INSTITUTE FOR DEFENSE ANALYSES

Comparing M&S Output to Live Test Data: A Missile System Case Study

Dean Thomas, *Project Leader*

Kelly M. Avery

March 2018

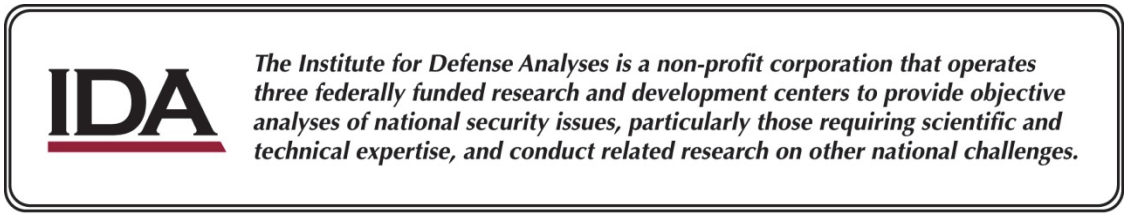
Approved for public release.
Distribution is unlimited.

IDA Non-Standard Document
NS D-9002

Log: H 2018-000104

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

Approved for public release; distribution is unlimited.



About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Task BD-9-2299(90), "Test Science Applications," 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
Review was conducted by Allison L. Goodman and Matthew R. Avery from the Operational Evaluation Division.

For more information:
Dean Thomas, Project Leader
dthomas@ida.org • (703) 845-6986

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

Approved for public release; distribution is unlimited.

INSTITUTE FOR DEFENSE ANALYSES

IDA Non-Standard Document NS D-9002

Comparing M&S Output to Live Test Data: A Missile System Case Study

Dean Thomas, *Project Leader*

Kelly M. Avery

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Comparing M&S Output to Live Test Data: A Missile System Case Study

Dr. Kelly Avery

Institute for Defense Analyses

DATAWorks 2018

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

The Outline: What am I going to talk about?

- The System
- The M&S
- The 3-Phased Test Approach
- Designs and Associated Analyses for Each Phase
- The Evaluation

Note: All data presented are either transformed or notional.

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

The System

So what are we testing?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Goal is to plan an efficient operational test of a missile upgrade

Surface to surface, long range, precision missile

New proximity sensor to increase area coverage

Lethality is the primary measure of effectiveness

Short timeline and limited resources

Modeling and Simulation (M&S) is required to supplement live test data

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

The M&S

I hear these computer models can help me?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Lethality model incorporates both the missile and the target

Given a missile burst point, the model:

1. Generates a fragment distribution
2. Flies fragments to target
3. Determines damage to target components
4. Assesses target loss of function

This process can be replicated many times to generate a probability of kill for a given target and set of input conditions.

Model must be validated before its output can be used in the evaluation of missile effectiveness

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

The Test Design

How do I figure out if this thing works and the model is right?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Phased test approach incorporates multiple venues and data types

1. M&S Data – simulated missile, simulated targets

2. Panel Data – real missile, non-operational targets

3. Live Fire Data – real missile, real targets




Designs for each environment should support both system characterization and M&S validation

Approved for public release; distribution is unlimited.

Different (and multiple) validation analysis techniques are planned for each phase

1. Explore the M&S itself
   - Sensitivity and variation analyses
   - Statistical emulation and prediction
2. Compare M&S to panel data
   - Exploratory data analysis
   - Statistically compare distributions
   - Model live vs. sim taking into account all other factors
3. Repeat #2 for live fire data

Think about the analysis you want to perform before you begin the test design process

Approved for public release; distribution is unlimited.

Design & Analysis Phase 1: M&S Data

First things first…how does the M&S behave?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Design

Goal: Ensure M&S input and output relationships and associated variations make sense.

Response variables:

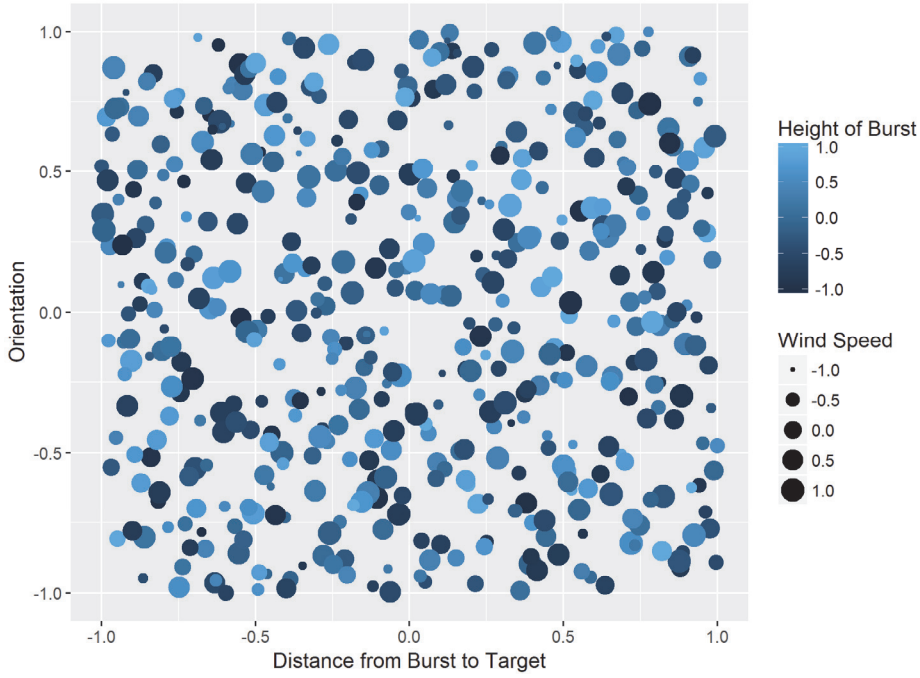
- All M&S outputs

Controllable Factors:

- All M&S inputs

Design:

- Space Filling with Replicates



** Data are notional*

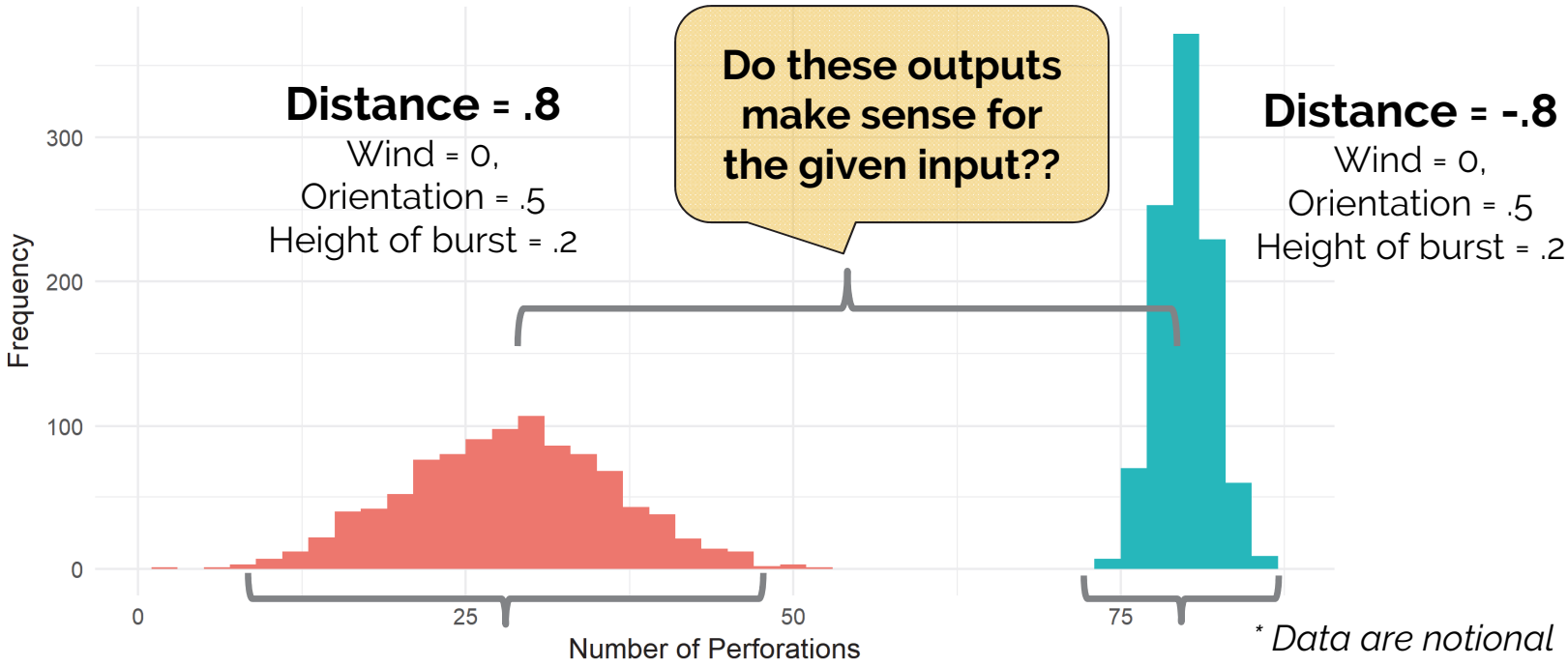
Cover the entire M&S space with the DOE

Analysis

Replicate to explore the behavior of **Monte Carlo** variables

Perform **sensitivity** analyses

Generate **prediction** models for future spot checking



Understanding variation is key

Design & Analysis Phase 2: Panel Data

Our missile put holes in metal plates…

now what do I do?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Designs

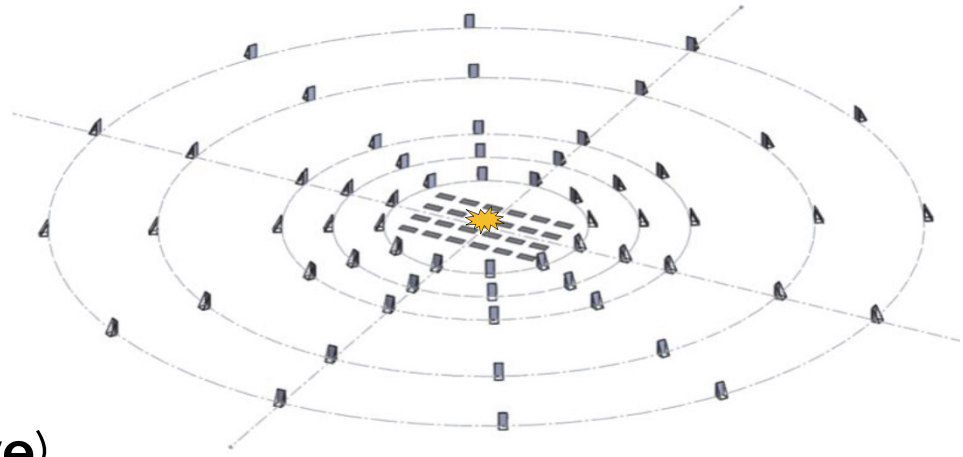
Goal: Determine whether M&S fragment bursts match actual bursts

Response variable:

- Number of perforations

Controllable Factors:

- Distance to target, orientation (angle)



Design:

- 60 point full factorial (**Live**)
- 100 replications of each of those 60 points (**Simulation**)

Continuous or count metrics provide more information than binary metrics

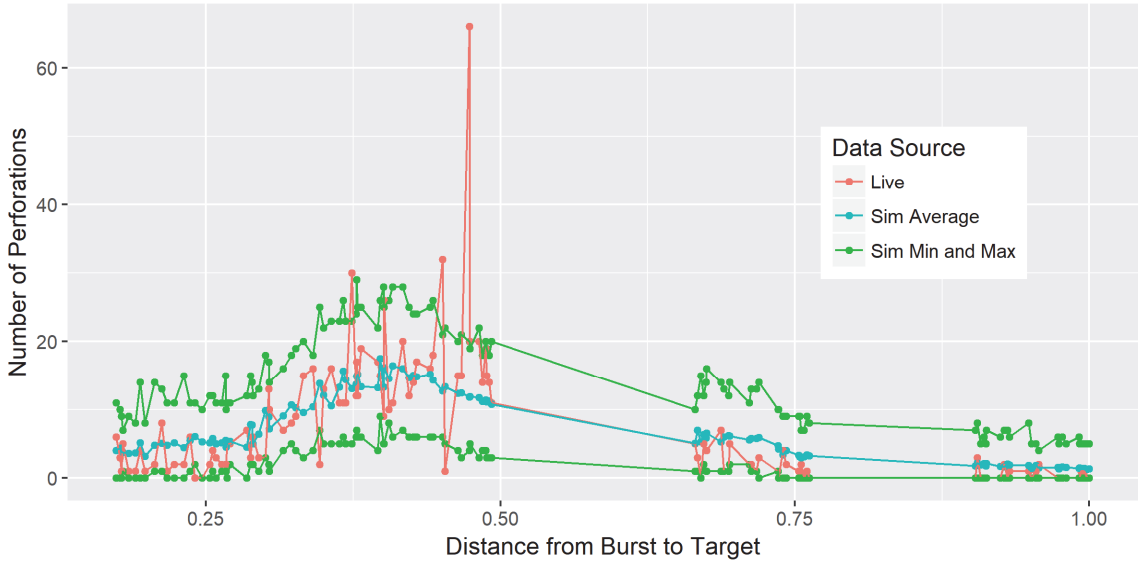
Exploratory analysis

M&S replications form a distribution, but only the average, min, and max values were reported.

Clear relationship between Range and the Number of Perforations.

Not much going on with Orientation.

A few live shots exceed simulation min and max.



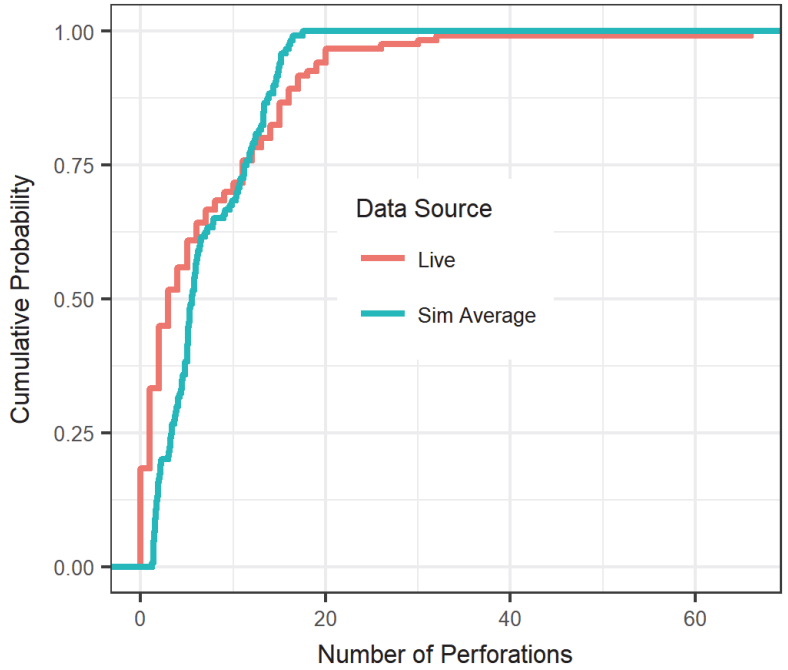
* *Data has been transformed and all values are notional*

A simple statistical look

The **Kolmogorov-Smirnov (KS) test** quantifies differences between two samples of data (in this case, live and M&S).

If the test is rejected, the two samples are highly unlikely to have come from the same distribution.

Caution: The traditional KS test does not account for the effects of factors.



This KS test rejects the null hypothesis (p-value < .01).

Thus, the live data as a whole is statistically significantly different than the average simulation data.

* *Data has been transformed and all values are notional*

A rigorous modeling approach

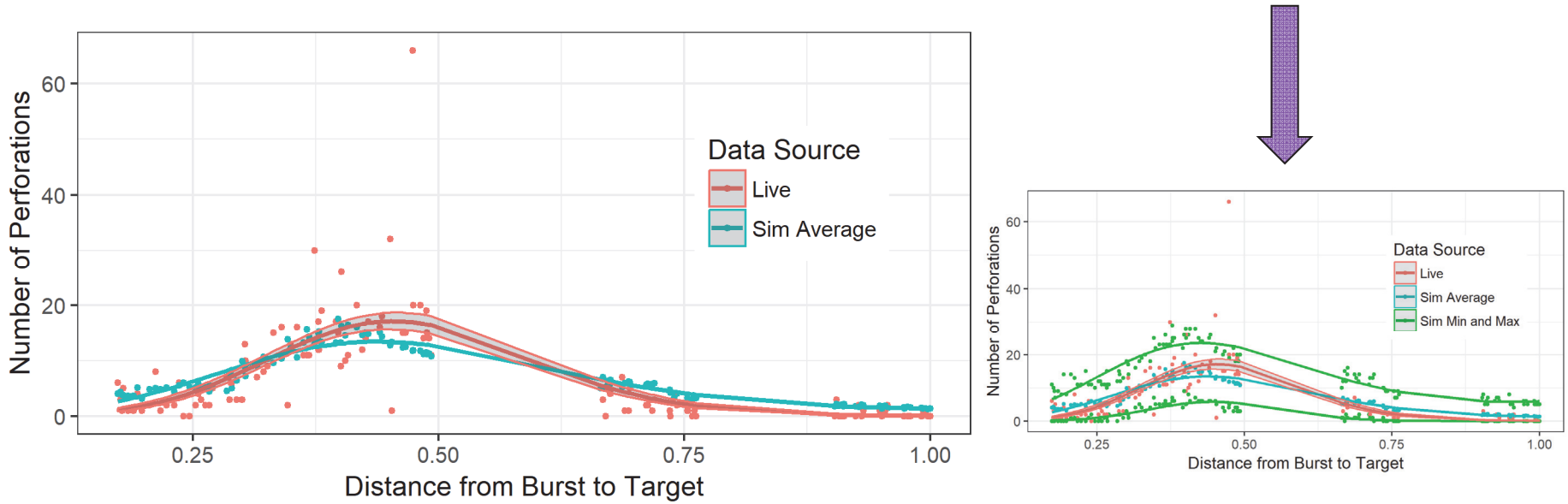
Poisson Regression models count data over several factors.

- Uncertainty intervals can be added to model estimates.

If live and sim are statistically matching, 95% of blue dots should fall into the gray band.

- Only about 20% of blue dots are in gray band.

However, the gray band is contained within the max and min bounds…



* *Data has been transformed and all values are notional*

Design & Analysis Phase 3: Live Fire Data

The M&S can model fragment bursts, but what about lethality against real targets?

P.S. I only have 5 missiles to answer this question…

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Designs

Goals: Cover the operational space of interest and determine whether M&S accurately predict target loss of function.

Response variable:

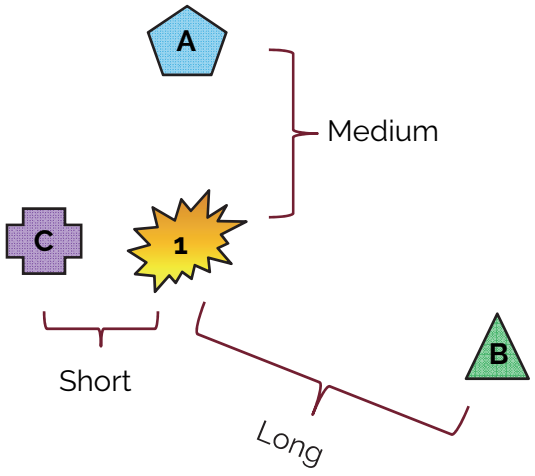
- Number of hits to critical components

Controllable Factors:

- Distance to target, orientation, target class

Design:

- An optimal design is best for the **live** test design since we have a limited number of missiles and targets at our disposal.
- Whatever we do in the live environment we can replicate one or more times in the **simulation.**



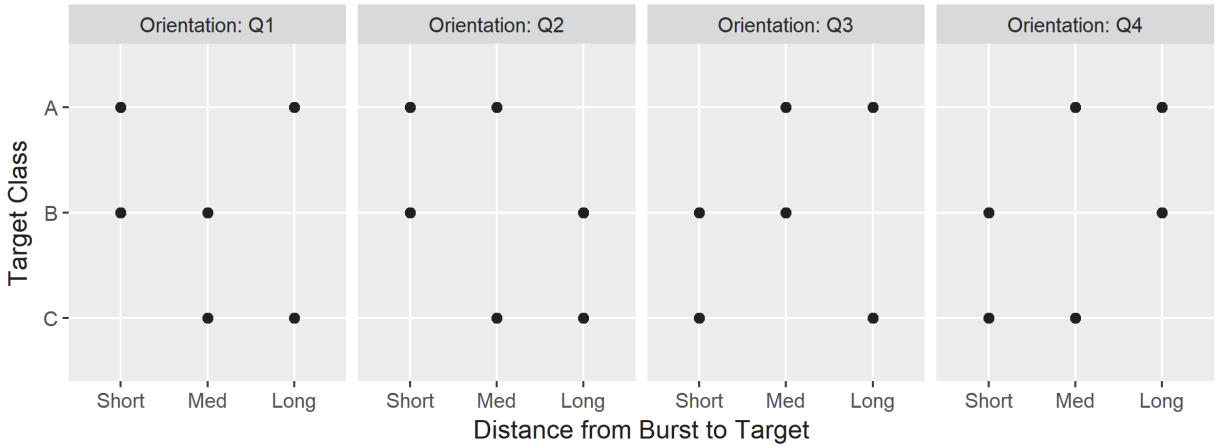
Approved for public release; distribution is unlimited.

Using multiple targets per shot can ensure my live test spans the operational space…

5 missiles with 3-6 targets/shot provides 24 total data points!

These points span the operational space of interest.

Power is also sufficient for detecting differences between live and sim, all main effects, and interactions with source.



Distance	Target Class	Orientation
Short	B	Q3
Short	A	Q2
Short	C	Q4
Medium	A	Q2
Long	C	Q3
Short	B	Q4
Long	A	Q1
Medium	B	Q3
Short	B	Q1
Short	C	Q3
Long	B	Q4
Medium	C	Q2
Long	C	Q2
Medium	B	Q1
Short	B	Q2
Long	C	Q1
Medium	C	Q4
Medium	A	Q4
Long	A	Q4
Medium	C	Q1
Long	A	Q3
Short	A	Q1
Long	B	Q2
Medium	A	Q3

x 2
(replicate in simulation)

Approved for public release; distribution is unlimited.

…but ignoring missile-to-missile variability is risky

Since each missile shot generates several data points, we technically have a blocked design!

Power drops and the ability to estimate factor effects could completely disappear if variability in missiles exists and needs to be estimated.

Spread points out as best as possible to avoid an analysis disaster, and quantitatively test for inter-missile variability in the analysis.

Distance	Target Class	Orientation	Missile
Short	B	Q3	1
Short	A	Q2	1
Short	C	Q4	1
Medium	A	Q2	2
Long	C	Q3	2
Short	B	Q4	2
Long	A	Q1	3
Medium	B	Q3	3
Short	B	Q1	3
Short	C	Q3	3
Long	B	Q4	3
Medium	C	Q2	3
Long	C	Q2	4
Medium	B	Q1	4
Short	B	Q2	4
Long	C	Q1	4
Medium	C	Q4	4
Medium	A	Q4	4
Long	A	Q4	5
Medium	C	Q1	5
Long	A	Q3	5
Short	A	Q1	5
Long	B	Q2	5
Medium	A	Q3	5

x 2
(replicate in simulation)

Approved for public release; distribution is unlimited.

Possible analysis

Assuming that missile behavior was consistent enough to combine data across runs…

We can take a similar approach as for the panel data and perform Poisson regression to highlight differences and risk areas across the factor space.

* *Data are notional*

Evaluation

Do the differences really make a difference?

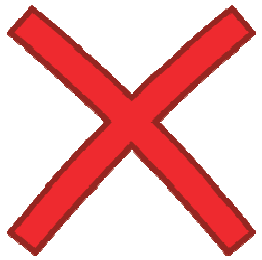
Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

The results in this case are not clear cut

Statistical tests suggest significant differences between average M&S values and actual live data.

- M&S tends to over-predict the mean perforation at the extremes and under-predict in the middle of the range.



However, in the vast majority of cases, live data points fell within the min and max range of the simulation.

So, does the M&S do a good enough job of simulating the outcome?

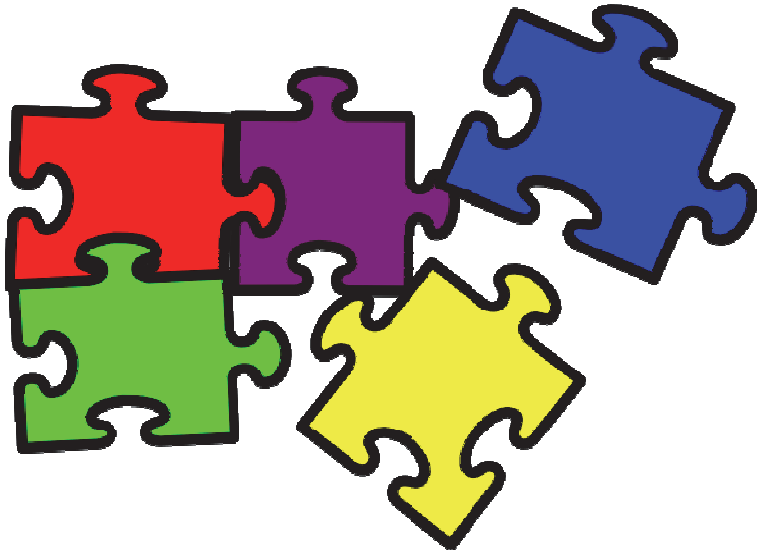
- Maybe….
- Ability of the missile to kill a target may not be affected by these differences between M&S and test results.
- Subject matter expertise along with additional data analysis can provide more insights.

Approved for public release; distribution is unlimited.

Statistical analysis is just part of the puzzle

Analysts/statisticians typically don't make validation and accreditation decisions.

But we can and should inform them by providing the decision-maker with information about M&S performance across the input space and identifying risk areas.



Approved for public release; distribution is unlimited.

Conclusions

Approved for public release; distribution is unlimited.

Testing is hard! But…

Well-thought-out designs facilitate collecting as complete a data set as possible and ensure we learn *something* about the entire operational envelope.

Careful statistical analysis that incorporates all factors ensures we get the most information from limited data.

M&S accreditation is not a simple yes/no decision, and analysts are well-equipped to inform a more nuanced assessment that is ultimately more useful to the warfighter.

Approved for public release; distribution is unlimited.