modeling projected or mature DIVAD capabilities. These results indicate significant sensitivity in performance but not much change in overall results. Finally, the Carmonette reports showed only slight sensitivity to different levels of attacking enemy forces at the beginning of the battle.

The importance of a scenario that encompasses more than the single type of weapon system was clearly demonstrated in the Carmonette scenarios that varied the DIVAD's capabilities between those that were current and those of the mature weapon. The results, if accurate, indicated that even though the gun's performance improved, there was little change in overall air defense performance. These results could not have been developed except by using a simulation in which the DIVAD was but one element of the defense system, demonstrating the need for the appropriate context in which effectiveness questions can be posed.

The COMO Stinger battery-coolant-unit simulation included sensitivity analyses of visibility. They were accomplished by changing only the Stinger team's visibility for several of the air defense-threat combinations. The quality of this effort would have been greatly improved, however, by multiple runs. Sensitivity analyses for some other weapon-system models used in the COMO have also been performed. We found a documented example of sensitivity analysis performed on the COMO HAWK surface-to-air missile model.

The COMO battery-coolant-unit study, however, is an excellent example of developing scenarios that could provide a comprehensive view of the simulation's response under a broad range of alternatives. The major weakness of the 11 scenarios is that only one replication of each one was made. The insensitivity of results among the scenarios, relative to the battery requirement, suggests that additional replications were probably not needed. Nevertheless, it is poor procedure to ignore the need for some measure of variance, especially since there is no evidence that earlier analyses developed any measure of variability of the model's results. Even when the variability of results has been estimated, the need for multiple replications of a scenario must be carefully considered.

The questions raised with regard to a model are formulated so that answers can be developed by experimenting with parameters and scenarios. Scenario testing is essentially what we equate with results. Valuable information that contributed to credibility was developed in the ADAGE, the Carmonette, and the COMO by varying parameters and testing alternative scenarios.

Validation

Validation is the process of determining that a model is an accurate representation of, or agrees with, the real-world system being modeled. Validation includes comparing simulation results to results from the actual system or from other models, historical data, and operational testing. In the context of our framework, we are interpreting validation narrowly as the process of developing confidence in the simulation results by comparing them with results from other sources. (We have summarized our case study results for validation in table 5.5.)

In reviewing simulations related to the DIVAD, we found several examples of validation efforts for engineering simulations that were planned and conducted as part of the simulation development. The role of the Army Materiel Systems Analysis Activity in validating an engineering model of the DIVAD developed by Ford Aerospace demonstrated the interrelatedness of verification and sensitivity analysis with validation. The effort identified validation as a purposeful function within the decision-making process; and it described problems that can be expected when a system's data are collected for validation purposes. In addition, its documentation made extensive use of graphs that contributed to the analysis of the results.

In contrast to the validation efforts for engineering models, we found that validations of operational-effectiveness simulations are not planned for or conducted routinely but, rather, are undertaken when individuals or an organization questions a disparity in results between similar models or between the model and real data or even between the model, perceptions, and impressions. Validation of the operational-effectiveness simulations in our case studies was undertaken to address the questions or issues that arose. For example, a comparison of the modeling of the DIVAD by the ADAGE and Carmonette was requested because of the substantial variance in results reported for the two models.

In another situation, an undersecretary of Defense wanted the Army and Air Force to jointly review their models, the COMO and SORTIE, to understand the substantially lower attrition of U.S. aircraft against Warsaw Pact air defense compared to Warsaw Pact aircraft against the air defense of the North Atlantic Treaty Organization. In another example, the COMO configuration management board, questioning whether the simpler COMO simulations yielded information similar to that of the more complex ones, requested a study that would corroborate the output of simpler and more complex Patriot simulations. The results of this study lent credibility to the COMO integrated air defense model, which uses simpler weapon-system models. In the following discussion, we explain why some of these simulations were important to our case studies.

When we made our review, no formal validation efforts had been performed on the ADAGE. Because the ADAGE modeled combat at the division level, test data were generally not available for performing validation efforts, since tests are not conducted at that level. Perhaps because the Carmonette has been in existence so long, it has come to be viewed as a "standard" against which to validate other models rather than as a model requiring validation itself. We did, however, identify one Carmonette validation effort reported in 1975, when the Army Concepts Analysis Agency compared the results of the Carmonette with a tank warfare field experiment. The use of both the Carmonette and ADAGE to model the DIVAD was really an attempt at validation because original ADAGE results had not been well received in some circles. Part of the justification originally given for using the Carmonette to analyze the DIVAD was to provide insight into certain key parameter values used in the ADAGE.

Because early efforts showed Carmonette results that diverged from ADAGE results, a combat development study plan was adopted in January 1984. It established a study advisory group and described the tasks and responsibilities necessary for correcting model, scenario, and data problems discovered in the ADAGE and Carmonette models. These problems included correcting the Carmonette's model of the DIVAD, modeling the DIVAD directly instead of the ZSU-23, and including the DIVAD's primary mode of operations. The effort was to go beyond an examination of inputs and outputs and provide a description and evaluation of each simulation, covering structure, scenario, inputs, data usage, and outputs. Its purpose was to give insight into how a simulation affects the perceptions of a system's performance and combat effectiveness. The uncertainty regarding the modeling of the DIVAD was great enough to cause concern that results for other systems such as the AAH helicopter and M1 tank might be affected if decisionmakers lost faith in the ADAGE and Carmonette.

The considerable concern about the credibility of the disproportionately heavy losses in the ADAGE attributable to enemy aircraft was reflected not only in the minutes of the study advisory group but also in our discussions with DOD personnel. Evaluating the legitimacy of this concern is difficult. One comparison of the results from the ADAGE, Carmonette, and other models, for example, showed that much greater damage was attributable to enemy aircraft in the ADAGE than in any of the other models. However, this comparison included battles of different lengths (from 30 minutes to 7 days) and different coverage (from battalion to theater), so that direct comparisons are problematic. Moreover, the Carmonette did not include ground damage by fixed-wing aircraft.

This comparison did, however, serve as the basis for further analyses for the 1984 update in which adjustments were made for consistency of inputs and the scenarios were made more comparable with respect to size and duration of battle. Results from a segment of the ADAGE battlefield were compared to the Carmonette results. Results from the other model were also normalized to establish a comparison base. Results from the adjusted scenarios showed that the damage attributable to enemy aircraft in the ADAGE was basically comparable in the other models and, in fact, somewhat conservative. In similar analyses for the 1985 comparative analysis, the ADAGE and Carmonette comparative results diverged. In the meantime, however, several changes had been made to the Carmonette, principally the addition of fixed-wing aircraft and changes in enemy infrared countermeasures. The comparative analysis report cited different modeling of suppressing or aborting helicopter missions and differing levels of battle, but the precise source of the new divergence could not be pinpointed, since several changes had been made at one time. Consequently, there are still unanswered questions about which of the two models, if either, produces the more believable results.

In reviewing the ADAGE and Carmonette by comparing results, we also made the following observations:

- The ADAGE per-raid attrition results are close to the Carmonette raid results.
- Carmonette analysts admit substantial problems in measuring the effectiveness of various air defense systems because of the size of the model, the number of air defense units, and the design and operation of the enemy fixed-wing aircraft.
- While the Carmonette used live-fire test results from Fort Hunter Ligget to help in modeling the DIVAD, there is some concern as to whether these tests were fair to the DIVAD, because test conditions at Hunter Liggett did not match European battle conditions very well.
- Attempts to crossvalidate the Carmonette against another model, using both to design thermal pinpoint-firing operational tests, were unsuccessful. The results from the two models were inconsistent, and both sets of results were inconsistent with respect to the operational test results.
- The cost and operational-effectiveness update based on the Carmonette tended to support the ADAGE conclusion that the DIVAD was the preferred weapon.

In our opinion, there are still enough unresolved issues, both here and as discussed in chapter 4, to raise questions about whether either the

ADAGE or Carmonette has been "validated" as a model for studying air defense. The use of both the ADAGE and the Carmonette to study the DIVAD did lead to improvements in both models, but more work still needs to be done.

The COMO is an extraordinarily dynamic simulation system. With the addition of new weapon models, differing levels of detail, multiple scenarios, and variations for different computers, addressing the issue of validation is a complex question. Validation efforts should be directed at the several levels of simulation at which the COMO model is run: the generic large-scale air defense scenario; the detailed simulation of a particular weapon system; and simplified, faster-running versions of weapon systems that can be substituted in some applications for more detailed ones.

We did not find documented evidence of a validation effort specifically for the Stinger weapon simulation, but we did find evidence of validation of the overall COMO model that lends credibility to any effort in which the COMO is used. We also found a validation effort for the Patriot that was sufficiently successful to lead to similar weapon-system validations.

The validation effort that addressed the overall COMO model came about as a result of Army and Air Force interest in jointly understanding the reasons for major differences in attrition estimates. This analysis was in a sense a validation of two models using the overall similarity of results as the measurement device. The great disparity in results between the COMO and the SORTIE suggested initially that either one or both models had serious failings. In May and June 1980, Air Force and Army evaluators met to review the scenarios, input data, structure, and assumptions of the two models and each group adjusted its model to reflect agreed-upon standard conditions and assumptions. The results, which were overall measures of attrition, indicated that the models are in good agreement when simulating similar conditions. The original differences were primarily attributed to different estimates of system effectiveness and differences in aircraft attack philosophies, goals, and doctrine. This resolution lent credence to each model and suggested factors likely to be modified in military planning. The emphasis was on the selection of input data that accurately represented operational conditions. Test procedures for sensitivity analysis were present in the form of limited variation of individual factors and determining their effect on results. This practical effort addressed the credibility issues of interest to staff at high levels of the Department of Defense.

The validation for the Patriot simulations used with the COMO was a rigorous evaluation effort. The quantity of input data was kept at an experimental level rather than extensive. Most of the processing was limited to single or several aircraft against a single Patriot battery. Thus, the results could be closely analyzed and an explanation for differences could be determined. As substantial as such validation efforts must be, they provide a degree of credibility that is probably not matched by any other weapon-system simulations with the COMO.

Attempts to validate the ADAGE and Carmonette simulation results with each other and with those from other models have not been completely successful. Differences between simulations make comparison quite difficult and, while some results have been basically comparable, under other conditions they have diverged. In one major comparison between the ADAGE and Carmonette, there was a reasonable correspondence between results that did not carry over to a second major comparison. There was no evidence of the use of historical data. Operational data were used as input to the Carmonette but not comparatively.

The comparison between overall simulation results demonstrated in the COMO-SORTIE analysis did not approach the rigorous standards of the Patriot analysis, but such analyses become less feasible as the compared simulations encompass a larger and more complex environment and represent more divergent modeling approaches. There was no documented evidence of validation for the Stinger battery-coolant-unit model.

Independent Validation Efforts

In the validation efforts described in chapter 5, work was performed by the organization developing the model or one organizationally related. The Army Material Systems Analysis Activity provided an independent review of the Ford DIVAD gun model but had previously participated in writing its design specifications. In the air battle attrition validation effort, the Air Force and Army operated their own simulations (the SORTIE and the COMO). The ADAGE and Carmonette were modified or corrected by their respective organizations but under the review and direction of the study advisory group. These instances approximated independent validations that may enhance a model's credibility. We identified two others, not part of our case study analyses, that are noteworthy because of the efforts made to ensure impartial evaluation.

In 1983, the Center for Naval Analyses prepared a report for the Navy's Harpoon (an antiship missile) project office to assist in selecting one or more models that represented the state of the art in evaluating the Harpoon's performance. The center was asked to provide a detailed comparison of six widely used models that had been developed by various naval laboratories and contractors. The models were developed for different applications, and as the scenarios were made more complex, some models were not applicable. Some of the models ran on mainframe computers and others on minicomputers. Comparisons appeared to be for overall similarity of results, with some comparison of specific events between models. The data that were produced did not allow statistical analyses of important model details such as flight paths, for example. When disagreements were found, the center attempted to find the cause and gave the developers of the models the opportunity to make corrections and rerun the scenario. When differences appeared to be the result of the modeling approach or basic assumptions, the apparent causes and results were documented. One especially interesting outcome was that several of the models gave results for a many-on-many scenario that were nonintuitive and would not have been suggested by the one-on-one scenarios.

In the second case, Sandia National Laboratories had developed a model, the SANDEMS, for analyses related to a surface-to-air missile. A committee of users, representing various naval commands, laboratories, and contractors recommended that the model undergo validation so that its results would have more credibility with the Navy and be formally accepted for the Navy's AEGIS project.

While there is no formal process of models review at the Sandia laboratory, the developers had subjected the SANDEMS to an informal validation and verification. The independent reviewers agreed, however, that it was not a full-scale validation of all the aspects of the model. The only documentation available was preliminary, partial, and in some cases made up of obsolete subroutine descriptions and the program listing.

The Applied Physics Laboratory, a member of the committee, was asked to undertake the validation effort. RCA Corporation, also a member, provided assistance. The laboratory selected personnel who had extensive experience in the development and use of naval surface-to-air missile engagement models. They developed a detailed checklist as a framework: (1) steps in validation (purpose of model, completeness, realism, correctness of data and computations, flexibility for expansion), (2) structural description of the model, (3) scenario capability, (4) model output, (5) nuclear effects, and (6) modeled processes. Test runs for some scenarios agreed with accepted results from other models. For other scenarios, the analysts relied on examining the logic of the model because there were no other models with broad Navy acceptance to which they could compare the results.

The review identified the limitations of the model in terms of what it represented and what it failed to represent or represented only partially. The review also described the conditions that could and could not be validly addressed because of the limitations. Early in the assessment effort, the analysts noted that "The items judged to be critical for SANDEMS validity will depend on the intended purposes of the model," thus recognizing that validity depends on context.

Summary

We recognize that verification is substantially integrated with the programming process and that documentation of the process has been sparse, even though the documentation of the programmer's product may be quite complete. We think that simulation users are entitled to some knowledge of the verification efforts in a simulation's development and that such information strengthens the credibility of simulations. It can be incorporated into existing documentation. We have seen that when questions about credibility are raised, other analysts brought in to assess a simulation do perform their own verification efforts. We take this as evidence of the importance of verification and the need for some recording of it.

In our case studies, the longer-running, more complex simulations were evaluated with fewer simulation runs. If this represents a tendency to treat the results of one or a few runs of a complex model as "true" estimates, we see the potential for substantial questions about credibility. If the true values of simulation are not known, then one really never knows the degree of comparability between the model and the real world or between the model's results and the results from other models.

We note that the number of simulation runs is not the sole criterion to be used in developing these estimates. Various statistical measures have been and are being developed to increase the efficiency of the estimating process, but the basic issue—confidence that the simulation results are an accurate reflection of the model's underlying values—is important and requires recognition. Analysts working with very large, long-running simulations such as the COMO should try to develop or identify methods in which confidence levels based on results from individual model components can be incorporated into the total estimation process, so that less demand will be made on computer resources for a given level of confidence.

The sensitivity testing of parameters and the testing of scenarios was treated effectively in the ADAGE, Carmonette, and COMO. In fact, the apparent need is to integrate parameter and scenario work with the work of determining the true estimates for a simulation. When the true underlying results are not determined, then the simulation results are open to question. Analysts have a firmer foundation upon which to discuss the results of variations in parameters and changes in scenarios when the underlying information requirements have been developed.

Validation is an appealing and potentially powerful method of raising a simulation's credibility. For all its attraction, however, it does not appear to have been used in the ADAGE, Carmonette, and COMO as a matter of course. Instead, validation efforts were initiated when questions were raised about credibility. We found some COMO weapon simulations in which validation contributed to credibility. Validation based on model-to-model comparisons, typified by the COMO Patriot simulation work and the Harpoon comparison, contributes importantly to a model's credibility and should be performed routinely, not merely in an ad hoc effort to respond to questions or criticism.

Supporting Material for Chapter 6

Support for Design, Data, and Operations

Institutional practices can help ensure that credible simulations are established and maintained. Two such practices that we found in reviewing the ADAGE, Carmonette, and COMO were configuration management and the use of oversight and review groups. (We have summarized support structures for the design, data, and operations of our case studies in table 6.2.)

Configuration Management

The 1982 TRADOC regulation entitled "Management: TRADOC Models" provides guidance on managing models with considerable attention to the control functions. It states that "only one agency designated by HQ [Headquarters] TRADOC will be responsible for the configuration management and development of software and data base maintenance of each model." That agency may provide a model to other TRADOC agencies, but the receiving agency's changes in the model are to be only for internal use and must be coordinated with the responsible agency. Changes made for internal use are not to leave the receiving agency and not to incorporate routines that change the nature of the model. All other changes are to be made only by the responsible agency.

This regulation further designates TRADOC service schools (like the Army Air Defense Artillery School) responsible for developing configuration control and improving, operating, and maintaining models that permit the evaluation of two-sided military engagement in which a single function of combat is considered in detail. The schools are to assist all other users in the development, improvement, and operation of their specific functional battlefield "modules" included in other combat arms and agency models at all levels. The ADAGE and COMO, for example, are models for air defense.

While the regulation indicates that the Army Air Defense Artillery School would be assigned management responsibilities, ADAGE is controlled by the Army Material Systems Analysis Activity (the original developer), and the COMO is assigned to the U.S. Army Missile Command. Not only does this differ from the current regulation but it also puts the functional operational-effectiveness models under the control of the agencies developing weapon systems, which gives the appearance of a possible conflict of interest. TRADOC Systems Analysis Activity is designated to develop and operate force-on-force combat development models at the battalion level, such as the Carmonette and others. It is responsible for the Carmonette but other versions of the Carmonette are used by other organizations such as the Concepts Analysis Agency.

For all TRADOC models, the regulation designates the deputy chief of staff for doctrine responsible for ensuring that doctrine, future concepts, and threat are properly portrayed. Other organizations are responsible for establishing and maintaining the actual data used in the simulations. Although the modelers obtained their input from these sources, the information was not always current, complete, or compatible with a model's data requirements. For example, no organization had an updated scenario that could be used for the ADAGE analyses of the DIVAD. Similarly, the probability-of-kill information had to be manipulated before it could be used in the Carmonette analyses of the DIVAD.

Of the three simulations we reviewed, only the COMO had an established interagency group that focused on configuration control. In 1980, a COMO models management board was established that developed a baseline COMO III software ensemble, produced a management plan, established working configuration control, supervised documentation, and monitored the development of new weapon models. The board's plans include improving command, control, and countermeasures and establishing a formal hierarchy of COMO models.

In August 1985, a group representing most COMO users convened at Kirtland Air Force Base to share information and develop strategies regarding the COMO's "standardization" and future use, especially given the newly developed transportable version. They discussed the need for a model resource group that would assist individual groups by jointly determining the need for improvements and who should be responsible for them. This group would also be responsible for preventing the uncontrolled proliferation of COMO operating systems. They made it clear that the COMO should be sufficiently uniform that outputs will be reliably similar, regardless of the computer on which a simulation is run.

The establishment of the COMO's model resources group in 1986 and model management board suggest that organizations do recognize the need for the management and coordination of major modeling efforts. With the use of the COMO extending to many Army, Navy, and Air Force units and to NATO and its allies, the need for such coordination is obvious. The extent to which multiple applications of the COMO will be effectively managed or coordinated by these oversight groups is still not clear, but that they recognize the need for and have attempted to coordinate a higher level of oversight or management are important initial steps.

Oversight and Review

TRADOC establishes study advisory groups to monitor the progress of its significant studies, and weapon-system program offices appoint system simulation working groups to oversee engineering simulations. The Army Air Defense Artillery School analysts conducting the initial DIVAD cost and operational-effectiveness analysis acknowledged in their 1977 report that the study advisory group provided an open forum for discussing disagreements and that overall the group was helpful and provided great assistance. The study advisory groups for the 1984 cost and operational-effectiveness update and the 1985 comparative analysis were especially active in directing the reconciliation of disparities in the ADAGE and Carmonette results. For example, when the Carmonette was first used to analyze the DIVAD for the update in late 1983, its initial results and those of the ADAGE led to different appraisals of the DIVAD on the battlefield. This situation, along with other identified or apparent errors, omissions, and anomalies in the models plus inconsistencies in the scenarios warranted a detailed review of both models before they were used further in analyzing the DIVAD.

To give insight into how a model affects perceptions of a system's performance and combat effectiveness, the deputy chief of staff for combat developments established a study advisory group to correct problems with the models, scenarios, and data. The following study objectives were established:

- Describe and evaluate the ADAGE and Carmonette models structures, scenarios, inputs, data usage, and outputs.
- Identify errors, omissions, and problems associated with the models, including coverage of data, scenarios, structure, and any other questionable factors or characteristics.
- Prioritize corrections by severity of problem, level of difficulty, time, and resources required for each correction or improvement or change to the models.
- Make changes where feasible within established deadlines and review them prior to production runs.
- Update the cost and operational-effectiveness analysis according to the run designs approved by the study advisory group and evaluate the results.

We believe that while study advisory groups provided a quality-control check for the simulations used in a specific study, their involvement was limited to short-term issues. The membership of the three study advisory groups for the DIVAD studies was not always the same, further limiting their ability to focus on long-term problems, such as defining validation requirements and working with operational test organizations to obtain the necessary data.

For the engineering simulations for several of the Stinger weapons, system simulation working groups were established to define validation requirements and to review and approve validation data. They were chartered by the Stinger program office and included representatives from Army laboratories, test and evaluation groups, users of simulations, and the contractor. They usually met three or four times a year, as necessary. They coordinated closely with the test integration working group, which had broader representation, to ensure that test data necessary for validation were obtained. Many of the persons representing the different organizations on the system simulation working groups have remained the same, giving some continuity to their oversight functions.

Documentation

For the three case study simulations, we found varying levels of documentation. The Carmonette had very little documentation. The ADAGE and COMO documentation were more complete. (We have summarized our review of the documentation in table 6.3.)

The TRADOC Systems Analysis Activity provided us with an executive summary describing key features of the Carmonette and list of input elements. These were helpful for a general understanding but incomplete for answering many of our questions. We also found examples of the problems created by the lack of documentation in the use of simulations in the DIVAD update. When that cost and operational-effectiveness analysis was being conducted, the necessity and utility of computer documentation, particularly for the Carmonette, became apparent when analysts at the Army Air Defense Artillery School attempted to understand the apparent disparity between the outcomes of the Carmonette and ADAGE. The school's analysts repeatedly expressed their concern that a reasonable understanding of the internal function of the Carmonette would be extremely unlikely in the absence of documentation. Although the TRADOC Systems Analysis Activity analysts were cooperative in providing detailed answers to specific questions, deficiencies were uncovered only through a serendipitous discovery.

Concern about this lack of documentation for the Carmonette was also expressed by various persons at higher Army levels, including the general appointed chairman of the study advisory group for the 1984 update. At the first meeting, the Army Material Systems Analysis Activity gave a detailed briefing on the ADAGE, describing formulas, and

TRADOC Systems Analysis Activity presented general information on the Carmonette, addressing broad capabilities. The chairman stated specifically in the minutes that the lack of documentation on the Carmonette was disappointing. Although this dissatisfaction was noted, we did not find that the TRADOC Systems Analysis Activity was directed to improve its documentation.

The ADAGE was better documented. However, the update analyses resulted in changes that would require commensurate changes in the documentation in order to keep it current, but the most current documentation provided to us is dated September 1978, prior to the update. The Army officials whom we interviewed did not mention any major problems with the adequacy of the ADAGE documentation. We were told that the two principal users, the Army Material Systems Analysis Activity and the Army Air Defense Artillery School, communicated frequently about the ADAGE's functioning and changes.

TRADOC analysts identified the disadvantages of documenting all existing models as time consuming or costly if done by contract with civilian firms. Documenting future models adds to the cost of a model's development and maintenance. Other analysts noted that a significant and probably unaffordable level of effort would be required to provide even minimum documentation for the current Carmonette simulation.

The COMO has been documented with more than 90 reports. Early documentation was produced in the late 1960's and early 1970's at the technical center of the Supreme Headquarters of the Allied Powers in Europe. Documentation has been produced since the late 1970's, primarily by or for the Army Missile Command in developing and validating individual weapon system models and improving the COMO program within which the simulations are run.

We reviewed the main documentation produced for the Stinger and found it comprehensive and detailed. It was essentially a combination of the analysts' and programmers' manuals. It assumed a substantial knowledge of the total COMO system. We did not obtain a briefing document on the Stinger simulation, which was described to us as a combination of a gross overview and user-analyst charts. It is not known to what extent the differences between the COMO III models that are operated at various facilities have been documented. Our discussion with the developer of the new transportable version led us to believe that such documentation is limited and informal. There is no listing of such material in an index of COMO documents. Validation documents produced for the

Patriot and Hawk simulations were comprehensive reports greatly useful for understanding the validation efforts and the strengths and limitations of the models, but no such documents were available for the Stinger.

Reporting

We reviewed the major reports containing the Carmonette and ADAGE results for the DIVAD cost and operational-effectiveness analyses. For the COMO, we reviewed the Stinger battery-coolant-unit usage study report and report on the validation of the Patriot model. We looked for information that explained a simulation's purpose, its theoretical basis, and its capabilities and limitations. (We have summarized our findings in table 6.4.)

The 1977 DIVAD Report

The purposes of the initial DIVAD cost and operational-effectiveness analysis reported in 1977 were to determine if there was a mission need for low-altitude air defense of mobile combat forces deployed near the forward edge of battle and to evaluate the merits of systems that might fill this need and to recommend one. The report addressed 11 study objectives ranging from gathering basic data describing enemy threat systems and defending forces through developing alternatives to the cost-effectiveness justification of any proposed new gun. Essential elements of the analysis were also clearly stated.

The 1977 report clearly established the theoretical importance of studying air defense in a division context and considered the support of combat forces at the forward edge of battle and the defense of critical assets in the rear. The report clearly stated its underlying assumptions. The report included reduction in damage to division assets as a measure of effectiveness, thus addressing an essential function of air defense. It explicitly stated that the Campaign is an expected-value submodel and described the model's logic and the relationship between the Incursion and Campaign submodels. This report also described how the air-to-air and ground battle results were integrated into the overall calculation of battle results. The report indicated that ground battle damage was generated from sources external to the ADAGE but did not describe those sources in detail. This is important, since the ADAGE's portrayal of the ground war was controversial. The portrayal of the ground battle was clearly labeled an external element to the ADAGE, but implications resulting from errors in the ground battle portrayal, such as the effect on measures of effectiveness, were not reported.

The 1984 Carmonette Report

The purpose of using the Carmonette in the cost and operational-effectiveness update reported in 1984 was to determine the operational effectiveness of the DIVAD gun in a European, combined arms, main battle area. The background statement for the report indicated that analysis using the Carmonette was to investigate the probability of all ground units participating in air defense, but this objective was not reported. Three objectives were addressed: (1) the determination of the operational effectiveness of the DIVAD, (2) the determination of the characteristics of the DIVAD necessary to achieve successful engagements, and (3) the determination of the potential contribution of other friendly ground forces to the air defense role. Essential elements of the analysis were not clearly delineated.

The Carmonette 1984 report did not describe the theoretical basis for the analysis very well. An understanding of the ground war, a major portion of the Carmonette, seems to have been taken as a given, since the ground war was hardly discussed. The main report neglected to mention the magnitude or implications of the notable differences—presented in an appendix—between the air defense burst-fire submodel used in the Carmonette and a similar model used by the Army Material Systems Analysis Activity. The report mentioned the modification of the Carmonette to reflect realistic play of the gun, but it described the DIVAD's modes of operation rather than the Carmonette's modeling of these features. Although it did discuss how the Carmonette classified and prioritized targets, there was only minor mention of firing doctrine.

Visual acquisition was discussed and the 3-kilometer limitation for visual detection and the substitution of forward-looking infrared sensors were mentioned. However, the report made a recommendation about equipping the DIVAD with a forward-looking infrared sensor when the DIVAD's performance without that capability had not been studied because of limitations in the Carmonette.

The report explicitly listed the following limitations:

- fixed-wing aircraft are not addressed,
- high- and medium-altitude air defense systems are not addressed, and
- DIVAD radar signature disclosing the location of friendly forces is not addressed.

An implicit limitation is computer time as reflected in the number of replications for determining the stability of the results. Clearly unstable results were accepted and not enough information was presented to evaluate their extent or effect.

The 1984 ADAGE Report

Since the Army Air Defense Artillery School has not yet issued a formal report delineating the results of the ADAGE modeling for the 1984 update, we reviewed a draft of the proposed report. The stated purpose of the update analysis was to

- determine if the DIVAD gun was still cost effective and operationally effective and
- analyze various force structures, including alternatives in order to recommend the preferred air defense artillery weapons.

The purpose of the simulation was to quantify changes in losses to the division from fixed-wing aircraft and helicopter attack caused by variations in the gun's performance in different operating modes with different performance capabilities based on observed performance and forecasted capability. The report updated the enemy air threat capability and posture. The ADAGE's primary purpose of studying air defense in the context of a combat division was stressed again, and the protection of forward combat units and the interdiction of fixed-wing aircraft and helicopters flying by the main battle area to the division's rear were also stressed.

The 1984 report described the model's logic less rigorously than the 1977 report. While the later report described the relationship between the Incursion and Campaign submodels, it did not describe in detail how air-to-air and ground battle results were integrated into the overall calculation of battle results. Changes to the ADAGE model made after the original report were described, including changes to correct problems that were uncovered and to improve the efficiency of the model's processing capability. To address criticisms, the report contained a section reconciling the ADAGE's results with results produced by the Carmonette and other TRADOC models of air defense, showing the ADAGE comparing favorably with the other models.

The 1984 update was not as explicit as the original report in describing the underlying assumptions of the basic analysis. However, it included analyses of alternative air defense artillery force structures, and their assumptions and limitations were clearly stated. A feature of the update not present in the first report was a description of nonquantifiable areas that were indicated as further supporting the need for the DIVAD.

The 1985 Carmonette Report

In October 1985, TRADOC Systems Analysis Activity published its report on the DIVAD comparative analysis, the purpose of which had been simply stated as to compare the DIVAD to selected alternatives by means of the Carmonette model. There were four objectives:

- obtain the most current information on the performance data for each air defense weapon-system alternative considered in the analysis;
- determine the operational effectiveness of each alternative;
- compare the operational effectiveness of the DIVAD and its near-term alternatives at two different levels of daytime visibility; and
- analyze the effects of a particular alternative as a replacement for the DIVAD in the air defense force.

The assumptions stated in the report clearly indicated that the Carmonette analysis was limited to the study of battalion air defense. The report also indicated that command, control, and communications and IFF were not to be played in the analysis.

The comparative analysis report gave only a cursory description of the theoretical foundations of the Carmonette model. Limitations from the previous Carmonette analysis were still not adequately addressed: high- and medium-altitude air defense systems were not mentioned, the problem of the DIVAD radar signature's disclosing friendly forces was not addressed, and the report coverage of enemy fixed-wing aircraft was misleading. The report implied that the Carmonette model included a submodel for handling attacking fixed-wing aircraft, whereas results for fixed-wing aircraft engagements with most air defense weapons were based on data from another computer model. (The Carmonette now has a fixed-wing submodel that it did not then have.) Nothing was mentioned about computer time, but the analysis was again based on few replications, and clearly unstable results were accepted for some important factors.

The following quotation from the comparative analysis report casts considerable doubt, in our opinion, on the adequacy of the Carmonette for studying air defense alternatives:

"The force effectiveness shows little discrimination between the air defense alternatives because of the small number of air defense units at the battalion level, the ineffective RED [that is, enemy] fixed wing capability played in this scenario, because a number of other ground systems made up some of the difference in kills against RED helicopters and the relatively ineffective RED helicopters that were attempting to acquire and engage BLUE [that is, friendly] units in a defensive posture." (Infantry System Division, Army TRADOC Systems Analysis Activity, Sgt. York Comparative Analysis (White Sands Missile Range, N.M.: October 1985), p. v. Underscoring added.)

The 1985 ADAGE Report

As with the 1984 update, no official reports of the ADAGE comparative analysis were issued and our comments here apply only to a November 1985 draft report. Its purpose and objectives were essentially the same as those for the Carmonette report. The objective that differed was to use the ADAGE to conduct a parametric analysis of the DIVAD to determine both its sensitivity to variations in performance parameters and how much these would have to be degraded to make the DIVAD no longer preferred among the alternative weapon systems being compared.

The report contained a simple description of the ADAGE model that was not nearly as complete as the description in the original analysis. Although the report clearly emphasized the importance of a division context for studying the effectiveness of an air defense weapon and assets preserved as a measure of effectiveness, the results tended to concentrate on the effectiveness of protecting forward combat units rather than the whole division. This resulted from a direction from the Army's deputy undersecretary for operations research to simulate test results from follow-on evaluation tests conducted at Fort Hunter Liggett, which played only a battalion task force along the forward edge of battle. Another limitation of the ADAGE mentioned in this report was its inability to portray mission aborts, a feature that was modeled in the Carmonette. Finally, the ADAGE report discussed a limitation that was also true for the Carmonette analyses— the inability to simulate human factors displayed at the Fort Hunter Liggett follow-on evaluation.

The ADAGE modelers tried to reconcile their division-level results with the Carmonette's battalion-level results, but the inconsistencies could not be completely reconciled. The report attributed the differences to two basic sources:

- differences in the models, especially the suppression of enemy helicopters and mission aborts present in the Carmonette but not in the ADAGE, and
- the ADAGE's portrayal of a 2-day division battle compared to the Carmonette's portrayal of a 25-minute battalion battle whose outcome was very dependent on the scenario.

The ADAGE report concluded by recommending that the DIVAD should be bought.

The COMO Reports

The purpose of the Stinger battery-coolant-unit usage simulation was to assist the Stinger project office in determining the appropriate number of battery coolant units that should be available to the Stinger team in wartime. The report clearly developed the rationale for the scenarios that were used, and it identified limitations stemming from both the computer and the model. The description of the model was detailed enough to allow an analyst to examine its operation fully. However, the DIVAD and threat aircraft models were not discussed in equivalent detail. Thus, the main limitation of the report was the implicit acceptance by the analyst, and necessarily the user of the results, that the DIVAD simulation was valid and that the overall results would not be biased by an inaccurate DIVAD model. In fact, the use of the unit appeared to be noticeably different when the DIVAD was part of the scenario, suggesting that the DIVAD had an effect on its usage. Fortunately, these differences were small in the aggregate and would not appear to affect the decision options available to the Stinger project office.

Our discussion of the Patriot validation was based on a document entitled Benchmarking the COMO III Baseline Patriot with Simpler COMO III Generic Models. The purpose of the document's analysis was clearly stated. Differences from varying theoretical approaches and practical implementation were clearly explained and guidance was provided as to whether differences could be reconciled by modifications and corrections or were essentially a result of the theory under which the particular model was developed. The document is an outstanding example of reporting on comparison and validation.

Summary

The third area of concern in the assessment framework we sketched in table 2.1 deals with the institutional practices used to establish and maintain the credibility of a simulation's results. We looked for the quality-control mechanisms used for the ADAGE and Carmonette in the DIVAD cost and operational-effectiveness analyses and for the COMO in the Stinger analyses. Our observations cannot be generalized beyond these three cases. We found several examples of efforts to improve a simulation's credibility as well as problems that arose when such efforts were not made.

The lack of documentation appears to be the greatest problem. The frustrating experience of trying to use the Carmonette without sufficient documentation for the DIVAD is one example.

Oversight groups for the Stinger appear to have been explicitly concerned with validity, defining validation requirements in advance, and working with developmental testers to ensure that test data would be available. The study advisory groups for the DIVAD also appeared to be concerned about credible results but focused more on assumptions, design, and inputs and on comparing the results with those of other simulations. They appeared neither to define validation requirements systematically in advance nor to work with the operational test groups to obtain data for validation.

Most of the reports that we reviewed explicitly described major omissions in the simulations. The Carmonette report did not, however, include some of the lesser limitations that may not be severe by themselves but cumulatively may seriously damage credibility.

Appendix V

Comments From the Department of Defense

Note. GAO comments supplementing those in the report text appear at the end of this appendix.

OFFICE OF THE SECRETARY OF DEFENSE

WASHINGTON, D.C. 20301-1700

OPERATIONAL TEST AND EVALUATION

6 OCTOBER 1987

Mr. Frank C. Conahan
Assistant Comptroller General
National Security and International
Affairs Division
U.S. General Accounting Office
Washington, D.C. 20548

Dear Mr. Conahan,

This is the Department of Defense (DoD) response to the General Accounting Office (GAO) Draft Report "DoD SIMULATIONS: Lack of Assessment Procedures Threatens Credibility of Results," dated July 2, 1987 (GAO Code 973195/OSD Case 7336).

See comment 1

The DoD has reviewed the draft report for technical adequacy and for application to current or future weapon system acquisitions. The report is generally technically correct. However, because this is a large and complex subject, the scope and focus of the report needs to be better defined, otherwise the report will be misleading with respect to the use of simulations in the acquisition process. Technical corrections and additions are noted in the enclosure.

With minor modifications the DoD concurs with the recommendations and is continuing efforts in this important area.

See page 15

In the findings, the GAO makes the following statement: "...one limitation of this approach (case study method) is that it prevents generalizing from the findings...." The DoD review indicates this was done, however. DoD comments are that the findings and recommendations may be valid with respect to most high level force on force models. The models that are used for simulations below the "war gaming" level need to be considered. Without these additional models/simulations being considered, unexpected and, in many cases, incorrect predictions will result.

The DoD finds the examination of three models and two weapon systems is neither a large nor broad enough sample size for the extrapolation suggested. Final acquisition decisions are based on many factors. DoD Directive 5000.3 indicates the use of simulation data is a subset of the test portion of the decision process. The part the three simulations played in the acquisition decisions of the two programs evaluated by the GAO was not addressed in the report. The 14 factors for evaluating simulations are a useful tabulation, but must include the caveat to take into account the final use (from an acquisition standpoint) of the simulation results. The GAO factor checklist is inappropriate to evaluate a simulation model's credibility separate from its people, application and input data. It is disappointing that the report does not have a consistent scope and focus. Models are only tools. A consistent and useful focus of the report would be the quality of workmanship which depends on the combination of qualified people, model application, input data choice and the assessment of the results.

The report recommends that the DoD provide formal guidance for the use of models/simulations in the acquisition process. The Military Operations Research Society (MORS) has devoted a great deal of attention to this area. The monograph on Military Modeling, for example, gives useful guidance on the subject of verification and validation. The state of the art is still evolving and is still subject to discussion and contention. These aspects, therefore, may not be ready for treatment by formal regulation. The 14 GAO factors have been forwarded to the Services for review and evaluation.

The Defense Systems Management College (DSMC) includes this area in the Test and Evaluation Management Course. The manual for this course includes a chapter on the use of simulations in test and evaluation. This course will be offered starting this December. The OT&E Commanders' Conference also included a review of this subject.

The detailed DoD comments on the report findings and recommendations are provided in the enclosure.

JOHN E. KRINGS
Director

Enclosure

GAO DRAFT REPORT - DATED JULY 2, 1987

"DOD SIMULATIONS: LACK OF ASSESSMENT PROCEDURES THREATENS CREDIBILITY OF RESULTS"

(GAO CODE 973195) OSC CASE 7336)

DOD RESPONSE TO GAO DRAFT REPORT

* * * * * * * * *

FINDINGS

o <u>FINDING A: Simulations</u>. The GAO noted that multi-billion dollar acquisition decisions for major weapon systems should, ideally, be based on testing the operational performance of weapons under conditions that replicate actual combat; however, as weapon systems have become more complex and expensive, the practicality of subjecting them to the necessary number of such tests has diminished. The GAO observed that one alternative has been to turn to the use of simulation models to provide additional information regarding performance. The GAO noted that simulations can be, and often are, used throughout the life cycle of a weapon system and are frequently used in connection with other analytical methods and field experimentation. The GAO found that the overriding advantage of simulation is perhaps the opportunity to investigate questions and problems that would otherwise not be addressed and to investigate them systematically with numerous replications under controlled conditions. On the other hand, the GAO noted that simulation has disadvantages. The GAO explained that, because a model is only an approximation, not the equivalent, of a real system, inaccurate assumptions about a weapon or its environment may cause the results of a simulation to diverge from reality. The GAO concluded that although simulations are useful tools, they are always approximations to reality and, therefore their credibility--the level of confidence that a decision-maker should have in their results--is open to question. (pp. 1-2 Executive Summary; 1-1 to-1-6/GAO Draft Report)

Now pages 1 and 10-12

<u>DOD RESPONSE</u>: Concur.
<u>It should be noted, however, that acquisition decisions are based on many factors and simulation results are considered to be a subset of developmental and operational tests. The Army used the three simulation models (used to evaluate the GAO factors) because of the battlefield complexity as well as the complexity of DIVAD and STINGER. Many levels of simulations are used by DOD in the acquisition of weapon systems. These simulations range from subsystem simulations to complex weapon systems on to wargame simulations to determine weapons systems requirements.</u>

o <u>FINDING B: Factors In Assessing A Simulation's Credibility</u>. The GAO observed that various procedures have been proposed to permit reasoned judgment concerning the credibility of simulation results. Based on a review of the literature and consultations with the simulation experts, the GAO developed a framework of 14 factors that the GAO concluded are necessary to address in an attempt to assess the credibility of a simulation. The GAO reported that these factors fall under three broad areas of concern regarding simulations--(1) theory, model design, and input data, (2) model validation, and (3) model management and reporting. The GAO observed that severe limitations in any of these three areas would lead to doubts about the credibility of a simulation, but for different reasons. As examples, the GAO cited:

-- problems with the first area of theory, model design or input data would pose questions about the basic integrity of the simulation's internal structure;

-- little or no evidence in the second area of model validation would leave a user with insufficient proof of the extent to which the simulation represents reality;

-- the absence of efforts in the third area would create doubts that good practices had been used to assure quality in the first two areas, that the continuing integrity of the model is assured, and that critical limitations had been properly disclosed.

Now pages 17-22 and 60

The GAO concluded that the overall framework is feasible to apply and will, at least for operational effectiveness simulations, provide a structured, useful way to review the credibility of simulation results. (pp. 2-1 to 2-11, pp. 8-1 to 8-2/GAO Draft Report)

See comment 2

DOD RESPONSE: Partially concur.
From a textbook point of view the three areas of assessment and the 14 factors certainly should be considered in the development and use of simulations. As a practical measure, however, large scale simulations of complex systems or large force simulations do not easily lend themselves to the total level of validation suggested. Consistency of results also is not always an indicator of good results. The DoD considers these shortcomings in the use of simulation data in its acquisition decisions. In addition, people, simulation application, and type of input data also need to be considered.

o Finding C: The Case Study Simulations. The GAO reported that in order to examine general purpose models, consideration was restricted to models that had the capability of simulating several types of weapons. The GAO reported that it judgmentally selected two Army antiaircraft defense systems: the portable, shoulder-fired, infrared surface-to-air STINGER missile and the division air defense gun (DIVAD), a surface-to-air radar-guided gun on a tracked vehicle. For these two weapon systems, the GAO selected three simulations; the COMO III model for the STINGER and the Carmonette and Air Defense Air-to-Ground Engagement (ADAGE) models for the DIVAD. The GAO reported key features of the simulation models, as follows:

-- the ADAGE model is a functional simulation used to study the relative effectiveness of combinations of air defense weapons in a division;

1

-- the Carmonette is an event-sequenced, combined-arms combat model that simulates small-unit, ground combat involving the actions of individual soldiers and weapons; and

-- the COMO III, used primarily for studies of tactical air defense, is a Monte Carlo simulation in which particular submodels are aggregated to simulate a specific air defense environment.

The GAO concluded that the case study method was the most plausible for illustrating application of the framework. The GAO further concluded, however, that one limitation of this approach is that it prevents generalizing from the findings beyond the three cases. (pp. 1-9, pp, 3-1 to 3-10/GAO Draft Report)

Now pages 14 and 23-28.

DOD RESPONSE: Concur.

o FINDING D: The Credibility Of Selected Simulations Based Upon Theory, Model Design and Input Data. The GAO noted that since the credibility of simulation results is relative to the intended use, the results from a model that is misapplied will not be credible, even though the model itself is sound. The GAO found that, in almost all instances, the case study simulations had some limitations. With regard to matters of the weapon-target engagement after detection had occurred, the GAO found that the evidence indicated that all three simulations had considerable capability. The GAO noted that this also appears to be the case for all three models in simulating important aspects of measures of effectiveness. The GAO concluded that for some of the other areas, the effort required to remove them might be relatively minor, while for others, much more work could be required. The GAO further concluded that, in a few cases, an effort to fix the model may not really be the appropriate response and instead, using a different model might be more appropriate. (p. 4-1, pp. 4-11 to 4-12/ GAO Draft Report)

Now pages 29 and 39

DOD RESPONSE: Concur.
The GAO reports elsewhere that the three simulations were useful. They were, however, only a small set of the tools used to provide inputs to the decision process. It should be recognized that the Army does have a mechanism to update the simulations as more information becomes available, and does endeavor to use the proper simulation for each task, if such a simulation exists.

o FINDING E: The Match Between The Theoretical Approach And The Questions Posed. The GAO found that the ADAGE, designed as a functional air defense model, was, in general, a reasonable choice for estimating the effectiveness of the DIVAD. In contrast to ADAGE, the GAO noted that Carmonette was designed to answer broad tradeoff questions which go beyond the issues of an air arms model. The Carmonette is generally not well-suited to answering the kinds of questions that were posed about the DIVAD. Specifically, the GAO found that the model attempts to portray an overall ground battle with limited air war features, but it is not focused on individual weapon systems. The GAO further found that, in general, the COMO III model was properly matched to the questions asked and was based on a scenario that represented official policy at the U.S. Army Air Defense Artillery School (USAADASHC). The GAO concluded that, while Carmonette has sound theory for a combined armed analysis, its approach is not the most appropriate for decisions regarding competing air defense weapons. The GAO further concluded that ADAGE and COMO were designed with such decisions in mind. (pp. 4-2 to 4-3; p. 3, Appendix II/GAO Draft Report)

Now pages 29-30 and 68-69

DOD RESPONSE: Concur.
It must noted, however, that functional models do not lend themselves to use in combined forces examinations; therefore, any comparison is not appropriate.

o FINDING F: Coverage Of Operational Measures Of Effectiveness. The GAO found that both the ADAGE and the CARMONETTE simulations provide for the coverage of protection of critical assets to some degree, although the former emphasized protection of assets, whereas the latter emphasized measures of aircraft attrition. The GAO further found that, although the COMO III simulation concentrated on both attrition-type measures and logistics measures, it was more limited in its ability to use preservation of assets deployed in the forward area as a principal measure of effectiveness because the ground war is not simulated. The GAO concluded that this condition threatens the credibility of the results of this simulation. While all the models address operational measures of effectiveness, the GAO concluded that ADAGE appears to relate its measures more closely to the ultimate missions of air defense--protection of assets--while the other models stress loss-exchange ratios. (pp. 4-3 to 4-4/GAO Draft Report)

Now pages 30-31

DOD RESPONSE: Concur.
While this is a correct assessment of the proper utilization of the three simulation models, the distinction, is the subjective weighting of results. The models are otherwise equivalent.

o FINDING G: Portrayal Of The Weapon System's Immediate Environment. The GAO observed that, with some limitations, the ADAGE and the COMO III models were capable of simulating the weapon systems' immediate environment across the five attributes of battle (i.e., level of battle, length of battle, targets, deployment/movement, and terrain). The GAO found that both were strong in characterizing the desired size of the battle and the full range of targets. Specifically, the GAO reported that the ADAGE model simulated longer battles but was limited by its uniform and static deployment of weapons. On the other hand, the GAO reported that COMO III portrayed a shorter battle with STINGER weapons that, while realistically deployed, did not move, a limitation for man-portable systems for which movement provides a form of defense, but at a cost of decreased operability. The COMO III and ADAGE

3

models used different approaches to portray terrain; however, the GAO noted that neither approach is obviously superior. The GAO found that Carmonette was more limited in its ability to portray the immediate environment than ADAGE or COMO III. As an example, the GAO cited that the small battalion size and short length of the Carmonette battle were inappropriate for the DIVAD weapon and the lack of fixed wing aircraft targets for most of the analyses contributed to an appropriate target set. While these limitations were partially offset by Carmonette's realistic portrayal of deployment, movement and terrain, the GAO concluded that these limitations of Carmonette's portrayal of the immediate environment threaten its credibility. The GAO concluded that all of the models are restricted and incomplete in some respect in their coverage of the surrounding combat arena. (pp. 4-4 to 4-6; p. 4, Appendix II/GAO Draft Report)

Now pages 31-33 and deleted

DOD RESPONSE: Concur.
As indicated by GAO (See Finding A), simulations are not exact replicas of systems or battles. The limitations are acknowledged and taken into account, as necessary. Two questions were of interest, however: Does the system do what it is supposed to do? Does it also provide a positive value in the battlefield when used with other systems?

o FINDING H: Broad-Scale Battle Environment. The GAO observed that any model of modern warfare should address the critical aspects of that warfare--in the air defense tactical areas this includes three dimensions--the air war, the ground war, and the interaction of the two. The GAO found differing approaches among the models in the coverage given to various aspects of modern warfare and how they interact with one another. The GAO further found that all of the models have serious weaknesses in the portrayal of at least one critical aspect of the air defense combat areas.

-- the ADAGE and COMO give inadequate consideration to the effect of ground war activities on air defense weapons and they do not completely portray the interaction of air and ground activities; and

-- the Carmonette's treatment of the air war is incomplete since it consistently failed to include fixed wing aircraft effects and only recently addressed these aircraft even in an indirect manner.

The GAO concluded that ADAGE and COMO strengths lie in the portrayal of ground activities. The GAO further concluded that all of the models have certain strengths in dealing with critical aspects of air defense weapons, but all of them also have serious weaknesses. (pp. 4-7 to 7-8; pp. 33-42, Appendix II/GAO Draft Report)

Now pages 35-36 and 83-87

DOD RESPONSE: Concur.
The strengths and weaknesses of the ADAGE, COMO III and Carmonette models are noted for the weapon systems reviewed. The seriousness of the weaknesses is related to the use of the results, which is appropriate. (See DoD response to Finding A).

4

o FINDING J: Mathematical And Logical Representatives Used To Depict Combat. The GAO observed that another critical area of concern in modeling weapon system operational effectiveness is how the theory and phenomena are mathematically and logically represented. The GAO found three areas of concern about ADAGE:

- its expected value approach used for modeling the engagement of a multiple air defense weapon against multi-plane attacks;
- its use of probability of participation of air defense weapons; and
- its apparent exaggeration of DIVAD survivability.

The GAO further found that Carmonette's basic mathematical formulations of fixed wing aircraft engagements are not much different from ADAGE; moreover, both of these models have other mathematical/logical problems, which though not as serious, nevertheless threaten the credibility of model results. The GAO reported that only COMO III appears to be free of serious mathematical/logical problems in the model structure. The GAO concluded that the three models, the ADAGE had the greatest number of mathematical and logical flaws, raising concerns about the credibility of its results. (pp. 4-3 to 4-9; pp. 42-49, Appendix II/GAO Draft Report)

Now pages 36-37 and 87-91

See comment 3

DOD RESPONSE: Partially concur.
The expected value approach is not intrinsically bad. Also, simulation models must, of necessity, limit replication to match run-time constraints, costs and expected input data.

o FINDING K: Appropriateness Of Input Factors. The GAO observed that, since the whole simulation can falter when input data are not clearly relevant, complete information about the data is necessary. The GAO found that all of the models use recognized data sources. The GAO observed that the Carmonette and COMO III models are relatively strong with respect to the appropriateness of data. In addition, the GAO observed that, in the earlier analyses, ADAGE and Carmonette modelers differed in the selection of input data and detection models for visual detection of approaching aircraft; however, the compromise position reached resulted in the use of data that did not properly describe DIVAD detection capabilities. The GAO further found that, with regard to data reflecting the real world, the ADAGE simulation had the most serious limitations because some of its data were outdated and some key values were too large to be accepted by knowledgable military officials. In addition, the GAO found that ADAGE modelers included weapon system characteristics as an integral part of the model rather than addressing them through an external data base, which is contrary to the Joint Forward Area Defense (JFAAD) Test Force suggested model requirements. The GAO noted that, while some of the Carmonette's early data problems were

5

corrected, the problems with disputed visual detection data remained. The GAO also found that the Carmonette and COMO III simulations were limited in data handling because of the extensive data tailoring required. The GAO concluded that ADAGE and Carmonette both experienced problems with obtaining appropriate data and these problems were related to the credibility of simulation results. The GAO further concluded that because Carmonette and COMO III require extensive data tailoring, the effects of data tailoring cannot be easily distinguished from model manipulations. (pp. 4-9 to 4-11, pp. 49-58, Appendix II/GAO Draft Report)

Now pages 37-38 and 91-95

DOD RESPONSE: Concur.
Data tailoring, however, is a useful tool in examining model or system sensitivity.

o FINDING L: The Credibility Of Selected Simulations Based Upon The Correspondence Between A Model and The Real World. The GAO found that the factors that address the credibility of the simulation (based upon correspondence between the model and the real world) are--(1) evidence of a verification effort, (2) evidence that the results are statistically representative, (3) evidence of sensitivity testing, and (4) evidence of model validation. The GAO observed that while analysts can never provide absolute guarantees regarding model credibility or output accuracy, several steps can be undertaken to determine whether a simulation is sufficiently close to representing the operation of an actual weapon system. These steps include analyses to produce evidence that:

-- the computer program operates in the manner intended by the simulation model's designers;

-- the output of the simulation is sufficiently representative of what the average model output would be over many runs;

-- the sensitive model parameters and alternative scenarios are properly accounted for by the simulation results; and

-- the simulation results are a sufficiently accurate representation of what the real world results would be.

In general, the GAO concluded that efforts to directly validate simulation results by comparison to weapon effectiveness results derived by other means were very weak, requiring substantial work to increase credibility. The GAO further concluded that more enhancement of credibility could have been achieved by more intensive efforts to document the verification of the computer representatives of the real world, and to establish that the simulation results were statistically representative, or to directly validate simulation results by comparison to weapon effectiveness results derived by other means. In addition, the GAO further concluded that the strongest contribution to credibility probably came from efforts to test the parameters of models and to run the models with alternative scenarios. (pp. 5-1 to 5-2; p. 5-11/GAO Draft Report)

Now pages 40 and 46

6

DOD RESPONSE: Concur.
Simulations can also be used to examine trends. It should be recognized that real world data is not always available. As an example, treaties may preclude testing. In some instances these trends may be more revealing than real world data in assessing the effectiveness of a weapon system. In addition, there are cases where the model user deviates from the "real world" for good and sufficient reasons. The credibility of results to the new user can be biased by good documentation that provides strengths and limitations, as well as technical instructions. DOD-STD-2167 provides for documentation to be included in software design and is as applicable to simulation models as to weapons system design. Strengths and weaknesses that address non-valid uses would be more appropriate, when applied to large scale simulations.

o FINDING M: Evidence Of A Verification Effort. The GAO observed that verification refers to the process of determining if the computer program correctly represents the theory, model design, and input data. The GAO found that no documentary evidence of verification was available for either ADAGE or Carmonette. While there was no evidence of verification efforts on Carmonette, the GAO noted that it was informed that Carmonette had been subjected to extensive peer reviews. The GAO further found that it could not identify any verification of the COMO III STINGER simulation models or the variant that was developed for the battery coolant unit (BCU) analysis. The GAO concluded simulation users are entitled to some knowledge of the verification efforts that were involved in simulation development and that such information will strengthen the credibility of simulations. The GAO further concluded that lack of documented evidence of verifications presents a clear threat to the credibility of the simulations. (pp. 5-2 to 5-3; pp. 30-31, Appendix III GAO Draft Report)

Now pages 40-41 and 113-14

DOD RESPONSE: Concur.
The more a model is used, the more it is subject to peer review. The real problem is documentation, which is a limitation throughout the modeling world. On simulations used for the first few times, this finding may be an accurate statement, but depends on the expertise of the user. The statement becomes less and less accurate as validated data is provided. Long term credibility lies with consistent data, validated with actual measured data. DoDD 5000.3 requires a verification and validation effort.

o FINDING N: Evidence That The Results are Statistically Representative. The GAO observed that the credibility of simulation results is enhanced when users are assured that simulation outputs are representative of how the model will perform during repeat runs. The GAO observed that, within ADAGE, the incursion submodel is the only model using the Monte Carlo modeling technique. The GAO noted that, according to analysts who worked with ADAGE, each incursion scenario had been run five hundred times and the resultant mean was within one or two percent of the true mean at the 98 percent confidence level. The GAO further found that, for Carmonette, analysts addressed the statistical representatives

7

factor, but with only limited success. As an example, the GAO cited that the COMO III STINGER simulation made only one run for each scenario, and the report did not address the statistical representativeness of the results. The GAO also found that the COMO III/STINGER simulation appeared not to have addressed the need for developing statistically representative model values, which the GAO concluded represents an outright threat to the credibility of the simulation. The longer-running, more complex simulations were evaluated with fewer simulation runs; therefore, the GAO concluded that if this represents a tendency to treat the results of one or a few runs of a complex model as being the "True" estimates for the simulation model, this situation has the potential to create substantial credibility questions. (pp. 5-3 to 5-5a, p. 5-10; p. 31, Appendix III GAO Draft Report)

Now pages 41-42 46 and 113

DOD RESPONSE: Concur.
The statistical validity should always be addressed. Review groups are in place to accomplish this.

o FINDING O: Evidence Of Sensitivity Testing. The GAO observed that this factor addresses the need for simulation analysts and users to develop an understanding of how changes in key parameters affect the model results. The GAO observed that sensitivity testing identifies how changes in model parameters affect results in both direction and magnitude and provides the analysts expectations of model behavior. The GAO found that ADAGE analysts used both parameter testing and experimentation with alternative scenarios to examine simulation results for both small and major changes. The GAO further found that the credibility of both Carmonette and COMO III also benefited from the use of alternative scenarios and parameter testing. The GAO found that sensitivity testing was a factor, for all three simulations, which contributed to a strengthening of the credibility of the models. The GAO observed that the apparent need is to integrate parameter and scenario work with the work performed on determining the true estimates for the simulation. The GAO concluded that valuable information that contributed to simulation result credibility was developed in ADAGE, Carmonette and COMO III by varying parameters and testing alternative scenarios. (pp. 10-16, p. 32, Appendix III GAO Draft Report)

Now pages 103-06 and 114

DOD RESPONSE: Concur.

o FINDING P: Evidence Of Model Validation. The GAO observed that validation is the process of determining that a model is an accurate representation of, or is in agreement with, the real world system being modeled. The GAO observed that validations are not planned for or conducted routinely but are more likely to be performed when a disparity in results is found among different performance estimating methods. The GAO found no validation efforts had been performed on ADAGE or Carmonette using real world DIVAD data. (The GAO noted that this is not intended to suggest that there was no

8

attempt at validation.) The GAO further found no evidence of validation specifically for the COMO III/STINGER simulation; however, evidence was found of an effort to validate the general COMO model by comparing COMO results to those from an Air Force model called SORTIE. The GAO observed that the results suggest that model-to-model validation can marginally strengthen the credibility of a model, especially when comparisons with real world data are lacking. The GAO generally concluded that efforts to directly validate simulations results by comparison to weapon effectiveness results derived by other means are very weak and require substantial work to increase credibility. The GAO further concluded that validation based on a model-to-model comparison contributes substantially to model credibility and should be performed as a normal part of the simulation cycle. (pp. 5-8 to 5-11; p. 32, Appendix III/GAO Draft Report)

Now pages 44-46 and 114

DOD RESPONSE: Concur.
Validation indicates end use. It should be noted, however that prior to the validation, trends are usually more appropriate.

o FINDING Q: The Support Structures Established To Manage The Simulation Design, Data, And Operating Requirements The GAO observed that institutional practices or mechanisms can help to ensure that credible simulations are established and maintained. The GAO found two such practices in reviewing ADAGE, Carmonette, and COMO III were configuration management and the use of oversight and review groups. The GAO reported that each of the models had been assigned to an agency for management; ADAGE to the U.S. Army Materiel System Analysis Activity (AMSAA); Carmonette to the U.S. Army Training and Doctrine Command (TRADOC) Systems Analysis Activity; and COMO III to the U.S. Army Missile Command. The GAO noted that the TRADOC, which plays a role in both managing and using simulation models, illustrates configuration management support. The GAO further found that, in an effort to establish oversight and review at different levels, the TRADOC established Study Advisory Groups (SAGs) to monitor the progress of individual studies using models under TRADOC control. The GAO concluded that the Army has been at least partially successful in establishing mechanisms to maintain simulation models and to control their development and use. While formal control responsibilities were assigned for each case study model, the existence of several stakeholder groups with various roles to play may indicate an immature and still evolving structure; however, the GAO further concluded that the present mix may be appropriate as a permanent structure, which recognizes the diffuse Army interests in simulation. (pp. 6-1 to 6-4/GAO Draft Report)

Now pages 47-49

DOD RESPONSE: Concur.
The management procedures evidenced and reported by the GAO are appropriate to the continuing credibility of these simulation models and their modifications. However, as indicated earlier, these procedures will not ensure valid data.

9

o FINDING R: The Documentation Needed By Persons Using The Simulation Or Its Results. The GAO observed that well-documented simulation models inspire confidence that the models will be correctly used to address the types of issues for which they were designed. The GAO found that the ADAGE was a relatively well-documented model, at least through September 1978. The GAO further found that Carmonette is a relatively poorly documented model, which became evident during the Cost and Operational Effectiveness Analysis (COEA) Update Study, when analysts at the USAADASCH tried to reconcile disparities in the results produced by ADAGE and Carmonette. The GAO reported that concern about the lack of Carmonette documentation was also expressed by the Chairman of the Study Advisory Group charged with overseeing the COEA Update for DIVAD. The GAO found that extensive documentation exists for the COMO series of models. The GAO concluded that in the case of COMO III, and to a lesser extent for ADAGE, the documentation tends to strengthen the confidence of the user in the credibility of the simulation. On the other hand, the GAO concluded that the considerable lack of documentation for Carmonette detracts from the confidence that a prospective user might have in its credibility. (pp. 6-4 to p. 6-6, 6-10/GAO Draft Report)

Now pages 49-51 and 53

DOD RESPONSE: Concur.
(See DoD response to Recommendation 2).

o FINDING S Disclosure Of The Simulation Strengths and Weaknesses. The GAO examined several reports: (1) for ADAGE--the 1977 DIVAD COEA Report, the 1984 COEA Update and the 1985 DIVAD Comparative Analysis; (2) for Carmonette--the 1984 COEA Update; and the 1985 Comparative analysis; and (3) for COMO III--the STINGER Battery Coolant Unit Study Report, a validation report for PATRIOT missiles studies and documentation for the STINGER model. The GAO found that:

-- the ADAGE reports contained explicit statements of the study objectives and the strengths and limitations of the particular simulation;

10

-- the Carmonette 1984 COEA Update appears to make recommendations that are not well-supported by simulation results. Little or no attention is given to the theoretical basis of the analyses; and

-- the report on STINGER Battery Coolant Unit Study clearly developed the rational for the scenarios and identified limitations due to both the computer and the model. One limitation of the report was the implicit assumption that the submodel for another air defense weapon being simulated within COMO III was sufficiently credible and accurate and that the overall results would not be biased.

Now pages 49-53

The GAO concluded that, while reporting practices could be improved, they contributed to the credibility of all three simulations. (pp. 6-4 to 6-11/GAO Draft Report)

DOD RESPONSE: Concur.
Reporting practices contribute to the credibility of the results. Again, it should be noted that strengths and weaknesses need to be relative to the use of the simulation data.

o FINDING T: OSD LEVEL GUIDANCE. The GAO found no formal guidance specifically for simulation at the OSD level; however, related OSD-level regulations that included concepts which could be applied to computer simulations did exist. Specifically, the GAO reported that:

-- the need for information and the use of analysis to support weapon system acquisition decisions is stated in that some form of system effectiveness analysis, in conjunction with analyses of costs and other factors, shall be performed to support milestone decisions; and

-- the test and evaluation directive, DoDD 5000.3, states that the use of properly validated analysis, modeling, and simulation is strongly encouraged, especially during early development phases.

Now pages 54-55

The GAO concluded, however, that while the above directives encourage the use of validated simulations, they do not give guidance on prerequisites for sound simulations, on how to development them, nor how to validate them. The GAO noted that the Automatic Data Processing (ADP) and Information Resources Management (IRM) regulations may be applicable, at least in part, because simulations are run on computers; however, the GAO found that these directives focus largely on input/output processing and file structure. The GAO, therefore, further concluded that while directives or standards in this area are useful, they are inadequate to guide the development and maintenance of computer simulations. (pp. 7-2 to 7-3/GAO Draft Report)

11

DOD RESPONSE: Concur.
The DOD provided the 14 factors to the Services for review and evaluation on August 4, 1987. The Defense Systems Management College included the use of simulation results in the new course on test and evaluation management beginning on December 14, 1987. The OT&E Commanders' Conference held August 1987 reviewed this same subject. (See DoD response to Recommendation 1.)

o FINDING U: The Software Quality Issue. The GAO observed that, while the interest in software quality began with weapon system software applications, it may be generalized to all computer systems. The GAO found, however, no substantial interest in, or recognition of, the importance of the issue of software quality. The GAO recognized that some arguments can be made against designing, programming and testing software to satisfy the established quality standards for some simulations that are small, short-lived, limited purpose applications. On the other hand, the GAO pointed out there is a class of simulations that are long-lived, that develop a community of users, and are intensive consumers of computer resources. The GAO observed, therefore, that over their lifetime these simulation systems will be intensive users of manpower and computer resources and, additionally, their results may influence major decisions. The GAO concluded that more attention by management to the technical aspects of modeling, such as software quality, statistical analysis, and validation, would encourage the greater adoption of practices to assess and improve simulation credibility. (pp. 7-3 to 7-5, p. 7-11/GAO Draft Report)

Now pages 55-56 and 59

DOD RESPONSE: Concur.
The standard is scheduled to be issued by March 1988. The DOD Standard 2168 addresses this issue and is in final review.

o FINDING V: Service Level (Army) Regulations And Practices. The GAO found that at the Service level, the Army has issued regulations that address (1) the management of models within the context of the Army Model Improvement Program and (2) the management of studies and analyses, of which models are a component. The GAO reported that a major Army effort to develop a comprehensive hierarchical modeling system, reflecting the guidance of the Army Models Committee, was formalized in the issuance of the Army Regulation 5-11 (AR5-11), Army Model Improvement Program. (The GAO noted that the purpose of the program is to evaluate combat capabilities and determine resource requirements through an integrated system of models operating at the theater force, corps/division, and combined arms and support task force levels.) The GAO found AR5-11 to be the most detailed statement issued by the Army regarding modeling policy and practice among the documents reviewed. The GAO further reported that the TRADOC provides specific guidance on managing model, and on using and reporting on simulations as part of studies. The GAO further reported that, in addition to issuing regulations as a means of guiding studies and models, the Army has established various groups to address technical and management aspects of the studies and modeling process. The GAO concluded that overall, the Army analytical community appears to be concerned about the quality of its models and its responsibility to provide guidance for model management. (pp. 7-5 to 7-10/GAO Draft Report)

Now pages 57-59

12

DOD RESPONSE: Concur.

RECOMMENDATIONS

o RECOMMENDATION 1: The GAO recommended that the Secretary of Defense adopt or develop and implement guidance on producing, validating, documenting, managing, maintaining, using, and reporting weapon system effectiveness simulations. In the GAO view, this guidance should include a mechanism to routinely provide reviews of a simulation's credibility and, in this way, identify problems that need to be resolved. The GAO also suggested that the OSD should explore including a requirement for a statement regarding validation efforts to accompany simulation results. (p. 9/GAO Draft Report)

Now page 5.

DOD RESPONSE: Concur.
The GAO has addressed an area of concern to the DOD. As indicated earlier, the 14 factors provided were forwarded to the Services on August 4, 1987. The inclusion of simulation modeling and simulation results in the OT&E Commanders' conference and in the new DSMC course are additional evidence of the importance the DOD is giving to this area. Specific guidance from these initiatives will be considered. The DOD will provide specific inputs on this recommendation within 6 months. (See DoD response to Finding T).

o RECOMMENDATION 2: The GAO recommended that the Agencies responsible for managing the ADAGE, Carmonette, and COMO models explore the feasibility of and, where indicated, take actions to correct, the limitations the GAO has identified--especially in the validation area. (p. 9/GAO Draft Report)

Now page 5.

DOD RESPONSE: Concur.
The Army management process for the use of these models is continuing to see that the models are used properly and are updated and corrected when and where necessary. This action will be reported with the inputs on Recommendation 1.

13

The following are GAO's comments on the October 6, 1987, Department of Defense letter.

GAO Comments

1. Our response to DOD's letter is presented in chapter 8. We have also included in the final report additional information about our objectives in chapter 1 to address DOD's concerns about the scope and focus of our draft report.

2. We have pointed out in the report that the validation of simulations is a difficult problem, and we have only suggested that more efforts in this area might be taken. Achieving a "total level of validation" is not likely ever to be possible, but we believe incremental improvements can be made. We agree that the persons involved, a simulation's applications, and the type of data input also should be considered in assessing credibility, and we believe these are considered within our framework: persons under factor 12, simulation application under factor 1, and input data under factor 7.

3. We did not mean to imply that the expected-value approach is intrinsically bad. However, we point out several limitations that resulted from the use of this approach in the ADAGE Campaign submodel. DOD personnel and experienced models practitioners also pointed out the concerns that we raised. Our criticisms in this area were tempered by other statements in the report pointing out strengths of the ADAGE model. For example, we noted that the ADAGE's theoretical approach was appropriate for addressing decisions concerning competing air defense weapons.

Bibliography

Balci, O. Requirements for Model Development Environments, technical report CS83022-R. Blacksburg, Va.: Virginia Polytechnic Institute and State University, 1985.

————, and R. E. Nance. Formulated Problem Verification as an Explicit Requirement of Model Credibility, report CS830021. Blacksburg, Va.: Virginia Polytechnic Institute and State University, 1985.

Balci, O., and R. G. Sargent. "A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation Models." Communications of the ACM, 24:4 (April 1981), 190-97.

————. "Validation of Multivariate Response Models Using Hotelling's Two-sample T^2 test." Simulation, December 1982, pp. 185-92.

————. "Validation of Simulation Models Via Simultaneous Confidence Intervals." American Journal of Mathematical and Management Sciences, 4:3 (1984), 4.

Banks, J., and J. S. Carson II. Discrete-Event System Simulation. Englewood Cliffs, N.J.: Prentice-Hall, 1984.

Banks, J., D. M. Gerstein, and S. P. Searles. "The Verification and Validation of Simulation Models." School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 1986.

Battilega, J. A., and J. K. Grange (eds.). The Military Applications of Modeling. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology Press, 1984.

BDM Corporation. Forward Area Air Defense Test Facility Analysis and Definition: Manned Simulator and Computer Model Survey and Evaluation, final technical report BDM/A-84-486-TR. Albuquerque, N.M.: 1984.

Berg, D. J., and J. L. Elias. Catalog of War Gaming and Military Simulation Models, 8th ed. Washington, D.C.: Studies, Analysis, and Gaming Agency, Joint Chiefs of Staff, The Pentagon, 1980.

Bonder, S. "Issues Facing Model Developers—I." In S. I. Gass (ed.). Utility and Use of Large-Scale Mathematical Models, National Bureau of Standards special publication 534. Washington D.C.: U.S. Government Printing Office, 1979.

Brantley, L. CARMONETTE/TRASANA Artillery Module Documentation, TRASANA TD-32-82. White Sands Missile Range, N.M.: U.S. Army TRADOC Systems Analysis Activity, 1982.

Bronowitz, R., and K. Morton. Evaluation of Harpoon Acquisition Models, CNR 71. Alexandria, Va.: Center for Naval Analyses, 1983. SECRET

Brooks, F. C. "The Stochastic Properties of Large Battle Models." Operations Research, 13:1 (January-February 1965), 1-13.

Burns, F. Dynamic ECM as Modeled in the MICOM COMO III Ensemble. Huntsville, Ala.: SRS Technologies, 1984.

Davis, E. A., and D. K. Pace. A Review of SANDEMS. Laurel, Md.: Applied Physics Laboratory, Johns Hopkins University, 1983. CONFIDENTIAL.

DeMillo, R. A., and R. J. Martin. OSD/DDT&E Software Test and Evaluation Project, vols. 1-6. Atlanta, Ga.: Georgia Institute of Technology, 1983.

Diaz, S. A., C. D. Spruyt, and H. R. Wilhelm. The COMO III Interceptor Operations Model, technical memorandum STC TM-635. The Hague, Netherlands: SHAPE Technical Center, 1981.

Eichblatt, E.J., Jr. Performance Simulation in Support of Test and Evaluation. Point Mugu, Calif.: Pacific Missile Test Center, 1983.

Etheridge, E. W., L. G. Bryan, and M. J. Sanders. "Air Defense Analysis Review." Draft CAORA/TR-11/85, Studies and Analysis Directorate, Combined Arms Operations Research Activity, Fort Leavenworth, Kansas, 1985. SECRET.

Federal Computer Performance Evaluation and Simulation Center. Computer Model Documentation Guide, National Bureau of Standards special publication 500-73. Washington, D.C.: U.S. Government Printing Office, 1981.

Gass, S. I. Computer Model Documentation: A Review and an Approach, National Bureau of Standards special publication 500-39. Washington, D.C.: U.S. Government Printing Office, 1979.

———— (ed.). Utility and Use of Large-Scale Mathematical Models, National Bureau of Standards special publication 534. Washington, D.C.: U.S. Government Printing Office, 1979.

———— (ed.). Validation and Assessment of Energy Models, National Bureau of Standards special publication 616. Washington, D.C.: U.S. Government Printing Office, 1981.

————. "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis." Operations Research, 31:4 (July-August 1983), 603-31.

General Research Corp. COMO System Upgrade Report: COMO-T and Hawk Lashe Model, TIO-4345. Huntsville, Ala.: 1986.

————. COMO-T: The Transportable Computer Model System Guide, TIO-4347. Huntsville, Ala.: 1986.

George, R. Validation of Hawk Simulation. Huntsville, Ala.: Spectra Research Systems, 1984.

Happel, W. J. M. COMO III Program Description, STC TM-554, vol. 2. The Hague, Netherlands: SHAPE Technical Center, 1978.

Harris, G. L. Computer Models, Laboratory Simulators, and Test Ranges: Meeting the Challenge of Estimating Tactical Force Effectiveness in the 1980's. Fort Leavenworth, Kan.: U.S. Army Command and General Staff College, 1979.

Hoeber, F. P. Military Applications of Modeling: Selected Case Studies. New York: Gordon and Breach Science Publishers, 1981.

Horrigan, T. J. A Generating Function Technique for the Construction of Submodels of Combat Damage, STAG 69-1. Chicago: Caywood-Schiller Associates, 1969.

————. "Weapon Performance, Mathematical Models, and Combat Effectiveness—A Faulty Synthesis." Draft, Horrigan Analytics, Chicago, 1983. CONFIDENTIAL.

Hughes, W. P., Jr. (ed.). Military Modeling. Alexandria, Va.: Military Operations Research Society, 1984.

IEEE Computer Society. "IEEE Standard P982 Software Reliability Measurement." Draft, Institute of Electrical and Electronics Engineers, New York, 1985.

Infante, D. R. Phase I Division Air Defense Gun Cost and Operational Effectiveness Analysis, vols. 1 and 2. Fort Bliss, Tex.: U.S. Army Air Defense School, 1977. SECRET.

Ismari, D. L., and C. V. Rolli. "Improvements to and Validation of the TAC Disrupter Simulation for Joint Test Support." Journal of Test and Evaluation, 5:1 (January 1984), 47-52.

Kheir, N. A., and W. M. Holmes. "On Validating Simulation Models of Missile Systems." Simulation, April 1978, pp. 117-28.

Killian, T. L. COMO Frame and COMO Runtape Assembly Program. Huntsville, Ala.: SRS Technologies, 1985.

Law, A. M. "Statistical Analysis of Simulation Output Data." Operations Research, 31:6 (November-December 1983), 983-1029.

————————, and W. Kelton. Simulation Modeling and Analysis. New York: McGraw-Hill, 1982.

Lufkin, B. M. The Air Defense Modern Gun Effectiveness Model, technical report 360. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1982.

Mann, G. A. The Role of Simulation in Operational Test and Evaluation. Maxwell Air Force Base, Ala.: Air University Press, 1983.

Meredith, J. The Air Defense Gun Effectiveness Model (GEM), technical report 337. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1981.

————————, and R. Scheder. Air Defense Gun Study, interim note A-80. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1975. CONFIDENTIAL.

Meredith, J., et al. Evaluation of the Gun Low-Altitude Air Defense Fire Control Test Bed, technical report 149. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1977.

Meredith, J., et al. Air Defense Air-to-Ground Engagement (ADAGE) Simulation, vol. 2, The Incursion Model, technical report 227. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1978.

Meredith, J., et al. Air Defense Air-to-Ground Engagement (ADAGE) Simulation, vol. 3, Incursion Model Classified Data, technical report 227. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1978. SECRET.

Metzger, J. J. Air Defense Air-to-Ground Engagement (ADAGE) Simulation, vol. 1, The Campaign Model, technical report 227. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1978.

————————. Air Defense Air-to-Ground Engagement (ADAGE) Simulation, vol. 4, Update, technical report 227. Aberdeen Proving Ground, Md.: U.S. Army Materiel Systems Analysis Activity, 1978.

Office of Missile Electronic Warfare. Stinger-POST Missile Flight Simulation. 1979. SECRET.

Oren, T. I. "Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference." Communications of the ACM, 24:4 (1981), 180-89.

Pharr, O., and F. Burns. Benchmarking the COMO III Baseline Patriot with Simpler COMO III Generic Models. Huntsville, Ala.: SRS Technologies, 1985.

Pullum, L. The STINGER Weapon Deck for COMO III (Revision I), SRS SE-TR83-105. Huntsville, Ala.: Spectra Research Systems, 1983.

Rudolph, R. R., and R. L. Stadter. "CARMONETTE/TRIAD Modifications." Draft, Ketron, Baltimore, Maryland, 1985. CONFIDENTIAL.

Sargent, R. G. An Assessment Procedure and a Set of Criteria for Use in the Evaluation of "Computerized Models and Computer-Based Modelling Tools," RADC-TR-80-409. Griffiss Air Force Base, N.Y.: Rome Air Development Center, Air Force Systems Command, 1981.

Schellenberger, R. E. "Criteria for Assessing Model Validity for Managerial Purposes." Decision Sciences, 5 (1974), 644-53.

Shannon, R. E. Systems Simulation: The Art and Science. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Shubik, M., and G. Brewer. Models, Simulations, and Games—A Survey, R-1060-ARPA/RC. Santa Monica, Calif.: Rand Corp., 1972.

Sims, D. M., and L. E. Foster. Stinger Battery Coolant Unit Usage Study, technical report D-83-1. Redstone Arsenal, Ala.: U.S. Army Missile Command, 1983. SECRET.

SRS Technologies. COMO III Executive Summary. Redstone Arsenal, Ala.: U.S. Army Missile Command, 1985.

Stockfisch, J. A. Plowshares Into Swords—Managing the American Defense Establishment. New York: Mason and Lipscomb Publishers, 1973.

————. Models, Data, and War: A Critique of the Study of Conventional Forces, R-1526-PR. Santa Monica, Calif.: Rand Corp., 1975.

Taras, K. J. COMO Model of Patriot Acquisition, C-78-1. Redstone Arsenal, Ala.: U.S. Army Missile Research and Development Command, 1977. CONFIDENTIAL.

————. Roland COMO III Simulation, C-77-3. Redstone Arsenal, Ala.: U.S. Army Missile Research and Development Command, 1977.

————. Patriot COMO III Simulation Revision, C-79-4. Redstone Arsenal, Ala.: U.S. Army Missile Command, 1979. CONFIDENTIAL.

————, and J. C. Richardson, Jr. Aggregated Shorads Attrition Weapon Deck, C-77-1. Redstone Arsenal, Ala.: U.S. Army Missile Research and Development Command, 1977.

Taylor, J. G. Lanchester Models of Warfare, vols. 1 and 2. Arlington, Va.: Military Applications Section, Operations Research Society of America, 1983.

U.S. Army. Management: Army Studies and Analyses, Army regulation 5-5. Washington, D.C.: 1981.

U.S. Army. 1984 Weapon Systems. Washington, D.C.: 1984.

————. Stinger Team Operations, field manual 44-18-1. Washington, D.C.: 1984.

U.S. Army, Air Defense Artillery School. "DIVAD Gun COEA Update." Briefing charts, Fort Bliss, Texas, 1984. SECRET.

————. "Sgt. York Alternative Analysis." Briefing charts, Fort Bliss, Texas, 1984. CONFIDENTIAL.

————. "Sgt. York Comparative Analysis Executive Summary." Draft, Fort Bliss, Texas, 1985. CONFIDENTIAL.

————. "Sgt. York Comparative Analysis Report." Draft, Fort Bliss, Texas, 1985. CONFIDENTIAL.

————. "Sgt. York COEA Update." Draft, Fort Bliss Texas, no date. SECRET.

U.S. Army, Air Defense School. Division SHORAD Study. Fort Monroe, Va.: U.S. Army Training and Doctrine Command, 1974. SECRET.

————. SHORAD/MANPAD Force Structure Study. Fort Monroe, Va.: U.S. Army Training and Doctrine Command, 1979. SECRET.

————. STINGER Cost and Operational Effectiveness Analysis, vols. 1-6. Fort Monroe, Va.: U.S. Army Training and Doctrine Command, 1977. SECRET.

U.S. Army, Combat Developments Command. Concepts and Doctrine for Air Defense of the Division Area. Fort Belvoir, Va.: 1971. SECRET.

U.S. Army, Materiel Systems Analysis Activity. Joint Munitions Effectiveness Manual: Performance and Effectiveness Estimates for the STINGER Air Defense System. Aberdeen Proving Ground, Md.: Air Warfare Division, 1983. CONFIDENTIAL.

U.S. Army, TRADOC Systems Analysis Activity. Antihelicopter Study Operational Effectiveness Analysis Using CARMONETTE, TRASANA-TR-39-83, vols. 1 and 2. White Sands Missile Range, N.M.: 1983. CONFIDENTIAL.

————. Division Air Defense Gun—Update, TRASANA-COEA-11-84, vols. 1 and 2, addendum and addendum 1. White Sands Missile Range, N.M.: 1984. CONFIDENTIAL.

————. Sgt. York Comparative Analysis, TRASANA-COEA-25-85. White Sands Missile Range, N.M.: 1985. CONFIDENTIAL.

————. Sgt. York Comparative Analysis Using Results from CARMONETTE/TRASANA Simulation Model, White Sands Missile Range, N.M.: 1985. CONFIDENTIAL.

————. "Sgt. York Comparative Analysis Using Results from CARMONETTE/TRASANA Simulation Model: TRASANA Executive Summary." Draft, White Sands Missile Range, New Mexico, 1985. CONFIDENTIAL.

U.S. Army, Training and Doctrine Command. Management: TRADOC Models, TRADOC regulation 5-4. Fort Monroe, Va.: 1982.

————. Army Programs: Studies and Analysis Handbook. Fort Monroe, Va.: 1985.

————. Management: Army Programs—Studies Under AR 5-5, TRADOC regulation 11-8. Fort Monroe, Va.: 1985.

U.S. Department of Defense, Weapon System Software Development, MIL-STD-1679 (Navy). Washington, D.C.: 1978.

————. Software Development, MIL-STD-1679A (Navy). Washington, D.C.: 1982.

————. Army Model Improvement Program, Army regulation 5-11. Washington, D.C.: 1983.

————. Defense System Software Development, DOD-STD-2167. Washington, D.C.: 1985.

————. "Software Quality Evaluation." Draft MIL-STD-2168, Washington, D.C., 1985.

————. Test and Evaluation, DOD directive 5000.3. Washington, D.C.: 1986.

U.S. General Accounting Office. Advantages and Limitations of Computer Simulation in Decisionmaking. B-163074. Washington, D.C.: 1973.

————. Guidelines for Model Evaluation. GAO PAD-79-17. Washington, D.C.: 1979.

————. Models, Data, and War: A Critique of the Foundation for Defense Analyses. GAO PAD-80-21. Washington, D.C.: 1980.

Van Tilborg, A. M., et al. "Simulation of Multicomputer Radar Systems for Ballistic Missile Defense." Simulation, June 1982, pp. 205-13.

Zeigler, B. P. Theory of Modelling and Simulation. New York: John Wiley and Sons, 1976.

Glossary

Air-To-Air Exchange Ratio	The proportion of enemv aircraft to friendly aircraft expected to be destroyed in a large series of one-on-one air combat encounters.
Battery Coolant Unit	A component used to cool the electronics systems of the Stinger missile.
Benchmark	A critical comparison of the results of one simulation model with those of another.
Configured Encounter	An interaction between hostile forces in which the geometry of the situation is a component of the analysis of outcomes.
Data Tailoring	Making significant adjustments to raw data so that they can be used in a particular application.
Deterministic Model	A model that uses expected values rather than distributions.
Fixed-Wing Aircraft	All aircraft except helicopters.
Free Encounter	An interaction between hostile forces in which the geometry of the situation is not considered in the analysis of outcomes.
Hardware-in-the-Loop	A form of simulation that incorporates components of the actual weapon system.
High Resolution	Compared to low resolution, the consideration of large number of factors in simulation.
Intervisibility	Lines of visibility where terrain must be considered between a threat and a target.

Jinking	Making relatively small but abrupt movements in three dimensions, as might be expected from a helicopter trying to avoid enemy fire.
Low Resolution	See High resolution.
Man-in-the-Loop	A form of simulation that incorporates the human operator of the weapon system.
Model	A set of mathematical or logical relationships that describe how a system works and behaves.
Monte Carlo Simulation	Any simulation involving the use of random numbers.
Radar Signature	The characteristics of electromagnetic waves reflected from a target that has been subjected to a radar beam.
Replication	The repetition of a simulation using different random numbers.
Sensitivity Testing	Determining if a model behaves as expected when one or more input variables are changed.
Simulation	A computer program that imitates the operations of various kinds of real-world facilities or processes.
Stochastic Model	A model that uses random variables defined within a common sample space.
Structured Walk-Through	An organized procedure for reviewing the quality and accuracy of a computer program.

Validation

Determining that a model is an accurate representation of the real system by comparing the model's output to that of the actual system or substitutes for it.

Verification

Determining that the computer program prepared for a simulation model is performing properly.

DATE
FILMED
5 8