INSTITUTE FOR DEFENSE ANALYSES

Comparing Live Missile Fire and Simulation

Rebecca Dickinson (Presenter)
Pamela Rambow (Presenter)
Douglas Peek

Science of Test Workshop

4 April 2017

Approved for public release.

IDA Document NS D-8443
Log: H 17-000234

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

The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

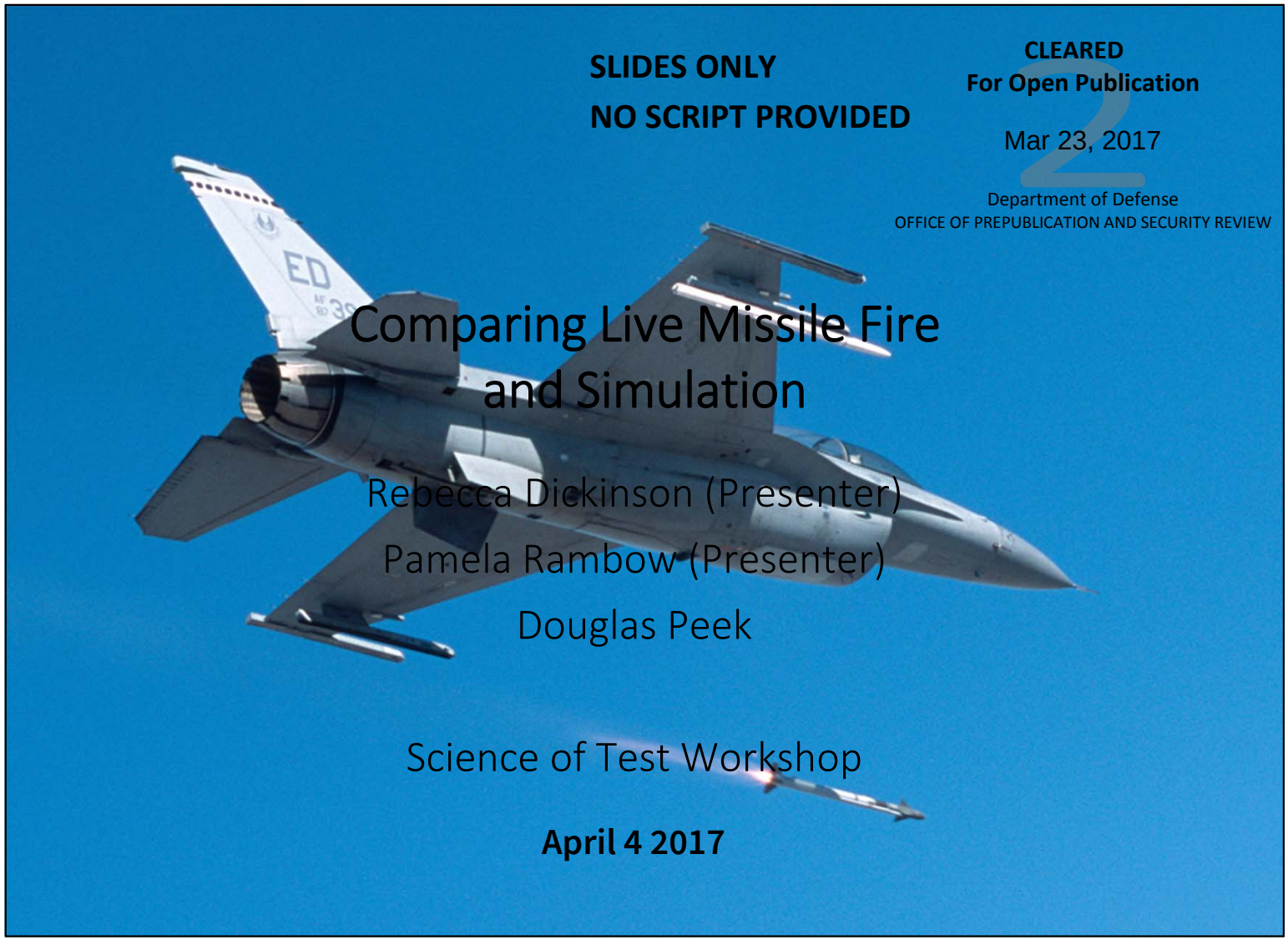
About This Publication

Modeling and Simulation is frequently used in Test and Evaluation (T&E) of air-to-air weapon systems to evaluate the effectiveness of a weapons. The Air Intercept Missile-9X (AIM-9X) program uses modeling and simulation extensively to evaluate missile miss distances. Since flight testing is expensive, the test program uses relatively few flight tests and supplements those data with large numbers of miss distances from simulated tests across the weapons operational space. However, before modeling and simulation can be used to predict performance it must first be validated. Validation is especially challenging when working with a limited number of live test data. In this presentation, we show that even with a limited number of live test points (e.g., 16 missile fires), we can still perform a statistical analysis for the validation. We introduce a validation technique known as Fisher's Combined Probability Test and show how to apply Fisher's test to validate the AIM-9X model and simulation.

For More Information:

Laura Freeman, Project Leader
lfreeman@ida.org, 703-845-2084

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



SLIDES ONLY

NO SCRIPT PROVIDED

Comparing Live Missile Fire and Simulation

Rebecca Dickinson (Presenter)

Pamela Rambow (Presenter)

Douglas Peek

Science of Test Workshop

April 4 2017

Overview

Live missile fire testing is extremely limited due to test constraints and cost – only 16 test events were executed.

Given the limited live test data, we can still perform a statistical analysis for the M&S validation (Fisher's Combined Test).

The M&S was used to supplement the live testing in the evaluation of the weapon system.

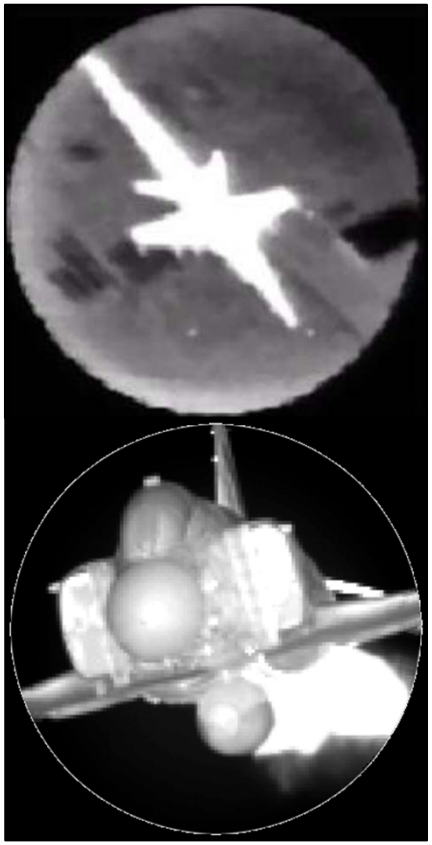
The Air Intercept Missile-9X (AIM-9X) is the principal U.S. short range air-to-air missile.

AIM-9X has an imaging infrared seeker with thrust vector propulsion and is capable of high off-boresight engagements

Performance requirements for AIM-9X are expressed in terms of Probability of Kill (P_K), which is calculated with a six-degree-of-freedom (6DoF) simulation that uses tactical code and simulated target and background environments.

The latest hardware version, Block II, with the latest software version, finished IOT&E in 2015.

- **Only 16 Live Missile Fire Shots were executed.**
- **A significant portion of the evaluation depended on M&S!**



Images from AIM-9X seeker

Cost and test constraints limit the number of live missile firings in the operational testing of the AIM-9X.

Modeling and Simulation (M&S) aides the operational testing of the AIM-9X by reducing the number of required live missile firing events and expanding the operational test space through increased virtual testing

> **BUT… the M&S must first be shown to accurately replicate live missile firings. In other words, we must validate the M&S.**

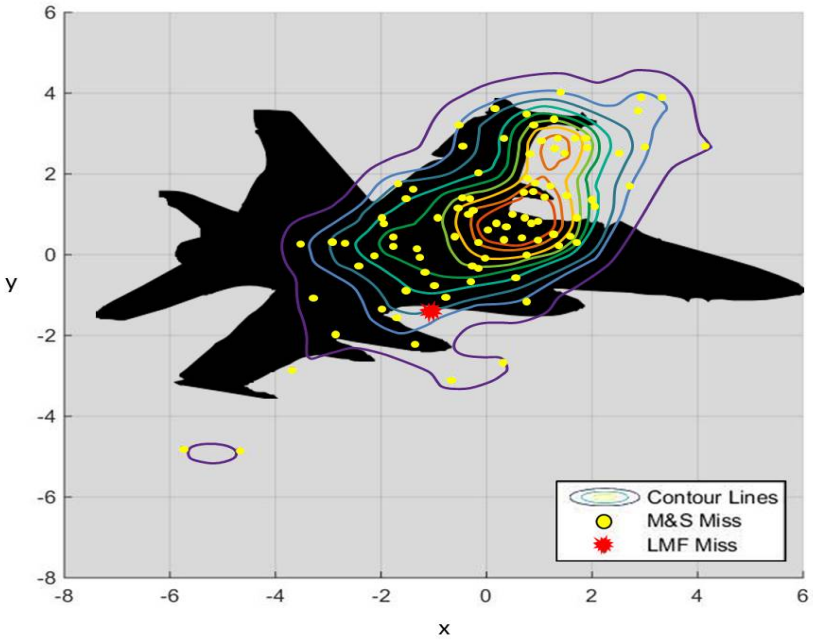
The AIM-9X operational test community used Fisher's Combined Probability Test to validate the simulations for the latest hardware version, Block 2, with 9.313 version software (finished IOT&E in 2015).

Once validated, the M&S is used to predict performance across the weapon space, thereby reducing program cost/time/ resources.

Tail Probabilities (P_{TAIL}) are used to validate the M&S

For each live missile firing, the M&S is typically run 50 to 100 times, producing one miss distance for each run (miss distance cloud).

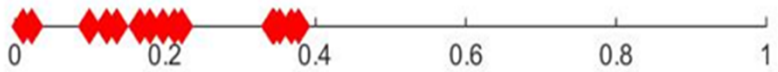
The P_{TAIL} for an individual shot is determined as the percentage of M&S runs that are farther away from the target than the LMF.



The P_{TAIL} for this single event is 23/93, or 0.25

Examples of P_{TAIL}

Case 1: P_{tail} values trend low; simulation miss distances are smaller than live missile firings. Simulation performance is better than reality.



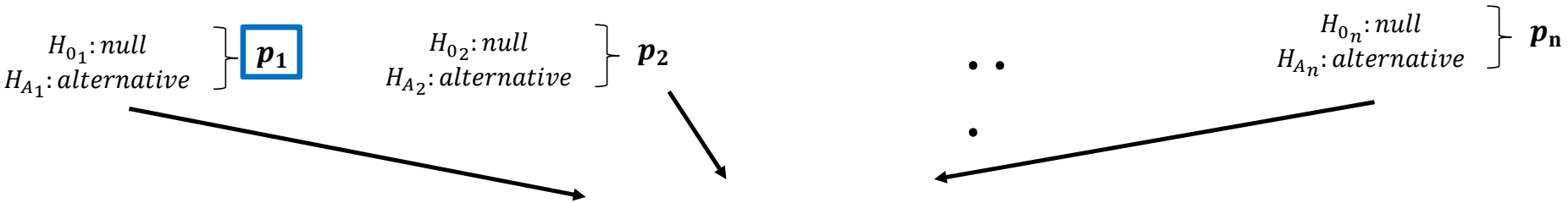
Case 2: P_{tail} values are evenly spread; simulation accurately predicts miss distance.

Case 3: P_{tail} values trend high; simulation miss distances are greater than live missile firings. Simulation performance is worse than reality.

Note: notional data is used in this briefing to illustrate the proof of concept

An Analysis of Analyses: Fisher's Combined Probability Test

Fisher's Test is a method for combining the results from several independent tests, each testing common hypotheses of interest



Definition of P_i

p_i is the p-value, which is the computed probability *under H0* of obtaining a result as rare as or rarer than (determined by H1) what was actually witnessed in the experiment. Rareness is measured consistent with the "directional" component(s) of H1 – e.g., one-sided or two-sided. Small *p*-values provide more evidence against H0.

Note: In our case, p_i is represented by P_{tail}.

H_{0_G}: **null**
H_{A_G}: **alternative**

$$X = -2\sum_{i=1}^{n} \ln(p_i)$$

Deriving the Test Statistic

Under any formulation of the null hypothesis, the $p_i's$ are uniformly distributed on the interval [0,1] and the transformed variables –ln(p) follow and exponential distribution…

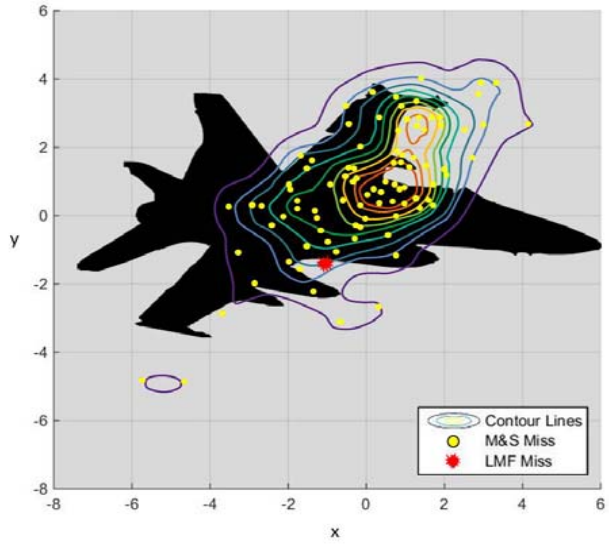
The test statistic X follows a Chi-Square distribution with 2n degrees of freedom

Applying Fisher's test to the Validation of AIM-9X M&S validation

The information from each test point constitutes as a single experiment

A SINGLE live missile fire miss distance

M&S predicted miss distances (typically 50-100)



Constructing the Hypothesis Test

$$H_0: Simulation\ Distribuiton = Live\ Fire\ Distribution$$

Option 1: Simulation predictions are conservative

$$H_{A_G}: Simulation\ Distriubtion < Live\ Fire\ Distribution$$

Option 2: Simulation predictions tend to be pessimistic

$$H_{A_G}: Simulation\ Distriubtion > Live\ Fire\ Distribution$$

Option 3: Guard against too optimistic or too pessimistic

$$H_{A_G}: Simulation\ Distriubtion \neq Live\ Fire\ Distribution$$

In the validation of miss distances each p_i is represented by a P_{TAIL}

$$H_{A_G}: Simulation\ Distriubtion < Live\ Fire\ Distribution$$

$$p_i = P_{TAIL}$$

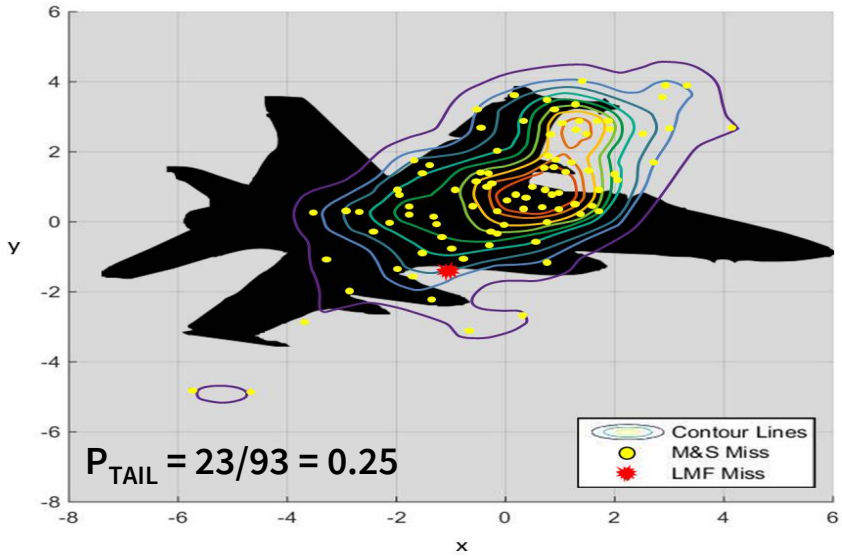
$$H_{A_G}: Simulation\ Distriubtion > Live\ Fire\ Distribution$$

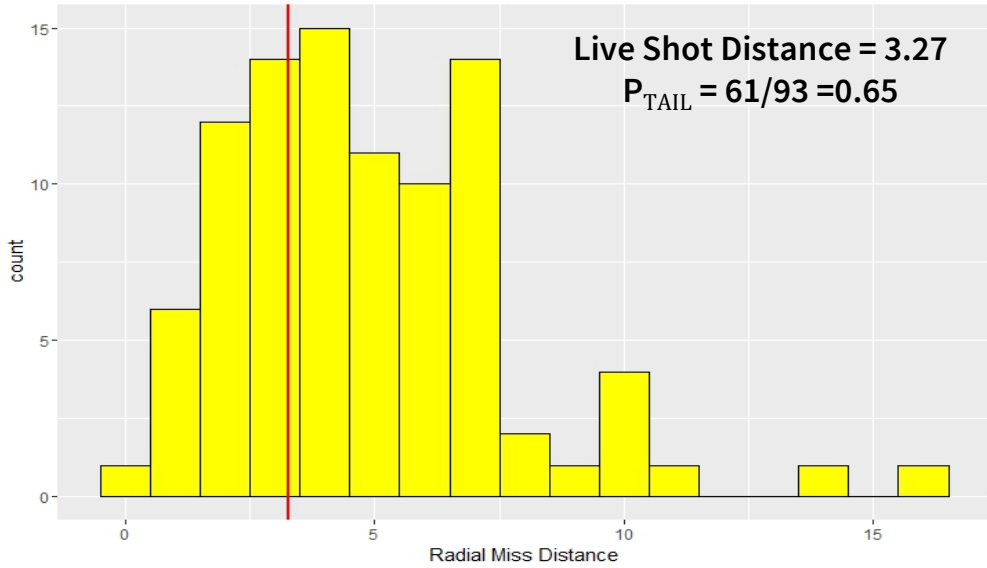
$$p_i = 1 - P_{TAIL}$$

$$H_{A_G}: Simulation\ Distriubtion \neq Live\ Fire\ Distribution$$

$$p_i = \begin{cases} 2(P_{TAIL}) & if \quad (P_{TAIL}) \leq 0.5 \\ 2(1 - P_{TAIL}) & if \quad (P_{TAIL}) > 0.5 \end{cases}$$

P_{Tails} can be calculated in a variety of ways





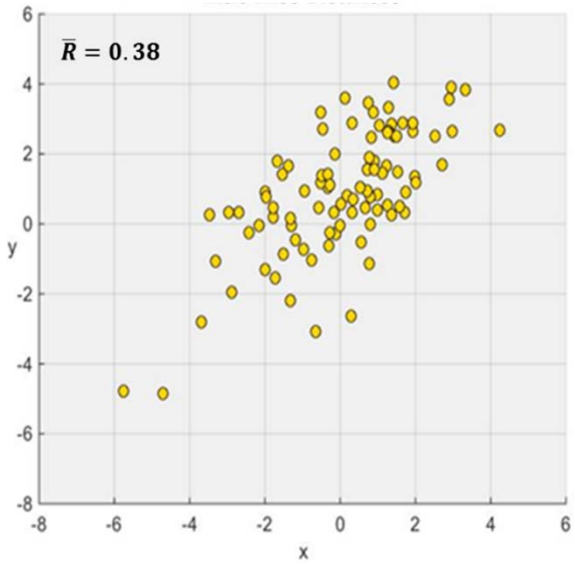
2-Dimensionally using Contours

1-Dimensionally using Miss Distance Quantiles

$$Radial\ Miss\ Distance = \sqrt{(x^2 + y^2)}$$

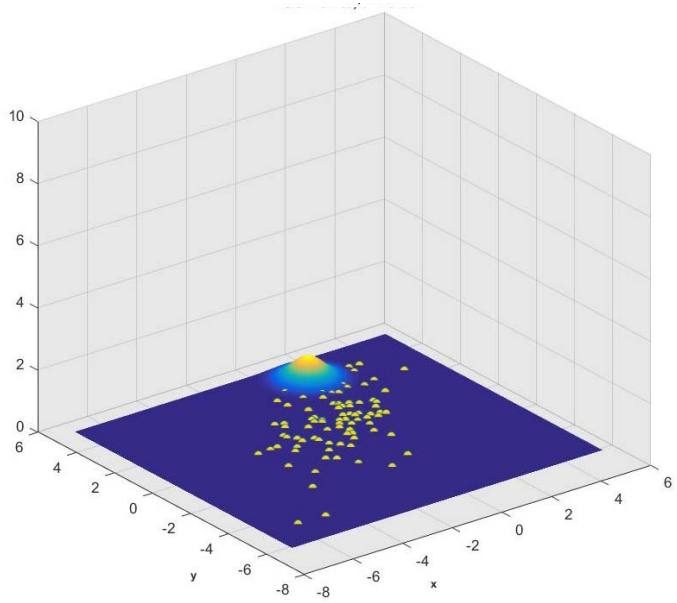
Create a 3D Density of the M&S Miss Distances

M&S Distance Points
$\sigma = 0.56$



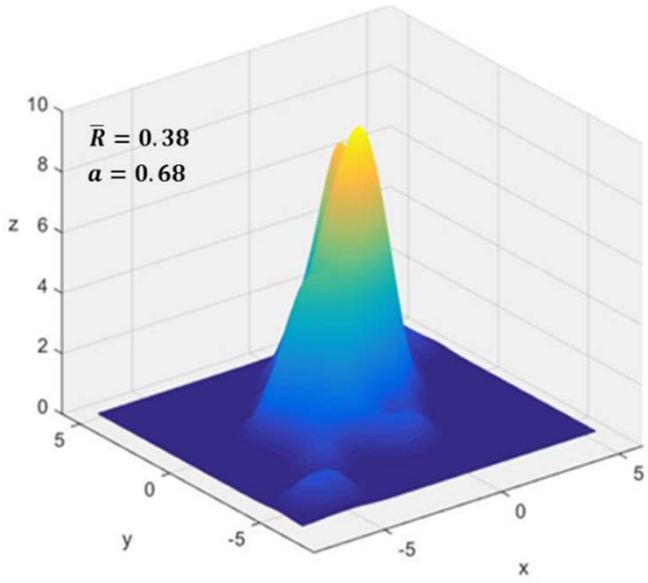
$$\sigma = \frac{1}{a} * \bar{R}$$

Gaussian Surface of Each Point,
$\sigma = 0.56$



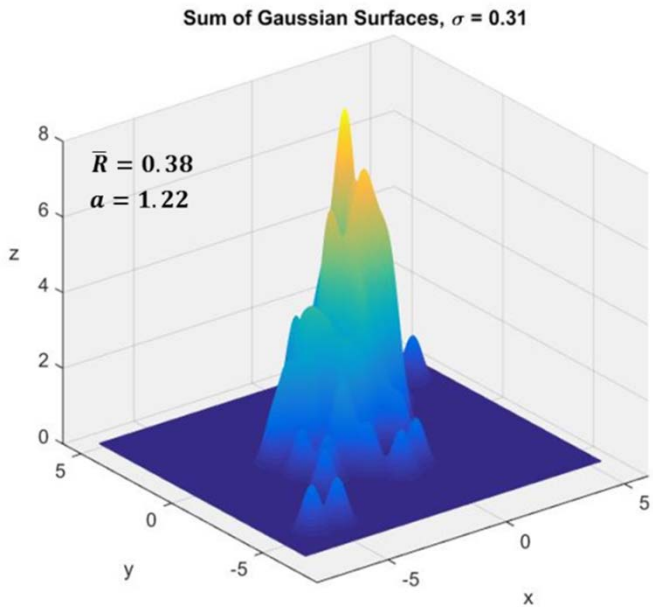
$$f_i(x,y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-[(x-\mu_{xi})^2+[(y-\mu_{yi})^2]}{2\sqrt{\sigma}}}$$

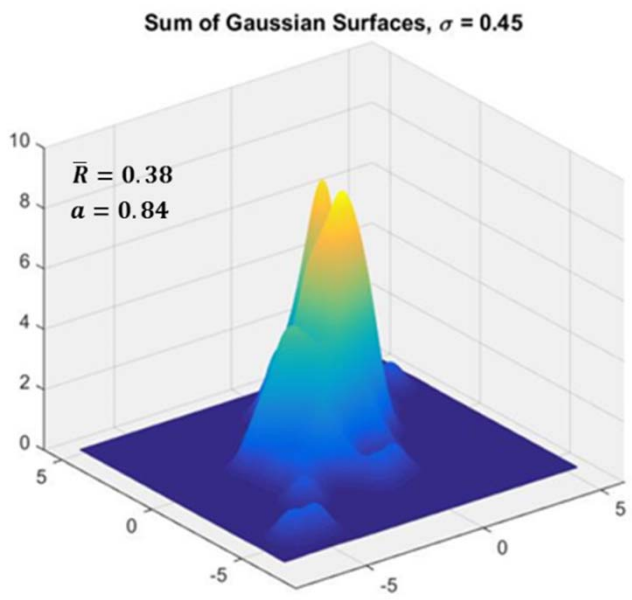
Sum of Gaussian Surfaces
$\sigma = 0.56$

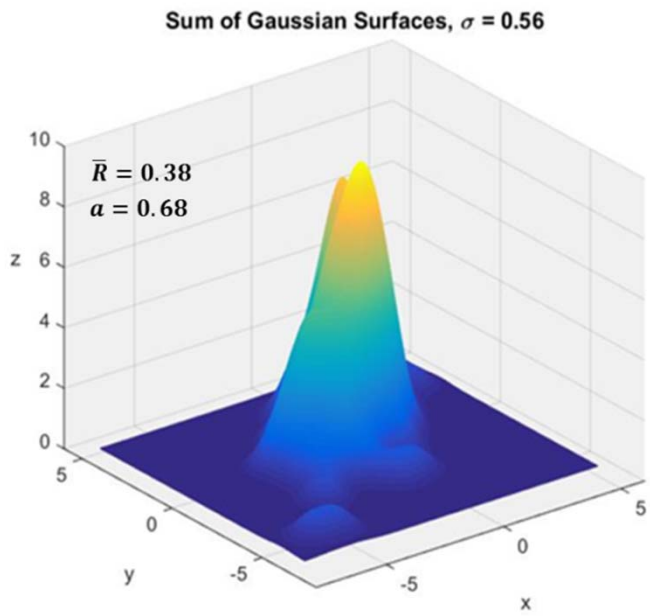


$$\sum f_i(x,y)$$

Create a 3D Density of the M&S Miss Distances







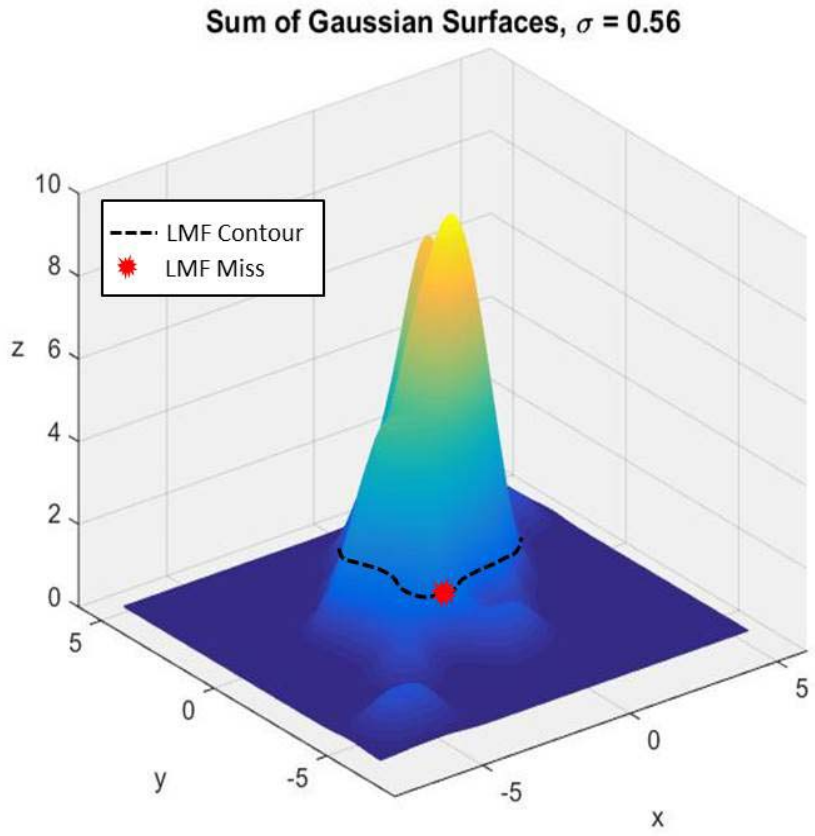
increasing σ

Note: notional data is used in this briefing to illustrate the proof of concept

Determine the Contour Line Corresponding to the LMF Miss Distance

1. **Determine the height of the LMF Miss on the surface of the density plot**

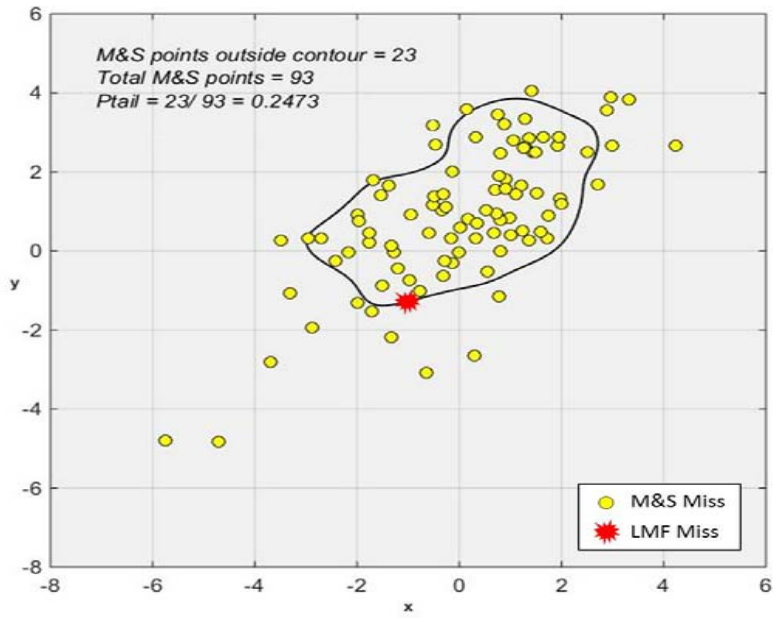
2. **The LMF Contour line is the intersection of the horizontal plane defined by the LMF Miss and the surface of the density plot**



Note: notional data is used in this briefing to illustrate the proof of concept

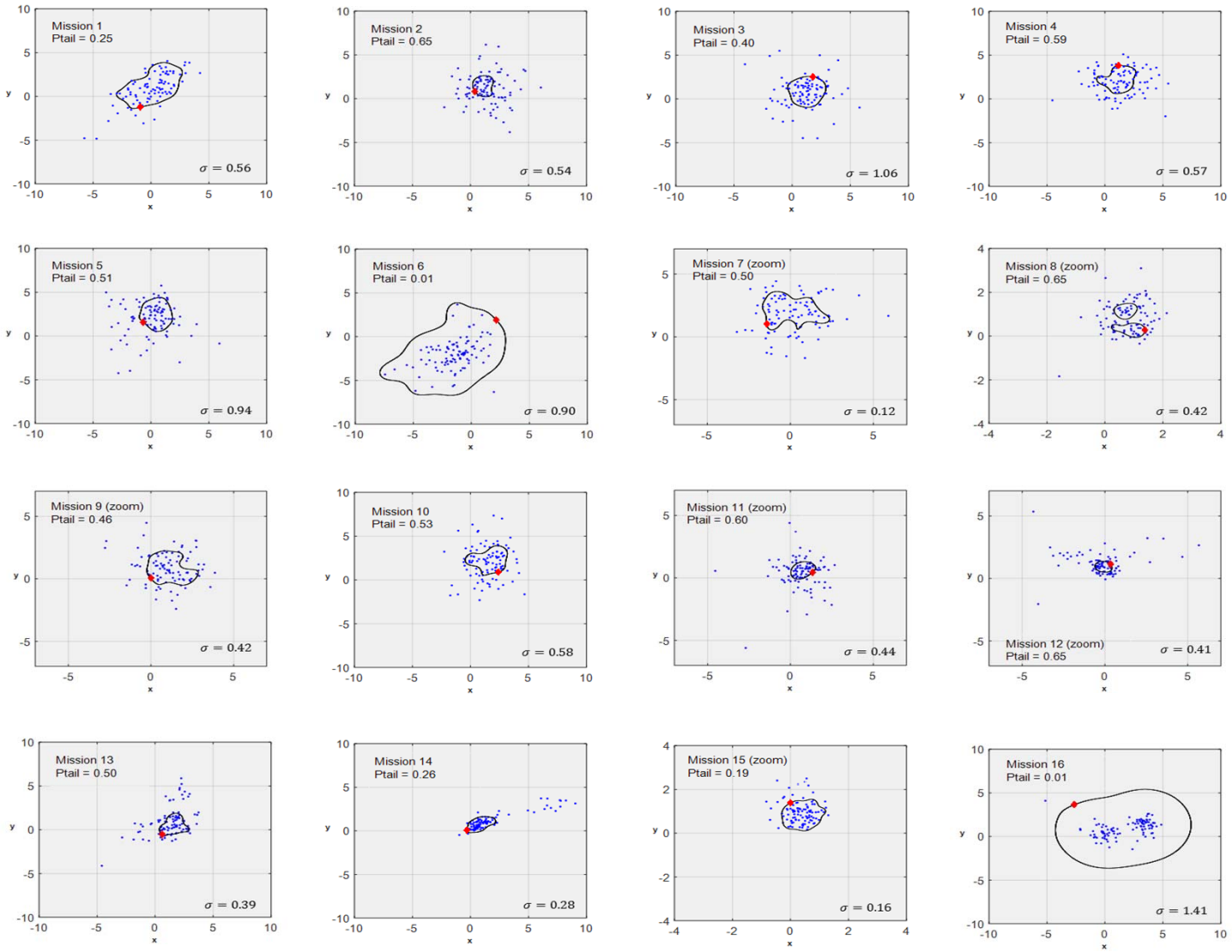
Calculate a P_{TAIL}

$$P_{TAIL} = \frac{M\&S\ Points\ Outside}{Total\ M\&S\ Points}$$

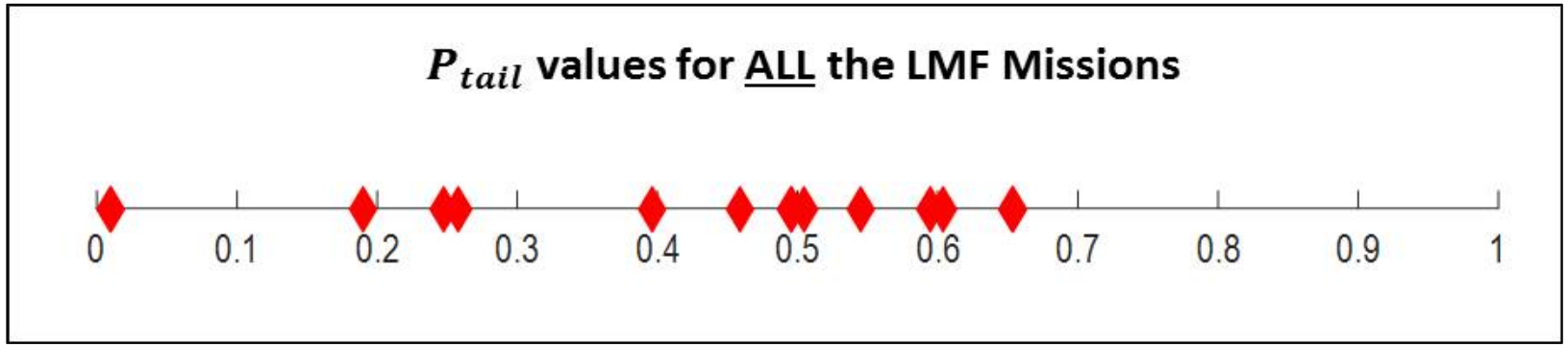


Example:

LMF Miss Distances from 16 Events, each of which as one Live Missile Fire miss distance accompanied by 90-105 Simulated Miss Distances



Summary of the P_{Tails}



$$H_{0_G}: Simulation\ Distriubtion = Live\ Fire\ Distribution$$
$$H_{A_G}: Simulation\ Distriubtion \neq Live\ Fire\ Distribution$$

Calculate the p_i for the two sided test

$$p_i = \begin{cases} 2(P_{TAIL}) & if \quad (P_{TAIL}) \leq 0.5 \\ 2(1 - P_{TAIL}) & if \quad (P_{TAIL}) > 0.5 \end{cases}$$

Does the M&S distribution match the Live Fire Distribution?

H_{0_G} : M&S Distribution = Live Distribution
H_{A_G} : M&S Distribution ≠ Live Distribution

(translate the P_{TAILS} for the two-sided hypothesis test)

$$X = -2\sum_{i=1}^{n} \ln(p_i) = 24.36$$

(follows a chi-square distribution with 2n = 32 degrees of freedom)

The p-value for the global hypothesis test is 0.83

CONCLUSION: the observed data are likely with a true null hypothesis (assuming a significance level of 0.05). We fail to reject the null hypothesis and conclude that the M&S data sufficiently represent the collection of Live Missile Fire data

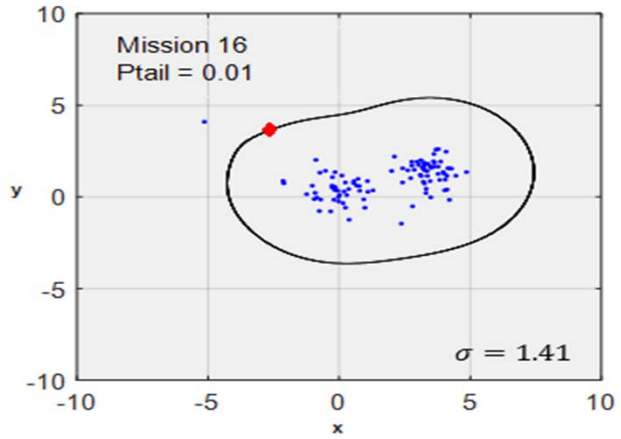
Limitations of Fisher's Test: One "outlier" can dominate completely

A <u>single</u> small p-value can lead to the rejection of the null hypothesis

$$\ln(\text{small number}) = \text{large number} \rightarrow \text{large test statistic} \rightarrow \text{rejct the null}$$

BUT rather than dismiss the usefulness of the test in this situation one should explore and discuss the "outlier" p-value

- **Do we know how and why it occurred?**
- **Should a single test data outcome invalidate the M&S?**

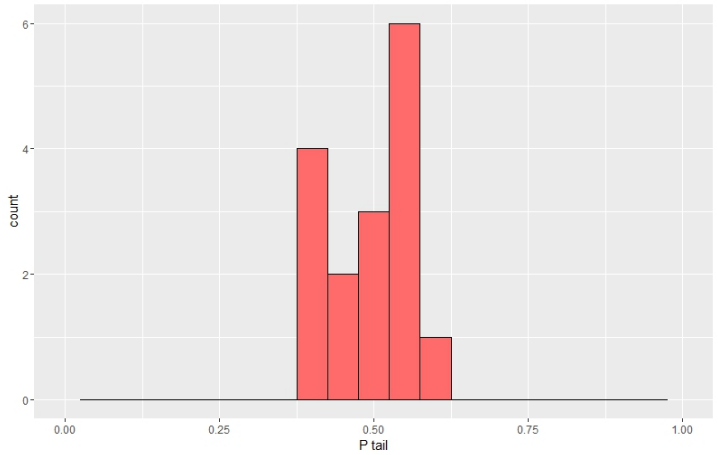


Why not just use a Goodness of Fit Test? (e.g. Kolmogorov-Smirnov)

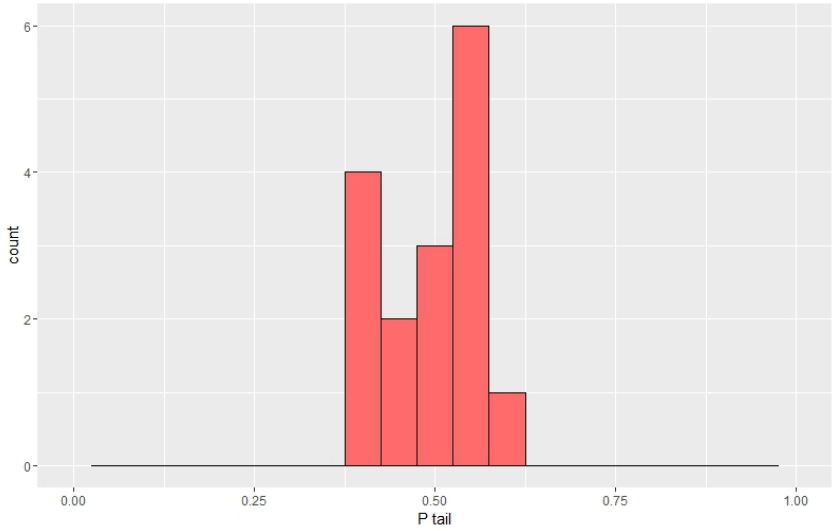
Under any formulation of the null hypothesis, the $p_i's$ are uniformly distributed on the interval [0,1] . . .

Consider the case where we have a clustered concentration of p-tails all in close proximity to 0.5.

Clearly not uniform…but do would we really want to reject the global hypothesis?



Fisher's Combined Test is completely insensitive to such circumstances



$$\ln\left(\frac{1}{2}\right) \leq 1$$

when $p_i = 0.50$

Fisher's Test (global hypothesis test)
(Two-Sided Test)

p-value = 0.889

KS Test (compare to uniform)
(Two-Sided Test)

p-value = 0.005

Note: notional data is used in this briefing to illustrate the proof of concept

Limitations of Fisher's Test (cont.)

Construction of contours requires some level of judgement

We validated that *miss distance* could not be shown to be biased

However, getting from miss distance to Pk requires one more step, a warhead lethality model, which is validated separately through another test program

Summary

Live missile fire testing is extremely limited due to test constraints and cost

Before M&S is used to further the understanding of the system, it must be validated.

Fisher's Combined Test can be used as a tool to validate AIM-9X simulations