



Advancing Statistical Engineering through a Department of Defense and NASA Partnership

Peter A. Parker

National Aeronautics and Space Administration

peter.a.parker@nasa.gov

Laura J. Freeman

Institute for Defense Analysis

lfreeman@ida.org

58th Fall Technical Conference, Richmond, VA

October 3, 2014



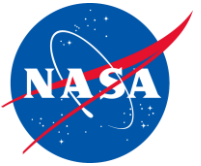
Introduction



- **DoD and NASA develop, acquire, evaluate, and operate some of the world's most advanced and sophisticated systems**
- **Infusing statistical engineering is promoting statistical thinking and the effective application of statistical methods**
 - **Generating better solutions to solve high-impact challenges**
- **To advance and broaden statistical engineering practice, a partnership was developed to share knowledge and experience**
- **In Feb. 2014, an interagency agreement between the Director, Operational Test & Evaluation (DOT&E) within the Office of the Secretary of Defense and the Chief Engineer of NASA was ratified**
- **Partnership is envisioned as catalyst to engage the broader statistical engineering community in industry, academia, and other government agencies.**



Outline



- **Background and Motivation of the Partnership**
- **Elements of the Agreement**
- **Statistical Engineering Case Studies**
- **Implementation Steps**

Highlight effective techniques to institutionalize statistical engineering



Statistical Engineering



Collaborative study and application of the links between statistical thinking, statistical tools, and discipline-specific tools

**Engineering statistical sciences to generate better,
more durable solution approaches**

A Collaborative and Methodological approach to:

- **Define test, research, and development objectives, quantitatively**
- **Specify and negotiate adequate resources**
- **Execute strategically and increase the likelihood of success**

It is Not

- **Replacement for good science and engineering**
- **Fix-all solution to all problems - very careful to not oversell**



Collaboration and Methods



Effectiveness requires a collaborative, team member role that requires

- Early involvement
- Education and/or introduction
- Discipline to identify the high-level objectives and impact

Primary methods included in our practice

- Statistical Thinking
- Decision Making, Hypothesis Testing, Inference
- Design of Experiments
- Response Surface Methodology
- Regression Modeling
- Reliability
- << Laura, please add/modify others...>>

Integrating collaboration with the methods (tools) maximizes benefits



NASA Background and Motivation



- **Early 1990s, response surface methodology was applied to computational studies, multi-disciplinary optimization**
- **Late 1990s, design of experiments was introduced to aeronautical ground testing, including wind tunnels and instrumentation**
- **Successes generated interest and further education from within the current workforce**

Impact of this early work

- **Opened the door for broader interest in industrial statistics**
 - **Community unknown to most of the engineering workforce**
- **Recognizing the potential, in 2005, a core team began forming within the Engineering Directorate at Langley Research Center**
- **In 2006, Statistical Engineering solved an Agency-level challenge**



Solid Rocket Motor Requirements



- Leverage heritage rocket motor design for a new application and performance

Objectives and Impact

- Estimate roll torque during live firing to size the reaction control system?
- Impacts requirements, cost (lifecycle), schedule, volume, mass (payload)

Statistical Engineering Applied

- System level perspective on apparatus, available theory and data
- Integrated design of experiments and response surface methods with force measurement expertise
- Measured roll torque, quantified uncertainty, reduced design risk

Embedded in standard test processes through training over 5 years

Impact: Agency level recognition through briefings and awards



NASA Strategy to Broaden Influence



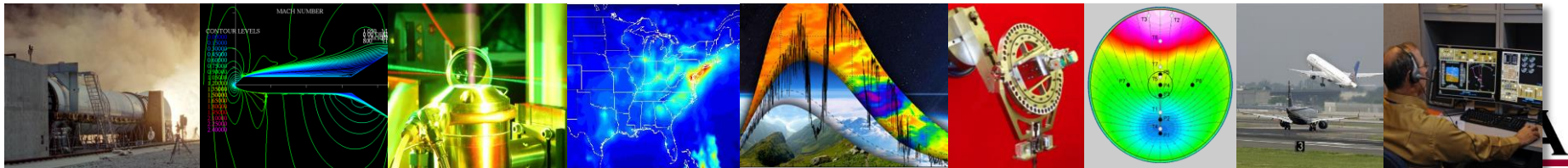
- Over the last 8 years, a **Statistical Engineering Team** is building a critical mass of practitioners and researchers
 - Expertise is sparse within the Agency, training is critical
 - Developed active university collaboration to augment skills and provide a pipeline for capability
- Organized the **NASA Statistical Engineering Symposium, 2011**
 - Cross-discipline: Engineering, Science, Statistics
 - Cross-Agency: DoD, NIST, NOAA, DOE
 - DoD: DOT&E, USAF, NPS
 - Keynote: Dr. Michael Gilmore
 - Cross-Community: Academia, Industry, Government
 - Leaders in Academic community on the topic
- Jointly planned 2013 (postponed) Meeting, hosted by DOT&E
 - Obtained NASA Chief Engineer's support



NASA Statistical Engineering Team

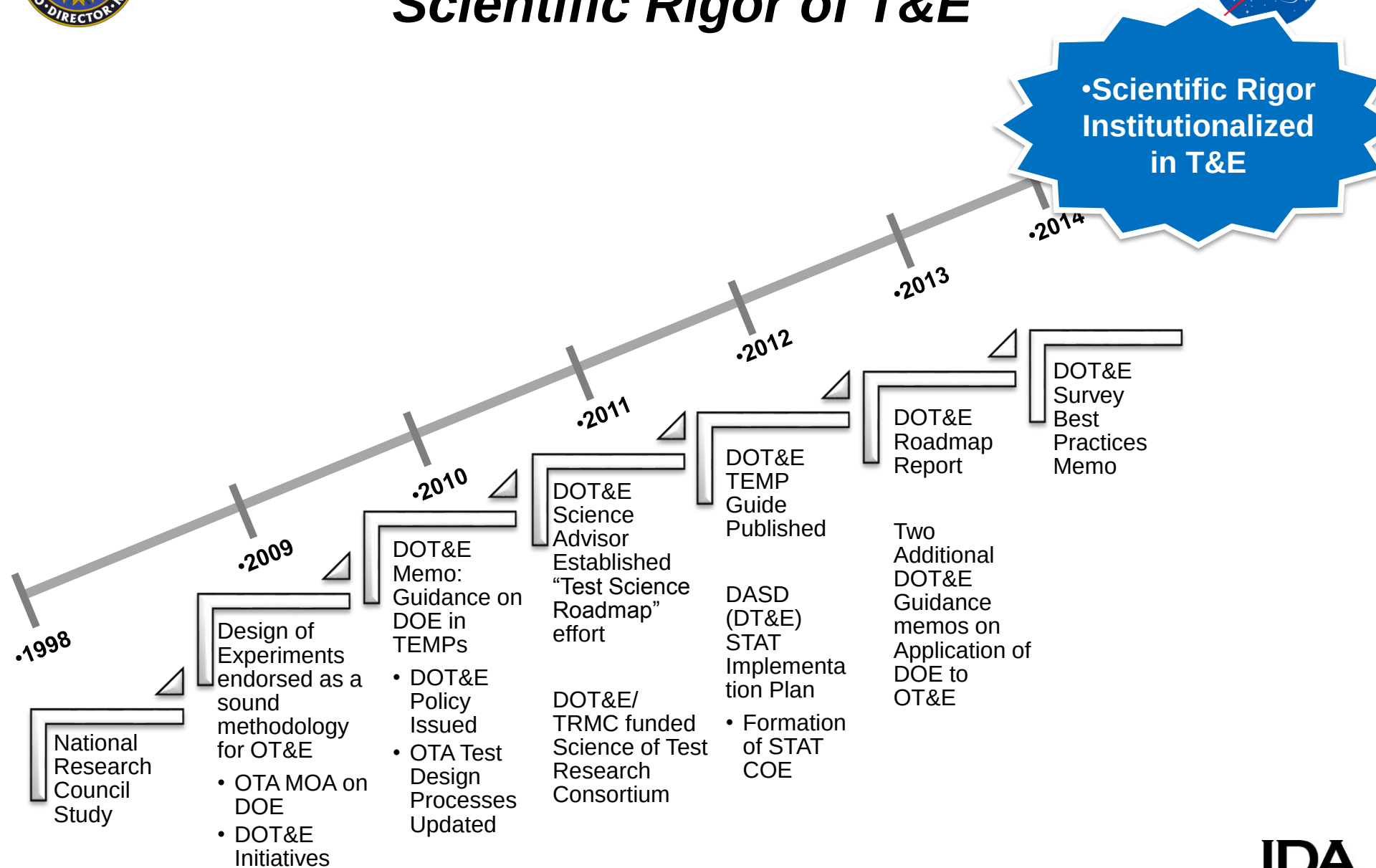


- **Team Members**
 - 6 full-time civil servants
 - 2 university collaborations
 - 4 graduate students
 - 3 contract support
 - ...and expanding
- **Members have obtained leadership positions**
 - Co-Investigator, Next Generation Air Transportation Systems
 - Deputy-Lead, LIDAR for Autonomous Rendezvous and Docking
 - Co-Lead, Engineering Complex Aerospace Systems
- **Span NASA's Missions: Aeronautics, Exploration, Science**





DOT&E Path to Increasing the Scientific Rigor of T&E

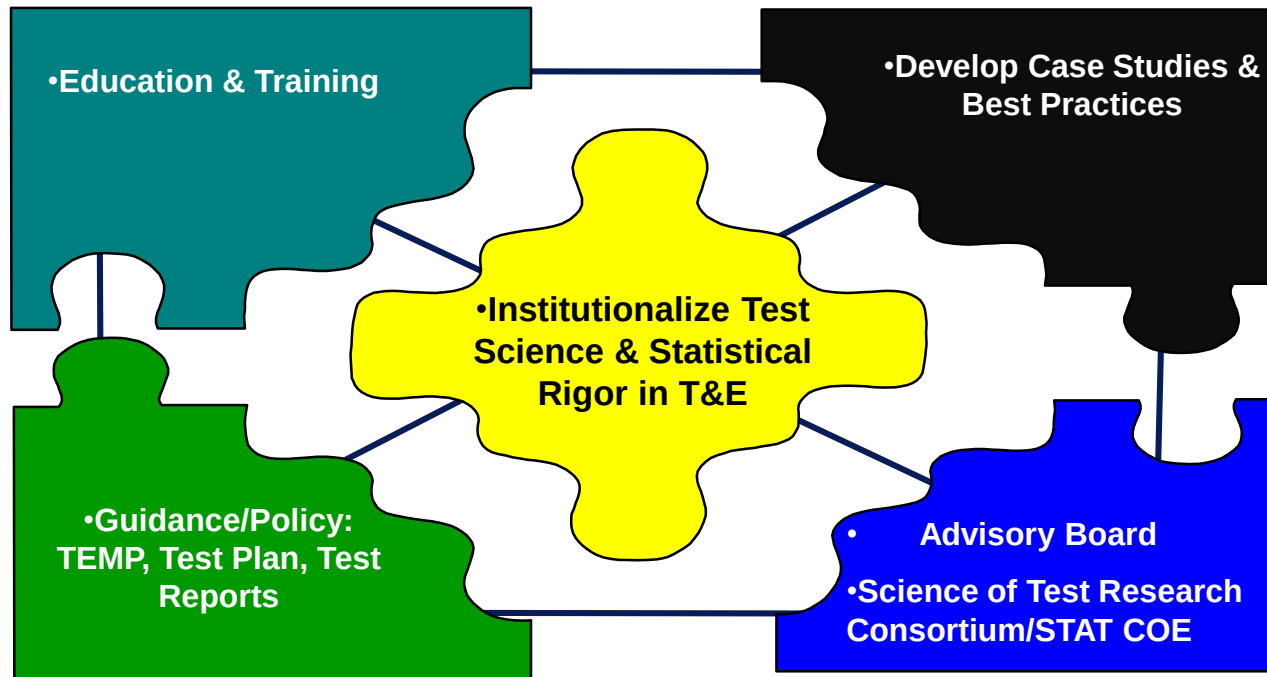




DOT&E Test Science Roadmap



- **Mission:** increase the scientific and statistical rigor of test and evaluation.
- **Multi-faceted effort:**



Statistical Engineering Agreement will **contribute across all areas!**



DOT&E Efforts to Institutionalize Statistical Rigor in T&E



- **DOT&E Test Science Roadmap – published June 2013**
- **DDT&E Scientific Test and Analysis Techniques (STAT) Implementation Plan**
- **Scientific Test and Analysis Techniques (STAT) Center of Excellence provides support to programs**
- **Research Consortium**
 - **Navel Post Graduate School, Air Force Institute for Technology, Arizona State University, Virginia Tech, Univ. of Arkansas, North Carolina State Univ.**
 - **Research areas:**
 - **Case studies applying experimental design in T&E.**
 - **Experimental Design methods that account for T&E challenges.**
 - **Improved reliability analysis.**
- **Current Training and Education Opportunities**
 - **DOT&E AO Training: Design, Analysis, and Survey Design**
 - **Air Force sponsored short courses on DOE**
 - **Army sponsored short courses on reliability**
 - **AFIT T&E Certificate Program**
- **Policy & guidance**
 - **DOT&E Guidance Memos**
 - **DOD 5000**
 - **Defense Acquisition Guidebook**



DOT&E's Need for Statistical Engineering



- DoD is improving in knowledge and practice of statistical methods; but application is **far from wide-spread**
 - *What is standard practice in industry is still viewed as advanced or novel*
 - *DOE employed for test design but analysis methods lag*
- DoD well-versed in engineering sciences for complex system development but not in statistics or robust data analysis methods
 - *Operational testing community has advanced recently because of high-level leadership;*
 - *Similar progress does not yet exist in the developmental testing or system development communities*
- Need the Statistical Community to engage with DoD
 - *Workforce is generally not statistically trained or statistically minded*
 - *Statistics community has the body of knowledge needed*
- Big problems, of national importance, to solve
 - Data rich environment
 - Need to improve the workforce
- Leaders in the field are needed on high-level advisory bodies (e.g., Defense Science Board)



Interagency Collaboration Developed



- **NASA and DOT&E began collaboration around 2004 through a series of propitious networking opportunities**
- **Informal collaboration grew, broadened, and provided mutual benefit**
 - **Reciprocal briefings that brought external perspectives**
 - **Recognized complementary strengths and weaknesses**
 - **Rallied around solutions to common challenges**
 - **Organic, fueled by passionate, dedicated individuals**
- **Collaborative activities expanded horizontally and vertically**
 - **Horizontally - across the organization and system lifecycle**
 - **Vertically – into leadership and grassroots practitioners**
- **Collaboration began on a National Conference in 2012**
 - **And, then sequestration hit ...**
- **Partnership was born out of adversity, ... made lemonade**



Statistical Engineering Partnership Operational Elements



Strategic framework to share leadership perspectives, foster practitioner collaboration, propagate training, and highlight accomplishments

Key Elements

- **Leadership seminars**
- **Tactical site visits**
- **Statistical engineering summit**
- **Collaboration on documents/handbooks**
- **Annual report to document benefits on the partnership**





Leadership and Collaboration



Leadership Seminars

- Virtual, Webinar, Open Access
- High-level leadership perspectives on Statistical Engineering
- Impact perspective of cases studies and best practices
- Disseminate organizational understanding of leadership perspectives

Tactical Practitioner Site Visits

- F2F meetings designed to provide direct access to leading research and practitioners across the organizations
- Opportunity to benchmark current state-of-the-art, facilitate increased awareness of best practices, and introduce new tools across our organizations



Centerpiece Meeting

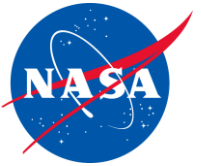


Statistical Engineering Summit

- **Forum for interaction among organizational and technical leadership, engineering and science subject-matter experts, and statisticians**
- **Collaboration among all these groups is essential for successful development, fielding and operational use of complex systems**
- **Designed to maximize practitioner benefit, tangible and immediate**
- **Summit program will incorporate**
 - **Training and tutorials**
 - **Case studies**
 - **Application-drive research**
- **Participation from industry and academia, as well as partner agencies**



Document to Institutionalize Gains

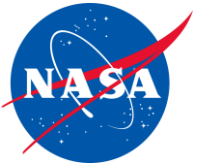


Best Practices Collaboration

- Development of guidebooks, best practices, and handbooks
- Building a body of knowledge directly applicable to partners
- External peer-review on the implementation of statistical engineering

Technical Reports

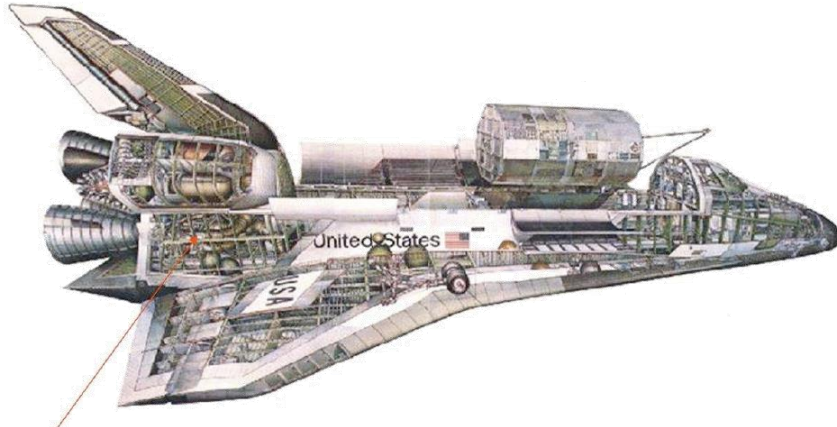
- Jointly prepared by the partner agencies
- Document the activities, impact, and benefits of the partnership
- Disseminate progress and successes



Statistical Engineering Case Studies



Space Shuttle Anomaly



GH2 Flow Control Valve Location

- STS anomaly required characterization for flight rationale

Original Question:

- Estimate probability of failure for a given orientation and velocity

- Prior testing was inconclusive, data-focused, rather than knowledge-focused
- Pass/fail response, rather than continuous
- Extensive collaboration to redefine objectives and response metrics was the key to success
- **Modeled depth of impingement as a function of orientation and velocity**





Rotorcraft Performance Research

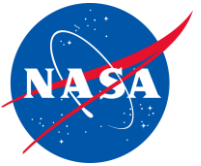


- **Wind tunnel characterization of Kiowa Warrior upgrades for computational model refinement**
- **DOE demonstration on a powered wind-tunnel model with a remote pilot**
- **Employed a factorial approach and response surface methods**
- **Discovered low drag condition at a previously unexplored condition in the design space**
 - **Impacts mission distance and duration**
- **Gained more information with a reduction in test resources**
 - **Minimizes facility utilization and resources**
 - **Accelerates learning**

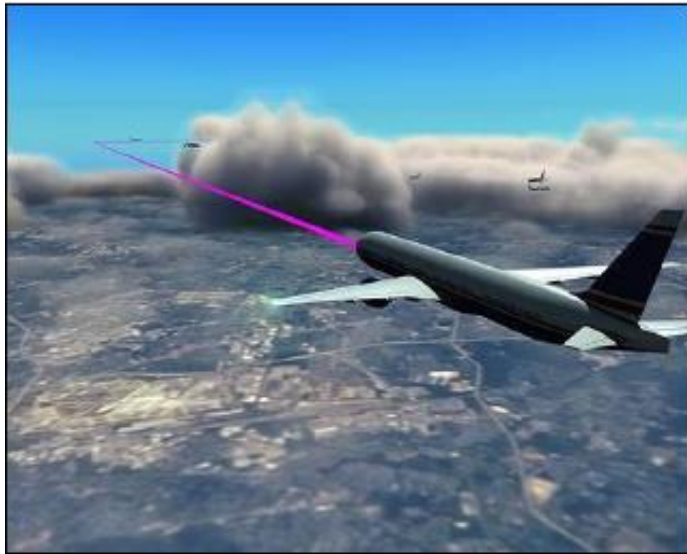




Airspace Systems Program



- Concepts, capabilities, and technologies for increases in the capacity, efficiency, and flexibility of the National Airspace System
- New insights into system performance, safety, efficiency, and pilot response



Statistical Engineering

- Extensive Human-in-the-Loop and simulation experiments in the Air Traffic Operations Laboratory (ATOL)
- Design of Experiments, uncertainty analysis, advanced regression modeling, hypothesis testing



Entry, Descent, and Landing

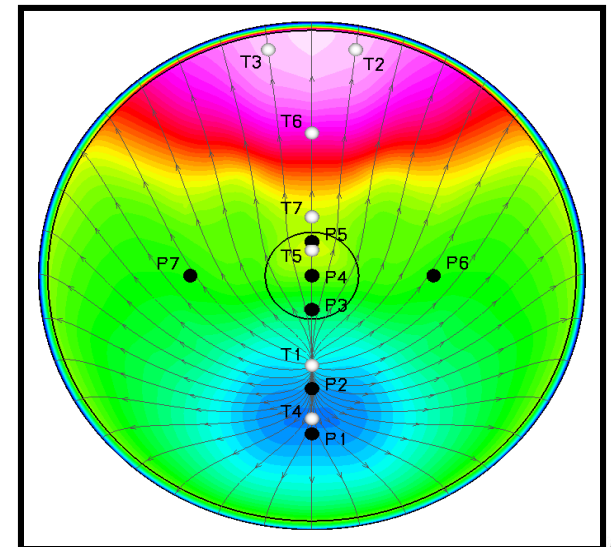


- Developing robust EDL technologies rely on improving Earth-based modeling capabilities
- Limited amount of flight data available and adequate test facilities



Statistical Engineering

- Measurement system characterization with response surface methods
- Integration of physical measurements and computational database for trajectory reconstruction





- **Operational Envelope Defined – 128 possible cases**
- **Test team identified factors and their interactions and refined them to identify the most important aspects of the test design**

[illegible]

Background Complexity	Threat	Formation Size	Location Confidence	Time of Day	Variant	Weapon
	Yellow	Yellow	Red	Green	Green	Yellow
		Red	Red	Green	Red	Red
			Green	Red	Green	Red
				Green	Green	Yellow
					Green	Red
						Yellow

	No significant interaction expected
	Significant interaction in one response
	Significant interaction in multiple responses



Proposed Design



- Test team used combination of subject matter expertise, and test planning knowledge to efficiently cover the most important aspects of the operational envelope
- Provided the data are used together in a statistical model approach, plan is adequate to evaluate JSF performance across the full operational envelope.
- Determined that 21 trials was the minimum test size to adequately cover the operational space
 - Ensures important factor interactions will be estimable
- Note the significant reduction to the 128 possible conditions identified.

			Variant - A								Variant - B							
			Category-B Threat				Category-C Threat				Category-B Threat				Category-C Threat			
			Low TLC		High TLC		Low TLC		High TLC		Low TLC		High TLC		Low TLC		High TLC	
			L	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H
2-Ship	Day	JDAM			1							1						
		LGB							1	1			1					
	Night	JDAM	1						1						1			
		LGB		1								1			1			
4-Ship	Day	JDAM					1						1					
		LGB			1			1									1	
	Night	JDAM		1								1						1
		LGB		1			1											



JSF Air-to-Ground DOE Summary



- **TEMP test design required 16 trials**
 - **Would have been insufficient to examine performance in some conditions**
- **Updated test design requires 21 trials but provides full characterization of JSF Pre-planned Air-to-Ground capabilities.**
- **New test design answers additional questions with the addition of only 5 trials:**
 - **Is there a performance difference between the JSF variants?**
 - **Do those differences only manifest themselves only under certain conditions?**
 - **Can JSF employ both primary weapons with comparable performance?**



Case Study: Submarine Detection Time



- **System Description**

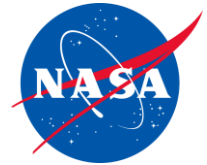
- Sonar system replica in a laboratory on which hydrophone-level data, recorded during real-world interactions can be played back in real-time.
- System can process the raw hydrophone-level data with any desired version of the sonar software.
- Upgrade every two years; test to determine new version is better
- Advanced Processor Build (APB) 2011 contains a potential advancement over APB 2009 (new detection method capability)



- **Response Variable: Detection Time**
 - Time from first appearance in recordings until operator detection
 - » Failed operator detections resulted in **right censored data**
- **Factors:**
 - Operator proficiency (quantified score based on experience, time since last deployment, etc.)
 - Submarine Type (SSN, SSK)
 - System Software Version (APB 2009, APB 2011)
 - Array Type (A, B)
 - Target Loudness (Quiet, Loud)



Submarine Detection Time: DOE Plan



		SSK		SSN	
		Quiet	Loud	Quiet	Loud
APB-11	Array A	12	12	6	12
	Array B	6	6	6	6
APB-09	Array A	8	8	4	8
	Array B	4	4	4	4

- **A full-factorial design across the controllable factors provided coverage of the operational space**
- **Replication was used strategically:**
 - **Allowed for characterization across different operator skill levels (randomly assigned)**
 - **Provided the ability to support multiple test objectives**
 - **Skewed to the current version of the system under evaluation (APB-11)**
- **Power analysis was used to determine an adequate test**
 - **Power was 89% detecting a 1 σ difference between APB versions – primary goal of the test**
 - **Power was > 99% for all other factor differences**
 - **Power was lower for APB due to blocking by day**



Submarine Detection Time: Data Collected



		SSK		SSN	
		Quiet	Loud	Quiet	Loud
APB-11	Array A	16	18	5	14
	Array B	10	5	6	3
APB-09	Array A	5	7	1	4
	Array B	3	1	2	0

- **Execution did not match the planned test design**
- **Test team used the DOE matrix at the end of the first round of testing to determine the most important points to collect next**
 - **Real time statistical analyses revealed that there was only limited utility in executing the remainder of the planned test**
 - **Analysis revealed that there was a significant difference in APB versions**
 - **Additionally all other factors considered were statistically significant due to larger effects than anticipated**



Submarine Detection Time: Analysis



- Advanced statistical modeling techniques incorporated all of the information across the operational space.
 - Generalized linear model with log-normal detection times
 - Censored data analysis accounts for non-detects
- All factors were significant predictors of the detection time

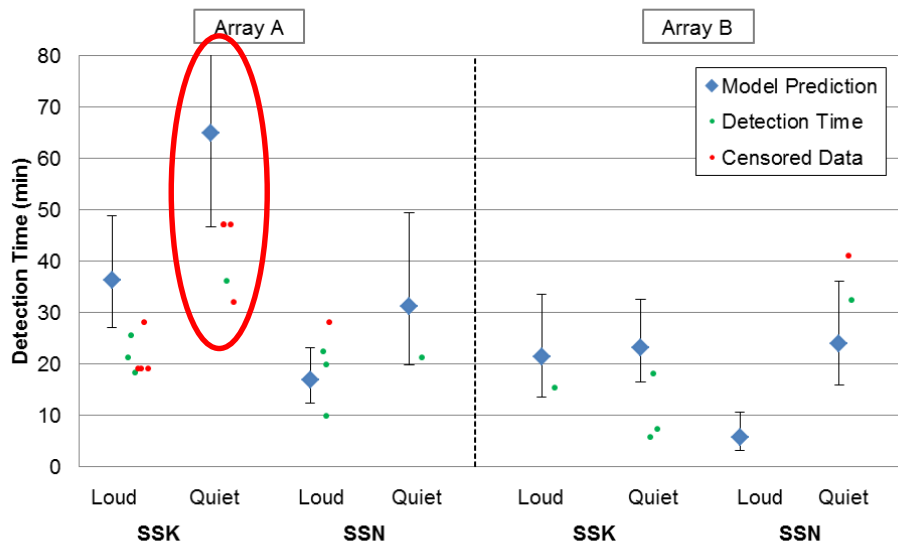
Factor/Model Term	Description of Effect	P-Value
Recognition Factor	Increased recognition factors resulted in shortened detection times	0.0227
APB	Detection time is shorter for APB-11	0.0025
Target Type	Detection time is shorter for SSN targets	0.0004
Target Noise Level	Detection time is shorter for loud targets	0.0012
Array Type	Detection time is shorter for Array B	0.0006
Type* Noise		0.0628
Type* Array	Additional model terms improve predictions. Third order interaction is marginally significant, therefore all second order terms are retained.	0.9091
Noise*Array		0.8292
Type* Noise*Array		0.0675



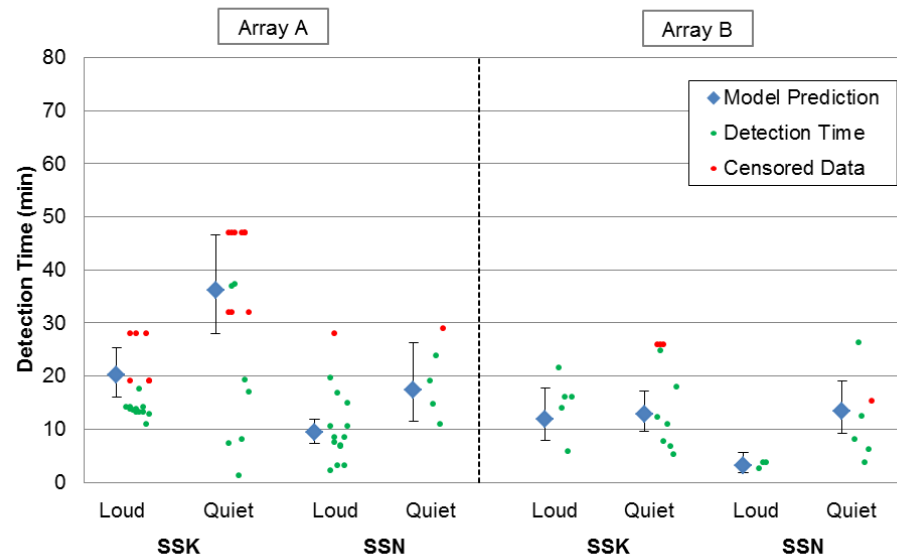
Submarine Detection Time: Results



APB-09 Median Detection Times



APB-11 Median Detection Times



- Median detection times show a clear advantage of APB-11 over the legacy APB
- Confidence interval widths reflect weighting of data towards APB-11
- Statistical model provides insights in areas with limited data



***Statistical Engineering Partnership
Near-term Implementation
and
Opportunities to Participate***



Upcoming Events



Leadership Seminar

- Kick-off Seminar December 2014
- Dr. Michael Gilmore, Director DOT&E
- “Efficient and Effective Test Programs and the Need for Statistical Rigor”
- Open access through Defense Connect Online (DCO)
- March – NASA Leader, June – Statistical Leader

Best Practices Collaboration

- Science of Test Knowledge Center
 - Online Resource
 - Provide guidance, best practices, tools (applications and code), video tutorials

Statistical Engineering Summit

- Targeting August 2015



Summary and Benefits



- **NASA and DOT&E have natural synergies, codified in partnership**
- **Grew organically, not mandated – strong support among participants**
- **Formal collaboration raises organizational awareness and provides a platform to disseminate benefits**

Partnership Benefits

- **Improve requirements specification to achieve high-level system objectives**
- **Expedite understanding of system capabilities through accelerated learning**
- **Efficient and effective test programs, that maximize information gain for minimal resources**
- **Improve quantification of risk, supporting better decision making**
- **Improve integration of test and evaluation activities over all phases of testing**