A Hybrid Approach to Combat Simulation Experimentation

Chris Willis, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
christopher.r.willis2.civ@army.mil

John Bayer, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
john.l.bayer.civ@army.mil

MAJ Jake Kelly, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
jacob.p.kelly4.mil@army.mil

Sam Anderson. CMSP
Maneuver Battle Lab
Fort Moore, Georgia
samford.d.anderson.ctr@army.mil

ABSTRACT

To ensure the validity of a warfighting experiment, the experimenter must be able to detect change, isolate its cause, and relate experimentation results to real-world (combat) operations. Constructive simulation experiments (SIMEXps) with multiple runs are ideal to detect change and to isolate its cause. However, because modeling complex events requires many assumptions, critics often question the applicability of constructive simulation results to operational situations. This is juxtaposed by command-and-control human in the loop (HITL) experiments in which military staffs receive real-time, simulated sensor inputs, make real-time decisions and direct simulated friendly forces against simulated threat forces. The use of actual military personnel and staffs allow this type of experiment to reflect warfighting decision-making better than purely closed loop constructive experiments thus increasing the applicability to real-world operations. However, when humans make decisions, variability increases, and changes are more difficult to detect. The Maneuver Battle Lab (MBL) uses a hybrid approach to SIMEXps to capitalize on the strengths of both the closed loop constructive and HITL approaches, while minimizing the weaknesses of each technique. The MBL uses operating force military role players (MRPs) to replicate staffs during HITL SIMEXps to ensure applicability to real-world operations. The tactics developed by these MRPs – during several replications per case – are then scripted into the simulation scenario and replicated in a constructive (or closed loop) manner for the required number of times to ensure statistical significance. The hybrid approach described in this paper provides a means for other organizations conducting military experimentation to increase the validity of their constructive experiments, while simultaneously increasing the real-world applicability of their experiments.

ABOUT THE AUTHORS

Chris Willis, an Army Civilian, currently serves as the Director of the Maneuver Battle Lab (MBL), Fort Moore, Georgia. He is a graduate of the United States Military Academy and holds a master's degree from Long Island University. Chris is a retired United States Army Infantry Officer. His assignments included service in Afghanistan, Iraq, and Korea. Chris has led virtual and constructive experimentation both locally at the MBL and distributed through the Army Persistent Experimentation Network (APEN). He is a Certified Modeling and Simulation Professional (CMSP) and holds a professional certificate in modeling and simulation from the University of Alabama, Huntsville.

John Bayer, an Army Civilian, currently serves as the Chief of the Modeling and Simulation Branch (MSB) for the Maneuver Battle Lab (MBL), Fort Moore, Georgia. He holds a bachelor's degree from the University of Phoenix. John has served as an Army experimentation simulationist for over 20 years. He has participated in virtual and constructive experiments at the current MBL, the former Mounted Maneuver Battle Lab (Fort Knox, Kentucky), and those distributed through the APEN. John is a CMSP.

MAJ Jake Kelly currently serves as a Project Officer within the Modeling and Simulation Branch for the Maneuver Battle Lab, Fort Moore, Georgia. He graduated from Stephen F. Austin State University and commissioned through ROTC as an Armor Officer. He also holds a Masters in Operational Studies from the U.S. Army Command and General Staff College. His assignments include service in the Sinai Peninsula, European RAF mission, and Korean RAF mission. He is a CMSP and has led live, virtual, and constructive experimentation at MBL and distributed through the APEN.

Sam Anderson, a Department of Defense Contractor, currently employed as a Senior Research Operations Analyst with Science Applications International Corporation for the Fort Moore, Maneuver Center of Excellence, Maneuver Battle Lab. Sam is a retired United States Army Officer. He served in Kuwait, Bosnia, Korea, and Iraq. Sam has conducted simulation experimentation on future force design for Army Brigade Combat Formations. His college education consists of a bachelor's degree from the University of Hawaii and a master's degree from Regents University. Sam is a CMSP.

A Hybrid Approach to Combat Simulation Experimentation

Chris Willis, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
christopher.r.willis2.civ@army.mil

John Bayer, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
john.l.bayer.civ@army.mil

MAJ Jake Kelly, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
jacob.p.kelly4.mil@army.mil

Sam Anderson, CMSP
Maneuver Battle Lab
Fort Moore, Georgia
samford.d.anderson.ctr@army.mil

INTRODUCTION

The Maneuver Battle Lab (MBL), in support of the US Army Maneuver Capabilities Development and Integration Directorate (MCDID), conducts live and constructive experimentation and analysis to support concepts, requirements, and integration for the Soldier, Brigade Combat Teams, and Cross Functional Teams (Soldier Lethality and Next Generation Combat Vehicles). The MBL's primary simulation tool is One Semi-Automated Forces (OneSAF), which is a composable, open source, medium resolution, constructive ground combat simulation. A government project officer (PO) leads the coordination and synchronization process to ensure that the experiment planning is synchronized and that the planning will result in an experiment that produces outcomes to answer the customer's experiment objectives. The warfighting experiment may result in an acquisition decision, changes to US Army concepts or doctrine, inform science and technology (S&T) investments, and/or inform the Joint Capabilities Integration and Development System (JCIDS). Some of the outcomes of recent experiments include:

- Informing the Optionally Manned Fighting Vehicle (OMFV) Request for Proposal (RFP) and Capability Development Document (CDD)
- Refinement to Next Generation Combat Vehicle (NGCV) Cross Functional Team (CFT) Robotic Combat Vehicle (RCV) program requirements
- Informing the Multi-Domain Operations (MDO) Concept
- Informing Army Functional Concept 2016-2028
- Assessing an artificial intelligence (AI) enabled Adaptive Command and Control Web (AC2W) designed to control swarms of drones
- Informing AC2W/Human Interface Requirements
- Informing AC2W strike requirements
- Assessing Modeling and Simulation (M&S) capability to model intelligent behaviors for swarm Unmanned Aerial Systems (UAS)

An important consideration for warfighting experiments is the distinction between statistical significance and operational relevance in the military context. There may be cases in which statistical differences are found between two alternatives but the relative difference between them has no operational impact. Oftentimes, insights around the operational relevance of the differences between alternative configurations are considered by subject matter expert (SME) input after the simulations have been completed. This approach can lead to potential inefficiencies relating to the number of replications that need to be conducted. As the MBL staff designs experiments to address desired outcomes that inform requirements as described above, they often struggle with balancing the establishment of statistical significance with real-world (combat) applicability of the new capability that is being assessed. This paper will address an experimentation method, known as the constructive hybrid method, that balances those two demands.

MILITARY EXPERIMENTATION

In *The Logic of Warfighting Experiments,* Richard Kass, states that the purpose of military warfighting experiments is to explore the effects of manipulating proposed warfighting capabilities or conditions (Kass, 2006). Conditions include terrain, threat, and weather. At the MBL, we expand on Kass' definition to include the manipulation of new warfighting concepts and new combat unit formations. Therefore, a military warfighting experiment at the MBL explores the effects of changing the variables of concepts, capabilities, unit formations and conditions.

Combat is an inherently dangerous endeavor; failure in the design and experimentation of Army weapons and combat vehicles could ultimately lead to the loss of American lives. To prevent the grave consequences of failing to deliver sound results for military experimentation, the MBL must prioritize the validity of its experiments. According to Kass (2006), there are four requirements for ensuring warfighting experiment validity. They are, 1) the ability to use a new capability, 2) the ability to detect change, 3) the ability to isolate the cause of the change, and 4) the ability to relate experimentation results to real-world (combat) operations.

Kass further states that most warfighting experiments can be grouped into one of four general methods: analytic wargame, constructive experiments, human-in-the-loop experiments, and field (or live) experiments. Each of these four methods has inherent strengths and weaknesses with respect to the previously discussed validity requirements (see Figure 1, below). Because no single experimentation method can satisfy all four requirements, a comprehensive campaign of learning requires multiple experimentation methods (Kass, 2006). Although the MBL conducts all four methods of experimentation, we will focus only on the constructive, closed-loop and human-in-the-loop (HITL) methods for the purposes of this paper.

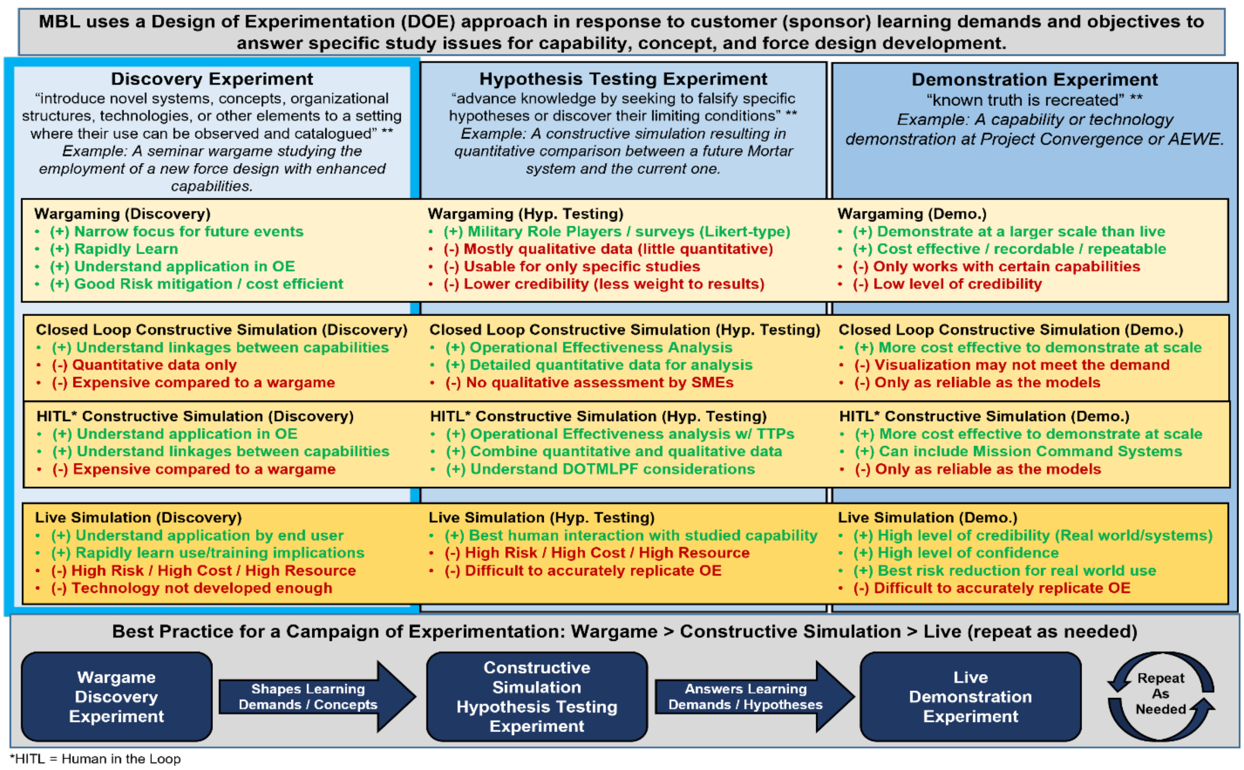

Figure 1 – Warfighting Experiment Techniques and Methods

STATISTICAL SIGNIFICANCE AND POWER

Assuming that change can be detected during an experiment, then determining the cause of that change is essential to valid experimentation. Because an experiment will only simulate a sampling of real-world data, the experimenter must establish that the detected change is not due to chance. Statistical significance helps quantify whether the results are due to a measurable change or to chance (Gallo, 2016). Statistical significance is established by the experimenter, but generally understood to be inside the 95% confidence interval of a sample.

According to the central limit theorem (CLT), for almost all populations the sampling distribution of the mean is approximately normal when the simple random sample size is sufficiently large. Statisticians generally agree that if the sample size is 30 or larger, then the distribution will be normal and the CLT will apply. However, if the distribution is skewed, then a larger sample size may be required for the CLT to apply (Brase & Brase, 2012).

Large sample sizes reduce the chance that variability (or noise) will impact on the chance that change is detected. An experiment's ability to detect change is known as power. Various statistical techniques exist to determine the sample size required to achieve specific statistical power levels. As sample size increases, so does statistical power (Kass 2006).

Statistical techniques, known as power analysis, can be used to estimate the probability of detecting change given a certain magnitude for a specified effect. A power analysis prior to data collection can be accomplished by estimating the sample size required for statistical significance. Following data collection, the true statistical power provided by the experiment can be determined (Kass, 2006). If sufficient power is not attained, then more replications may be required.

CONSTRUCTIVE SIMULATION

Constructive simulations are those that use simulated people operating simulated systems. However, real people can be allowed to stimulate (make inputs) to constructive simulations (MSCO, 2011). Constructive simulations normally involve two opposing sides that each make independent decisions based on their perception of the battlefield. Military command and control (C2) systems are stimulated by the simulation, which creates the common operational picture (COP) for both the friendly and enemy role players (see Figure 2, below) (SOC, 2023). For the purposes of this paper, constructive simulations are further subdivided into closed loop and human in the loop (HITL).

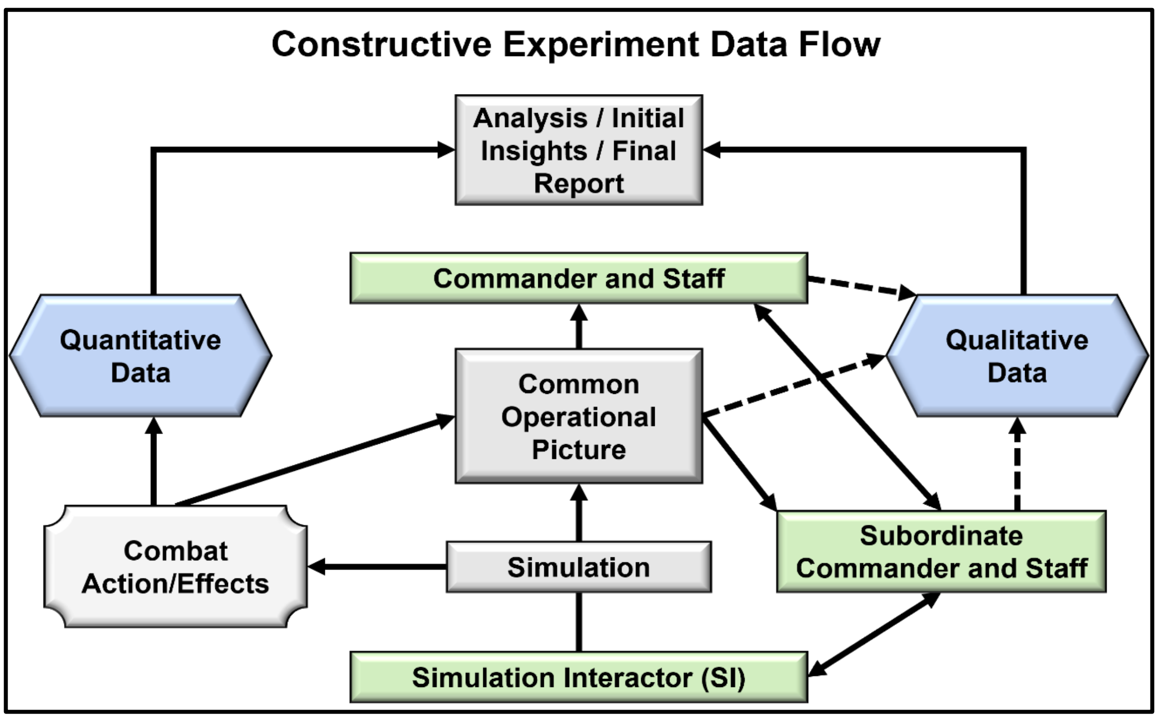

Figure 2 – Constructive Experiment Data Flow

Closed Loop Constructive Experimentation

Constructive simulations, in a closed form fashion, allow us to conduct the required number of replications to establish statistical significance. Additionally, constructive experiments provide statistically defensible evidence of improvements across a wide range of conditions (Kass, 2006). A typical methodology for conducting closed loop simulation experiments is shown in Figure 3, below.

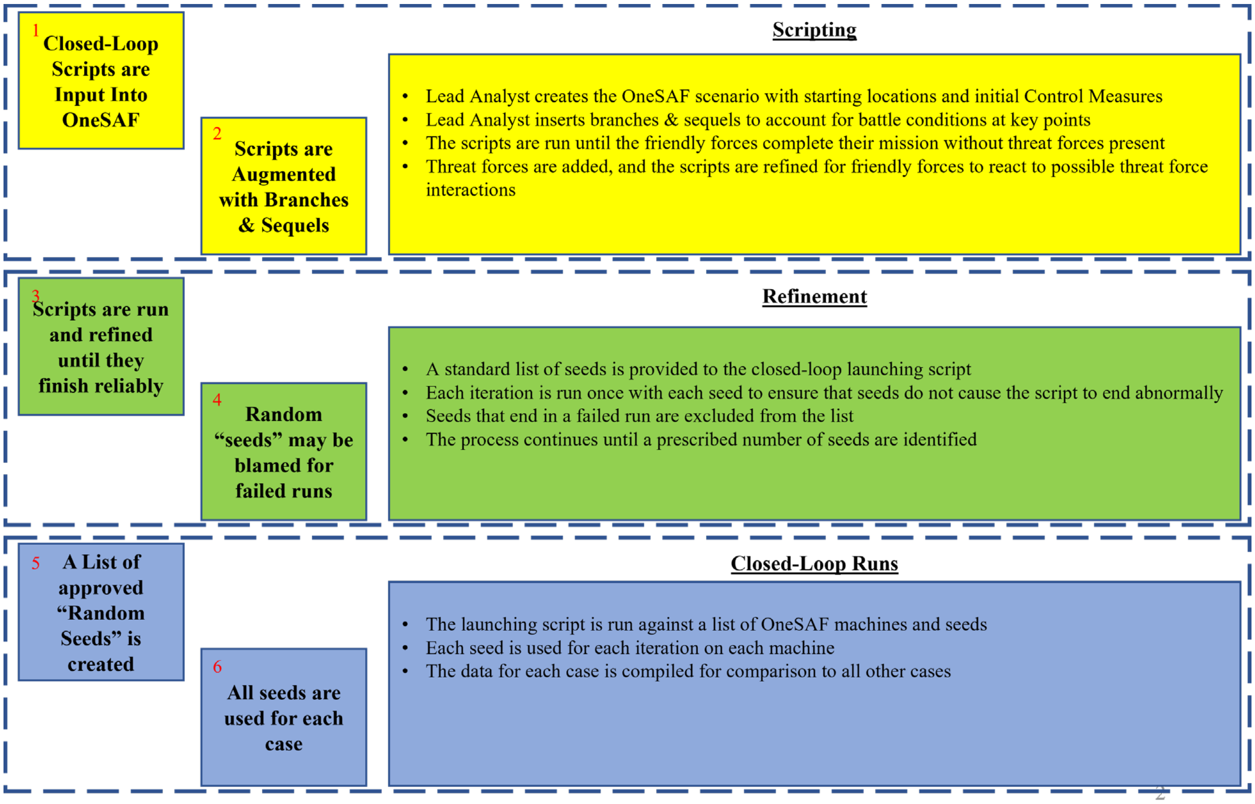

Figure 3 – Closed-loop Methodology

In a closed loop SIMEXp, no human intervention occurs in the play after experimenters choose the initial parameters and then conduct the simulation. Constructive simulations are a mainstay of warfighting experimentation and are conducted by practically all military analytical agencies and battle labs. Constructive simulations allow for repeatability of the same battle under identical conditions, while allowing for a controlled and systematic variation of variables. As previously discussed, this can include the insertion of a new weapon or sensor, employment of different concepts, doctrine, and/or tactics, or the encounter of a different threat. Closed loop constructive simulation experiments (SIMEXps) with multiple replications are ideal to detect change and to isolate its cause. However, because modeling complex events requires many assumptions, critics often question the applicability of constructive simulation results to real-world operations (Kass, 2006).

Furthermore, an important consideration for warfighting experiments is the distinction between statistical significance and operational relevance in the military context. There may be cases in which statistical differences are found between two alternatives but the relative difference between them has no operational impact. Oftentimes, insights around the operational relevance of the differences between alternative configurations are considered by SME input after the simulations have been completed. This approach can lead to potential inefficiencies relating to the number of replications that need to be conducted. An alternative approach would be to identify the operational relevance threshold, or desired effect size, for each metric of interest prior to running the simulation. This may allow a smaller number of replications for each metric to be calculated using statistical techniques, assuming a pilot set of replications was run to estimate the variance of each metric. However, this approach is predicated on the assumption that the magnitude of the absolute values generated by the simulation is comparable to those expected in the real-world. Anecdotal evidence indicates that combat simulations can produce trends consistent with real-world military scenarios although the order of magnitude of the results is likely to be different (Chau, Gill & Griefer, 2017).

Human in the Loop (HITL) Simulation Experimentation

To overcome the disadvantages of post hoc SME input, military simulation experiments often use a HITL approach. HITL SIMEXps are a blend of constructive experiments and field experiments. In a C2 HITL SIMEXp, a military staff receives real-time, simulated sensor inputs, makes real-time decisions, and directs simulated friendly forces against simulated threat forces. The use of operational force military role players (MRPs) and staffs allows this type of experiment to reflect warfighting decision-making processes better than purely closed loop constructive experiments. This increases the applicability to real-world operations. However, humans may play differently against computer opponents than against real opponents. Additionally, when humans make decisions, variability increases, and changes will likely be more difficult to detect (Kass, 2006).

HITL simulations are used extensively for training exercises in the military, but they can also be used for both concept and capability development simulation experiments. HITL experiments essentially use a simulation to stimulate military C2 systems. This allows the experimenter to garner both qualitative and quantitative data from the experiment. Qualitative data is obtained through after-action reviews (AARs), surveys, and interviews. Since combat is an inherently human endeavor, the input of human MRPs is vital to understanding the applicability of new concepts and capabilities to a combat formation. However, it is extremely resource intensive, in terms of both time and manpower, to conduct enough replications to establish statistical significance while using HITL SIMEXps. To garner both the qualitative data provided by a HITL experiment, and the statistical significance provided by a closed loop experiment, a hybrid approach should be used.

MANEUVER BATTLE LAB HYBRID APPROACH

The Maneuver Battle Lab (MBL) has developed a hybrid approach to SIMEXps to capitalize on the strengths of both the closed loop constructive and HITL approaches, while minimizing the weaknesses of each technique. The MBL uses operational force MRPs to replicate staffs during HITL simulation experiments to ensure the validity of operational employment and applicability to real-world operations. The tactics developed by these MRPs – during several replications per case – are then scripted into the simulation scenario and replicated in a closed loop manner for the required number of times to establish statistical significance. The current hybrid execution model has evolved as training and capability opportunities presented themselves.

Maneuver Battle Lab HITL Execution

In the past, the MBL exclusively conducted HITL experiments. During these HITL experiments, participants took great care to understand the new formation or equipment and how to employ it. MRPs would conduct the Military Decision-Making Process (MDMP) to develop tactics, techniques, and procedures (TTPs) that would leverage the changes to a tactical advantage. They would then conduct mission rehearsals in the simulation to ensure that the developed TTPs would provide the desired effect. The MRPs would then modify the TTPs until they were confident in their effectiveness. The MRPs would finalize their plans during pilot runs, adjusting for new conditions presented by an enemy threat MRP. The pilot runs also allow for standardization of communication and tactics between the MRPs and simulation interactors (SIs). The MRPs would then lock down their schemes of maneuver in the first iterations of record replications such that deviation among runs was kept to a bare minimum. The intent was to try to ensure that conditions among record replications were as consistent as possible. Once plans were locked in, deviations to the scheme of maneuver were approved by the experiment director. To prevent the MRPs from learning from their previous mistakes and modifying their actions among runs, the base case and alternative cases were executed in pseudo-random order. Random order run sequencing reduces the likelihood of artificially increasing variability in the experiment.

The MBL would then record approximately ten replications per base case and each alternative case by the end of the experiment. These results were the best effort given the time and human resources available for HITL experiment execution.

Limitations of HITL Execution at the MBL

As previously stated, time and human resource limitations constrained the number of HITL iterations to a relatively small number. Although the number of iterations was considered sufficient, they tended to be on the lower end of the scale. Conducting the required number of runs to establish statistical significance would have required investments that were untenable in terms of time, personnel, and money.

Furthermore, it is extremely difficult to prevent military professionals from attempting to correct errors that become obvious as the experiment progresses; also known as the learning effect. Small, almost imperceptible changes were often discovered after the completion of experiments. Removing these record runs, when discovered, reduced the number of viable replications, further complicating the establishment of statistical significance.

In 2012, prior to the MBL developing a hybrid approach, the MBL conducted an experiment focused on the utilization of a future weapon optic. During the experiment, the MBL had operational MRPs conduct missions within three different scenarios (Attack, Cordon and Search, and React to Enemy Ambush). Within each scenario they conducted three different use cases: Base Case with current optics, Alternate Case 1 with Rapid Target Acquisition (RTA) at a basic level, and Alternate Case 2 with RTA at an advanced level. The objective of the experiment was to compare the operational impacts of using an advanced weapons optic when operating over urban terrain, against hybrid/irregular forces. An important aspect of the future weapon optic is the capability to acquire and engage targets faster and at longer ranges than current optics. Therefore, to compare and assess the operational impact of the future optic, analysis focused on the performance of the optics and minimized the effects of other variables (such as specific weapon system, probability of hit, operator proficiency, tactics, etc.). To increase assurance that assessment outcomes were directly attributable to the specific optics, each optic was separated into subcomponents and analyzed as a function of range and time.

At conclusion, the MBL completed a total of 52 record runs with an average of 17 runs per case. While 17 runs per case is nearing the threshold for statistical significance, throughout the HITL record runs the operational MRPs and Threat Emulation Force (TEFOR) continued to make refinements and adjustments to their TTPs which added variability. An example of this learning effect noted in the experiment report occurred when the threat force adjusted and slowed their maneuver to limit the friendly force's target acquisition capability. Additionally, during execution of the record runs, six different adjustments were made within the simulation ranging from illumination levels to windowsill heights which added additional unintended variables throughout the simulation runs.

Due to the human induced adjustments and variables listed above, the analysis team had difficulty determining causality and providing quantifiable, statistically significant data for use in the final report. As the MBL continued to execute constructive experiments, unforeseen adjustments and variables in the simulation would almost always appear. Acknowledging and executing these necessary adjustments was instrumental in ensuring the experiment scenario was valid. However, before the completion of an experiment, the analyst had to ensure the data output had quantifiable integrity and statistical significance that can only be established through closed loop iterative runs. Without the use of multiple runs in closed loop, the analysis team is left with a dilemma of sifting through multiple variables – both from the simulation and from the MRPs and Threat Force learning effect – to determine whether, and to what, degree of influence those variables had on the result. While the HITL technique provided excellent SME tactical and technical inputs to the MBL's SIMEXps, the noise induced by the inevitable variability of human behavior required the development of a more robust technique.

Maneuver Battle Lab Hybrid Execution

To increase the statistical significance of its SIMEXps, while maintaining the real-world applicability provided by MRPs, the MBL developed a hybrid constructive simulation approach beginning in 2019. The experiments are still initially conducted as HITL, which allows for the MRPs to define and refine TTPs during execution, to evaluate the new organization or new capability, and to provide qualitative feedback to the analysts. Unlike the previously used HITL-only technique, MRPs are now encouraged to improve their performance during each replication until the results are as useful as can be expected. Additionally, the MRP feedback is collected as qualitative data through surveys, interviews, and AARs throughout the experiment process.

In the MBL, MRPs do not typically interact directly with the simulation. The MBL provides SIs, or professionals trained in the operation of the MBL's primary simulation tool OneSAF. The SIs take notes on the progression of the missions during HITL run execution. They are tasked with reproducing the final schemes of maneuver as directed by the MRPs. The notes are compiled into execution matrices by which a OneSAF Battle Master can script the battle from both the friendly and threat perspectives. The scripted scenarios are then run for the MRPs and evaluated to ensure that the results match the intent as provided (see Figure 4). This is perhaps the most important part of the hybrid process. The key advantage that the hybrid approach delivers over the closed loop technique is the inclusion of the MRPs' tactical expertise. The replay of the scenario for, and subsequent adjustments by, the MRPs ensures the tactical validity of the scripted behaviors for each experiment case. This is the last opportunity that a human can provide input into the closed loop scenario.

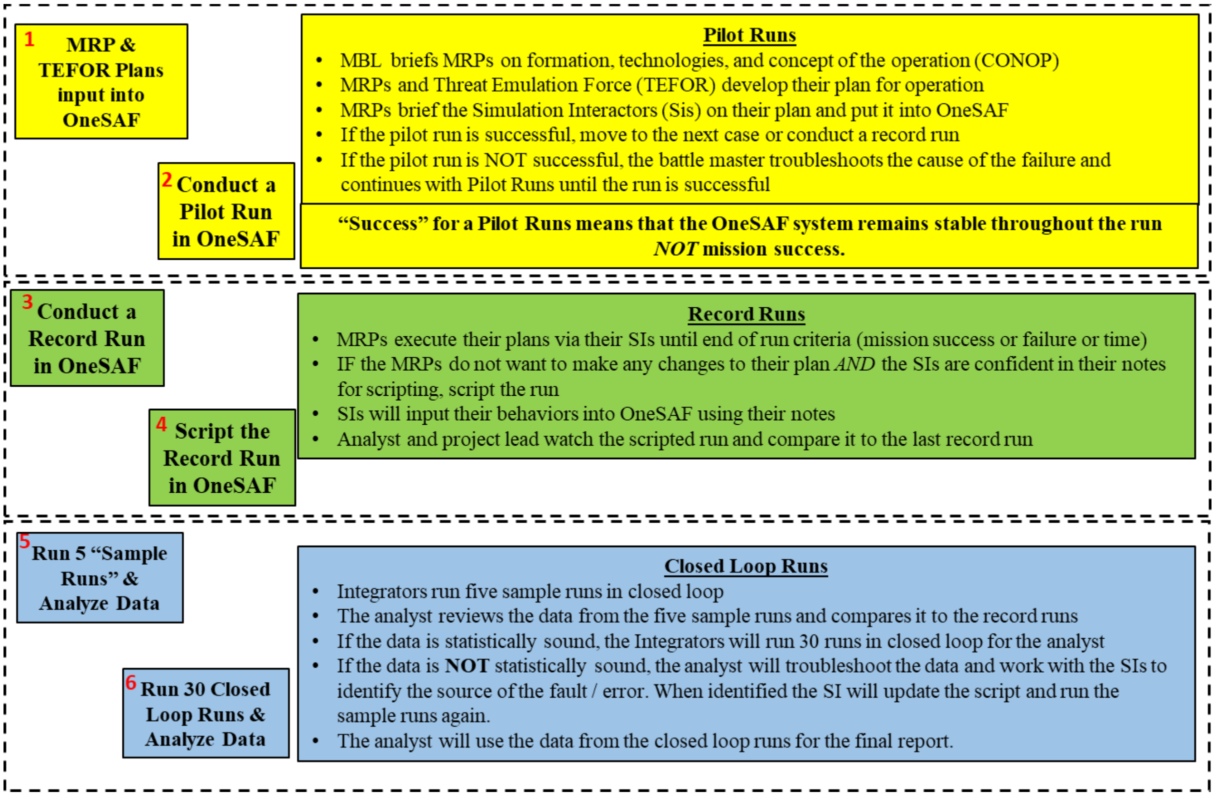

Figure 4 – MBL Hybrid Methodology

Once the scripted scenarios are finalized, they are replicated in closed loop fashion at least thirty iterations per case. If the power analysis determines that more replications are required, then the closed loop scenario is completed with as many iterations as required to establish statistical significance. The MBL data collectors have developed software to process the raw quantitative data provided by the OneSAF replications. Once this data is reduced, then it is provided to the experiment analyst in a summarized ordered report format. Using this data, the analyst evaluates it against measures of merit (MOMs) and determines if changes can be detected and if that change is statistically significant. This quantitative analysis will then be correlated with qualitative data from the MRPs, which will ultimately inform experiment findings. It is important to note that the experiment analyst also observes the HITL replication execution to truly understand why the MRPs made certain decisions and how they arrived at their qualitative feedback. The final report is then published following review by all stakeholders for accuracy of content.

An Example of the MBL's Hybrid Technique

The MBL completed its first hybrid constructive OneSAF SIMEXp in July 2020. The purpose of the SIMEXp was to determine the changes in mobility, lethality and survivability effects of a military combat unit employing different caliber indirect fire mortar munitions and capabilities against a formidable opposing force. The design for the experiment included a Base Case (BC) with current mortar capabilities, Alternate Case One (AC1) with an automated mortar capability, and Alternate Case Two (AC2) with an automated mortar plus a precision mortar round.

In contrast to a three-week HITL SIMEXp – that the MBL would normally have executed – the hybrid constructive SIMEXp was a four-week event. It included operational force MRPs, an opposing force commander, SIs, and an experiment control element. The first week focused on training and familiarizing the MRPs with technical systems and three experiment cases and culminated with pilot runs of each case. Weeks two and three comprised of HITL record runs and writing scenarios scripts for each case. Week four consisted of closed loop replication runs and data reduction. The SIMEXp replicated each case 38 times; eight times during the HITL phase and 30 times during the closed loop replication phase.

During execution of each HITL run, SIs wrote scripts for individual Soldiers, weapon systems, and platform capabilities. In addition to scripting, timings for specific OneSAF behaviors were segmented into time intervals. For example, after successful completion of the prescribed HITL runs for a case, the average simulation run time was 60 minutes. OneSAF behaviors and timings were then scripted for two intervals of 30-minutes each. The Battle Master then executed the first 30-minute simulation run interval at an accelerated time. This allowed MRPs with their assigned SIs to review the scripted scenario for discrepancies in both timing and in OneSAF behaviors. If corrections were required for the first 30-minute interval, then they were made, and the first 30-minute interval was re-executed and reviewed again by the MRPs and SIs until zero discrepancies were noted. The same process occurred for the second 30-minute interval. The Battle Master then executed the completed 60-minute simulation at an accelerated time with a final review and approval by the MRPs.

A key aspect of lethality is the range at which a friendly weapon can engage and damage/destroy the enemy (and vice versa). As the lead analyst for this experiment conducted a t-test (test differences between the means of two groups, or the difference between one group's mean and a standard value) to determine the statistical difference between cases (BC vs AC1, BC vs AC2, and AC1 vs AC2), they determined that no statistical difference existed between AC1 versus AC2 for the range factor in the HITL runs. However, when the analyst conducted the t-test for the closed loop runs, it was determined that AC2's increased range over AC1 was indeed statistically significant. The increased number of runs per case established the statistical significance required to make a sound analytic finding. Had the MBL not conducted the closed loop runs for this experiment, then inaccurate conclusions could have been made about this new mortar capability.

Ultimately, the MBL was able to determine that of the three experimental cases, AC2 outperformed the other two cases in terms of mobility, lethality, survivability, and capabilities of the employment of different mortar munitions. This conclusion was supported by both statistically significant quantitative results and further evidenced through human MRP qualitative input. This new hybrid methodology was so successful that the MBL has implemented it for all hypothesis test experiments since July 2020. This new technique has provided evidence for dozens of capability development decisions during the period since its inception.

Differential Experimentation using Hybrid Execution

A considerable benefit of hybrid execution is the ability to conduct differential experimentation, which is the ability to change a variable in the scenario and run it again while still being able to make comparisons to the original data set. For example, the MBL conducted a hybrid experiment in 2020 that compared an infantry platoon equipped with current rifles against two alternate cases with different new rifle capabilities. Following closed loop execution, the experiment lead determined that the second alternate case provided the greatest increase in both lethality and survivability. Subsequently, the experiment sponsor requested that the MBL consider a third new rifle capability. The MBL was able to inject this new rifle capability into the simulation, modify the script, and ultimately provide a third alternate case without building an entire new simulation experiment. This saved considerable time, money, and effort to both the MBL and to the experiment sponsor. This process will continue to allow the MBL to reuse previous experiments for the benefit of future considerations.

APPLICABILITY OF THE MBL HYBRID APPROACH TO OTHER ORGANIZATIONS

The MBL hybrid approach is a novel technique that potentially creates a fifth warfighting experiment method as described by Kass above. The five warfighting experiment methods would now be the analytic wargame, constructive experiments, human-in-the-loop experiments, hybrid (constructive/HITL) experiments, and field (or live) experiments. The MBL constructive hybrid approach described in this paper provides a means for other organizations conducting warfighting experiments to increase the validity of their constructive experiments, while simultaneously increasing the real-world applicability of those experiments.

ACKNOWLEDGEMENTS

The MBL hybrid approach has been developed over several years. U.S. Army Officers, Department of the Army (DA) Civilians, and the MBL's prime contractors from the Science Applications International Corporation (SAIC) have been instrumental in its development. It is a living process that continues to improve daily. The competence and professionalism of the MBL team allows this improvement to happen.

We would also like to thank our 'bird-dogs' Mr. Paul Butler and Dr. Tim Cooley for their feedback and for guiding us through this process. Finally, we want to thank Dr. Cindy Forgie, Major Sean Fraser, Cindy Dunn, Bill Miller, and Tom Yanoschik for their constructive criticism and feedback on this paper.

REFERENCES

Alberts, D. S., & Hayes, R. E. (2006), *Code of best practice: Campaigns of Experimentation*. Department of Defense (DoD) Command and Control Research Program (CCRP).

Alberts, D. S., & Hayes, R. E. (2002), *Code of best practice: Experimentation*. Department of Defense (DoD) Command and Control Research Program (CCRP).

Banks, J. (Ed.). (1998) *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John S. Wiley & Sons, Inc.

Brase, C.H. & Brase, C. P. (2012). *Understandable statistics* (10th ed.). Brooks/Cole CENGAGE Learning.

Chau, W. & Gill, A. & Griefer, D. (2017, December 3-8). 22nd International Congress on Modeling and Simulation, Hobart, Tasmania, Australia. http://mssanz.org.au/modsim2017.

Gallo, A. (2021, December 9). *A refresher on statistical significance.* Harvard Business Review. Retrieved April 1, 2022, from: https://hbr.org/2016/02/a-refresher-on-statistical-significance.

Kass, R. A. (2006), *The Logic of Warfighting Experiments.* Department of Defense (DoD) Command and Control Research Program (CCRP).

Modeling and Simulation Professional Certification Commission, http://Simprofessional.org, (accessed 11 May 2021).

Tolk, A. (2012), *Engineering principles of combat modeling and distributed simulation*. S. Wiley & Sons, Inc.

US Army Modeling and Simulation School. Simulation Operations Course (SOC) (2023). https://www.ms.army.mil/sp-div2/school/soc.html.

US Army Data Analysis Center (DAC), https://www.dac.ccdc.army.mil/, (last accessed 11 May 2021).

US Army Maneuver Capabilities Development Integration Directorate (MCDID). (2020). *Maneuver Battle Lab-Modeling and Simulation Branch face validation standard operating procedure (SOP).* Fort Moore, GA.

US Modeling and Simulation Coordination Office (MSCO). (2011). Glossary-Terms and Definitions, https://www.msco.mil/MSReferences/Glossary/TermsDfinitions (accessed 25 April 2022).