



CCDC DAC-TR-2020-044  
July 2020

# **Uncertainty Quantification and Sensitivity Analysis Methodology for the Advanced Joint Effectiveness Model (AJEM)**

by Craig D. Andres

### **DESTRUCTION NOTICE**

Destroy this report when it is no longer needed. Do not return it to the originator.

### **DISCLAIMER**

The findings in this report are not to be construed as an official Department of the Army position unless so specified by other official documentation.

### **WARNING**

Information and data contained in this document are based on the input available at the time of preparation.

### **TRADE NAMES**

The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software. The report may not be cited for purposes of advertisement.



CCDC DAC-TR-2020-044

July 2020

---

# **Uncertainty Quantification and Sensitivity Analysis Methodology for the Advanced Joint Effectiveness Model (AJEM)**

**by Craig D. Andres**  
*CCDC Data & Analysis Center*

<b>REPORT DOCUMENTATION PAGE</b>				<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE</b> July 2020		<b>2. REPORT TYPE</b> Technical Report		<b>3. DATES COVERED (From - To)</b> 1 January 2019 – 13 April 2020	
<b>4. TITLE AND SUBTITLE</b> Uncertainty Quantification and Sensitivity Analysis Methodology for the Advanced Joint Effectiveness Model (AJEM)				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Craig D. Andres				<b>5d. PROJECT NUMBER</b> JLF G-18-01	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Director U.S. Army CCDC Data & Analysis Center Warfighter & Futures Integration Division (ATTN: FCDD-DAS-LBS) 6509 Wayberry Road Aberdeen Proving Ground, MD 21005-5066				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  CCDC DAC-TR-2020-044	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Office of the Secretary of Defense/Director Operational Test and Evaluation /Live-Fire Test and Evaluation 1700 Defense Pentagon Washington DC 20301-1700				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> OSD DOT&E LFT&E	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The Advanced Joint Effectiveness Model (AJEM) is a joint forces model developed by the U.S. Army that is used in vulnerability and lethality (V/L) predictions for threat/target interactions. This complex model primarily generates a probability response for various components, scenarios, loss of capabilities, or summary conditions. Sensitivity analysis (SA) and uncertainty quantification (UQ), referred to jointly as SA/UQ, are disciplines that provide the working space for how model estimates changes with respect to changes in input variables. A comparative measure that will be used to characterize the effect of an input change on the predicted outcome is developed in this report. This measure will provide a practical context that stakeholders can better understand and utilize. We show graphical and tabular results using this measure.					
<b>15. SUBJECT TERMS</b> uncertainty, sensitivity, computational models, vulnerability, lethality, comparative measure					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UNCLASSIFIED	<b>18. NUMBER OF PAGES</b>  30	<b>19a. NAME OF RESPONSIBLE PERSON</b> Craig D. Andres
<b>a. REPORT</b> UNCLASSIFIED	<b>b. ABSTRACT</b> UNCLASSIFIED	<b>c. THIS PAGE</b> UNCLASSIFIED			<b>19b. TELEPHONE NUMBER</b> (include area code) (410) 278-6721

---

---

## Table of Contents

List of Figures .....	iv
List of Tables.....	v
1. INTRODUCTION AND PURPOSE.....	1
2. METHODOLOGY .....	2
3. ADDITIONAL DISCUSSION.....	8
4. ANALYSIS .....	12
5. RESULTS.....	17
6. CONCLUSION .....	18
7. REFERENCES AND DOCUMENTS .....	19
Appendix A – List of Acronyms .....	A-1
Appendix B – Distribution List.....	B-1

---

---

## List of Figures

Figure 1.	Rendering of the 62 aspects color coded for elevation .....	4
Figure 2.	Generic model depiction .....	5
Figure 3.	Fixed interval changes in estimation at various points on a typical S-curve .....	10
Figure 4.	The blue and green curves are the SREs using the standard normal CDF S-curve, while the red and orange curves are the SREs using the logistic S-curve. Red and blue curves are associated with the +25% input, while orange and green are the -25% input.....	11
Figure 5.	Example graphical output .....	13
Figure 6.	Example graphical summary output.....	14

---

---

## List of Tables

Table 1.	Comparison Measures by Some Examples .....	8
Table 2.	SRE Values Associated with Figure 3 .....	11
Table 3.	Example Output Table .....	13

---

---

## 1. INTRODUCTION AND PURPOSE

The Methodology Team in the Joint Software Engineering Branch of the U.S. Army Combat Capabilities Development Command Data & Analysis Center is charged with providing quantification methods for the sensitivity and uncertainty in the Advanced Joint Effectiveness Model's (AJEM's) response measures based on variations of model inputs. Sensitivity analysis (SA) and uncertainty quantification (UQ), referred to jointly as SA/UQ, are disciplines that provide the working space for this effort. AJEM is a joint forces model developed by the U.S. Army that is used in vulnerability and lethality (V/L) predictions for threat/target interactions. These model estimates are subsequently used for a variety of efforts, including use in higher level models. The quantification of the sensitivity and uncertainty of these V/L predictions requires a comparative measure that will help us characterize the effect of an input change on the predicted outcome.

Requirements for this methodology include the following:

1. Shall take into account customer perspective and how results will be contextualized or reported by them.
2. Measures of uncertainty/sensitivity shall be relevant but not over-complex.
3. A visual aid shall be developed to help report results.
4. Computing resources for the implemented method shall not be overburdened.
5. A developed application shall include capability for analysis for component, element, and system-level quantification.

Though this effort is aimed at the AJEM specifically, the resulting approach can be used for any modeling, and could be especially useful for any model that produces probability estimates.

---

---

## 2. METHODOLOGY

There are several challenges to overcome when modeling probability estimates. A general challenge is often the language of probability estimates. Though not always, we commonly report probability estimates as percentage, and in many statistical procedures we want to characterize estimates using confidence, prediction, or tolerance intervals with a designed level of confidence that is also reported in percentages. We would like to avoid referring to “percentages of percentages with respect to percentages”, which can easily lead to confusion. The use of adequately descriptive language that provides clarity and consistency is more critical than usual as confusion from the language can easily occur in this topic, especially when referring to models with probability estimates. Our subject model in this case is AJEM, but we may want to use a General Linear Model (GLM) or similar to characterize the sensitivity of our probability estimation model responses with respect to changes in inputs. So modeling the model responses can seem confusing. We may not entirely escape nested yet similar sounding terminology, but we will try to be clear where it is used.

Another challenge in using a standard GLM approach is the conditional nature of odds ratios. Odds ratios and log odds ratios are probably the most common way of comparatively analyzing probability estimates. It is common to use a GLM (specifically Logistic Regression or similar) to understand probability measures. Yet confidence intervals and other standard estimators are conditional to the value of the estimate. Simply put, the variance of a simple binomial response behaves somewhat normally between 0.2 and 0.8, but changes rapidly as the estimate approaches zero or one. This changing variance affects interval estimation in the same way. Add to this that these procedures can require that our true expected value is always greater than zero and less than one. Yet for AJEM and other models out there, a value of zero or one can sometimes be expected. For example, the expected probability that shooting a main battle tank with a BB gun will cause the tank to lose mobility is expected to be zero in practice. We want the capability to characterize where a change in input affects a change in prediction from zero/one to a non-zero/one outcome, or vice versa, and in many measures that leads to a division by zero.

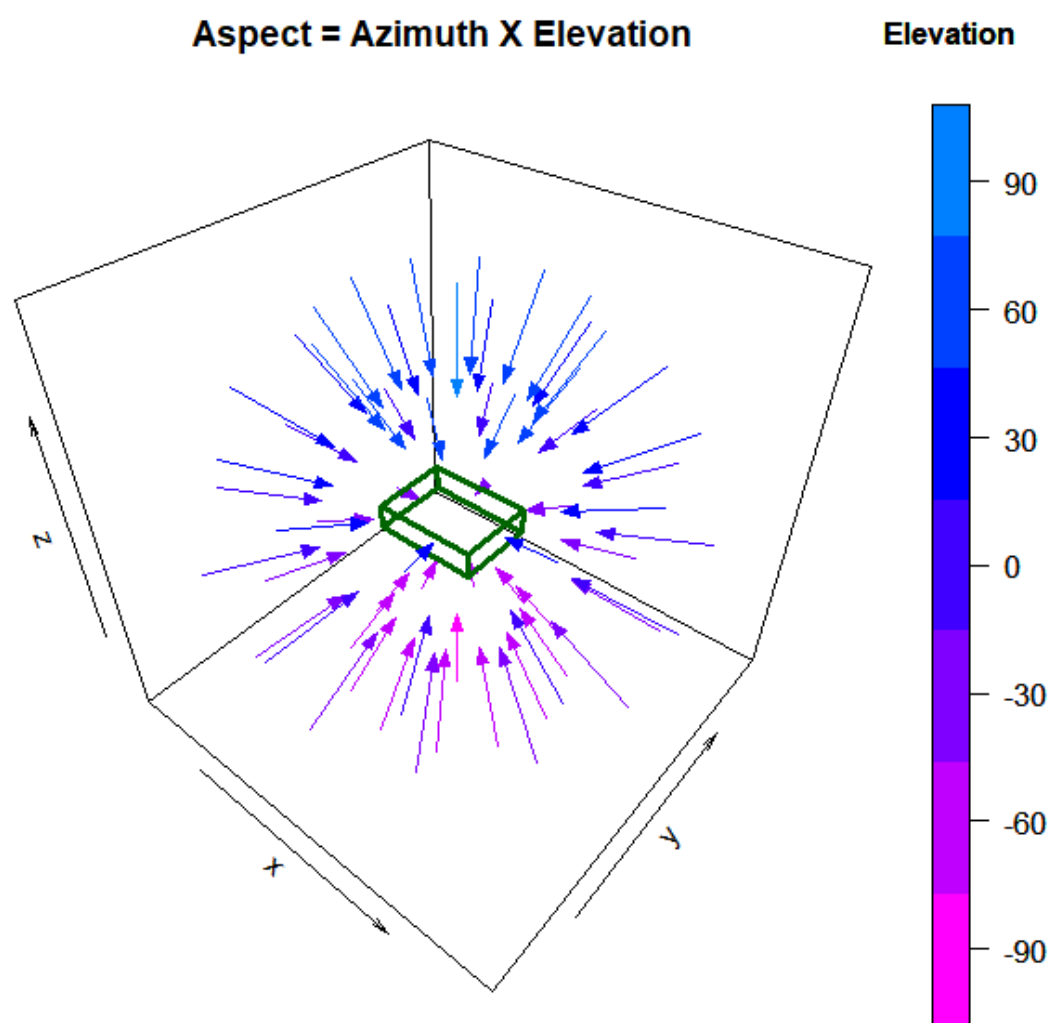
A challenge with AJEM is that a typical data set that we want to analyze is not independent and identically distributed (iid) with respect to how we want to characterize an overall estimate. A threat can hit a target from any aspect (angle and elevation), and each aspect has an expected probability of kill (Pk) or component damage (Pcd|h). So for each aspect, we can estimate a set of probability outcomes since the model is multivariate (we typically look at 4–6 outcomes, but there can be dozens of outcomes). At the aspect level, a set of runs of the model appear to produce iid results, often with a Gaussian spread, but estimates are expected to differ across aspects. We want to aggregate results across aspects so that we can characterize the possible outcomes for a given scenario. For our phase 1 effort of quantifying uncertainty in the model, we are using AJEM iterations to estimate results for a set of 62 chosen aspects (Figure 1) on four

---

---

or five defined kill types on two threat/target combinations. Note that the set of aspects are defined in  $30^\circ$  increments, but any set can be used, even a random set.

For each output estimated at a given aspect, the projected 2-D image from the aspect-based viewpoint of the target is meshed into a cell-by-cell grid pattern. Each cell of this mesh is modeled to be impacted by the threat at least once, either at a random location within the cell or a predefined location such as center. A level of vulnerability for each cell is determined with value on  $[0,1]$ . The vulnerability is summed across all cell estimates multiplied by cell size, and this sum is divided by the total presented area of the target for the given aspect. The resulting vulnerability estimate is one iteration of the AJEM model for a single aspect with the caveat that each cell has equal weight (i.e., hitting a cell has approximately uniform random probability). We can increase the resolution of the estimate by decreasing the cell size, or by increasing the number of shots per cell while using random placement. The cost of more resolution is more computation time, thus the classic tradeoff of accuracy versus cost.



**Figure 1. Rendering of the 62 aspects color coded for elevation**

The aspects we chose to use in phase 1 are pairwise angular combinations of elevation ( $-90^\circ$ ,  $-60^\circ, \dots, 90^\circ$ ) and azimuth ( $0^\circ, 30^\circ, \dots, 330^\circ$ ). There are 12 azimuths for each elevation except at  $-90^\circ$  and  $90^\circ$  where all azimuths result in the same vector ( $5 \times 12 + 1 + 1 = 62$ ). For each aspect we generated a set of probability estimates repeating that 100 to 1000 times (depending on resolution/computational burden). Standard descriptors produced in the AJEM output to characterize results included median, mean, min, max, standard deviation, and range. To characterize sensitivity of an output to an input, the input was changed by a defined ratio and another cycle of iterations were completed to produce another set of AJEM outputs. We then compare results between the two sets of results by aspect. We can also look at percentile type responses, and may follow that direction at a later date. For now the descriptors we settled on comparing were median, min and max values by aspect. Figure 2 shows our model as a black box.



**Figure 2. Generic model depiction**

Inputs:  $\{A\}$ , threat, target, aspect  $\geq$  Outputs:  $\{A\}$ ,  $P_k|h$ ,  $A_v$ ,  $P_A$

Inputs:  $\{B\}$ , threat, target, aspect  $\geq$  Outputs:  $\{B\}$ ,  $P_k|h$ ,  $A_v$ ,  $P_A$

The ideal measure for comparison has the capability compare the two sets of outputs/predictions and meets the requirements listed above, and we considered several candidate measures toward those goals. Important measures that do not adequately meet our SA/UQ assessment needs: odds, odds ratio, log odds ratio, and coefficient of variation (CV) (odds ratio and its log may be worth revisiting).

Without iid, much of our standard estimation procedures to provide a comparison of two results, such as prediction or confidence intervals are not valid, and yet we still want to be able to characterize the spread of a set of results (in our case, the 62 aspects).

Why symmetric relative effect (SRE)? The relative effect is an existing quantity used in epidemiology and has its roots in categorical data analysis. Dr. Alan Agresti's book of the same name states, "A difference in proportions of fixed size may have greater importance when both proportions are close to 0 or 1 than when they are near the middle of the range." The importance of this statement cannot be overstated in our world of modeling the probabilities of an outcome for a given scenario. The example he gives is that the difference between 0.010 and 0.001 may

---

---

be more noteworthy than the difference between 0.410 and 0.401. The delta P ( $dp = \Delta p = p_2 - p_1$ ) in both cases is  $dp = 0.001$ . However, these results in a V/L model would imply (for probability of kill  $P_k$ ) that our kill frequency is 10 times more often in the first case, but not much different in the second. We want to assess our uncertainty based on a measure that will help us determine how appropriate we believe our model estimate to be. The consideration of  $dp$  for SA/UQ purposes would lead us to conclude that a  $dp$  of 0.001 means the same for both of those scenarios. Clearly, we need a measure that gives a more relevant context.

Risk analysts examine relative increases or reductions in risk using specific terms for each direction of change. The general term “relative effect” is used to indicate either change with respect to the sign of the value (negative for reduction and positive for increase). A relative change is a fairly familiar and intuitive concept to professionals from most disciplines, and is used to characterize outcomes other than probability estimates.

What is meant by symmetric? For our analysis with respect to probability estimates, we want to treat that relative effect symmetrically about 0.5 on the interval  $[0,1]$ . Thus a change from  $p_1 = 0.2$  to  $p_2 = 0.1$  will be treated the same as a change from 0.8 to 0.9 (though this last change is negative). We see this symmetric nature reflected in variance calculation related to the binomial distribution variance which is  $V=np(1-p)$ . Our symmetric effect uses a stepwise function that changes the denominator from  $p_1$  to  $1-p_1$  if  $p_1 > 0.5$ . Of course at zero and one, these denominators become zero. A zero or one outcome, specifically for the base  $p_1$ , may be important to preserve, so a flag will be set if  $p_1$  is zero or one. The numerator of our ratio will be the difference  $dp = p_2 - p_1$ . If an outcome is  $p_1 = p_2 = 0$  or  $p_1 = p_2 = 1$ , then the comparison measure is set as zero (i.e., we have the same outcome). The measure logic is then:

SRE: Comparison of two outputs ( $p_1, p_2$ ) where  $p_1$  is the baseline or true value:

- **SRE( $p_1, p_2$ ) =**
  1.  $(p_2 - p_1)/p_1$  for  $p_1 < 0.5$  else.
  2.  $(p_2 - p_1)/(1 - p_1)$ .
  3. Special consideration for  $p_1 = 0$  or  $1$ :
    - If  $p_1 = p_2 = 0$  or  $p_1 = p_2 = 1$  then  $SRE = 0$
    - Otherwise use  $dp = p_2 - p_1$  and flag the occurrence
- Sign of SRE indicates a decrease(−) or increase(+) in effect.
- “Symmetric” so that our measure of influence is robust near the ends of  $[0,1]$ :
  - i.e.,  $SRE(0.8, 0.9) = -SRE(0.2, 0.1)$ .

- 
- 
- $p_1$  is usually a baseline response or a highly accurate response.
  - SRE result is the same if vulnerable area is used instead of  $P_k|h$ .
  - SRE is the same for both V/L perspectives.

The measure function is defined to be complete with respect to all possible outcomes. Yet we expect that parts 1 and 2 of the measure will be relevant for the vast majority of any such analysis because those will be the encounter scenarios we are most interested in. The main measures that were considered are shown in Table 1. Another measure initially considered was  $\sqrt{p(1-p)}$ , and related to the standard deviation of a binomial distribution, but this measure was not consistent when crossing 0.5. As mentioned earlier, the odds ratio and log odds ratio are a standard way of comparing to probability measures.

### 3. ADDITIONAL DISCUSSION

Over the next few pages, we examine how the SRE compares to odds ratio measures and how SRE can be influenced by various conditions. Keep in mind, this part of the discussion is not about error, but is interested in how input and responses can vary. The values in Table 1 give us some idea of how these two measures and SRE compare. These are really chosen points off of a 3-D surface. The column to the right is the SRE measure in percentage. Using percentage to report SRE is meant to help reduce confusion when working with probability comparisons. Notice that  $dp/p1$  is the same for all but the last three rows since the  $p1$  is less than or equal to 0.5, thus the same calculation as the first part of SRE. The odds ratio looks close to SRE, but the context for odds ratio is that a value of 1 means no difference. The log odds (which is the natural log of the odds ratio) has a no change reference point at 0 (like SRE), where negative indicates reduction and positive indicates an increase (also like SRE), but practical interpretation of the log odds is a challenge. The table mainly illustrates two changes across several values for  $p1$  ranging from close to zero to 0.5. A constant change of 0.1 for the first seven rows where the SRE shows a relative change that is larger as the  $p1$  nears zero. The next eight rows show measure values when  $p2$  is half of  $p1$  across those, so naturally the SRE will be  $-0.5$ . Our position is that the SRE measure is well associated with the other two measures, and gives us consistent, practical, and intuitive values.

**Table 1. Comparison Measures by Some Examples**

$p1$	$p2$	$dp = p2 - p1$	$dp/p1$	Odds Ratio	Log Odds	SRE	SRE (%)
0.01	0.11	0.1	10	12.24	2.50	10	1000
0.05	0.15	0.1	2	3.35	1.21	2	200
0.1	0.2	0.1	1	2.25	0.81	1	100
0.2	0.3	0.1	0.5	1.71	0.54	0.5	50
0.3	0.4	0.1	0.33	1.56	0.44	0.33	33
0.4	0.5	0.1	0.25	1.5	0.41	0.25	25
0.5	0.6	0.1	0.2	1.5	0.41	0.2	20
0.01	0.005	-0.005	-0.5	0.5	-0.69	-0.5	-50
0.05	0.025	-0.025	-0.5	0.49	-0.71	-0.5	-50
0.1	0.05	-0.05	-0.5	0.47	-0.76	-0.5	-50
0.2	0.1	-0.1	-0.5	0.44	-0.82	-0.5	-50
0.3	0.15	-0.15	-0.5	0.41	-0.89	-0.5	-50
0.4	0.2	-0.2	-0.5	0.38	-0.97	-0.5	-50
0.5	0.25	-0.25	-0.5	0.33	-1.11	-0.5	-50
0.6	0.4	-0.2	-0.33	0.44	-0.82	-0.5	-50
0.95	0.85	-0.1	-0.11	0.3	-1.20	-2	-200
0.99	0.89	-0.1	-0.10	0.08	-2.53	-10	-1000

---

---

For scenario encounters, a relative frequency can be as important in estimating mission success as actual underlying probabilities. For example, if we use a set of inputs into our model, and determine to expect a loss rate of 20% of vehicles in an encounter, and we have estimated that we need 30 to be functional to complete a mission, we would compensate for the expected loss. Using another set of inputs in the model could change that 20% so that now we can see how much a change of the inputs could affect that adjustment and adjust the strategy as needed.

For the AJEM V/L model there are questions that need practical answers that influence how it is used by analysts. Questions about how we set up runs for AJEM include the following:

- What size cells are good enough?
- How many shots per cell are enough using random location?
  - Should we use pseudo-random strategies?

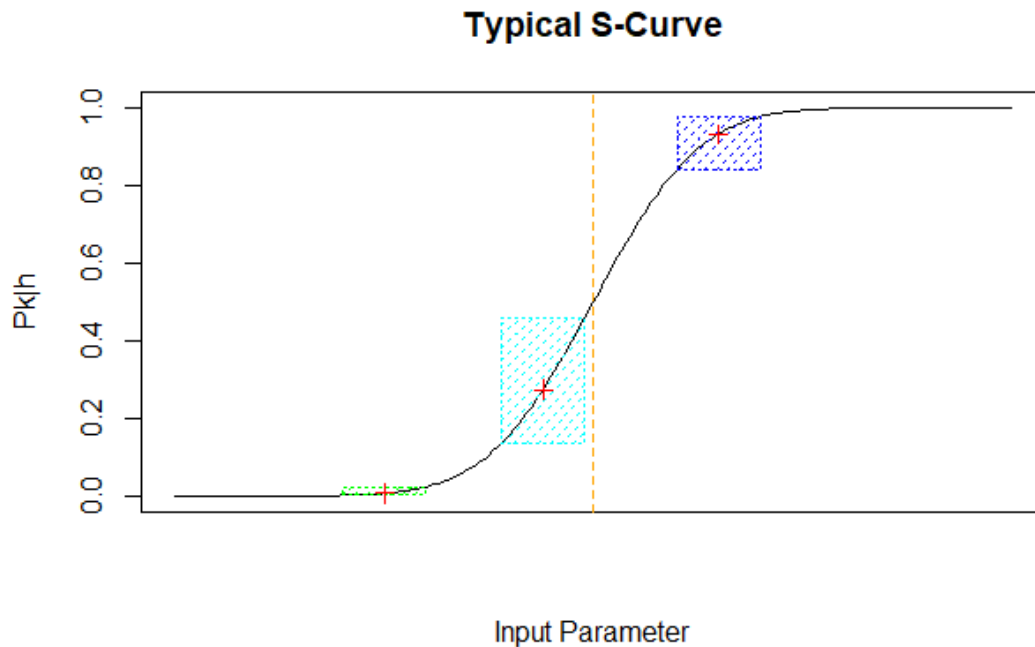
We also have questions concerning inputs:

- Which inputs have the biggest effect on responses?
- What inputs interact significantly?
- Can we generalize results? (maybe)

Any comparative measure above can help answer these questions. The SRE measure has a few advantages from a practical perspective:

- That SRE explains a relative change is easier for most people to comprehend than an odds ratio
- SRE measure can be used from a requirement perspective:
  - That is, a percentage of aspect points within a specification (threshold and/objective)
- SRE can also be easily incorporated with cutoff for very small or high values:
  - For example, rare events (e.g.,  $pk|h < 0.001$ ) may treat as probability of zero.
  - For example, very likely events (e.g.,  $pk|h > 0.999$ ) may treat as probability of one.
  - Customizable to user needs

For a sensitivity analysis, we aim to change our inputs by an amount or percentage and use the SRE measure to assess the change in model output. Each input has some amount of influence on the outcome, but the amount of influence can be very different. For example, if our penetration is insignificant (say 0.01 mm RHA with a Pk of or near zero), then increasing or reducing by 25% would not likely change the outcome. The points in the parameter space for our modeling, and the underlying distribution, can influence our comparative measure. We can illustrate the influence of an input on the model response by using an underlying S-curve. In Figure 3, we see the effect of a fixed change in an input parameter (rather than the probability directly) centered at three points on a simplified univariate probability response using a typical S-curve (Cumulative Normal). This simple probability response curve has a transition point (at the inflection point) where the change in the input value has a larger magnitude of relative change in the Pk|h. Using a fixed amount of difference in the input parameter, the SREs on the left and right side of the color coded boxes are reported in Table 1. Note that the fixed change to the left or right of a point is non-symmetric for SRE. This effectively shows that a fixed change in a parameter is conditional on the predicted value, and also shows whether it is an increase or decrease in the parameter.



**Figure 3. Fixed interval changes in estimation at various points on a typical S-curve**

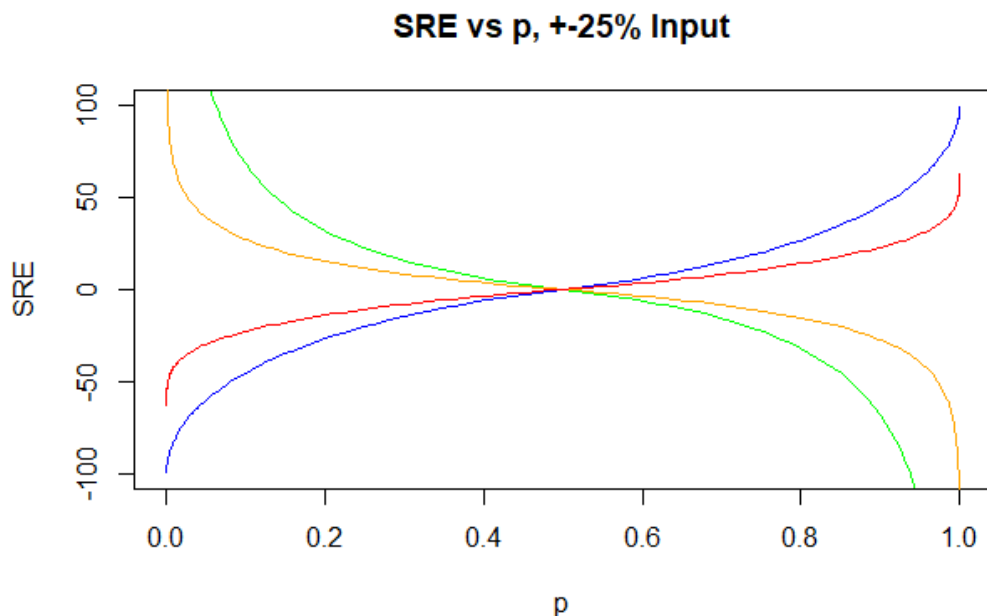
In Table 2, those effects are shown based on the three locations in the graph. This is more than an illustration of the model; this also illustrates what can happen in testing. The base values and the changes chosen (in modeling or testing) could have influence on the effect. Our SREs for the three points do differ, but they are all near the same magnitude. As always, analysis gets more challenging near zero and one. These table values tell us that we have consistency in the measure, for this underlying S-curve holds up even near zero and one. In testing, we would

prefer that our experimental design finds the parameter base somewhere in the sweet spot of interest (between 0.2 and 0.8 would be ideal). This would lead to better fidelity in the results for a given sample size and thus validation for the model for that parameter space. A more important ideal is to be near the expected reality of more common engagement scenarios.

**Table 2. SRE Values Associated with Figure 3**

Box Region	$P_k h$	Left SRE (%)	Right SRE (%)
Green	0.006	-78.3	266.4
Blue	0.274	-50.5	67.8
Purple	0.933	-137.5	65.9

In Figure 4, we use a similar perspective of increasing/decreasing an input variable using two different S-curves to get an idea how SRE may differ due to a change in the underlying simple univariate quantal response curve for a standard normal cumulative distribution function (CDF) and a logistic function. Just from these two examples, we can see that the SRE can differ by about a factor of two for two different inputs (or possibly the same input for two different systems). The potential benefit of SA/UQ efforts in general is to identify those differences and plan tests or develop models accordingly.



**Figure 4. The blue and green curves are the SREs using the standard normal CDF S-curve, while the red and orange curves are the SREs using the logistic S-curve. Red and blue curves are associated with the +25% input, while orange and green are the -25% input.**

---

---

## 4. ANALYSIS

Let us look at some example analysis using SRE. Initially we used mean, minimum, and maximum values from each aspect as the base data for analysis. We switched from mean to median because we felt the median is a better general descriptor of average across aspects, since they are not iid, and it made more sense to use the median of medians rather than a median of means. Median tends to be more robust for iid data than mean and is also a relevant descriptor for non-iid data. Median also works well with skewed data, which is often seen produced from our model. Our interest is in the spread of results across aspects when we look for trends. We used a threshold/objective approach in our graphical analysis of SRE using within 5% and 10% of the baseline as a way of keeping perspective on the effects of changing inputs. An early analysis looked into the question of how many shots per cell (SPC) are sufficient to get the level of accuracy desired. The main assumption being that more SPC give better resolution. We looked at 1, 4, and 16 SPC and compared to a baseline of 100 SPC. These equate to decreasing cell sizes of  $100 \times 100$ ,  $50 \times 50$ ,  $25 \times 25$ , and  $10 \times 10$ . Figure 5 shows results from an unspecified target/threat combination. Initially, we jittered the plots and used numbers coded to the elevation. This version takes advantage of ordering the points left to right by elevation (notice only 1 point on the left and right of each cluster for  $-90^\circ$  and  $90^\circ$ ). The azimuth is now coded from 1 to 4 in the points to indicate front, right, rear, and left aspects. Points are also color coded to the base probability,  $p_1$  (scale not shown). There are three clusters in each graph, each of which is related to the relative effect on different probability outcomes (these may or may not be related depending on the outcomes we are considering). Some simple examples of outcomes might be a probability of a mobility kill, firepower kill, and a combination of both (mobility or firepower). It is easy to see that there is a lot of information in each graph, which allows for greater data exploration. We immediately made a discovery that we did not expect. Using the 100 SPC as the “truth” set, we expected that our SRE values would be centered around zero with the spread decreasing as we used 1, 4, and 16 shots per cell (increasing resolution). Instead, we notice that, though the spread is certainly decreasing, the SREs are centered near 5%, 3%, and 0.5%, respectively. After investigating these and other results, we understood that the increase in SPC was the source of the bias. So our baseline was higher by about 5% on average. We used our center shot by cell size to help verify this. The center shot method increases resolution by reducing the cell size, thus increasing the number of cells in the grid, and there is only one shot per cell located at the center of the cell. Using the 100 SPC as a baseline for SRE generation with respect to the range of cell sizes, we saw that same bias across all cell sizes using the center shot method. For our first phase analysis, we then switched to using the 1 SPC set as the baseline where appropriate. From that, it was clear that the magnitude of the SRE increased in the negative direction as the SPC increased. We used 1000 runs in this scenario for the 1 SPC case, and 100 for the 100 SPC case. We were still able to observe the decrease in spread with increasing SPC. An example output table the application produces, and that would associate with a single pane of plots, is in Table 3. Note that the estimates for  $p_1$  and  $p_2$  are very small,

and this may be a case where an analyst might decide that they are too small to be of concern. It is at these low levels we can see sometimes very large SRE values. Note also that the SRE values in the table are reported in percentages, and the estimates are all in the  $[0,1]$  range.

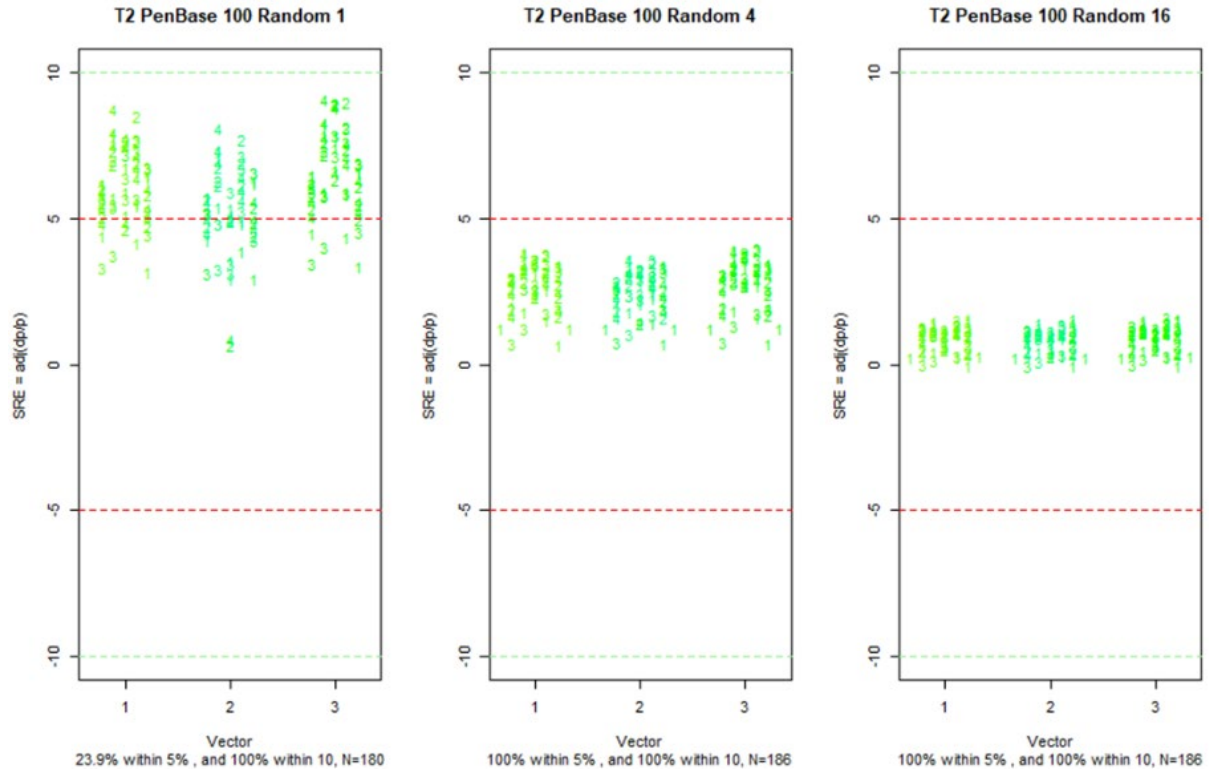


Figure 5. Example graphical output

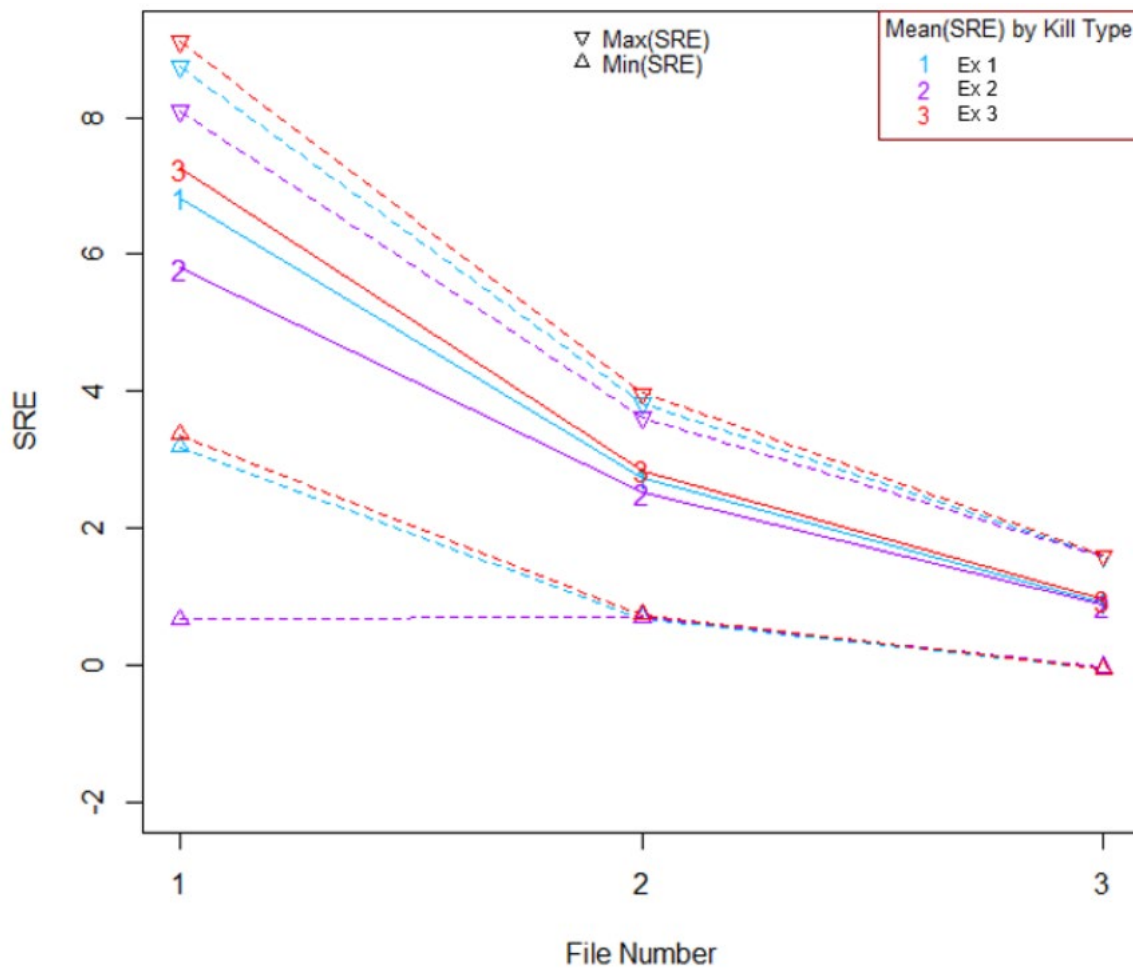
Table 3. Example Output Table

Code #	Kill Type	Within 5%	Within 10%	Median p1	Median p2	dp Δp	Median SRE	Min SRE	Max SRE	Min p1	Max p1	Min p2	Max p2
1	Function 1	40.3	62.9	0.17	0.12	-0.05	-30.7	-77.2	16.9	0.43	0.43	0.01	0.16
2	Function 2	54.8	71.0	0.47	0.44	-0.03	-6.4	-44.3	14.3	2.28	2.28	0.01	2.55
3	Component 1	59.7	87.1	0.26	0.28	0.02	6.1	-1.0	13.3	1.81	1.81	0.00	1.87
4	Mission 1	96.8	96.8	0.05	0.05	0.01	10.6	0.0	12.1	1.46	1.46	0.00	1.61

Note that the application automatically reports at the bottom of the graphs in Figure 5 the percentage of SREs within 5% and 10% of zero, and this is across all points shown in that pane. Table 3 gives a by-kill-type evaluation of this percentage. This was an early approach to assessing how much a change affected the outcome, and could be the basis for setting requirements for the robustness of our model analysis and our input data. For example, penetration is an input described by depth of penetration in a baseline rolled hardened armor. Model validators could use this context to describe how accurate test data must be to satisfy input validity. If we know that more than 90% of the SREs will be within 5% of zero when a

change in penetration is 15% or less, then we could set the testing requirement that test results have to be accurate within 15% to limit influence on the model results. The 5% and 10% are not fixed in stone, but rather are a commonsense easy-to-remember set of values that can be changed as needed by the user.

Because of the breadth of variables and changes our first phase produced, the amount of the above tables and graphs became dominate in the report being written. So to consolidate the same information, we added a summary graph and table for the main body of the report and placed the more detailed information in the appendix for reference. An example of the consolidated graph is in Figure 6.



**Figure 6. Example graphical summary output**

Figure 6 is a consolidation of the information from Figure 3, and the application produces an accompanying table (not shown). In that table, we wanted to point out the most extreme points and show the min and max SRE found at each aspect. Using Figure 4, the median of medians by aspect are displayed as an integer (which indicates kill type), the maximum SRE produced uses the down-pointing triangle (so the visual context is that all SRE values are between the min and

---

---

max). The kill types are color coded as well, and we can see from the graph that the absolute biggest change we observed was for kill type 3 on file 1 (which would associate to the details in the table not shown), with an SRE at about 9.1%. Our median effects range from about 5.9% to 7.2% for the conditions associated with file 1 with respect to the baseline data file. We can see the spread decreases somewhat consistently across, and the bias we observed in Figure 3 is still evident. So we have a somewhat smaller package to show the main results.

A natural question is “What is the worst case change we might see?” We quickly realized that using it is not as simple as reporting the minimum of minimum SREs and maximum of maximum SREs results across kill types and by aspects. Sometimes, the SREs are mostly positive, and sometimes they are mostly negative, but most of the time the SREs are centered at or near zero (which is not surprising). So a worst case would be the furthest point from zero. If a down-the-line user wants to know what the worst possible case is of increasing variable A, or decreasing variable B by some amount, then there are two perspectives we decided to report: the absolute furthest point for each data file analyzed across all kill types, and the worst case median value across all kill types. It was surprising that standard statistical language to cover these concepts did not seem to be in the main stream of terminology. It is possible there is something related to multivariate concepts, but we borrowed from the planetary sciences to describe these points of interest. The apogee is the furthest distance in orbit for an orbiting moon around its planet (the concept is extended to other conditions as well). Apogee is related to the word apex. The opposite of apogee is perigee. We named our worst cases for the two perspectives as the apogee and the apogean median.

Of course, if the worst case is defined, a natural question is what the best case would be, which can be more complicated. We do report the perigee (a closest value with respect to  $SRE = 0$ ). However, we do not recommend using “best case” language here. Let us illustrate with only two values. Suppose I have extreme SRE values of  $-12\%$  and  $4\%$ . The apogee of those two values with respect to zero is  $-12\%$ , definitely the farthest point from center 0. The perigee of the two extremes is  $4\%$ , but it is certain that there were many individual SRE values between  $-12\%$  and  $4\%$ , possibly very close to zero. So the perigee of  $4\%$  can be said to be the best of the worst case effects. Whether or not this can be applied in a useful/practical way can be determined by the user. That example also clarifies why minimum and maximum values are not equivalent to best and worst, as the  $-12\%$  is both the minimum and the worst case SRE.

The R code for our functional version of apogee and perigee on a data vector  $\times 1$  centered at  $z1 = 0$  is

```
APOGEE=function(x1,z1=0) return(x1[which.max(abs(x1-z1))])
```

```
PERIGEE=function(x1,z1=0) return(x1[which.min(abs(x1-z1))])
```

---

---

Apogee returns the maximum value of the absolute value of each value in  $\times 1$ , and restores the original sign. Perigee does the same for the minimum. In our application, the min and max SREs are put combined as the  $\times 1$  data vector to find the apogee and perigee values. The function is used on the vector of medians to find the apogean median and the perigean median. These are all real terms borrowed from astronomy. These values are marked for each data file number across all kill types and reported in the associated table in the current version of the application (not illustrated here). We are of course open to better terminology if it already exists.

---

---

## 5. RESULTS

Results are reported in the technical report. We do see a fair amount of consistency in the effect changing input variables and settings across the main kill types, with some larger differences in the characteristics of effects when looking at special or component-level probability estimates. Analysts and other users should be able to adapt the ideas to their needs and requirements. The 5% and 10% within bounds can be tailored to user needs, or not used at all depending on the analysis. Our discovery of the bias for multiple SPC runs has already led to improvement in code to address the bias. We will use the same analysis approach to verify that we have successfully addressed the bias. We can also answer questions such as “What is the effect of including spall versus no spall?” by framing the effects in terms of SRE.

---

---

## 6. CONCLUSION

Our SRE measure will be used for this and future SA/UQ efforts. It has the customizability needed. It can be explored and explained through graphs and tables. It can be used to set requirements and characterize the effective differences influenced by changes in methodology. Our intent is to build in equivalent tools and capabilities to our model database tools so analysts and other users can assess their efforts within the AJEM application tools.

---

---

## 7. REFERENCES AND DOCUMENTS

Agresti, A. (2002) *Categorical data analysis* (2nd ed.). John Wiley and Sons Inc. 132.  
<http://dx.doi.org/10.1002/0471249688>

Rothman, K. J. (2012). *Epidemiology: an introduction* (2nd ed.). University Press.

---

---

## **Appendix A – List of Acronyms**

---

---

2-D	two-dimensional
AJEM	Advanced Joint Effectiveness Model
CCDC	U.S. Army Combat Capabilities Development Command
CDF	cumulative distribution function
CV	coefficient of variation
DAC	Data & Analysis Center
GLM	General Linear Model
iid	not independent and identically distributed
Pcd h	probability of component damage
Pk	probability of kill
SA	sensitivity analysis
SPC	shots per cell
SRE	symmetric relative effect
UQ	uncertainty quantification
V/L	vulnerability and lethality

---

---

## **Appendix B – Distribution List**

---

---

## **ORGANIZATION**

U.S. Army CCDC Data & Analysis Center  
FCDD-DAS-LBS/C. D. Andres  
6509 Wayberry Road  
Aberdeen Proving Ground, MD 21005-5066

U.S. Army CCDC Army Research Laboratory  
FCDD-RLD-CL/Tech Library  
2800 Powder Mill Rd.  
Adelphi, MD 20783-1138

Defense Technical Information Center  
ATTN: DTIC-O  
8725 John J. Kingman Rd.  
Fort Belvoir, VA 22060-6218