INSTITUTE FOR DEFENSE ANALYSES

Aircraft Survivability Journal: Component Data Vector Methodology in Support of FUSL-AJEM Validation

David K. Grimm, Project Leader

Thomas H. Johnson
Lindsey D. Butler
Craig Andres
Julia Ivancik
Russ Dibelka

OED Draft

March 2024

IDA Product ID - 3002075

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

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-9-1833(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. John T. Haman 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 Product ID - 3002075

Aircraft Survivability Journal: Component Data Vector Methodology in Support of FUSL-AJEM Validation

David K. Grimm, Project Leader

Thomas Johnson
Lindsey Butler
Craig Andres
Julia Ivancik
Russ Dibelka

Executive Summary

This document proposes improvements to a modeling and simulation (M&S) validation technique called the Component Damage Vector (CDV) method. The CDV method is the proposed current state of the art for validating an intended use of the Advanced Joint Effectiveness Model (AJEM) using full-up system level (FUSL) test data. The CDV method is particularly important for evaluations of armored fighting vehicle (AFV) vulnerability because AJEM is the primary M&S tool used to supplement U.S. Code Title 10-mandated FUSL live fire to evaluate AFV vulnerability, and FUSL testing provides the most realistic vulnerability data prior to fielding the vehicle.

AJEM is a stochastic vulnerability/lethality (V/L) software suite that simulates the effects of fire munitions on targets. AJEM provides the Department of Defense with a standard platform for assessing the impact of munitions and the survivability of personnel, aircraft, missiles, and ground systems. AJEM is integral in all system acquisition phases, from research to production.

The CDV method facilitates a comparison between AJEM predictions and FUSL test observations of component damage. A ground vehicle has hundreds of components that are critical for the vehicle's functionality that may be damaged during a FUSL event, such as drivetrain and engine components, electronics and wiring, munitions and weapon systems, and many more. The comparison serves as a validation of AJEM predictions.

Our proposed improvements to the CDV method leverage existing statistical theory and exploratory data analysis to provide high-level validation results that span one or more FUSL test series. This approach stands in contrast to past implementations of the CDV method that focused on highly detailed, low-level results. The proposed improvements have three purposes: (1) provide a concise yet detailed validation assessment for a given FUSL test series; (2) discover high-level trends that cut across an entire FUSL test series, such as whether AJEM performed better for one type of threat versus another; and (3) compare the validation results of multiple FUSL test series. These are the most substantial proposed improvements to AJEM-FUSL validation within the last two decades.

As part of the improvements, we propose 11 metrics for assessing the discrepancies between AJEM predictions and FUSL test observations. The metrics originate from statistical literature and include Brier Score, Somers' Index, Spiegelhalter's statistic, and the Jaccard Index. The metrics are designed to compare binary observations (FUSL observations) to predicted probabilities (AJEM predictions) and are typically used to assess

the goodness of fit of logistic regression models. We compute each metric for each FUSL event in a test series. Given that a test series comprises numerous events, we display the distribution of each metric as a boxplot to visualize the metric's variability across all events in the test series. In this document we use notional data to illustrate five examples that showcase the benefits of these metrics.

In each example, we present the CDV results using a specific figure layout that accommodates the use of factors and levels to reveal trends in the metrics. For instance, in one notional example we use a Threat Type factor to reveal that AJEM provides better predictions for Indirect Fire threats, relative to Direct Fire threats, as shown in Figure 1. The choice of factor will likely be different for each program of record.

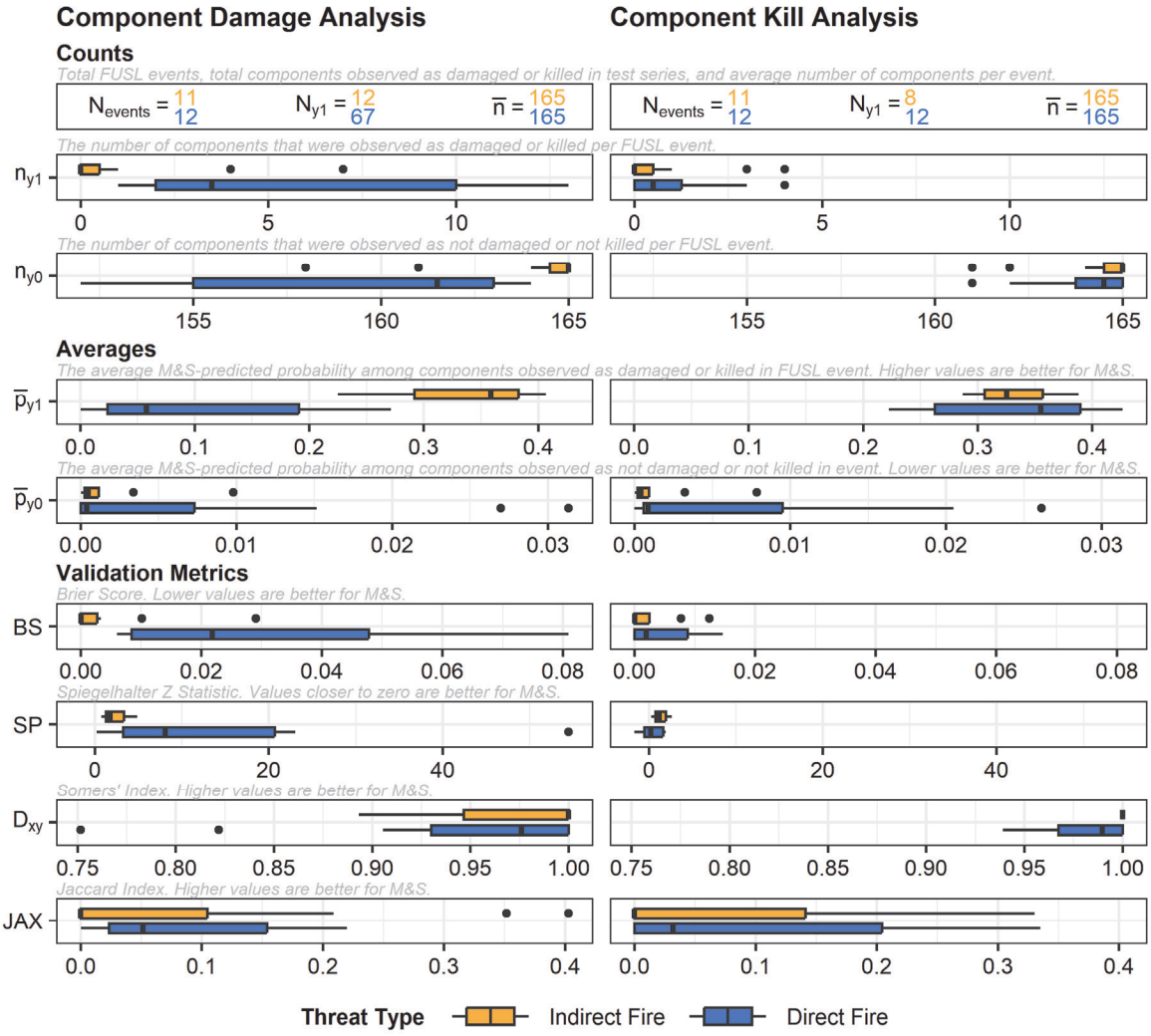

Figure 1. Example of CDV Results

Figure 1's averages suggest AJEM's superior performance with Indirect Fire threats, evident in higher $\bar{p}_{y1}$ for component damage and lower $\bar{p}_{y0}$ for damage and kill. These averages contribute to a better Brier Score, particularly noticeable in the non-overlapping interquartile ranges for component damage between Direct Fire and Indirect Fire. Because of small sample sizes and variation in CDV data, not all metrics align. Each metric offers a unique perspective on the AJEM predictions versus FUSL test results discrepancy. Notably, AJEM's better predictions for Indirect Fire prompt further investigation into its weaker Direct Fire predictions.

Future work could incorporate Loss of Function (LoF) data into the CDV method. A vulnerability assessment typically uses fault trees, deactivation diagrams, and subject matter expertise to extrapolate the effect that each killed component has on the core combat utilities of the AFV such as mobility, firepower, and communications. This extrapolated effect, called LoF, is quantified on a subjective scale that ranges between zero and one. The CDV method could use LoF data to weigh the importance of killed components in CDV analysis, which would cause killed components with high LoF values to have a greater influence on the validation metrics.

In summary, we propose improvements to the CDV method, including new metrics and a new figure layout. The objective of the CDV method is to compare AJEM output with FUSL test results to identify problems within AJEM that can be fixed with future investment. The overarching goal is to improve AJEM, which in turn improves AFV vulnerability evaluations, which in turn improves the survivability and combat effectiveness of the warfighter.

U.S. Marine Corps Photo

Component Data Vector Methodology in Support of FUSL-AJEM Validation

by Thomas Johnson, Craig Andres, Russell Dibelka,
Lindsey Butler, Dave Grimm, and Julie Ivancik

As the U.S. military's test and evaluation community increasingly relies on modeling and simulation (M&S) to complement the live testing of a wide range of ground, air, and sea systems, M&S validation has likewise become increasingly critical for ensuring system evaluations are credible. For the system-level evaluation of armored fighting vehicles (AFVs), the Army's Advanced Joint Effectiveness Model (AJEM) is used in combination with full-up system-level (FUSL) live fire (LF) testing to evaluate vehicle vulnerability. This article proposes improvements to a method—the Component Data Vector (CDV) method—that is used to validate AJEM. The method compares damage to components (e.g., driveshaft, engine, suspension, wiring, driver displays, weapons) sustained during FUSL testing to the predictions of that damage in AJEM. Though the vehicle application used herein is an AFV, the CDV methodology could also be extended to analogous vulnerability assessment tools used by the Navy and Air Force for other ground, air, and sea systems.

THE CDV METHOD

FUSL LF testing is an essential part of the vulnerability assessment of an AFV, wherein testers "attack" a fully loaded, combat-ready AFV with a wide variety of explosive mechanisms, including small- and mid-caliber munitions, shaped charge jets, artillery munitions, and underbody mines. To complement FUSL testing such as this, the Army, Navy, and Air Force have individually developed and maintained their own models— AJEM, the Advanced Survivability Assessment Program (ASAP), and the Computation of Vulnerable Area Tool (COVART), respectively— for assessing system-level vulnerability and lethality (V/L) for various combat systems. Despite the separate development of these models, they possess many similarities. They share libraries and modules, have similar inputs and outputs, and are composed of a collection of computationally inexpensive empirical models that have been calibrated to decades worth of test data and subject-matter expertise.

To determine the credibility of these and other models, a formal process of verification, validation, and accreditation (VV&A) is used. Validation, in particular, 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 [1]. The CDV method is a type of validation analysis.

AJEM and its predecessors have undergone extensive VV&A in the past. The Army has produced VV&A reports on 27 separate live fire test and evaluation (LFT&E) programs since 1998 [2]. These include major programs such as the Stryker, the 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 varying complexity, including tests conducted at the component, subsystem, and system levels. The CDV method focuses on validation using data from the highest level of testing that corresponds to the highest layer of AJEM, the FUSL LF test.

The method relies on component kill data because these data are rich, observable in FUSL testing, and directly comparable to AJEM output. Each component on the AFV (e.g., driveshaft, engine, suspension, wiring, driver displays, weapons, etc.) has an associated pair of data, with one value corresponding to a binary outcome indicating whether the component was observed in the FUSL event as killed or not and the second value corresponding to AJEM's prediction of the probability that the component was killed. Given that an AFV has many components, a FUSL event results in a *vector* of pairs, with each pair in the vector corresponding to a single component.

The CDV method was first applied to AJEM validation in 1998 as part of the Bradley Fighting Vehicle program and was applied most recently (in 2022) as part of the Armored Multi-Purpose Vehicle program. 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, the CDV method underscored another known limitation of AJEM— an inability to predict component damage caused by certain secondary threat effects, such as ricochet. The CDV method thus helps identify, quantify, and prioritize problems

within AJEM so that follow-on development efforts can improve the underlying algorithms.

In the past, the method has used a variety of different analysis techniques and result presentations. Many techniques focused on low-level results by providing extensive detail on each component damaged in each FUSL event [3–5]. However, the unique aspect of the improvements proposed herein, which complement past efforts, is that they focus on high-level results. This focus addresses two main goals: (1) providing a concise validation assessment for an entire FUSL test series, and (2) revealing overarching trends across the test series.

CDV DATA

As with system-level testing in other fields, FUSL test data are typically in short supply because of the high cost of producing such data. The most common data that FUSL testers collect pertain to the state of the crew, exterior armor, and critical components. The CDV method focuses exclusively on critical component data.

A critical component is defined as any component that, when killed, results in a loss of function (LoF) for any of the relevant metrics (mobility, firepower, etc.) for the vehicle. A specialized group within the integrated product team with the responsibility of identifying the critical components on the vehicle is staffed by soldiers, maintainers, and engineers who have specific knowledge about the effects of damage on vehicle operation and experience repairing damaged vehicles. This group identifies critical components by considering important design features, the layout of the reliability fault tree, corporate experience with similar test programs, battlefield observations, and the importance of a given component relative to various missions.

A straightforward approach to quantifying the effects to the vehicle caused by a threat engagement is to assess the kill status of each critical component. After each FUSL event, the test team inspects the critical components on the vehicle, attempts to power up the vehicle to diagnose the components' residual functionality, and holds a damage assessment meeting to review notes and video footage. The team's evaluation culminates with an assessment of each critical component on a binary scale: killed or not killed. A critical component is defined as killed if the component was physically harmed and suffered a loss of functionality, necessitating repair or replacement to restore functionality prior to the next FUSL event.

In AJEM, component kill is defined as a probability. AJEM is a stochastic simulation that incorporates numerous sources of uncertainty that produce a nondeterministic output. Modelers typically conduct 1,000 AJEM iterations per shot, and from iteration to iteration, the components that AJEM predicts to be killed vary. In each iteration, the component kill data are binary outcomes, but they become probabilities when averaged across the 1,000 iterations, yielding a single probability of kill for each component on the vehicle.

Given these definitions, we denote component kill as follows. Let y_i denote the component kill outcome observed in FUSL testing for the ith component, where y_i equals 1 or 0 if it was killed or not killed, respectively; $i = 1,2, \ldots, n$; and n is the number of critical components on the AFV. Additionally, let p_i correspond to AJEM's predicted probability that the ith component was killed. Given a FUSL event, one may collect the following vector of data: $(y_1, p_1), (y_2, p_2), \ldots, (y_n, p_n)$. CDV analysis synthesizes the vector of paired data into metrics that summarize the discrepancy between the FUSL observations and AJEM predictions.

To further illustrate the format of CDV, consider the following simple example. Figure 1 shows an AFV with five critical components. The left portion of the figure displays AJEM predictions, while the right portion shows FUSL test observations. Results from this simple example are then tabulated in Table 1.

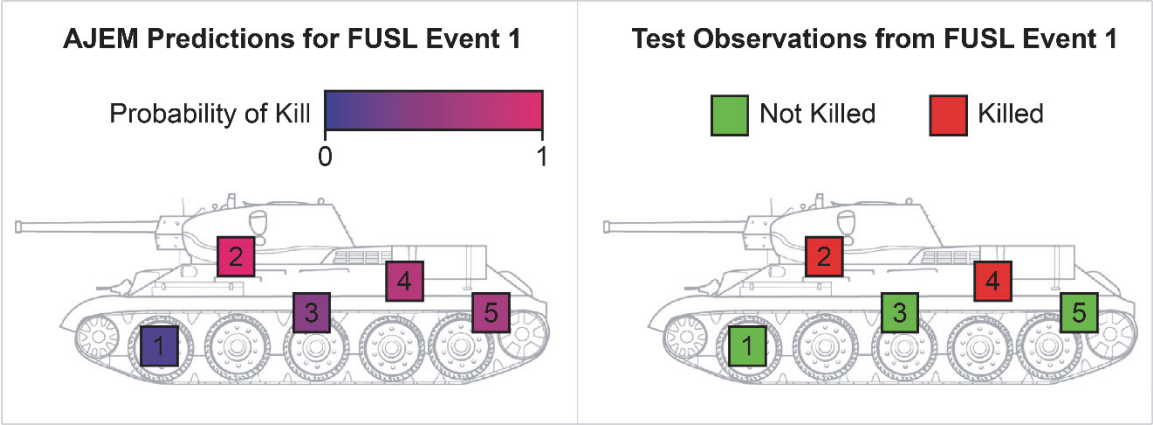

Figure 1. Example of Component Kill Data

Table 1. Example of CDV Data for a Single FUSL Event

Identification		Component Data Vector	
Component	FUSL Event	p	y
1	1	.05	0
2	1	.94	1
3	1	.39	0
4	1	.65	1
5	1	.53	0

In practice, however, the AFV will have many more critical components, and the CDV data will typically include data from numerous FUSL events as part of a comprehensive test series. This results in the generalized data format appear in Table 2. Additional columns, not shown in the table, may also provide information pertaining to factors and their levels, component descriptions, test notes, and AJEM settings. Data of this format serve as the input to the forthcoming CDV analysis.

Table 2. Example of CDV Data for a FUSL Test Series

Identification		Component Data Vector	
Component	FUSL Event	p	y
1	1	.05	0
2	1	.47	0
⋮	⋮	⋮	⋮
n	1	.65	1
1	2	.88	1
2	2	.12	0
⋮	⋮	⋮	⋮
n	2	.44	1
⋮	⋮	⋮	⋮
1	3	.87	1
2	3	.99	1
⋮	⋮	⋮	⋮
n	N_{events}	.95	1

PROPOSED IMPROVEMENTS TO CDV ANALYSIS

CDV analysis quantifies the discrepancy between AJEM-predicted and FUSL-observed component damage. A defining feature of this analysis is the underlying data being compared—a vector of data pairs of predicted probabilities and binary test outcomes. A comparison involving this type of data is not uncommon in machine learning and statistics.

For instance, an appropriate framework for conducting this comparison is called Predicted Probabilities Validation (PPV) [6, 7]. PPV is most commonly used to validate logistic regression models, but it applies to more complex models, too. PPV synthesizes a vector of paired data into one or more validation metrics that describe the discrepancy between the predicted probabilities and binary test outcomes.

We recommend the computation of numerous metrics and organize them into three groups: Counts, Averages, and Validation metrics. The Counts, which appear in Table 3, address details related to FUSL test results, the scope of FUSL testing, and the size of the vector of paired component data. The Averages, which appear in Table 4, summarize the conditional distribution of AJEM predictions (conditioned on whether the components were observed as damaged or not). The Validation metrics, which appear in Table 5, originate from PPV literature and summarize the discrepancy between AJEM predictions and FUSL test outcomes.

Table 3. Counts

N_{events}	N_{events} is the total number of FUSL events in the FUSL test series.
N_{y1}	N_{y1} is the total number of components that were observed as killed throughout the FUSL test series. It is computed once per FUSL test series.
n	n is the total number of components that were considered in the analysis for the FUSL event. It is computed once per FUSL event.
$\bar{n}$	$\bar{n}$ is the mean number of components considered in the analysis per FUSL event. It is computed once per FUSL test series.

N_{events}	N_{events} is the total number of FUSL events in the FUSL test series.
n_{y1}	n_{y1} is the number of components that were observed as killed in a FUSL event. It is computed once per FUSL event.
n_{y0}	n_{y0} is the number of components that were observed as not killed in a FUSL event. It is computed once per FUSL event.

Table 4. Averages

$\bar{p}_{y1}$	$\bar{p}_{y1}$ is the average AJEM-predicted probability of component kill among components that were observed as killed in a FUSL event. It is computed once per FUSL event.
$\bar{p}_{y0}$	$\bar{p}_{y0}$ is the average AJEM-predicted probability of component kill among components that were observed as not killed in a FUSL event. It is computed once per FUSL event.

Table 5. Validation Metrics

BS	BS is the Brier Score [8]. This score is the mean squared error of the AJEM-predicted probabilities relative to the FUSL binary outcomes. It is computed once per FUSL event as $$\text{BS} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \quad .$$ BS ranges from 0 (i.e., no error) to 1 (i.e., maximum error). Smaller values of BS generally indicate a better agreement between AJEM predictions and FUSL outcomes. A BS less than .25 indicates that the AJEM-predicted probabilities are better than if all predicted probabilities were equal to .5 (i.e., a coin flip).
SP	SP is the Spiegelhalter Z Statistic [9]. It is derived from a decomposition of the Brier Score. It is commonly associated with the hypothesis test, $H_0: E(y_i) = p_i$. This hypothesis states that the expected FUSL outcomes are equal to the predicted probabilities, which is often referred to in short as "perfect calibration." SP ranges from – ∞ to + ∞. Extreme values of SP tend to lead to small p-values and rejection of the null hypothesis. Thus, values of SP closer to zero indicate better AJEM predictions. $$\text{SP} = \frac{\sum_{i=1}^{n}(y_i - p_i)(1-2p_i)}{\sqrt{\sum_{i=1}^{n}(1-2p_i)^2(p_i)(1-p_i)}}.$$
D_{xy}	D_{xy} is Somers' metric [10]. This metric describes the correlation between the AJEM predictions and FUSL outcomes. D_{xy} ranges from -1 (test data exhibits bad agreement with M&S) to 1 (test data exhibits good agreement with M&S). Higher values indicate strong correlation, such as when AJEM produces high predicted probabilities for components that were observed in FUSL as killed and low predicted probabilities for components that were observed in FUSL as unkilled. It is computed once per FUSL event as $$D_{xy} = 2\left[\frac{\bar{R}(p \mid y = 1) - \frac{n_{y1}+1}{2}}{n - n_{y1}} - \frac{1}{2}\right],$$ where $\bar{R}(p \mid y = 1)$ is the mean rank order of predicted probabilities for components that were killed, and n is the number of components considered in a given FUSL event.
JAX	JAX is the Jaccard Index [11]. This metric is often derived from a two-by-two confusion table. The four cells of a confusion table indicate the number of AJEM iterations in a given FUSL event that were True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), where $$\text{TP} = \sum_{i=1}^{n} p_i y_i \ , \qquad \text{FN} = \sum_{i=1}^{n} 1 - p_i y_i \ , \qquad \text{FP} = \sum_{i=1}^{n} p_i(1 - y_i) \ ,$$ and

	$$\text{JAX} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} = \frac{\text{TP}}{1 + \text{FP}} \ .$$ JAX is a measure of the rate at which AJEM correctly predicts killed components. It ranges from 0 to 1. Values closer to 1 indicate better AJEM predictions of component kill. A benefit of this metric is that it is insensitive to True Negatives. That is, it provides AJEM no credit for correctly predicting unkilled components. This is a desirable characteristic because it is obvious in many cases that components will be unkilled when such components are far away from the location of impact on the vehicle.

EXAMPLES

To illustrate CDV analysis, we present two different examples involving notional data pertaining to a generic AFV. The first example illustrates a basic application of CDV analysis to a notional FUSL test series. The second example augments the first by using factors and levels to reveal high-level trends in the notional data.

The analytical approach in these examples aligns with exploratory data analysis. Each metric is computed once for each FUSL event in the test series. These metrics are then presented as boxplots, depicting the spread in the metrics across FUSL events. The left and right hinge of the boxplot corresponds to the 25th and 75th percentile of the computed metrics; the line in the middle of the box is the median; and the whiskers extend to the farthest computed metrics but no farther than 1.5 times the interquartile range. Metrics beyond the whiskers are considered outliers and are plotted as dots. Given the small sample sizes and rare-event nature of CDV data, we did not pursue parametric modeling or statistical inference. However, the exploratory data analysis presented in these examples may provide the insight to motivate such pursuits in the future.

Example 1

Here, we provide CDV results pertaining to a notional FUSL test series. The Counts that appear in Figure 2 indicate that the FUSL test series comprises 23 FUSL events; the total number of components that were observed as killed throughout the test series was 79, and the average length of the component kill vector (the number of critical components on the AFV) per FUSL event was 165.

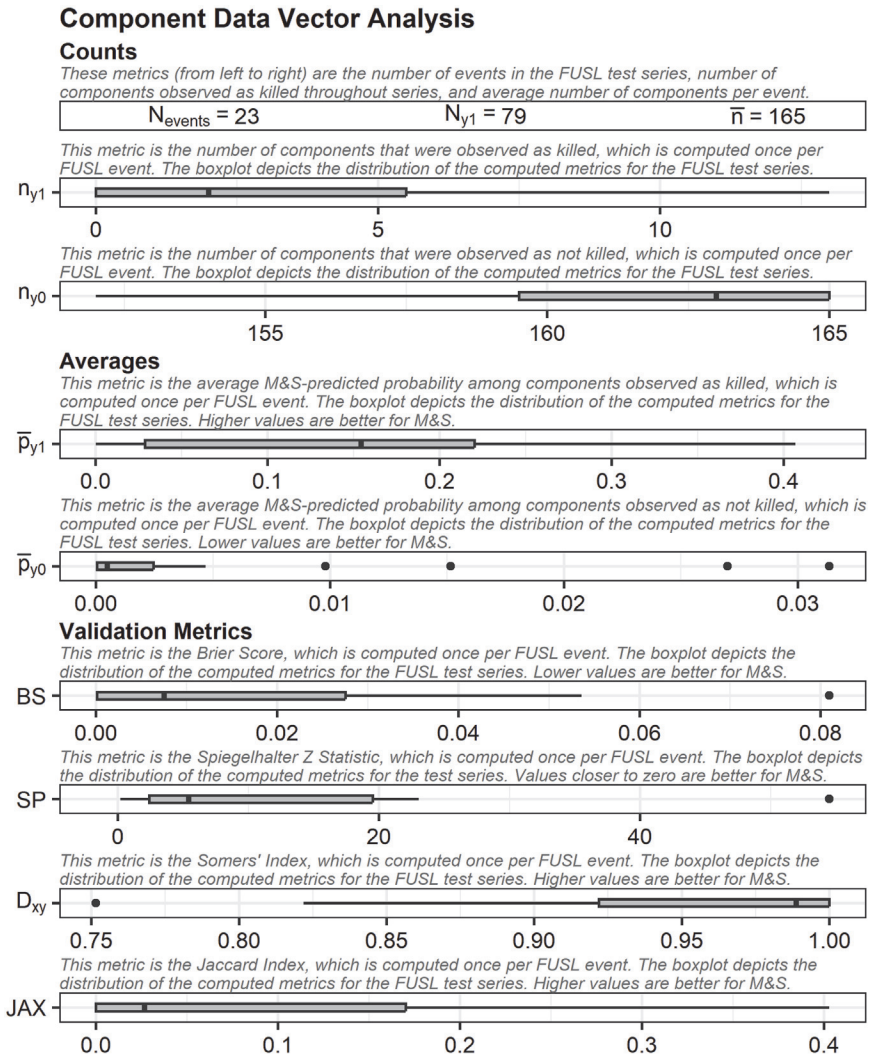

Figure 2. CDV Results for Example 1

The Counts also show boxplots of the number of components that were observed as damaged or killed per FUSL event. The maximum number of components observed as damaged or killed in a FUSL event was 16. Likewise, the minimum number of components observed as not killed in a FUSL event was 149.

The Averages in Figure 2 indicate that the median of the average AJEM-predicted probability of component kill per FUSL event, among components that were observed as killed (denoted as $\bar{p}_{y1}$), was .16. An ideal outcome for AJEM is a value of $\bar{p}_{y1}$ close to 1. Meanwhile, the median of the average AJEM-predicted probability of component kill, for components that were observed as not killed (denoted as $\bar{p}_{y0}$), was approximately .001. An ideal outcome for AJEM is a value of $\bar{p}_{y0}$ near 0.

The Validation metrics in Figure 2 indicate that the median of Brier Scores was .004. The median of the Somers' index was .990. By traditional rules of thumb, these results indicate good agreement between AJEM and FUSL outcomes.

Example 2

Assessing the metrics in an absolute sense, as in Example 1, can leave much to be desired, given that threshold requirements for these metrics are not set in practice. This issue can be alleviated to some degree by assessing the metrics in a relative sense, by grouping the computed metrics using factors and levels, as is common in Design of Experiments [12].

In Example 2, the metrics computed per FUSL event are grouped by the type of threat that was used in each event of the notional FUSL test series. Here, we assume that all 23 FUSL events involved a Direct Fire or Indirect Fire weapon engagement. Threat Type is referred to as an independent variable or *factor*, which has two categorical *levels* (Indirect Fire and Direct Fire). The purpose of this grouping strategy is to discover whether AJEM predictions were better for one level compared to the other.

The Counts in Figure 3 show that, among the 23 FUSL events in the FUSL test series, 11 events involved an Indirect Fire threat, while 12 events involved a Direct Fire threat. Figure 3 also shows that Direct Fire caused more damaged and killed components than Indirect Fire.

The results appear to indicate that AJEM performed better for the Indirect Fire threats. This is evident in $\bar{p}_{y1}$, which is higher for Indirect Fire, suggesting AJEM was better at predicting components that were observed as killed in events involving Indirect Fire threats. It is also evident in the Brier Score and Spiegelhalter statistic, which are lower for Indirect Fire with nonoverlapping interquartile ranges. In practice, this result could motivate follow-on work to improve AJEM predictions relative to Direct Fire threats.

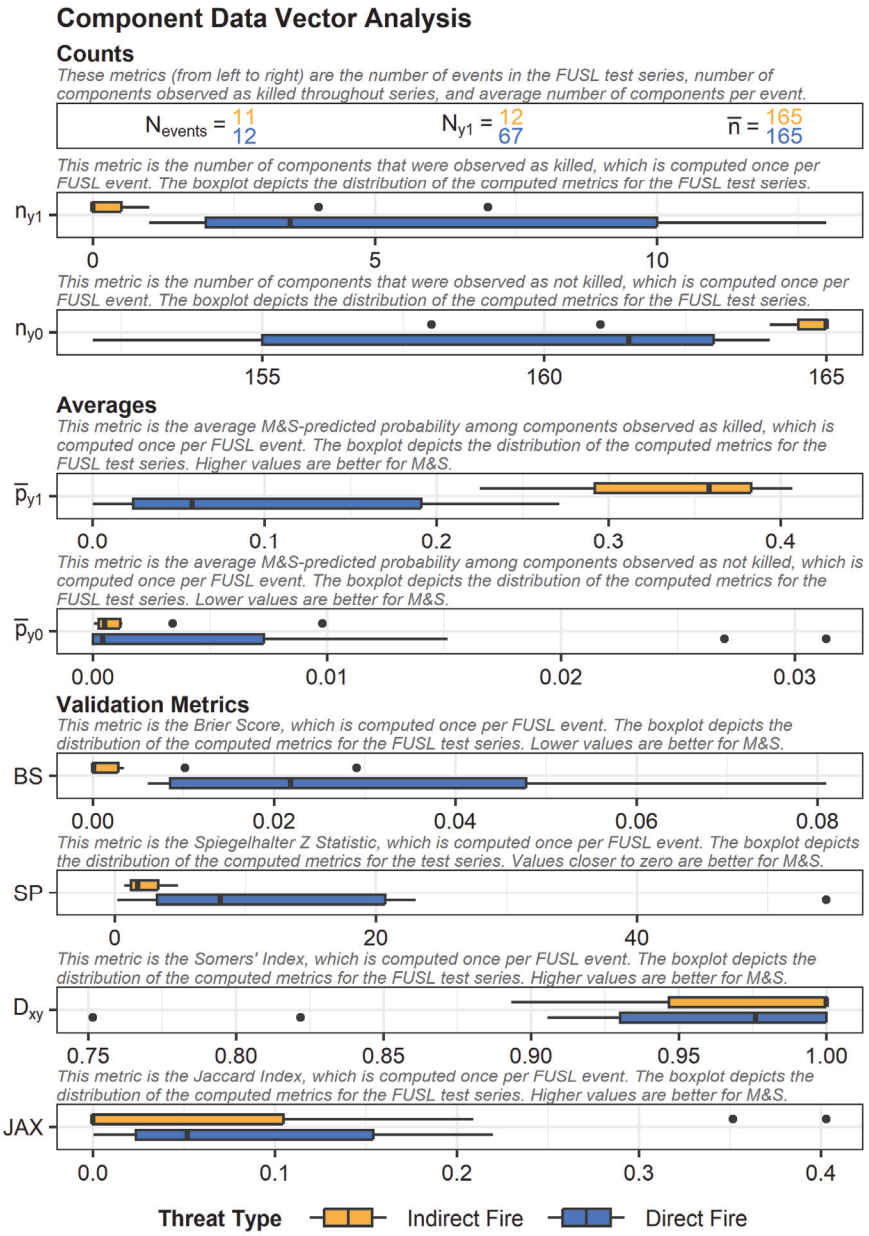

Figure 3. CDV Results for Example 2

CONCLUSIONS

As shown, improving the CDV method can improve vulnerability evaluation M&S, which in turn can improve the survivability and combat effectiveness of the warfighter and his/her weapon systems. The application of the CDV method presented herein leveraged statistical theory and exploratory data analysis to augment existing AJEM-FUSL validation techniques by focusing on high-level results. Example 1 illustrated a detailed yet concise validation assessment for a FUSL test series, while Example 2 illustrated trends in results across the threat type factor.

Future applications of this work could use other factors instead of threat type. For instance, FUSL events could be grouped in many ways, including by threat mechanism, vehicle variant type, or engagement geometry. In addition, as mentioned previously, while the focus of this work was on AJEM and AFVs, future endeavors could extend the CDV methodology to analogous vulnerability assessment tools, such as the Navy's ASAP and Air Force's COVART.

References

[1] Headquarters, U.S. Department of the Army. "Verification, Validation, and Accreditation of Army Models and Simulations." Department of the Army Pamphlet 5-11, 30 September 1999.

[2] Dunn, J. "Baseline Accreditation Report for the Advanced Joint Effectiveness Model Version 2.54." July 2022.

[3] Baker, W., R. Saucier, T. Muehl, and R. Grote. "Comparison of MUVES-SQuASH with Bradley Fighting Vehicle Live-Fire Test Results." ARL-TR-1846, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, November 1998.

[4] Deitz, R., R. Saucier, and W. Baker. "Developments in Modeling Ballistic Live-Fire Events." Paper presented at the 16th International Symposium on Ballistics, San Francisco, CA, 23–28 September 1996.

[5] Tonnessen, L., A. Fries, L. Starkley, and A. Stein. "Live Fire Testing in the Evaluation of the Vulnerability of Armored Vehicles and Other Exposed Land-Based Systems." Appendix A, IDA Paper P-2205.

[6] Harrell, F. Jr., K. Lee, and D. Mark. "Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors." *Statistics in Medicine* 15.4, pp. 361–387, 1996.

[7] Harrell, F. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Vol. 3, New York: Springer, 2015.

[8] Brier, G. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*, vol. 78, issue 1, 1950.

[9] Spiegelhalter, D. "Probabilistic Prediction in Patient Management and Clinical Trials." *Statistics in Medicine,* vol. 5, issue 5, pp. 421–433, 1986.

[10] Somers, R. "A New Asymmetric Measure of Association for Ordinal Variables." *American Sociological Review*, pp. 799–811, 1962.

[11] Jaccard, P. "The Distribution of the Flora in the Alpine Zone." *The New Phytologist, vol. XI, no.*2: pp. 37–50, February 1912.

[12] Johnson, T., J. Haman, H. Wojton, and M. Couch. "Design of Experiments (DOE) in Survivability Testing." *Aircraft Survivability*, Summer 2019.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX-03-2024	OED Draft	

4. TITLE AND SUBTITLE

Aircraft Survivability Journal: Component Data Vector Methodology in Support of FUSL-AJEM Validation

5a. CONTRACT NUMBER

HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Thomas H "Tom" Johnson (OED)
Lindsey D Butler (OED)
Craig Andres, Army, DEVCOM DAC
Julie Invancik, DOT&E
Russ Dibelka, Army, DEVCOM DAC

5d. PROJECT NUMBER

BD-9-1833(21)

5e. TASK NUMBER

183321

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
730 East Glebe Road
Alexandria, VA. 22305-3086

8. PERFORMING ORGANIZATION REPORT NUMBER

3002075

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)

DOT&E

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

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

Project Leader: David K. Grimm

14. ABSTRACT

M&S validation is 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 and improves upon one of the primary methods that analysts use to validate AJEM, called the Component Damage Vector (CDV) Method. The CDV Method compares vehicle components that were damaged in FUSL testing to simulated representations of that damage from AJEM. In the past, the CDV Method has employed a variety of different analysis techniques and results presentations. Many focused on low-level validation results, detailing each component that was damaged in each FUSL event. The unique contribution of this report, which complements past CDV efforts, is that it focuses on high-level results. This has three purposes: (1) to provide a pithy, yet detailed, validation assessment for a given FUSL test series, (2) to discover high-level trends that cut across an entire FUSL test series, such as whether AJEM performed better for one type of threat versus

15. SUBJECT TERMS

Armored Multi-Purpose Vehicle (AMPV); Full-Up System Level (FUSL); Ground Combat Vehicle (GCV); Modeling and Simulation (M&S) Validation; Statistical Analysis; Verification, Validation, and Accreditation (VV&A); vulnerability assessment

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