INSTITUTE FOR DEFENSE ANALYSES

DATAWorks 2023: CDV Method for Validating AJEM using FUSL Test Data

David K. Grimm, Project Leader

Lindsey D. Butler
John T. Haman
Thomas H. Johnson
Kerry N. Walzl

April 2023

Public release approved. Distribution is unlimited.

IDA Document NS D-33402

Log: H 2023-000052

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305

IDA

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task BD-09-1833(21) "Ground Combat," for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The IDA Technical Review Committee was chaired by Dr. V. Bram Lillard and consisted of Dr. Breeana Anderson, Dr. James Sheldon, and Dr. James Thome from the Operational Evaluation Division.

For more information:
Mr. David K. Grimm, Project Leader
dgrimm@ida.org • (703) 575-1431

Dr. V. Bram Lillard, Director, Operational Evaluation Division
vlillard@ida.org • (703) 845-2230

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-33402

DATAWorks 2023: CDV Method for Validating AJEM using FUSL Test Data

David K. Grimm, Project Leader

Lindsey D. Butler
John T. Haman
Thomas H. Johnson
Kerry N. Walzl

Executive Summary

As the Test and Evaluation community increasingly relies on Modeling and Simulation (M&S) to supplement live testing, M&S validation has become critical for ensuring credible weapon system evaluations. System-level evaluations of Armored Fighting Vehicles (AFV) rely on the Advanced Joint Effectiveness Model (AJEM) and Full- Up System Level (FUSL) testing to assess AFV vulnerability. This poster reviews one of the primary methods that analysts use to validate AJEM, called the Component Damage Vector (CDV) Method. The CDV method compares components that were damaged in FUSL testing to simulated representations of that damage from AJEM.

FUSL LF testing is an essential part of the vulnerability assessment of an AFV in which testers engage a variety of explosive mechanisms — including small and mid-caliber munitions, shaped charge jets, artillery munitions, and underbody mines — at a fully loaded, combat-ready AFV. Congress mandates this testing occur as part of the final phase of testing that leads to full-scale production of a new AFV.

AJEM is the primary M&S tool used in evaluations of AFV vulnerability. AJEM supports a FUSL test program in variety of ways by simulating a threat engagement's effects to the vehicle and the crew. AJEM helps guide shot-line planning by identifying areas of uncertainty and particular potential weaknesses of a vehicle design. AJEM can provide insights into conditions not tested. AJEM incorporates the Operational Requirement-Based Casualty Assessment (ORCA) model for assessing crew vulnerability.

Determining an M&S' credibility relies on verification, validation, and accreditation (VV&A). Verification is the process of determining that a model or simulation implementation and its associated output accurately represent the developer's conceptual description and specifications. Validation is the process of determining the degree to which a model or simulation and its associated outputs are an accurate representation of the real world from the perspective of the intended uses of the model. Accreditation is the official certification that a

model or simulation and its associated data are acceptable for use for a specific purpose.

AJEM currently has a robust VV&A program and has undergone extensive VV&A in the past. The Army has produced one or more VV&A reports on 27 separate LFT&E programs since 1998. This includes programs such as the Stryker M109 Family of Vehicles and the Joint Light Tactical Vehicle. The Army conducts validation analyses at every layer of AJEM using data from tests of various complexity, including the component, sub-system, and system levels. This poster focuses on validation using data from the highest level that occurs at the highest layer, the Full-Up System Level LFT.

The CDV Method uses component damage as the data to facilitate validation because it is information-rich, observable in FUSL testing, and directly comparable to output from AJEM. Component damage data exists in pairs, where each pair comprises two values: (1) a binary outcome, indicating whether the particular component was observed in FUSL testing as damaged or not, and (2) a probability, indicating AJEM's prediction that the particular component was damaged. A single FUSL event results in numerous damaged components, which can be realized as a vector of paired data.

The CDV Method was first applied to AJEM validation in 1998 as part of the Bradley Fighting Vehicle program, and most recently as part of the Armored Multi-Purpose Vehicle program in 2022. In the former application, the CDV Method led to the discovery that AJEM underpredicts the number of components damaged from certain threats. In the latter application, the CDV method underscored a known limitation of AJEM; it cannot predict component damage due to certain secondary effects, such as ricochet. The CDV Method identifies problems within AJEM so that follow-on efforts can improve the underlying algorithms.

The analysis involved with the CDV Method has varied throughout the last couple of decades. Recently, DOT&E, IDA, and the Army collaborated to improve the analysis. The purpose of this poster is to showcase the latest concepts from that collaboration.

Presented by **Tom Johnson**

with Dave Grimm, Lindsey Butler, Kerry Walzl, and John Haman

Institute for Defense Analyses

Problem Statement

As the Test and Evaluation community increasingly relies on Modeling and Simulation (M&S) to supplement live testing, M&S validation has become critical for ensuring credible weapon system evaluations. One of the most common M&S used for assessing Armored Fighting Vehicle (AFV) vulnerability is the Advanced Joint Effectiveness Model (AJEM). This poster addresses an important research topic: **what method should be used to validate AJEM using Full-Up System Level (FUSL) Live Fire (LF) test data?**

What is M&S validation?

Validation is the process of determining the degree to which a model or simulation and its associated outputs are an accurate representation of the real world from the perspective of the intended uses of the model.

What is FUSL LF Testing?

FUSL LF testing is an essential part of the vulnerability assessment of an Armored Fighting Vehicle (AFV) in which testers engage a variety of explosive mechanisms — including small and mid-caliber munitions, shaped charge jets, artillery munitions, and underbody mines — at a fully loaded, combat-ready AFV. Congress mandates this testing occur as part of the final phase of testing that leads to full-scale production of a new AFV.

What is AJEM?

The Advanced Joint Effectiveness Model (AJEM) is the primary M&S tool used in evaluations of AFV vulnerability. AJEM supports a FUSL test program in variety of ways by simulating a threat engagement's effects to the vehicle and the crew. AJEM helps guide shot-line planning by identifying areas of uncertainty and particular potential weaknesses of a vehicle design. AJEM can provide insights into conditions not tested. AJEM incorporates the Operational Requirement-Based Casualty Assessment (ORCA) model for assessing crew vulnerability.

The Army conducts validation analyses at every layer of AJEM using data from tests of various complexity, including the component, sub-system, and system levels. This poster focuses on validation using data from the highest level that occurs at the highest layer, the Full-Up System Level LFT.

Why is AJEM-FUSL-validation important?

FUSL LF testing provides the most operationally realistic vulnerability data prior to fielding the vehicle. Lower-level data (e.g., component or sub-system data) are more plentiful, but less realistic, and fail to capture the synergistic effects of an end-to-end engagement.

The Component Damage Vector (CDV) Method

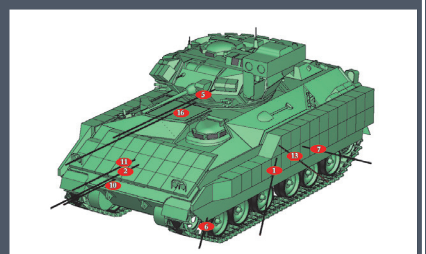
A statistically rigorous method for validating AJEM using FUSL LFT Data

The CDV Method involves the following steps:

Step 1) For each FUSL LF threat engagement (enumerated in picture below)…

a. **Engineers and the test team inspect the vehicle and record which components were damaged. The outcome is binary for each component (damaged/not damaged).**

b. **Modelers use AJEM to predict the probability that each component was damaged. The outcome is a probability (continuous between 0 and 1) for each component.**



This picture shows the engagement geometry of each FUSL event

Step 2) Format FUSL/AJEM data as follows to prepare for analysis.

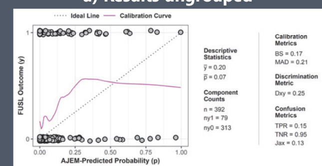
FUSL Event	Component #	p (AJEM)	y (FUSL)	Threat Type	Threat Severity	Vehicle Type
1	1	.74	1	Direct	Threshold	Command
1	2	.25	1	Direct	Threshold	Command
⋮	⋮	.27	0	Direct	Threshold	Command
1	N	.01	0	Direct	Threshold	Command
2	1	0	0	Indirect	Objective	General
2	2	.95	1	Indirect	Objective	General
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	N	0	0	Indirect	Objective	General
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	1	.05	0	Direct	Objective	Medical
16	2	.01	0	Direct	Objective	Medical
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	N	.25	0	Direct	Objective	Medical

These data were randomly generated for illustration purposes

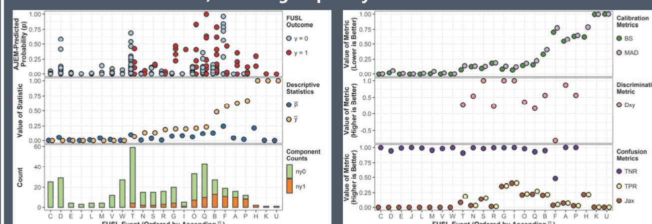
Step 3) Analyze data using PPV statistics (see right panel for details)

Step 4) CDV results:

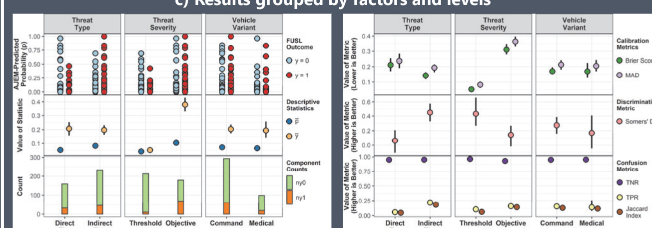
a) Results ungrouped



b) Results grouped by FUSL event



c) Results grouped by factors and levels



These results were randomly generated for illustration purposes

Predicted Probabilities Validation (PPV)

The defining feature of the CDV Method is the data being compared – data pairs of predicted probabilities and binary test outcomes. A comparison of this type of data is not uncommon in Machine Learning and Statistics.

For instance, Predicted Probabilities Validation (PPV) provides an established framework of analyses for this data. PPV is commonly used to validate Logistic Regression Models, but is also often applied to more complex models, such as AJEM.

PPV Statistics

PPV provides numerous statistics and metrics for assessing M&S predictions, which are often organized by **Calibration**, **Discrimination**, and **Confusion**.

Calibration addresses the degree of which AJEM's predicted probabilities are too high or too low compared to the binary test outcomes. This report includes two metrics for assessing Calibration: the Brier Score (BS) and Mean Absolute Difference (MAD).

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \quad , \qquad MAD = \frac{1}{N}\sum_{i=1}^{N}|p_i - y_i|$$

Discrimination describes AJEM's ability to discriminate or differentiate between damaged and undamaged components. Somers' D_{xy} Metric conveys discrimination on a continuous scale between zero and one. Higher values indicate good discrimination, such as when AJEM produces high predicted probabilities for components that were damaged, and low predicted probabilities for components that were undamaged. $\bar{R}(p \mid y = 1)$ is the mean rank order of predicted probabilities for components that were damaged, n_{y1} is the number of damaged components, and N is the total number of components.

$$D_{xy} = 2\left[\frac{\bar{R}(p \mid y = 1) - \frac{n_{y1}+1}{2}}{N - n_{y1}} - \frac{1}{2}\right]$$

A **Confusion** table provides additional insight about AJEM predictions, and is often used to describe the performance of a classification model. The cells of the table report true positives (TP) and false positives (FP), and false negatives (FN) and true negatives (TN). These values are then used to compute confusion metrics: the True Negative Rate (TNR), the True Positive Rate (TPR), and the Jaccard Index (JAX).

$$TNR = \frac{TN}{TN + FP} \quad , \qquad TPR = \frac{TP}{TP + FN} \quad , \qquad JAX = \frac{TP}{TP + FN + FP}$$

For more information about PPV see:

Harrell, Frank E. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Vol. 608. New York: Springer, 2001.

Haman, John T.; Johnson, Thomas H.; Grimm, David K.; Walzl, Kerry N.; and Butler, Lindsey D., 2022. Predicted Probabilities Validation. IDA Document D-33156.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
01-04-2023

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
DATAWorks 2023: CDV Method for Validating AJEM using FUSL Test Data

5a. CONTRACT NUMBER
HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
BD-09-1833(21)

5e. TASK NUMBER
183321

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Dr. Thomas H. Johnson; Dr. John T. Haman; Dr. Kerry N. Walzl; and Dr. Lindsey D. Butler

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
730 East Glebe Rd.
Alexandria, Virginia 22305-3086

8. PERFORMING ORGANIZATION REPORT NUMBER
D-33402

H 2023-000052

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington D.C. 20301

10. SPONSOR/MONITOR'S ACRONYM(S)
DOT&E

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
This OED Draft has not been approved by the sponsor for distribution and release. Reproduction or use of this material is not authorized without prior permission from the responsible IDA Division Director.

13. SUPPLEMENTARY NOTES
Project Leader: Mr. David K. Grimm

14. ABSTRACT
As the Test and Evaluation community increasingly relies on Modeling and Simulation (M&S) to supplement live testing, M&S validation has become critical for ensuring credible weapon system evaluations. System-level evaluations of Armored Fighting Vehicles (AFV) rely on the Advanced Joint Effectiveness Model (AJEM) and Full-Up System Level (FUSL) testing to assess AFV vulnerability. This report reviews one of the primary methods that analysts use to validate AJEM, called the Component Damage Vector (CDV) Method. The CDV method compares components that were damaged in FUSL testing to simulated representations of that damage from AJEM.

15. SUBJECT TERMS
Modeling and Simulation (M&S) Validation; Modeling and Simulation (M&S); Statistics; Live Fire Test and Evaluation (LFT&E)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Mr. David K. Grimm
Unclassified	Unclassified	Unclassified	SAR	7	19b. TELEPHONE NUMBER *(Include area code)* (703)575-1431