Application of
Fisher's Combined Probability Test
To the Validation of the AIM-9X Missile Model

U.S. Army Conference on Applied Statistics
Napa Valley, CA
October 2003

Arthur Fries
Institute for Defense Analyses
afries@ida.org

ACKNOWLEDGMENTS & DISCLAIMER

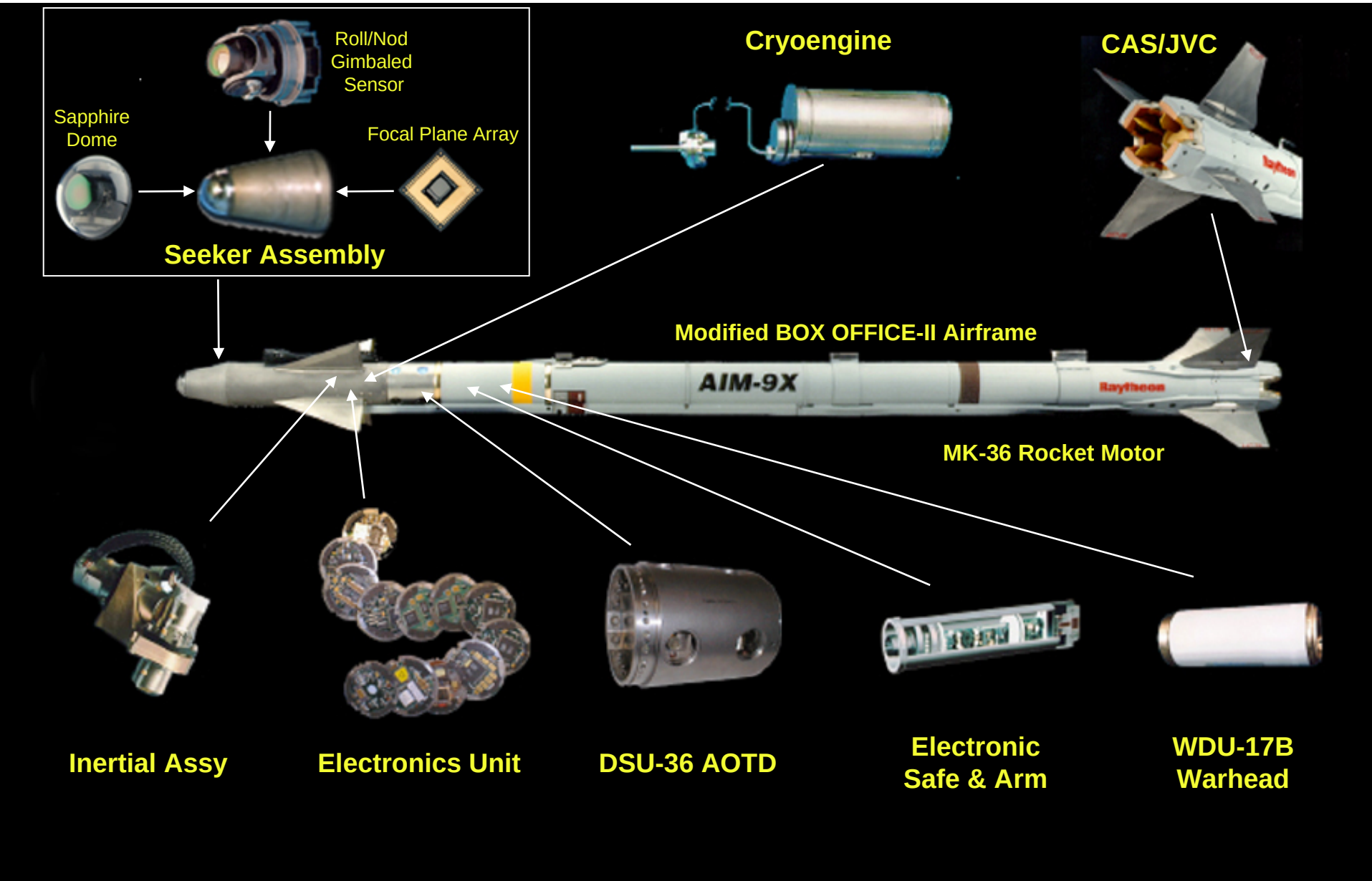
Portions of Arthur Fries' research were undertaken at the Institute for Defense Analyses (IDA) under IDA Central Research Project C9016, under IDA Professional Development Funding, and as part of various tasks sponsored by the Office of the Director of Operational Test and Evaluation (ODOT&E) in the Office of the Secretary of Defense (OSD) within the U.S. Department of Defense (DoD). The author gratefully acknowledges helpful technical review comments and inputs provided by A. Rex Rivolo and Robert R. Soule, IDA, numerous supporting analyses conducted by LeAnna Guerin, IDA summer intern, and briefing material graciously provided by Denise Beattie and Leonard Vance from Raytheon.

The contents of this presentation and the views therein are solely those of the author. No official endorsement by IDA, DoD, or ODOT&E is intended or should be inferred.

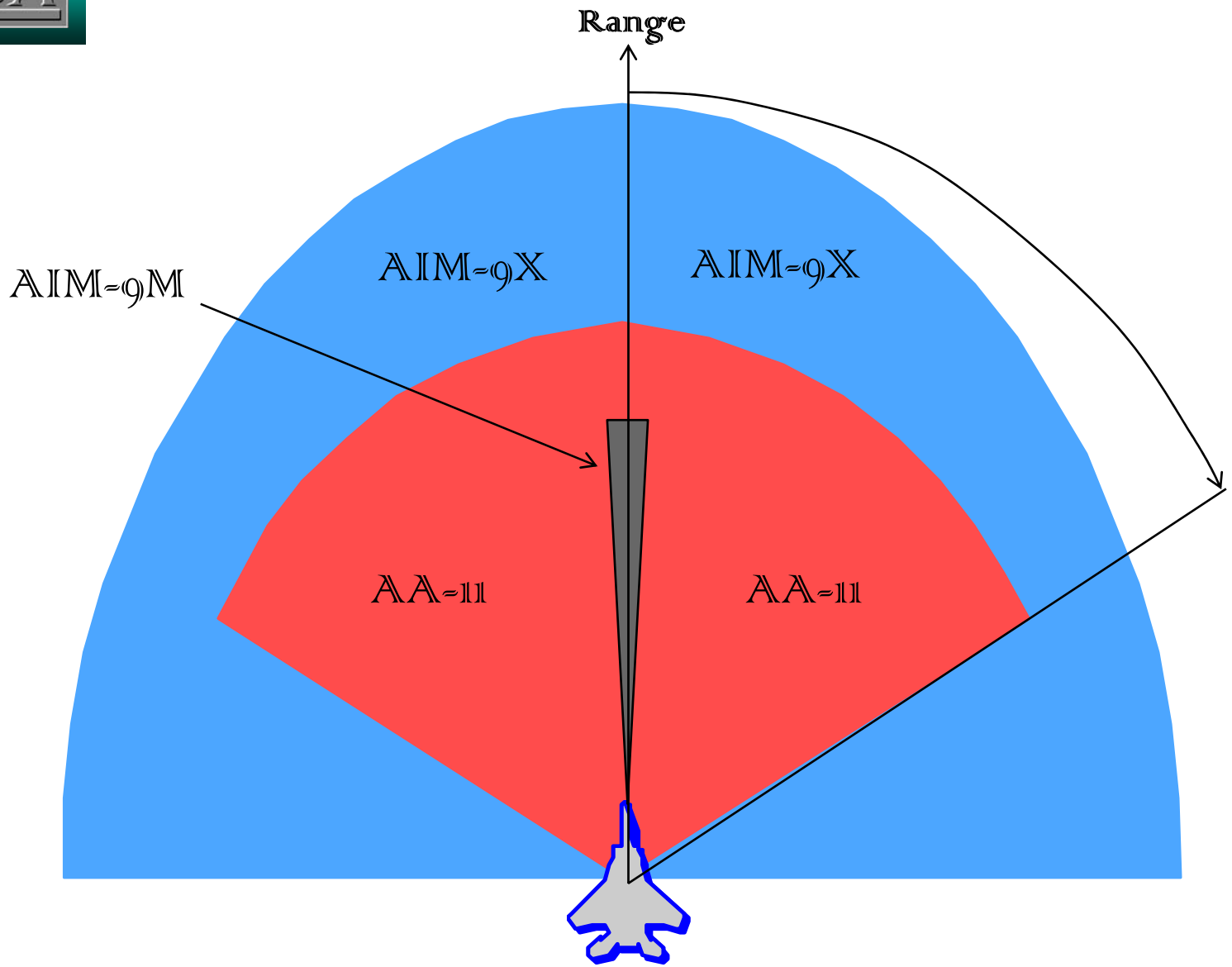
OUTLINE

- **AIM-9X**
  - System Description
  - VV&A Considerations
    - » How much to test?
    - » Where to test?
    - » How to analyze?
- **Fisher's Combined Probability Test**
  - Connection to M&S
  - Methodological considerations
    - » Accommodating measurement errors
    - » 1-sided or 2-sided testing?
    - » Other meta-analysis approaches
    - » Goodness-of-fit procedures
  - Statistical power
- **Application to AIM-9X**
- **Other Challenges**

AIM-9X: System Description



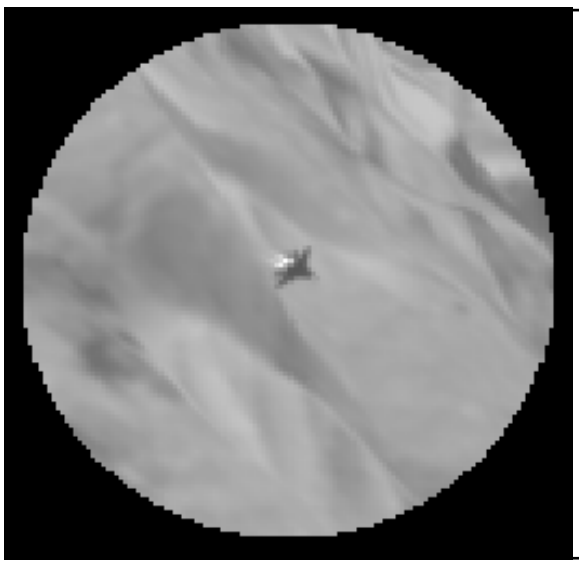
AIM-9X Launch Envelope (notional)

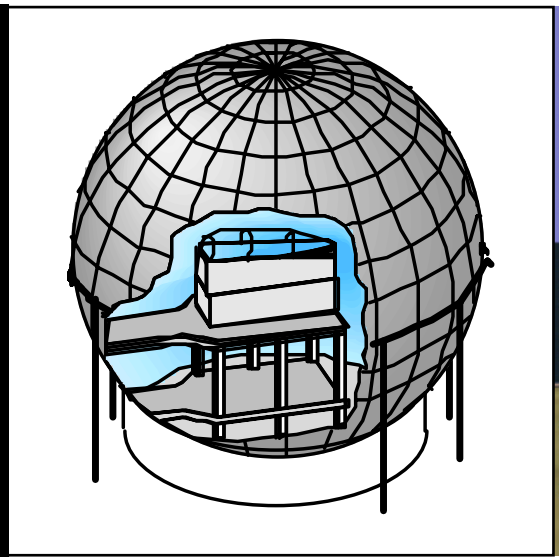


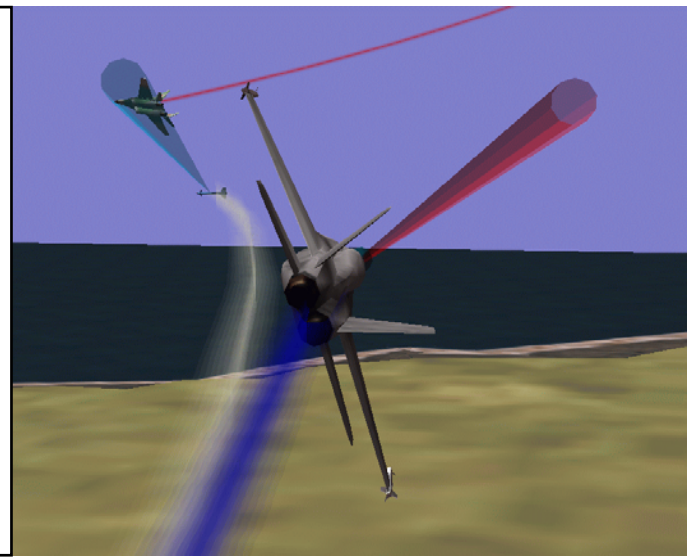
Modeling & Simulation Based Acquisition Program

- **AIM-9X Relies on Extensive Modeling & Simulation**
  - Improve the quality of flight tests
  - Reduce costs

- **Simulation Is Used to Predict All Levels of System Performance**
  - Weapon System, Missile System, Seeker, Kinematic
    - » Pk Performance Verified by Simulation Only
    - » Guided Flight Tests Used for Simulation Validation
  - Integrated Plan With Acquisition, DT, & OT Communities







Why Simulation Based Acquisition?

Simulation Based Acquisition (SBA) can lead to dramatic reductions in weapon system testing

Example: AMRAAM vs AIM-9X Full Scale Development

AMRAAM (1985-1990)	
Control Test Vehicle Flights*	29
Development Guided Flights	79
Operational Test Flights	17
Total Guided Flights	96

AIM-9X (1999-2002)	
Seperation Control Test Vehicle Flights*	22
Development Guided Flights	16
Operational Assessment Flights	5
Operational Test Flights	22
Total Guided Flights	43

Key Elements of Simulation Credibility

- **Software Accuracy**
  - Correctness of implementation
  - Accuracy of any data manipulations
    - » Unit conversions, coordinate transformations, etc.
  - Accuracy of software documentation

Measured by "verification"

- **Output validity**
  - Comparison with the "real world" (properly defined)
  - Accuracy of simulation inputs

Measured by "validation"

- **Software Stability**
  - Adequate management and control of M&S software, embedded data, and software documentation

"Configuration Management"

- **Data Accuracy**
  - Accuracy of any embedded data
    - » Physical constants, system parameters, etc.
  - Accuracy of any input data

"Data Quality"

Accreditation activities ensure robust assessments of simulation credibility in all these dimensions

WHAT IS "VALIDATION"?

- **Broad VV&A Context (AR 5-11, Management of Army Models and Simulations, 1997)**
  - "Validation is the process of determining the extent to which the M&S adequately represents the real-world from the perspectives of its intended use."

- **Limited Statistical Context**
  - Quantify consistency between observed data (real world or test) and predicted outcomes
  - Provoke detailed follow-on discussions and analyses

- **No "One-Time" Validation**
  - Same model for different purposes
  - Continuous comprehensive evaluation

AIM-9X Flight Test Matrix

	DTIIB/C				OTIIA					DTII B/C					DTIID			
	Guided Flight Sequence																	
System Simulation Aspect Validated	e1	e2	e3	e4	o1	o2	o3	o4	o5	p1	p2	p3	p4	p5	p6	p7	p8	p9
Target Signature (**L**ow, (**-**)med, **H**i)	-	-	-	-	L	-	-	-	-	-	-	H	-	-	-	-	-	-
Target Terminal Aspect (**N**ose, **T**ail, **P**lanform, **B**eam)	B	P	P	P	N	B	N	B	P	B	P	B	P	B	P	N	P	T
Closing Velocity (**L**ow, (**-**)med, **H**i)	L	-	-	-	-	-	-	-	-	-	-	-	L	-	H	-	L	L
OBA ((**-**)Low, **H**igh(vignetting))	-	-	H	H	H	-	H	-	H	-	H	H	H	-	-	-	-	-
CM (**m**idcourse,**t**erminal w/ **A**,**B**,**C**,**AC** or (**-**) none	-	-	-	mB	-	-	-	mB	mB	B	B	-	A	AC	-	mA	C	
CM - Lag**(-),** Lead (**+**), (**-**)n/a	-	-	-	+	-	-	-	+	+	+	(-)	-	(-)	+	-	(-)	+	+
Altitude (**L**ow, (**-**)**M**ed, **H**igh)	-	-	-	-	-	-	-	-	-	-	-	-	-	L	H	H	L	H
Background type: (blue sky [**bs**], **d**esert, **c**loud, **s**ea, **h**orizon)	bs	d	c	d	bs	bs	bs	d	d	d	bs	h	d	s	bs	bs	d	bs
Contrast: positive(**+**), washout(**0**), neg washout**(-)**	+	+	+	+	+	+	+	+	+	0	+	+	+	+	+	+	(-)	+
Range (**L**ow, (**-**)Med, **H**igh)	-	-	-	-	H	L	L	-	-	-	-	L	-	H	H	-	-	H
Target Maneuver ((**-**)None, **I**n-plane, **O**ut of plane)	-	-	I	I	-	I	-	O	I	-	I	-	I	O	O	I	-	-
Initialization (**R**adar, **H**elmet)	R	R	R	H	H	R	H	H	H	R	H	H	H	R	R	R	H	R
Dome heating ((-)Moderate, **H**igh)	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	H	H	-
Afterburner (**O**n, On **M**idgame, On **T**erminal, (**-**)None)	-	-	M	-	-	-	-	-	M	-	M	O	M	O	-	-	T	T
	CL	WS	WS	CL	CL	EG	EG	CL	CL	CL	CL	CL	CL	CL	EG	EG	CL	CL
	30 Jun99	1 Sep99	16 Dec99	31 Mar00	21 Apr00	23 May00	25 May00	28 Jun00	21 Jul00	12 Jun00	26 Sep00	17 Nov00	21 Dec00	5 Apr01	May-01	Aug-01	Sep-01	Oct-01

EDM DT shot

EDM OT shot

PRM DT shot (DTIIB/C)

PRM DT shot (DTIID)

PRM DT Assist

IMPLEMENTATION CONSIDERATIONS

- **How Much To Test?**
- **Where To Test?**
- **How To Analyze?**

HOW MUCH TO TEST?

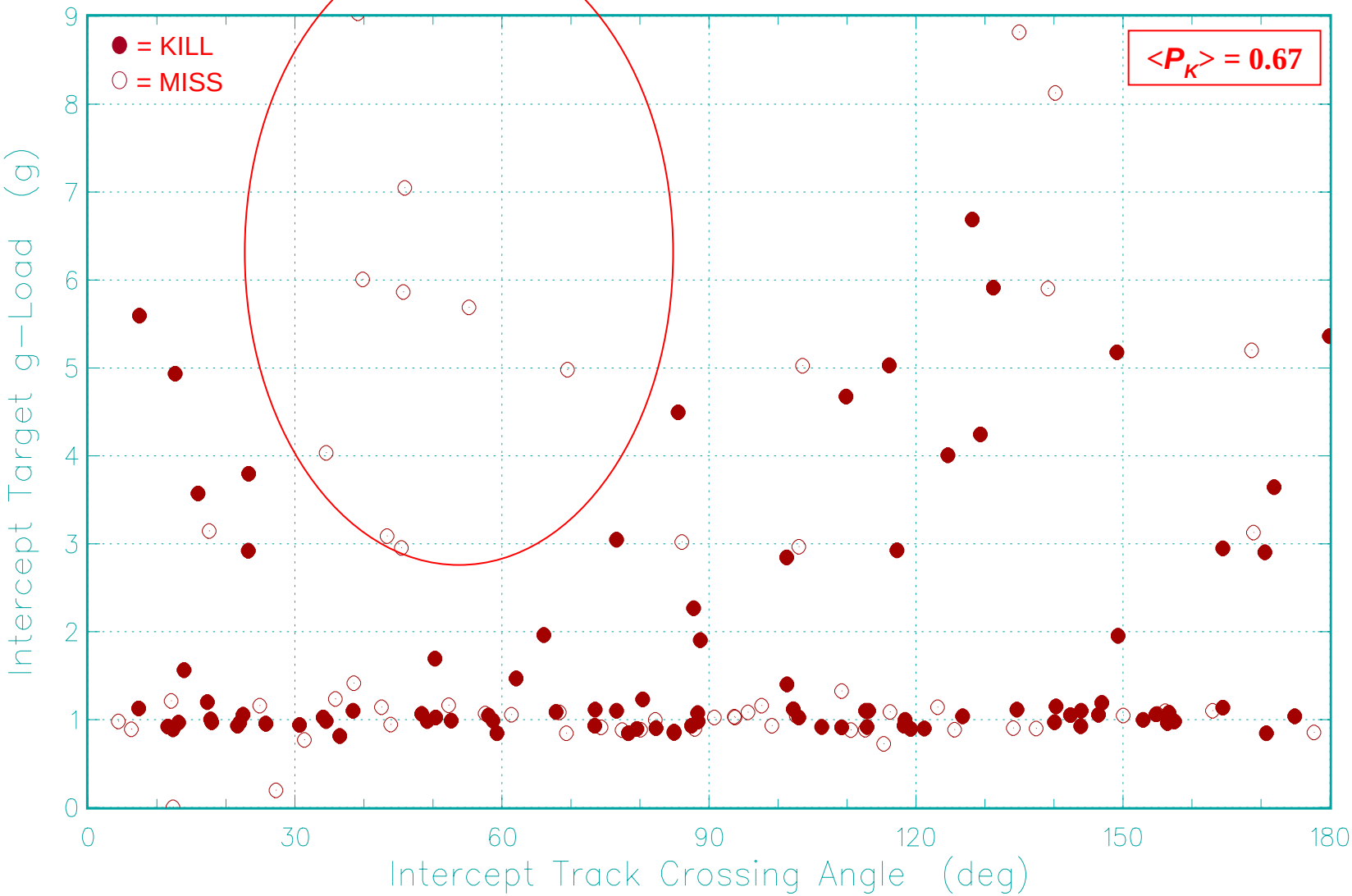
- **Whatever You Can Get**
- **20 – 40 ??**
- **Simulation-Based Investigation**

WHERE TO TEST?

- **One Point Replicated**
- **Few Points & Less Replication**
- **“Many” Points & No Replication**
  - Uniformly distributed
  - Concentrate on disparate “pockets”
    - » Based on M&S insights
    - » Based on physics insights
  - Include specific combinations
    - » Likely
    - » Extreme

M&S ALONE MAY BE INADEQUATE

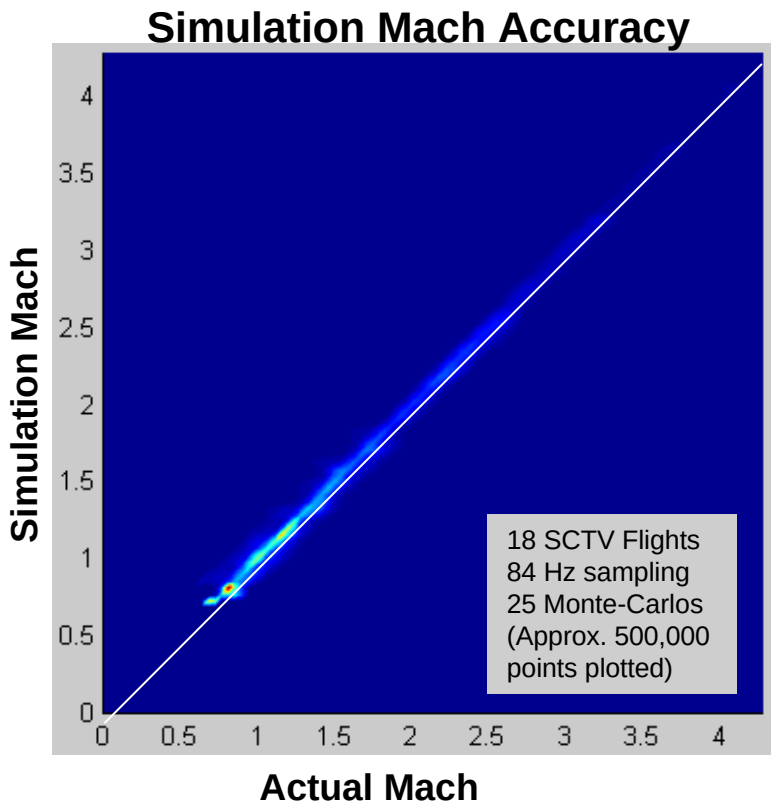
(hypothetical example, not AIM-9X)

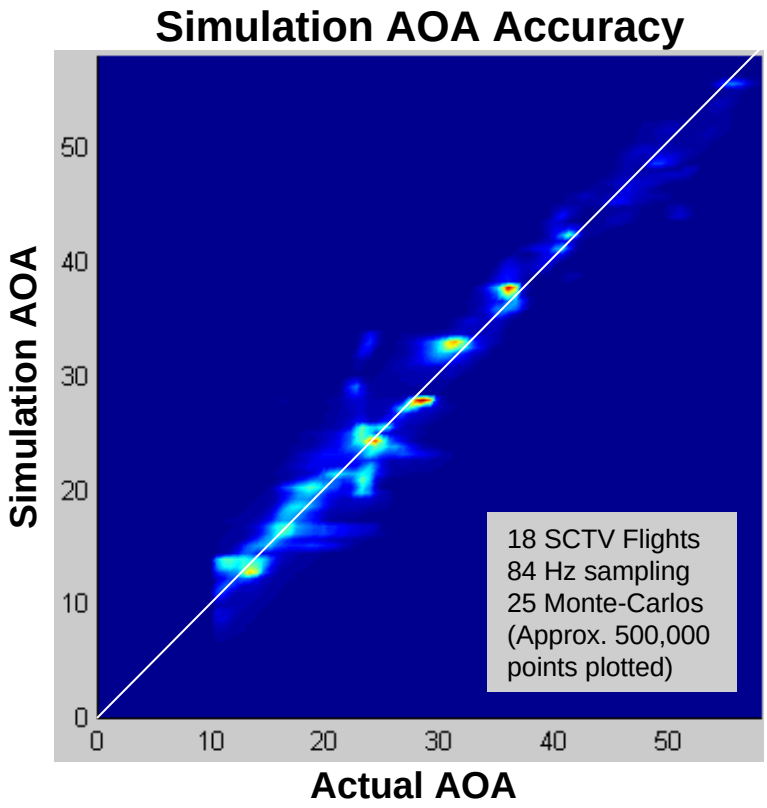


HOW TO ANALYZE?

- **Focus On End-Game Miss Distances**
  - *Old Method*: Look at the actual vs predicted miss distances and use our judgment
  - *Better Method*: Determine if actual miss distances are statistically plausible when compared to simulation predictions

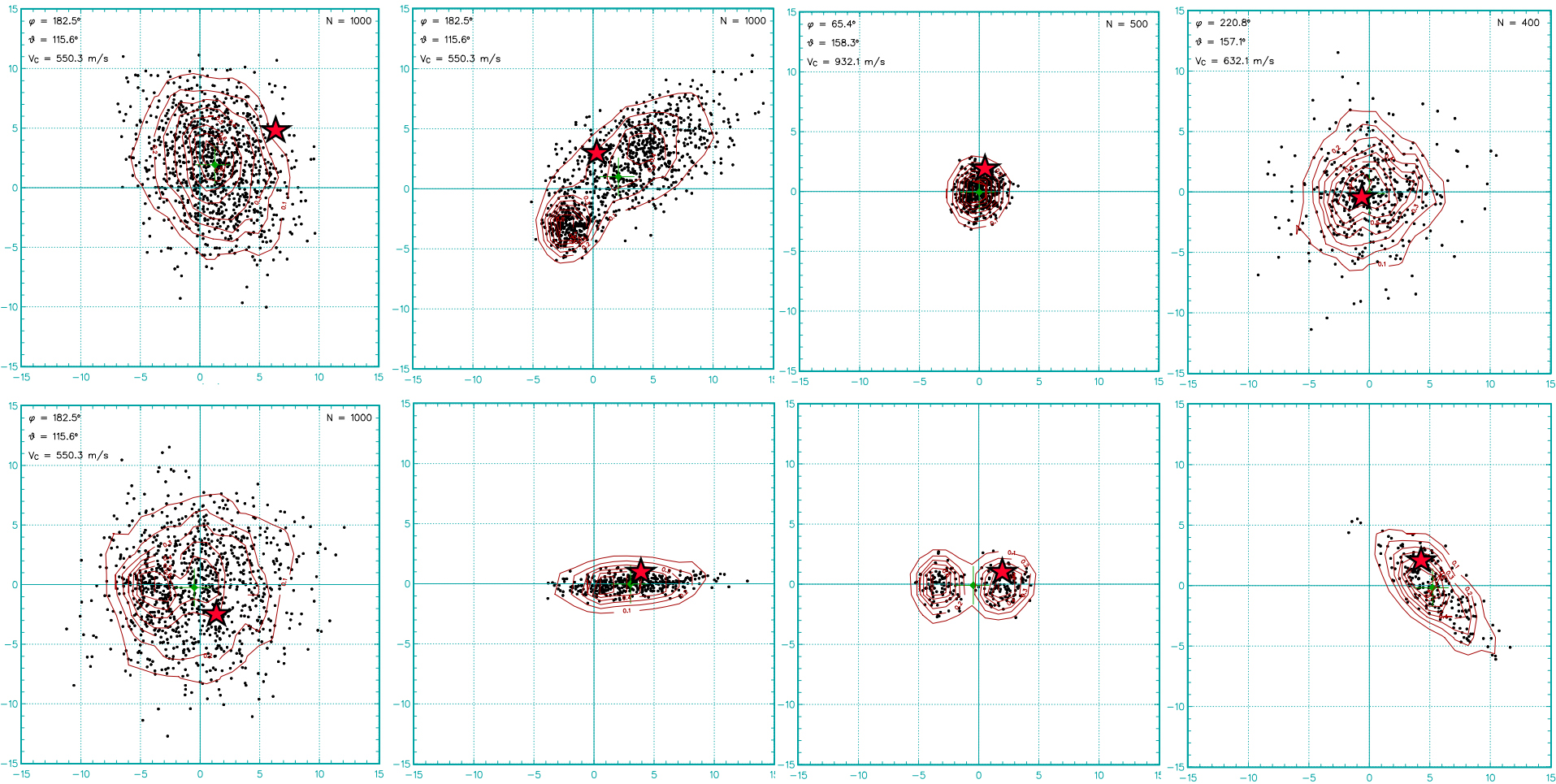
Excellent Kinematic Match





- **Actual vs simulation values of Mach and AOA are plotted for all 18 SCTV flights**
- **Excellent overall 1:1 correlation is evident**

HOW TO ANALYZE MISS DISTANCES?



FISHER'S COMBINED PROBABILITY TEST

- **Meta-Analysis Technique**
- **Combines Information From Different "Experiments"**
  - Same H_0
  - Different settings
  - Small sample sizes
  - Limited statistical significance

FISHER'S STATISTIC

$$F = -\sum_{i=1}^{N} 2\log_e(P_i)$$

Single Experiment P-Value

H_0: P_i is uniformly distributed in the interval [0 - 1.0]

⬇

Implies

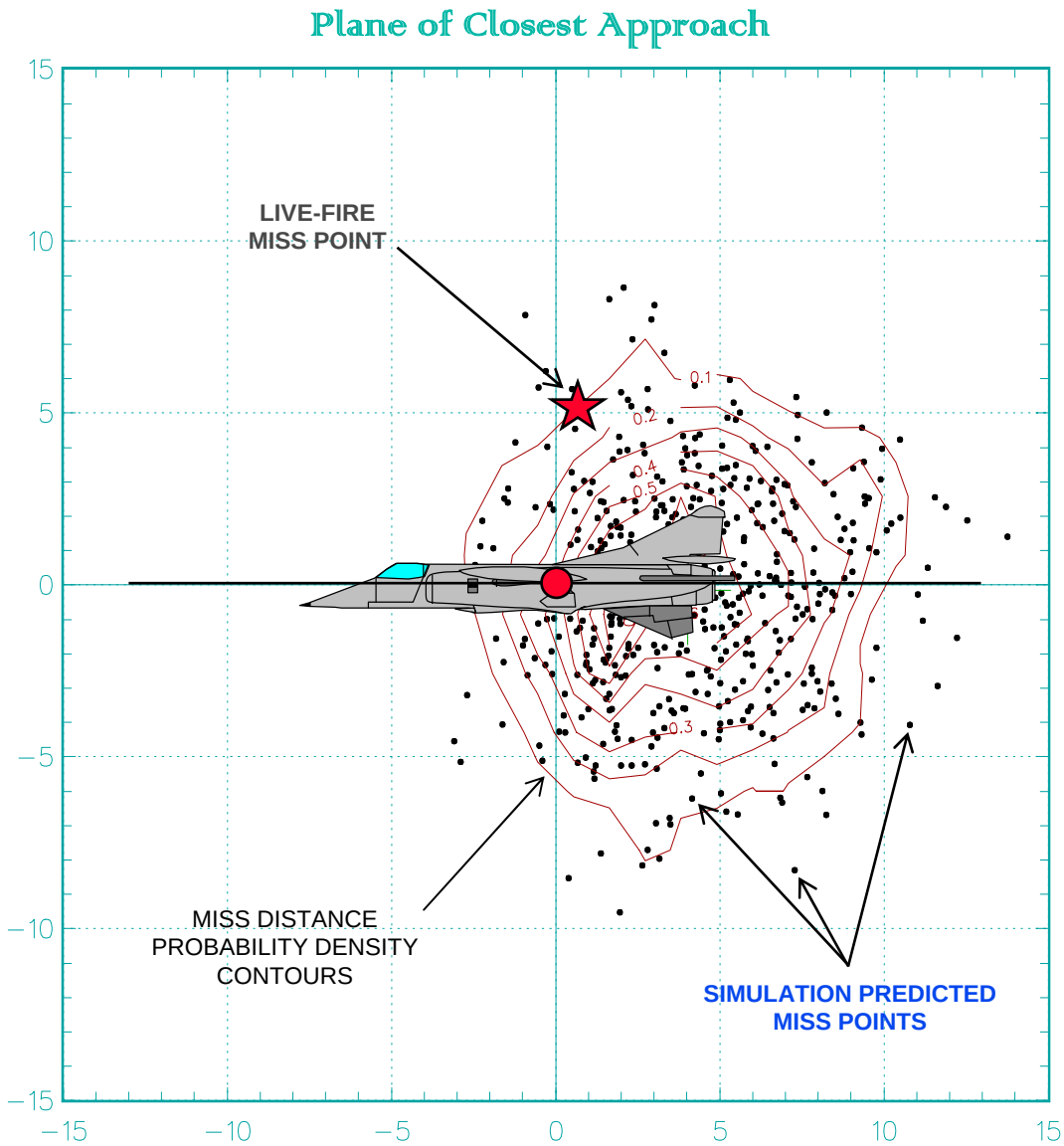
⬇

F is χ^2 distributed with $2N$ Degrees of Freedom

MISS DISTANCE MEASURE OF AGREEMENT

Definition: $\mathbf{P_i}$ (Tail Probability). Fraction of prediction points outside the tangent point contour.

$$P_i = 0.16$$



FISHER'S STATISTIC

$$F = -\sum_{i=1}^{N} 2\log_e(P_i)$$

Single Tail Probability

H_0: P_i is uniformly distributed in the interval [0 - 1.0]

⬇

Implies

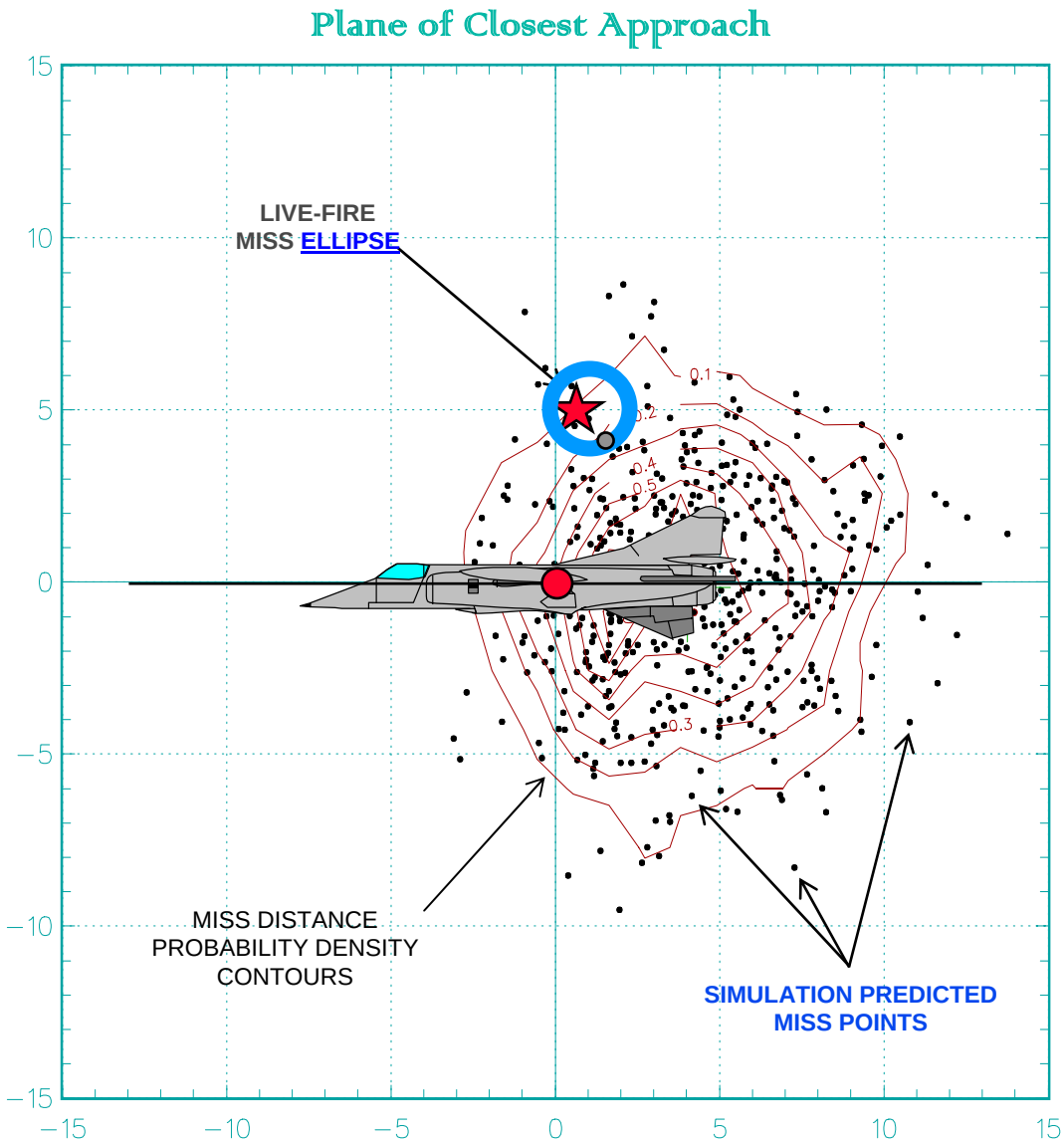
⬇

F is χ^2 distributed with $2N$ Degrees of Freedom

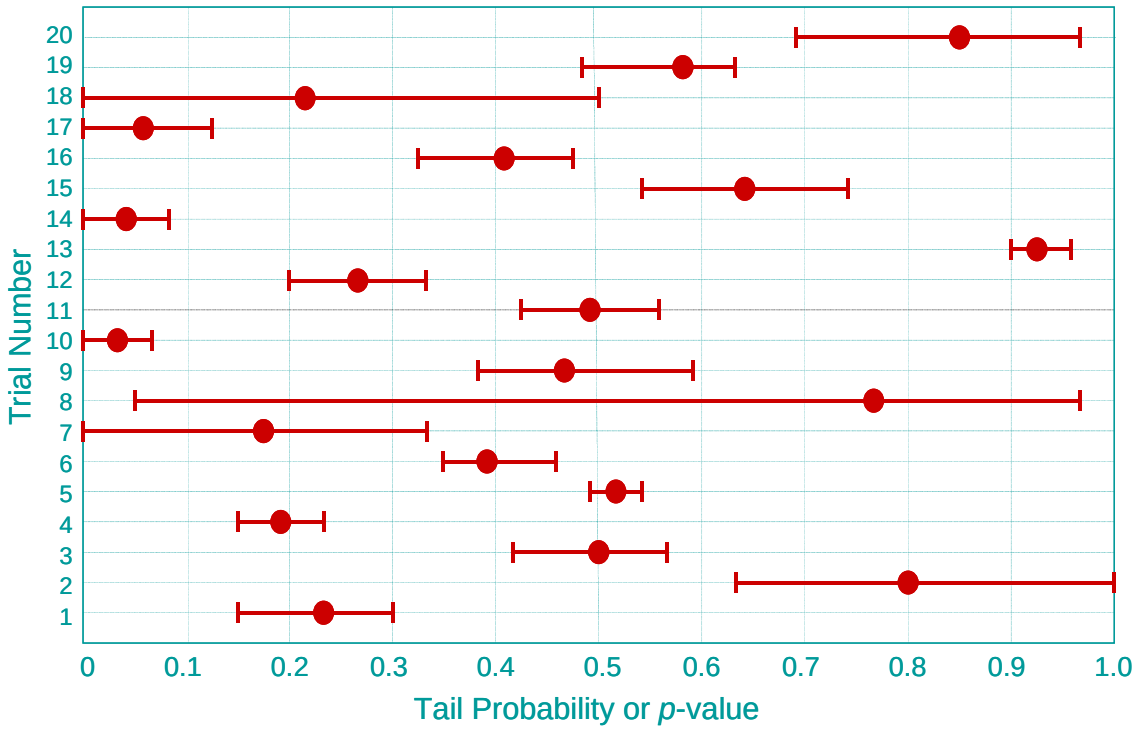
MISS DISTANCE MEASUREMENT ERRORS

Definition: $\mathbf{P_i}$ (Tail Probability). Fraction of prediction points outside the tangent point contour.

$P_i = 0.02$ - 0.28

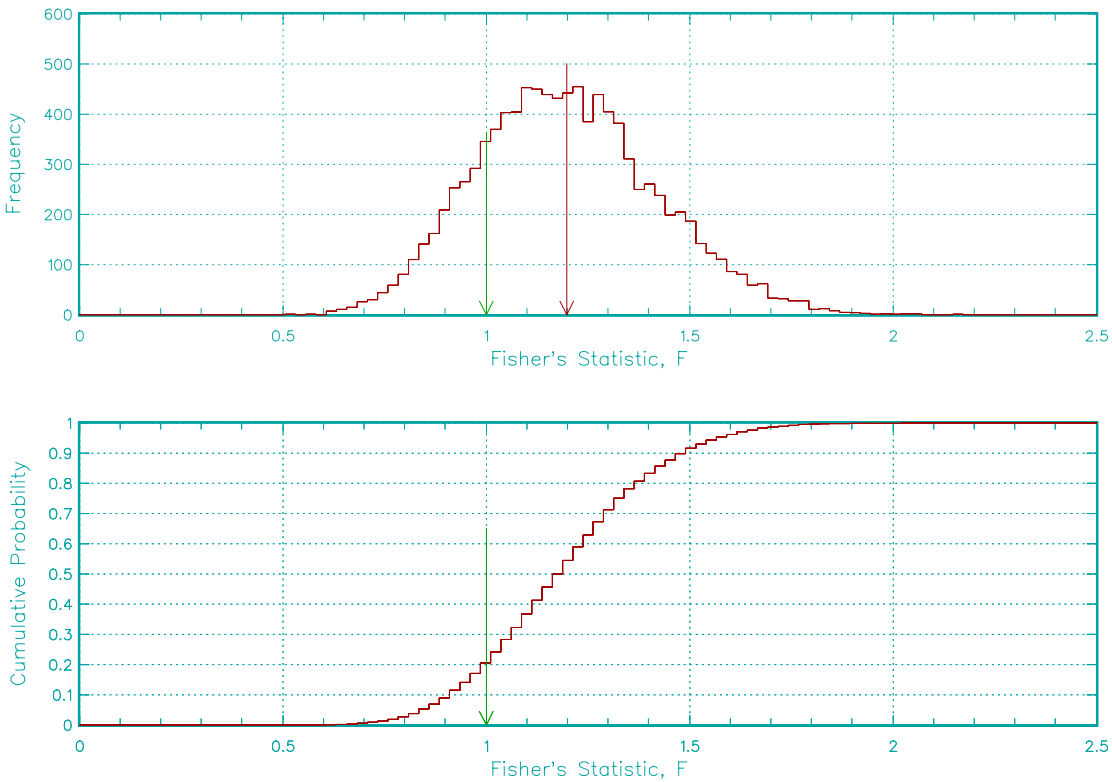


ENCOMPASSING MEASUREMENT ERRORS



Derived Tail Probabilities (*p*-values)

BOOTSTRAPPED FISHER STATISTICS



Distribution of 10,000 *F*-Values and Cumulative Probability

WHICH TAIL(S)?

- **Far-out ↔ Optimistic model**

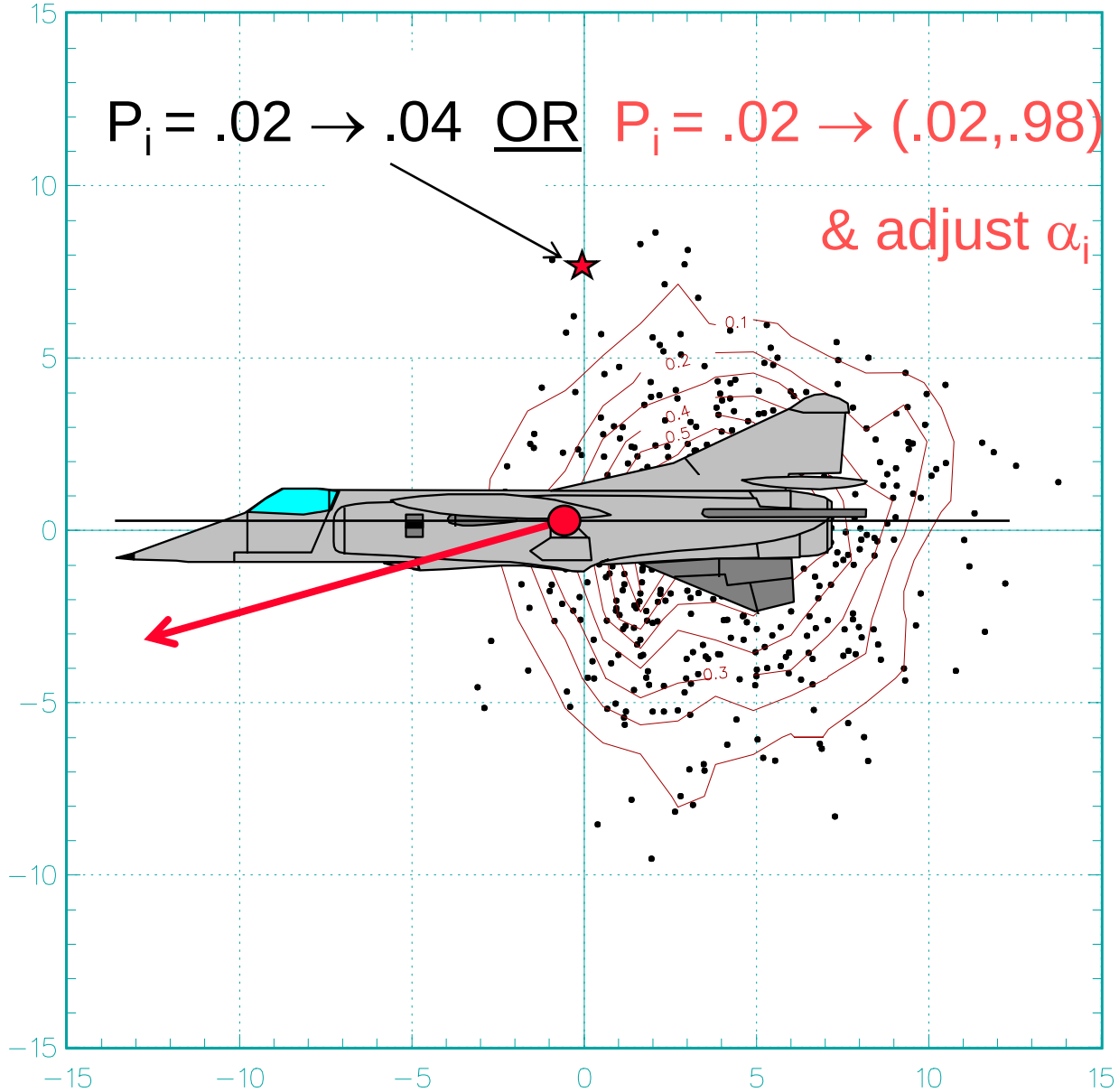
  (big misses are "bad")

- **Close-in ↔ Pessimistic model**

  (small misses are "bad")

- **Both tails ↔ Both concerns**

CONVERSIONS TO 2-SIDED P_i



EXAMPLE CALCULATIONS

- **Simulation Model Optimistic?**
  - $X = 2$ [- ln(.12) - ln(.37) - … - ln(.39)] = 14.4
  - X_O << $\chi^2_{16}(0.05) = 26.3 \Rightarrow$ NO!
- **Simulation Model Pessimistic?**
  - $X = 2$ [- ln(.88) - ln(.63) - … - ln(.61)] = 15.2
  - X_P << $\chi^2_{16}(0.05) = 26.3 \Rightarrow$ NO!
- **Simulation Model "Bad"?**
  - $X = \max(X_O, X_P) << \chi^2_{16}(\underline{0.025}) = 28.8 \Rightarrow$ NO!

META-ANALYSIS APPROACHES

- **<u>One</u> "Outlier" Test Result & Fisher ⇒ "Invalid"**
- **Other <u>Tempered</u> Meta-Analysis Procedures**
- **Endorse Classical Fisher Methodology**
  - Encourages investigation of "outliers"
    - » Test anomaly?
    - » Physical causes represented in model?
  - Promotes discussion
    - » Concept of "invalid"
    - » Statistical significance" vice "practical significance"

GOODNESS-OF-FIT PROCEDURES

- **P_i's Consistent With A Uniform Distribution?**
  - Fisher procedure weights small values heavily
    - » $-\ln(p) \uparrow \infty$ as $p \downarrow 0$
  - Other standard goodness-of-fit tests
    - » Maximum difference in cdf's
    - » Averaged difference

STATISTICAL POWER

- **When the Simulation Model Is "Good"**
  - $S \equiv T$, for all factor combinations
  - With high confidence, should not reject H_0
  - Should attain nominally prescribed α's
- **When the Simulation Model Is "Bad"**
  - $S \neq T$, for some/all factor combinations
  - With high confidence, should reject H_0
  - More "powerful" procedures reject more often

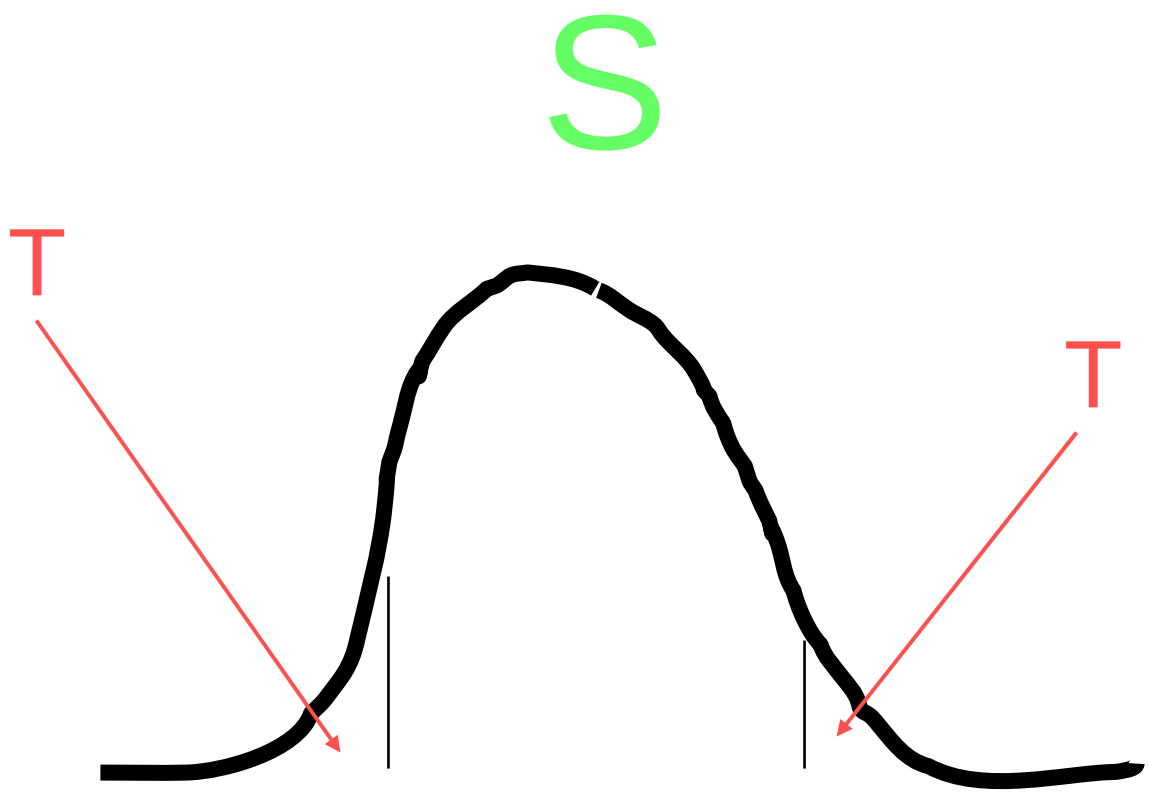
MONTE CARLO ANALYSES

- **Assume "w.l.o.g." $S_i \sim n(0,1)$**
  - Each S_i is "known"
- **Consider $T_i \sim [n(\mu_i,\sigma_i^2)]^{**}\rho_i$**
  - Change mean, variance, shape
  - *t*-test (*Z*-test) optimal for $\mu_i = \mu$, $\sigma_i = \rho_i = 1$
- **Statistical Characterizations**
  - *t*-test (2-sample paired)
  - Fisher
  - Goodness-of-fit

FALSE REJECTIONS

- **By construct, all procedures attain pre-set α's**
- ***t*-test relies on "known" $S_i \sim n(0,1)$**
  - Could generalize to any $n(\mu_i,\sigma_i^2)$
  - But not to "irregular" prediction distributions
- **Other procedures rely merely on "known" S_i**
  - Appropriate for any prediction distribution

VALID REJECTIONS - CASE 1

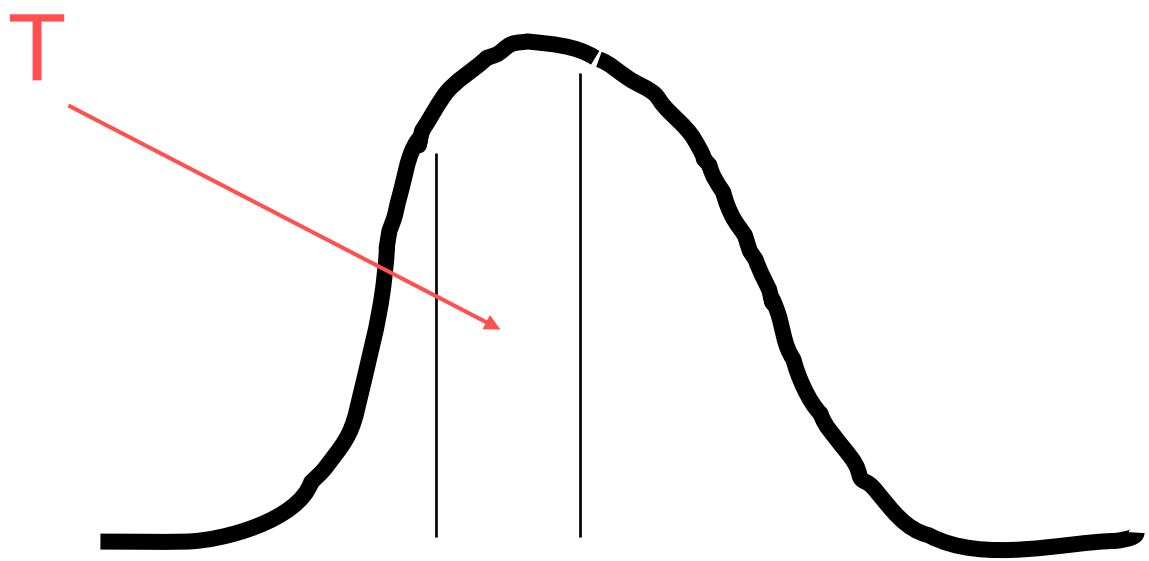


Fisher: Powerful for T in the tails of S

$(\pm \mu, + \sigma^2, + \rho)$

VALID REJECTIONS - CASE 2

S



Fisher: Weak for T in “middle” of S

$(-\sigma^2, -\rho)$

POWER - CASE 1

- **Fisher Dominates**
  - –Fisher >> t-test
  - –Fisher > ≈ GOF's

- **Exception: Shift μ only**
  - –t-test is "optimal"
  - –Fisher is relatively efficient
    - » $S \sim \mathbf{n}(0,1)$, $T \sim \mathbf{n}(0.5, 1)$, N = 25
      - *t*-test 0.88 0.78 0.67 0.40
      - Fisher 0.85 0.77 0.65 0.39

POWER - CASE 2

- ***S* ~ n(0,1), N = 25, α = 0.05**
  - » *T* ~ **n**(0, 0.6^2), Fisher = 0
    - GOF 0.55-0.83
    - t-test 0.05
  - » *T* ~ **n**(0.4, 0.6^2), Fisher = 0.04
    - GOF 0.21-0.38
    - t-test 0.41

- **GOF >> Fisher**

- **GOF > ≈ t-test**

“VALIDATION” CONCERNS

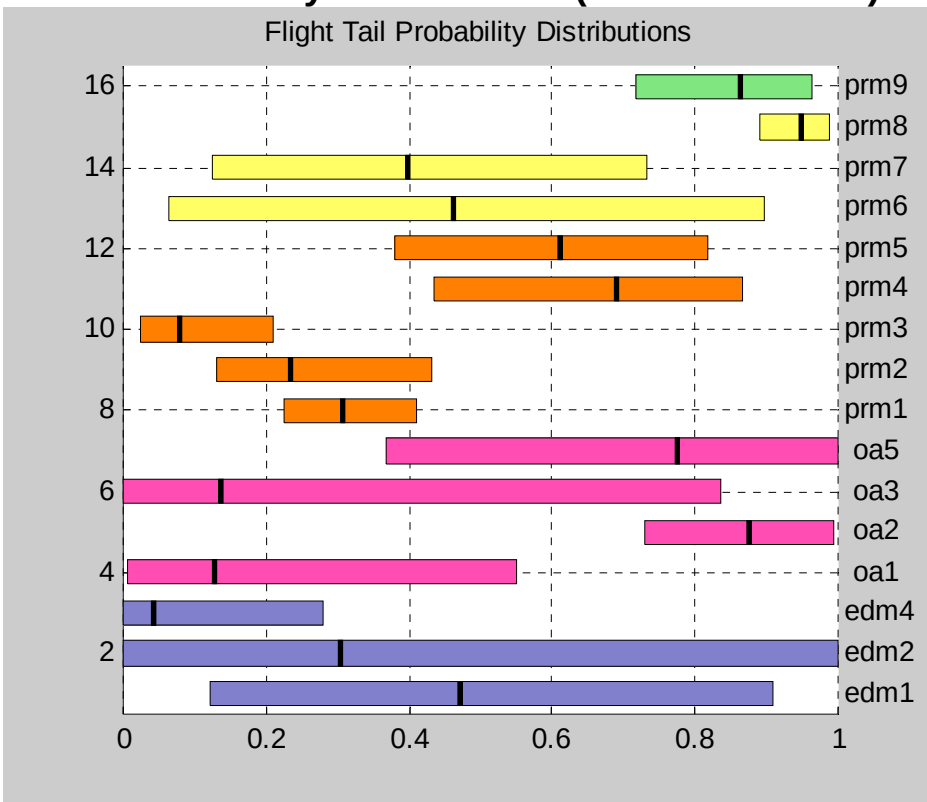
- **Case 1 >> Case 2**
  - –Optimistic Model!!
  - –Pessimistic Model??

LIKELIHOOD OF CASE 2?

- **1-Dimensional Model**
  - –Accurate μ's
  - –Overestimation of σ's
- **Multi-Dimensional Model**
  - –Annuli

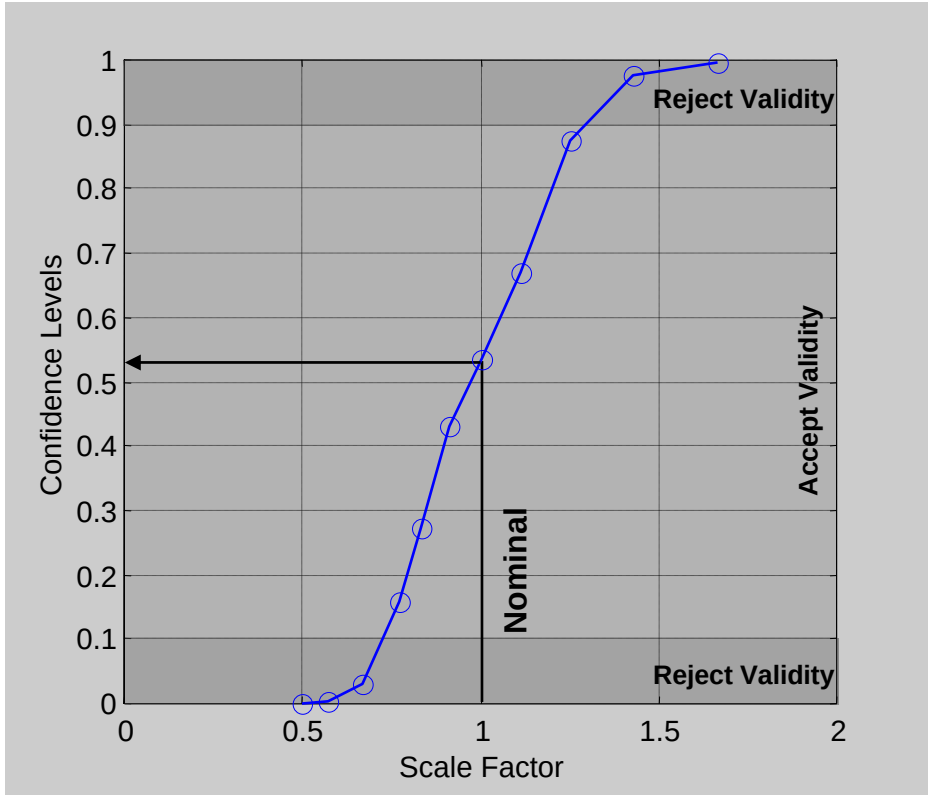
APPLICATION TO AIM-9X

Tail Probability Distributions (w/ uncertainties)



Tail probabilities distributed uniformly from 0 to 1 indicate simulation miss distance is valid

Summary Fisher Miss Distance Results



A Confidence level between 0.1 and 0.9 for nominal miss distance scaling factor (SF = 1) indicates sim miss distance prediction is valid

AIM-9X M&S VALIDATION HISTORY

- **Don't Need To Do It**
- **Can't Do It**
- **Can Only Test At One Design Point**
- **Will Do It!**
  - Implement rigorous managerial procedures & controls
  - Continually inform all T&E organizations
  - Utilize Fisher methodology
- **Awards From DoD Modeling & Simulation Office**
  - Government M&S award for 2001
  - Contractor M&S award for 2001
- **Recent Flight Test Anomaly**

OTHER CHALLENGES

- **Non-Point Predictions**
  - Curves Generated
  - No Single Primary Point Of Emphasis
- **Deterministic Models**
  - Single Prediction For Any Set Of Inputs