ANTI ARMOR ADVANCED TECHNOLOGY DEMONSTRATION (A2ATD) EXPERIMENT 1

Susan Harkrider
U.S. Army Simulation, Training, and Instrumentation Command
12350 Research Parkway
Orlando, Florida 32826-3276
USA

William Yeakel
U.S. Army Materiel Systems Analysis Activity
Aberdeen Proving Ground, Maryland 21005-5071
USA

ABSTRACT

Historically, the analytical community has used constructive models such as JANUS and Combined Arms and Support Task Force Evaluation Model (CASTFOREM) to do analysis for the acquisition process. These types of models do not fully represent the impacts of human interaction with the system or their effect on combat effectiveness of the system during the early phases of research and development. Both the training and the research and development communities have used real time man-in-the-loop Distributed Interactive Simulation (DIS) for several years. However, the full potential of DIS as an evaluation tool to support material acquisition decisions has not been realized.

The Anti Armor Advanced Technology Demonstration (A2ATD) will function as a vehicle to meet several BDS-D exit criteria. A2ATD is a joint DA/DoD program initiated with the goal of maturing DIS as a credible evaluation tool to support acquisition decisions. The purpose of the A2ATD is to develop and demonstrate a verified, validated and accredited DIS capability to support anti-armor weapon system virtual prototyping, concept formulation, requirements definition, effectiveness evaluation, and mission area analysis on a combined arms battlefield at the Battalion Task Force or Brigade level.

This paper describes the preparations made and the conduct of A2ATD Experiment 1. A top level comparison of live simulation versus virtual simulation is also addressed.

ABOUT THE AUTHORS

Susan Harkrider is employed as an Electrical Engineer in the Engineering Directorate at the U.S. Army Simulation, Training, and Instrumentation Command (STRICOM). Ms. Harkrider is the Systems Engineer for the A2ATD program, which supports the Army's Battlefield Distributed Simulation-Developmental (BDS-D) program. She holds a B.S. in Engineering from the University of Central Florida.

William Yeakel is employed as the technical team leader for the A2ATD program at the U.S. Army Materiel Systems Analysis Activity. After undergraduate his undergraduate studies, Bill was commissioned in the U.S. Army as an infantry officer. Later, he worked for Aetna Financial Services, Lukens Steel Company and operated his own businesses for a period of 14 years. He started his civil service career at the U.S. Army's Ballistic Research Laboratory, transferred to AMSAA, and has been employed there since 1987. Mr. Yeakel holds bachelors degrees in Mathematics, German and Philosophy from Towson State University and Gettysburg College, a masters degree in Statistics, and has completed the coursework for a Ph.D. in Operations Research from the University of Delaware.

ANTI ARMOR ADVANCED TECHNOLOGY DEMONSTRATION (A2ATD) EXPERIMENT 1

INTRODUCTION

Historically, the analytical community has used constructive models such as JANUS and Combined Arms and Support Task Force Evaluation Model (CASTFOREM) to do analysis for the acquisition process. Such models do not fully represent the impacts of human interaction with the system or their effect on combat effectiveness of the system during the early phases of research and development. Both the training and the research and development communities have used real time man-in-the-loop Distributed Interactive Simulation (DIS) for several years. However, the full potential of DIS as an evaluation tool to support material acquisition decisions has not been realized.

The Simulation Network (SIMNET) proved the DIS concept with manned simulators for unit training. SIMNET has been used for both training and acquisition applications. The SIMNET used for acquisition applications has been replaced by the Battlefield Distributed Simulation - Developmental (BDS-D) simulation. There are BDS-D sites at the Battle Labs in TRADOC and Research and Engineering Development Centers in AMC. These facilities include collections of manned simulators with first generation image generators, semi-automated forces (SAFOR) work stations, Planned View Displays (PVDs) for observing the battle area graphically, Stealth displays for observing the battle area from any vantage point visually, plus systems for data collection and analysis.

The BDS-D simulation and its created synthetic environment represent the current state of the art in DIS. Upgrades to the environment, simulators, data analysis tools, and verification, validation, and accreditation (VV&A) are required to make BDS-D simulation a viable tool for supporting acquisition decisions. The BDS-D Advanced Technology Demonstration (ATD) is upgrading the environment and has taken the first step in VV&A of semi-automated forces. In addition, simulators have been and will be developed with improved image generators and hardware, requiring VV&A.

The Anti Armor Advanced Technology Demonstration (A2ATD) will function as a vehicle to meet several BDS-D exit criteria. A2ATD is a joint DA/DoD program initiated with the goal of maturing DIS as a credible evaluation tool to support acquisition decisions. A2ATD's purpose is to develop and demonstrate a verified, validated and accredited DIS capability to support anti-armor weapon system virtual prototyping, concept formulation, requirements definition, effectiveness evaluation, and mission area analysis on a combined arms battlefield at the Battalion Task Force or Brigade level.

The A2ATD technical objectives are:

a. Demonstrate DIS as an evaluation tool and verify, validate, and accredit simulators used in A2ATD experiments, semi-automated forces, and the BDS-D simulation.

b. Develop, demonstrate, and document techniques/analytical tools to evaluate the causes of simulation outcomes.

c. Demonstrate the linkage of constructive models such as JANUS to DIS.

d. Demonstrate upgraded virtual prototypes (M1A2, Bradley Fighting Vehicle, NLOS, LOSAT) and virtual prototypes to be developed (JAVELIN, Comanche, Apache, Hunter, ITAS).

Simulator and SAFOR VV&A and development of analytical tools to support the evaluation of causes of simulation outcomes were initiated in FY93 to provide the foundation for a series of experiments in FY94 - FY96. The first A2ATD experiment replicated two M1A2 Initial Operational Test and Evaluation (IOT&E) battles conducted at Ft. Hood, TX during the autumn of 1993 and the winter of 1994.

The simulators in the A2ATD experiments employ second generation computer image generators, i.e., with more processing and visual capability than their (SIMNET) predecessors. Visual images are rendered based upon multiple level environments such as day, weather and obscurants, 30 meter source terrain, infrared, radar, radio.

A2ATD EXPERIMENT ONE: IOT&E VERSUS THE VIRTUAL ENVIRONMENT.

The A2ATD M1A2 IOT&E experiment replicated two M1A2 IOT&E vignettes from the force-on-force (FOF) segment of the field test conducted at Ft. Hood, TX. The two vignettes selected were Blue Hasty Attack Battle and Blue Hasty Defense Battle.

The Blue forces were composed of fourteen M1A2 Abrams Main Battle Tank system entities for both vignettes. Of the fourteen M1A2 systems, four were represented by manned simulators in the virtual environment, while the remaining entities were controlled through Modular Semi-Automated Forces (ModSAF). All Red tank systems were represented by ModSAF entities. In the Blue Hasty Defense vignette, the Red force was composed of 26 tanks and 1 BMP in the attacking force. For the Blue Hasty Attack vignette, the Red forces were composed of 4 tank systems and 3 BMPs.

SIMULATION NETWORK COMPONENTS.

Experiment 1 was conducted on a DIS local area network (LAN) at the Mounted Warfare Test Bed (MWTB), Ft. Knox, KY. Voice (radio simulation) communications were effected through a hard-wired CB radio network. Non-voice inter-communication among the elements of the simulation network was routed through the ethernet. Components of the network included the four manned M1A2 simulators, a stealth vehicle, a simulation manager, the ModSAF Red and Blue Commanders' workstations, and data loggers. The stealth vehicle is a viewing platform consisting of three video displays that provide the observer a panoramic view of the virtual battlefield from any chosen location in the database, viewing in any direction, at any altitude. The simulation manager is a computer, connected to the simulation network, which provides exercise/entity control. The data loggers recorded the network protocol data unit (PDU) traffic during the experiment trials and forwarded the logs to the DIS analytical tools (DISAT).

The various devices connected to the LAN communicated via DIS PDUs, specifically DIS version 2.0.3 PDUs as defined in the IEEE Standard, Standard for Information Technology - Protocols for Distributed Interactive Simulation Applications, Version 2.0, Third Draft. The enumerations used are defined in Enumeration and Bit Encoded Values for Use with Protocols for Distributed Interactive Simulation Applications, 4th Draft.

ENTRANCE CRITERIA FOR EXPERIMENT 1.

Before any of the experiment trials were conducted, entrance criteria had to be satisfied. The entrance criteria were established in five areas (M1A2 simulator, ModSAF, DIS analytical tools, network, and experiment 1 planning) to assess readiness to execute a credible experiment. At the beginning of simulator development, the development personnel were provided standard Army algorithms. The algorithms provided methodology for simulation and assessment of main gun delivery accuracy, target acquisition, rates of fire, vulnerability against threats (direct and indirect fire), and mobility. M1A2 simulator development was completed and simulator performance was VV&A'd. The VV&A process provided assurance that the simulator had sufficiently modeled the M1A2 tank to allow for a credible experiment.

A visual database of Ft. Hood was developed for the simulators, stealth, ModSAF, and the simulation manager. The effort to minimize differences in terrain databases was con-siderable. In particular, there were several iterations addressing z-value and line-of-sight (LOS) correlation between ModSAF and the manned simulators. The correlation effort was expanded to include the CASTFOREM Ft. Hood terrain representation.

The battlefield was predominantly populated by ModSAF entities, therefore ModSAF also underwent extensive VV&A prior to the experiment. The ModSAF developers were provided the same set of standard Army algorithms for weapons system effectiveness. The major difference in the implementation of these algorithms for ModSAF versus the manned simulators is the error contributed to the task by the human in the loop. For simulators, the man in the loop provides a set of errors and the source is not incorporated into the simulator code. Conversely, ModSAF must include an estimate of this human error contribution. The U.S. Army Materiel Systems Analysis Activity (AMSAA) provided subject matter experts to the ModSAF developers to address these modeling issues.

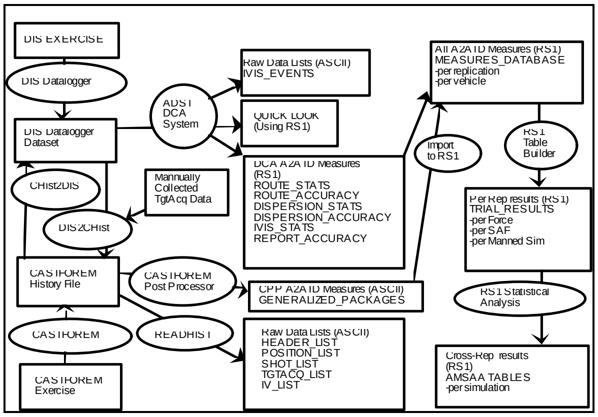

Figure 1. A2ATD Analysis Flow

VV&A check tests were performed exercising ModSAF physical algorithms for delivery accuracy, target acquisition, vulnerability, rate-of-fire, and mobility. These tests were designed to verify the implementation of standard Army algorithms and methodology and validate the system performance as consistent with the M1A2 system expectations.

Behavioral evaluation of ModSAF to assure conformance to Army armor tactics and doctrine sufficient to participate in the IOT&E replicated battles was completed by TRADOC Analysis Center at White Sands Missile Range (TRAC-WSMR). ModSAF was benchmarked for experiment 1 by comparisons of data gathered from the live simulation at Ft. Hood and from the results of pre-experiment CASTFOREM IOT&E replications.

To support the analysis process, DIS analytical tools (DISAT) were specified and developed for the experiment. The DISAT is a set of software tools that read and interpret the DIS PDUs logged on the data logger media, and among other things, translate them into CASTFOREM history files. The tools reduce DIS PDU information into DIS event histories, and export the information in a form that is compatible with the CASTFOREM post processor. DISAT provide utilities to convert recorded DIS data into a CASTFOREM history file and analysis routines to support VV&A activities, such as Loss Exchange Ratios (LER), Fractional Exchange Ratio (FER), and killer-victim score-boards. This data supported the analysis of the Measures of Performance (MOP), Effectiveness (MOE), and Behaviors (MOB). The ability to compare results from the BDS-D outcomes and CASTFOREM outcomes directly, without format changes, proved to be a significant benefit to the analysis effort. (See Figure 1)

The network was exercised to satisfy sufficient throughput requirements and careful attention was focused on the representation of various entities on different architectures. Differences in DIS enumerations were identified and corrected.

Troop orientation and training were conducted at the MWTB prior to the first official trial. Troops had the opportunity to practice with the simulators during pilot trials designed to be similar to, but not identical to the trials in the experiment.

Once the entrance criteria was satisfied, the A2ATD experiment ran 48 trials over a twelve day period. Each trial consisted of the replication of one of the live IOT&E vignettes. Twenty-four trials were run for each battle, twelve in BDS-D followed by the same vignette fought in ModSAF only. The scenario and platoon locations were randomized at each trial.

M1A2 SIMULATOR AND ModSAF VV&A SUMMARY.

Verification and validation tests were performed on the M1A2 simulators and ModSAF to ensure suitable functionality of each in the following areas: form, fit, and function; direct-fire vulnerability; target acquisition; main gun delivery accuracy; mobility; and timelines. Discrepancies in actual and expected algorithm output were analyzed to determine the source of the discrepancy. Code and/or input data structures were corrected until the performance of the simulators and ModSAF were in agreement with known performance data for the M1A2 Abrams system. V&V was completed and a detailed briefing identified each discrepancy, the corrective measure and the results of the retest before the simulators, ModSAF and BDS-D network were accredited for the experiment.

VV&A responsibilities for the manned simulator belonged with AMSAA. VV&A responsibilities for ModSAF were shared by AMSAA and TRAC-WSMR, with TRAC-WSMR conducting the V&V of behaviors. Behavioral deficiencies that could not be corrected in time were evaluated for recommended manned intervention or other workaround to adequately represent armor tactics and doctrine.

EXPERIMENT 1 COMPARISONS

EXPERIMENT ANALYSIS CYCLE.

The A2ATD experiment analysis cycle is shown in Figure 2. BDS-D simulations are run with verified, validated and accredited ModSAF and manned simulators (equipped with second generation computer image generators) in a synthetic environment using DIS 2.0.3+ standards. BDS-D simulation runs were made with and without simulators to provide the basis for comparing ModSAF and simulator behaviors with simulation outcomes. For the first experiment, these outcomes were also compared to the outcomes of the M1A2 IOT&E for the select vignettes replicated. Additionally, BDS-D simulation outcomes were compared to CASTFOREM outcomes to delineate differences and expose the reasons for those differences. This same process has been used for A2ATD experiment 2, and will be utilized in all future A2ATD experiments. CASTFOREM algorithm changes and runs were made in an attempt to bring the results into better agreement with the BDS-D simulation.

CASTFOREM.

Combined Arms and Support Task Force Evaluation Model (CASTFOREM) is a TRADOC accredited, stochastic, constructive force-on-force combat simulation which has been employed in Army acquisition analysis for many years. The employment of CASTFOREM to pre-experiment analysis provided the capability for refinement of input performance and scenario data. It also provided a reasonable benchmark for the performance of ModSAF throughout the experiment.

ANALYSIS APPROACH.

After each trial of experiment 1, the soldier participants were shown video of their performance and were asked to provide missing data such as which crew member first detected, recognized, and/or identified each battlefield entity that was encountered. At the same time, computer systems engineers transferred the electronic data to the MWTB analysis center. A "quick look" summary report was provided in less than half an hour after each trial.

At about the same time, the data logged files were copied to magnetic tape and electronically mailed to AMSAA via the internet. Since the data and the experiment were unclassified, data transfer

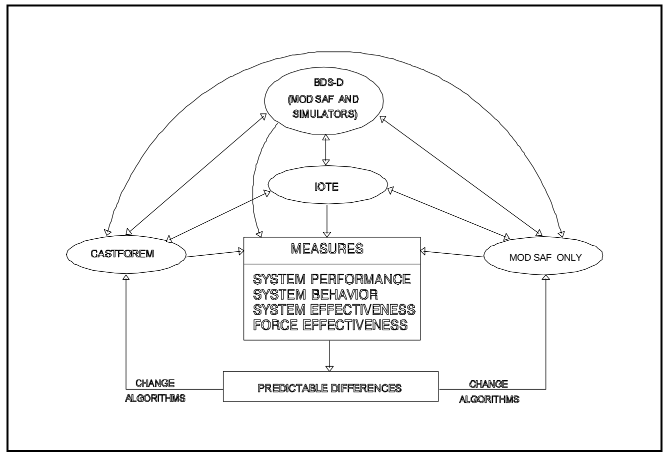

Figure 2. A2ATD Analysis Cycle

and data storage were not a problem. The CASTFOREM team at AMSAA received the raw datalogged files and reformatted files, readable by CASTFOREM, within 24 hours of experiment trial. At that point the post processor was engaged and analysts were presented with mountains of data.

CASTFOREM POST PROCESSING ANALYSIS.

The CASTFOREM post processor was developed by TRAC WSMR modelers to take raw study results and present data and summary statistics in an intelligible format. This format has gained considerable acceptance within the analysis community and is both well understood and thoroughly tested. It seemed reasonable therefore to format the captured PDU data so it could be provided as input to the accepted CASTFOREM post processor. The final data analysis product of the BDS-D runs would therefore be packaged in this familiar format. Direct comparisons could be made among BDS-D, CASTFOREM, and IOT&E data, and differences could be investigated easily as a result of this design decision.

Several charts were prepared from the reduced data, allowing for seemingly innumerable graphical comparisons of CASTFOREM versus BDS-D versus IOT&E. The detailed analysis is available in AMSAA Technical Report No. 569, Anti-Armor Advanced Technology Demonstration; Experiment 1, published in June 1995

Over 10 million data elements were collected in experiment 1 and the majority were never intended for a CASTFOREM post processor. Therefore, AMSAA chose an automated and statistically sound procedure to identify the significant MOE's, MOP's, and MOB's.

STATISTICS, STATISTICS AND MORE STATISTICS.

As discussed, an objective of experiment 1 was to compare a number of simulations, identify the differences among them, and determine the causes of those differences. This must be understood in the context of the A2ATD experiments, which are intended to demonstrate the usefulness of one of these simulation (BDS-D) as a tool to support the acquisition process. The other simulations being compared are ModSAF, CASTFOREM and the IOT&E live simulation. ModSAF is primarily of interest because it is a major component of BDS-D, and may be the cause of some differences. CASTFOREM and live simulation are current tools used in the acquisition process, so an understanding of the differences between these and BDS-D is needed in order to understand the utility of BDS-D and what advantages or disadvantages it might present.

That there are differences among these simulations is not surprising. The more interesting question is whether these differences significantly impact the utility of the simulations as acquisition tools. Experiment one analysis addresses differences in the performance of military forces as represented in the various simulations.

To evaluate this performance, we chose a number of measures of merit (MOM's), typically used for evaluations in support of acquisition decisions. The simulations were exercised to generate estimates of these measures. The values obtained were examined and compared using descriptive statistics such as graphs and tables. Formal statistical methods were used to confirm that perceived differences were statistically significant, and could not be explained by the stochastic nature of simulations. One could argue that any failure to detect statistically significant differences may be due to small sample sizes, rather than to agreement between simulations. Certainly the smaller the sample size, the greater the difference between simulations must be to be detected. The choice of sample sizes for this experiment was a compromise between a desire for large sample size and the limited resources available to produce them, A similar compromise is made whenever one of these tools is used in support of the acquisition process, and the sample sizes selected are typical of what can be expected for these applications. Each scenario was replicated 21 times in CASTFOREM, which is the number typically used for these runs. The selection of 21 is not serendipitous but is derived from observed convergence of the simulation. Each scenario was replicated only once in the IOT&E, which was an application of live simulation to the acquisition process. The number of replications used for BDS-D and ModSAF will typically fall between these two extremes. Thus, any difference that cannot be detected with these sample sizes would not be detected in a typical application. In other words, differences that are not statistically significant do not have a significant impact on the utility of the simulation as an acquisition tool. The converse is not necessarily true, a statistically significant difference may or may not be significant from the standpoint of model utility. The first step in the analysis is to look for statistically significant differences between the various simulations.

STATISTICAL ANALYSIS

There are many conditions (terrain, weather, force structure, etc.) which affect results in each of the simulations. There is a wide (but ill defined) range of conditions over which we would like to compare simulation performance. Each scenario represents a look into the behavior of the simulations in a small region within this wide area of interest. For Experiment 1, we had two such observations into simulation performance, the clarity of each observation was limited by sample size. In order to obtain as much information as possible from these data, we used a two stage statistical analysis, described below. This process was carried out for each measure of merit (MOM) and for each pair of simulations. The overall process is a hypothesis test. The null hypothesis is that the distributions of MOM estimates produced by the two simulations are the same.

FIRST STAGE ANALYSIS

The first stage analysis generates a statistic which under the null hypothesis above, will have a uniform distribution on the interval [0,1], i.e. (U[0,1]). To obtain this, the MOM values (for the particular MOM being considered) are computed for each replication of each of the two simulations being compared. When comparing CASTFOREM to IOT&E, for example, this would result in 22 data points, 21 from CASTFOREM and 1 from the IOT&E. These values are combined and ranked. Then the ranks corresponding to data from one of the simulations are summed. The result is the Wilcoxon Rank Sum Statistic, W. Our first stage statistic, T, is obtained from W by the transformation

$$T = F(W) - U \times d(W), where$$

F is the cumulative distribution function of W (under the null hypothesis)
d is the density function of W, and
U is a random number from a U[0,1] distribution.

In general, if X is a statistic with a continuous distribution with cumulative distribution F, then F(X) is distributed U[0,1]. If X has a discrete distribution, as does the Wilcoxon statistic, then the distribution of F(X) is a discrete approximation to U[0,1]. The subtraction of U*d(W) "smoothes" out the irregularity of F(W) to give an exact U[0,1] distribution (if the null hypothesis is true). Values of T close to zero or one indicate poor agreement between the simulations.

SECOND STAGE ANALYSIS

The second stage analysis looks at two or more first stage statistics which are related (i.e., the same measure but for different scenarios) but independent. These statistics form a two or higher dimensional random variable which, under the null hypothesis, has a uniform distribution over a two or higher dimensional hypercube. A random vector near the boundary of this hypercube indicates disagreement with the null hypothesis. To test for nearness to the boundary, we use a likelihood product, defined by the formula

$$L = \prod_{i=1}^{n} (2 \times \min(x, 1 - x_i))$$

where n is the dimension and x_i is the ith component of the random vector. The ith term of the product is the likelihood that a random value drawn from a U[0,1] distribution will be at least as near to the boundary of the interval [0,1] as x_i. We use the product as a measure of nearness to the boundary of the hypercube. If this value is too small, the null hypothesis is rejected. We will refer to this test as the Likelihood Product (LHP) test.

In order to make use of the LHP test, one must be able to obtain values from the cumulative distribution of the test statistic, L. This is given by the formula

$$P = L \times \sum_{i=1}^{n} \left\{ \left(\frac{1}{(i-1)!} \right) \times [-\ln(L)]^{(i-1)} \right\}$$

To find a value of L which corresponds to a given value of P, we start with an initial estimate, L_0 , of $P/2^n$ and iterate using the formula

$$L_{j+1} = P \times \left[\sum_{i=1}^{n} \left\{ \left(\frac{1}{(i-1)!} \right) \times [-\ln(L_j)]^{(i-1)} \right\} \right]$$

until an answer of sufficient accuracy is obtained.

Second stage analysis was used to combine results from the two scenarios, as well as, to examine questions which involve related subsets of the data from one scenario. For example, suppose we have data from BDS-D runs for both day and night runs with three and four manned simulators. Designate the four resulting populations as

A - day runs with 4 SIM's
B - day runs with 3 SIM's
C - night runs with 4 SIM's
D - night runs with 3 SIM's

To address the question of whether having only three SIM's makes a difference, compute first stage statistics comparing populations A to B and C to D. Combine results using the LHP test to look for consistent differences between three and four SIM estimates across day and night conditions.

FORCE AND SYSTEM LEVEL STATISTICS.

In order to maximize information obtained from the test data we looked at MOM's at both force and system level. A force level MOM is computed for the entire force under study (i.e., the number of shots fired by Blue forces). System level MOM's are computed for individual fire units (i.e., the number of shots fired by a particular entity).

Analysis of force level statistics is straight-forward. Computer the value of the MOM for each replication with each simulation. To compare values obtained by two different simulations, take the two populations of MOM values and computer first stage statistics as described above. In our Experiment 1 data analysis, we generally stopped at this point, although one could use second stage analysis to combine results from the two scenarios and obtain an overall assessment of model agreement with regard to this particular force level MOM.

Analysis of system level statistics is complicated by the fact that MOM values obtained for the various fire units are, in general, correlated. If first stage statistics were computed for each fire unit, these values would also be correlated, and therefore could not be used for the second stage analysis. To avoid this problem, the data are decorrelated before analysis. This is done as follows:

a. Organize the data into a matrix, M. The columns of M correspond to the fire units and the rows of M correspond to replications of the scenario in the various simulations.

b. Obtain the singular value decomposition (SVD) of the variance matrix, var(M), the variance/ covariance matrix of M. Using the SVD, a real matrix A can be written as

$$A = t(V)\ S\ U$$

where V and U are Unitary matrices and S is a diagonal matrix.

c. Computer the decorrelated data matrix,

$$D = MV$$

For each pair of simulations to be compared, first stage analysis con now be done on each column of D. This produces a vector of statistics that can be processed, using second stage analysis, to give an assessment of simulation differences for this pair of simulations, scenario, and system level MOM. There is a close relationship between the force and system level statistics. In fact, the force level MOM can be obtained from M by adding the columns together. Clearly, these two views of the data are not independent. The force level data can be thought of as a projection of the multi-dimensional system level data onto a single dimensional subspace. Because it focuses entirely on this single dimension, the force level test is more sensitive to differences in that one direction, but it ignores any differences in the other directions. The system level test, on the other hand, can detect differences in any direction. Singling out the force level MOM's is justified by the fact that they are typically the measures used when these models are employed for acquisition related applications.

SUMMARY

Results from the first A2ATD experiment were used in an attempt to validate virtual and constructive simulations against a live simulation, the M1A2 IOT&E. Results from the virtual simulations were also used as an additional validation for portions of constructive models, ModSAF and CASTFOREM.

To that end, the A2ATD program established entrance Criteria in five areas:

- M1A2 manned simulator VV&A;
- ModSAF VV&A;
- DIS analytical tool development;
- Simulation network operations, and
- Experiment 1 planning.

The ability of Experiment 1 hardware and software to meet the entrance criteria was evaluated to assess readiness to execute a credible experiment.

In this paper, we will share results from one of the two-by-two simulation comparisons as an example of the analysis results. Hence for the purposes of this paper, we will not discuss the comparison of BDS-D with ModSAF only, or live with constructive; instead we will share some of the results of the live versus BDS-D analysis, since that has generated most interest to date.

LIMITATIONS

The use of M1A2 IOT&E results as a representative benchmark for comparison with other simulations was limited by a number of factors:

a. Only one replication of each battle was available for comparison. The IOT&E battles examined had a relatively small force structure (e.g., a Blue Company versus an attrited Red Brigade in the Hasty Defense scenario) and, with only a single iteration, the outcome must be considered a random result from among a wide range of possible outcomes. The best that can be expected is to determine if the IOT&E battle outcome is within the range of possible outcomes from other types of simulations that have larger sample sizes. To use the IOT&E results as a firm benchmark is inappropriate -- it would be necessary to have multiple replications of these small live vignettes to resolve the issues concerning statistical confidence intervals.

b. The information obtained from the IOT&E lacks adequate detail necessary for a thorough analysis. Instrumentation problems were responsible for a large number of unknown target acquisitions and engagements in the live IOT&E. These shortfalls in the live simulation source data leave a number of data gaps dealing with the number of false engagements and weapon system performance critical to the analysis of force-on-force results. It is unclear whether these data gaps are truly unknown engagements, attributable to instrumentation problems, or a combination of both.

c. For safety reasons, a filter was placed upon the laser range finder which caused it to be inaccurate for making range estimates to engage targets at ranges greater than 1000 meters. Therefore, range estimates were based upon visual range observations. This contributed to unrealistically long-range engagements during the IOT&E. In order to match the conditions that occurred in the IOT&E, the opening range of engagement for the main tank rounds were extended from 3 to 4 km in BDS-D, ModSAF and CASTFOREM. The resultant unrealistic play of the M1A2s in the IOT&E laser-based RTCA system led to inappropriate firing doctrine in the field trials. When no RTCA response was received from an attempted engagement (i.e., no RTCA firer-target pairing), the M1A2 crew tended to move to a new target rather than engage the same target with a subsequent round.

COMPARISON of LIVE to VIRTUAL

The Hasty Attack scenario showed close correlation between BDS-D and IOT&E results. Differences found were related to discrepancies in munition performance for long-range engagements (greater than 3 km). These long range engagements resulted in significant kills in BDS-D while IOT&E had no kills.

The Hasty Defense scenario showed significant differences between BDS-D and IOT&E results. These differences were driven by BDS-D overplay of attacking Red force target acquisition capability. This, in combination with the lethal Red tank missile and favorable Line-of-Sight (LOS) opportunities, led to a high rate of Blue losses. As a result of the high loss rate, the Blue commander varied his tactics in subsequent runs to increase the survivability of his forces.

In both scenarios, the M1A2 manned simulators exhibited behavior characteristics similar to those observed for M1A2s during IOT&E. These characteristics include survivability drills and continual adjustment in defensive or over-watching postures to minimize exposure. Additionally, the BDS-D M1A2 gunners acquired most of the targets (as in IOT&E) even with the commander having an independent thermal viewer.

CONCLUSIONS:

Qualitative validation of one simulation with another requires a firm benchmark with an adequate level of detail and confidence in the data upon which to base the comparisons. One of the lessons learned from this A2ATD experiment was that caution must be exercised I using live quantitative results as a basis for validating virtual and/or constructive simulations. It is recognized that IOT&E's were not designed or implemented with the expressed purpose of providing the detailed information necessary to function as a live simulation benchmark.

Given this recognition, the IOT&E was used as a benchmark for this A2ATD experiment, with the understanding of the limitations involved and the assumption that the IOT&E battles selected were representative of typical battles. The A2ATD experiment 1 demonstrated that IOT&E scenarios could be replicated in the virtual simulation environment. However, this replication required special attention to weapon system performance and calibration of the IOT&E data based upon conditions that occurred in the field trials.

The fact that the BDS-D trials did not show agreement as well as CASTFOREM is not a surprise. After all, the CASTFOREM simulation is fully scripted and should by definition follow the flow of battles in precisely the same way as the live simulation. However, based upon M1A2 manned simulator performance, an argument can be made that the BDS-D M1A2 manned simulator can more closely represent M1A2 combat performance than even the M1A2 played in a field trial. The BDS-D crew is not constrained by real-world safety considerations, or by the engagement /lethality limitations associated with current live field-trial instrumentation. Additionally, the simulators can use certified M1A2 performance data. Since BDS-D commanders were allowed to fight the battles as they chose, deviations from the exact tactics did, in fact, occur. If IOT&E scenarios were replicated with this flexibility, there is strong evidence that similar variability would be expected.

Although, the balance of the A2ATD experiments will examine these issues more closely, it is reasonable to assert that based upon the results of Experiment 1 that BDS-D could be used as credibly as live simulation to represent real battles. This is, of course, dependent upon careful experimental design, good visual representation of moving models and the environment, performance data certification, and careful attention to verification, validation and accreditation of all components.