OED D-5286 non-std. H14-000995

Design of Experiments for In-Lab Operational Testing of the AN/BQQ-10 Submarine Sonar System

Dr. George Khoury – Task Leader
Dr. Justace Clutter
Dr. Laura Freeman

Approved for public release; distribution is unlimited.

Outline

- Background
- Test Motivation
- Test Design
- Test Results
- Conclusions
- Recommendations

A-RCI System Description

- **Acoustic Rapid COTs Insertion (A-RCI) Sonar System**
  - Collection of hardware and software tools to understand the acoustic environment and the threat space surrounding the submarine
  - System is independent of acoustic arrays and only provides processing
  - Deployed on all US submarines
- **Multi-Mission**
  - ASW, ASuW, MIW, INT
- **Acquisition Model (Spiral Development)**
  - Advanced Processor Build (APB)
    - Every 2 years (Odd years)
    - Brings new software algorithms and improvements
  - Technical Insertion (TI)
    - Every 2 years (Even years)
    - Increases system processing capability
    - Does not include new arrays

Image Link: https://www.flickr.com/photos/bwjones/3552816442
License: Creative Commons (CC BY-NC-ND 2.0)
https://creativecommons.org/licenses/by-nc-nd/2.0/legalcode
Image slightly cropped

ASW Metrics

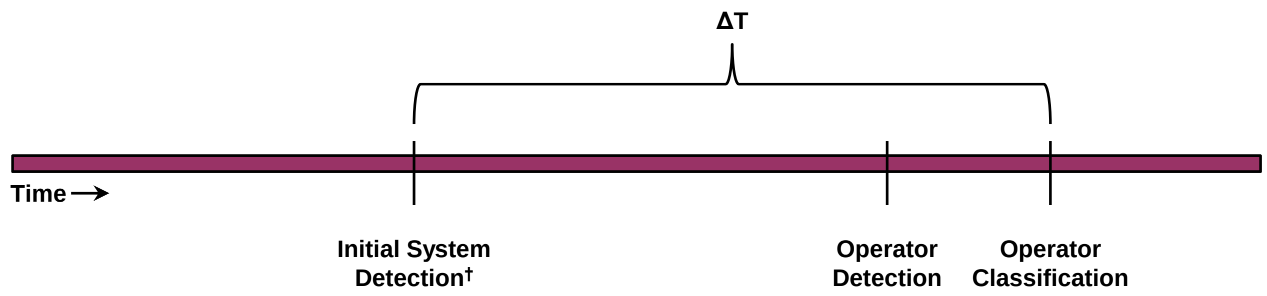
- **ASW Performance is measured through a number of different metrics, each focused on a different element of the ASW mission**
  - Search
    - » [Old] Search Rate – time it takes to find a threat
    - » [New] ΔT – time it takes the operator to detect a contact
  - Localization
    - » P_L – probability the contact will be localized at a specific range
  - High Density Contact Management (HDCM)
    - » $P_{D(PEZ)}$ – probability that a contact will be detected prior to being XX minutes from penetrating the exclusion zone
    - » $P_{D(EZ)}$ – probability that a contact will be detected by the operator within XX minutes of it showing up on the sonar system if it is inside the exclusion zone

- **ΔT will be the focus of this study**
  - Goal of the A-RCI program is to improve the processing of acoustic information and reduce the time it takes for an operator to detect and classify a contact once it appears on the display

- **ΔT represents only one component of the overall performance of the system, but it is appropriate to assess the initial detection phase of the ASW mission**

ΔT Metric Overview

- **ASW performance is defined in the CDD as the difference between the time the contact first appears on the display† and the time the operator classifies the contact (ΔT)**



- **CDD defines two thresholds depending on the signal processing method**
  - Median(ΔT) ≤ XX Minutes for Broadband Processing
  - Median(ΔT) ≤ XX Minutes for Narrowband Processing

†: This is determined during post test analysis

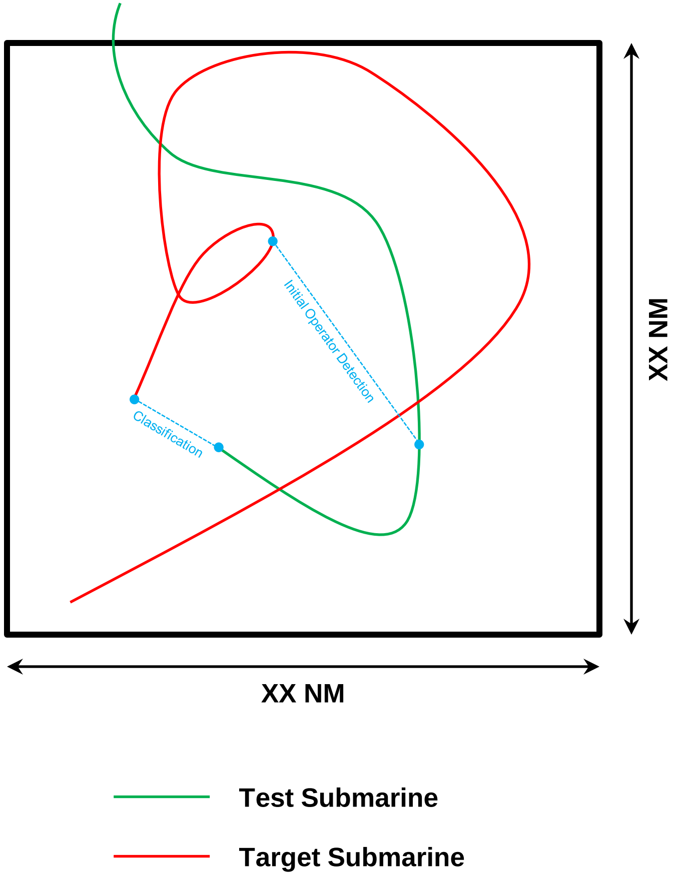
Approved for public release; distribution is unlimited.

Outline

- Background
- **Test Motivation**
- Test Design
- Test Results
- Conclusions
- Recommendations

Legacy Testing for ASW Performance

1. Establish a bounded test area
2. Place target submarine inside the area and instruct it to act as if it was on patrol
3. Have system under test enter the area and search for the target
4. Test ends when the system under test detects and classifies the target or when an upper limit on time is reached, typically 4 to 6 hours
5. Repeat to get required number of runs



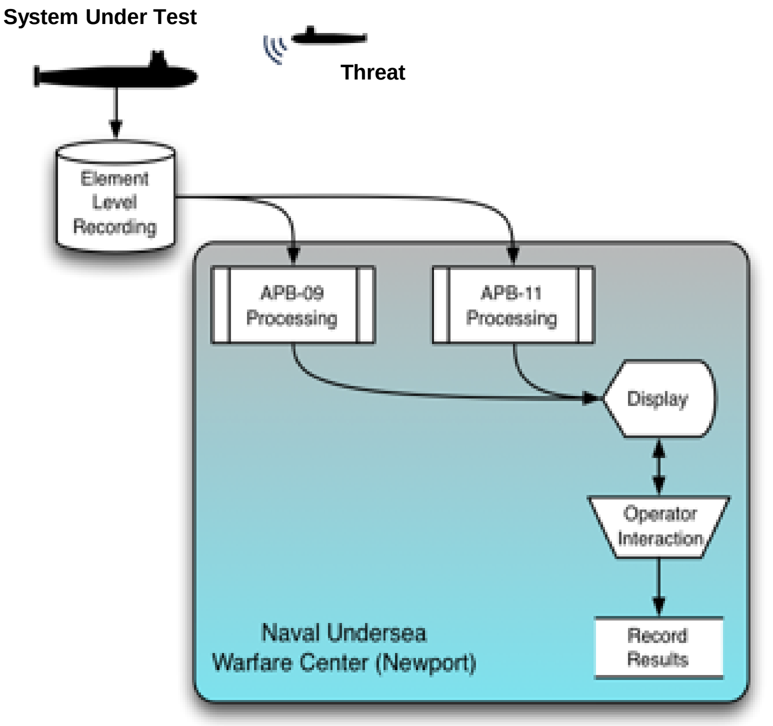
Legacy Testing for ASW Performance

- **Pros**
  - Real Data
    - All acoustic data has full multipath and environmental effects
  - Good Ground Truth
    - We can use the navigation data from both platforms to produce a high quality bearing and range dataset between the two platforms
  - Entire Sonar Team
    - Able to observe the interactions between the individual sonar operators
  - Allows for End-to-End Prosecution
    - Can observe an interaction from initial detection through torpedo launch which covers the search, localization, and attack ASW phases

- **Cons**
  - Target Familiarity
    - Operators learn the characteristics of the only target available which could bias the measured results
  - Limited Sample Size
    - Only see one set of operators vice a good statistical sample
    - Length of test make large sample sizes cost prohibitive
  - Difficult to Compare Results
    - Low sample sizes lead to larger uncertainties
    - Environmental and target variability lead to additional uncertainties that are almost impossible to control

ASW Operator in the Loop Testing

1. **Have forward deployed units record interactions, at the element level, with actual threat submarines. Or, use element level recordings from operational testing events.**
2. **Reprocess the element level recordings through either the APB-09 or the APB-11 processing system.**
3. **Have a single operator using a towed array station search for the threat target.**
4. **Test ends when the operator detects and classifies the threat.**
5. **Repeat to get required number of runs.**



Note: This is an extension of a current testing program being maintained at NUWC Newport

ASW Operator In the Loop Testing

- **Pros**
  - Real Data
    - All acoustic data has full multipath and environmental effects
  - Better Sampling of Operator Proficiency
    - Different operators with a distribution of proficiencies are used for each set of runs
  - Completely Controls for Environment and Target Variability
    - Using the same recordings with the same targets and environments on the two different systems
- **Cons**
  - Single Operator
    - Using only a single operator does not capture the effects from having an entire sonar team (cueing, other supporting info)
  - Inability to Maneuver
    - There is no way for the operator to request a course change to improve the signal, determine which side the contact is on, or gather ranging information
  - Limited Ability to Assess Other ASW Metrics
    - Difficult to determine initial detection range or localization metrics due to a lack of good reconstruction (one-sided only)
  - Operator Vigilance is not Accounted For
    - Sonar operators typically operate for 6 to 8 hours at a time, their ability to detect a target on the system is likely a function of this

Approved for public release; distribution is unlimited.

Outline

- Background
- Test Motivation
- **Test Design**
- Test Results
- Conclusions
- Recommendations

Test Design

- **120 Run D-Optimal Split Plot Test Design**
  - Response: Continuous [Lognormal distribution]
  - Categorical Factors: 4 2-level
  - Continuous Factors: 0

- **Planning Constraints**
  - Coordination with NUWC L4 Testing Efforts
    - » The NUWC L4 test program is a long running program to assess operator performance, OIL testing is a logical extension of this testing
    - » Generated asymmetric design with a preference for APB-11 data
    - » Limited the number of times the APB version was changed to 4
  - Number of Operators
    - » Focused on using not more than 20 operators to remain consistent with NUWC L4 planning efforts.
    - » Assumed that each operator can process 6 datasets ($6 \times 20 = 120$)
  - Datasets
    - » Limited number of datasets acquired from OT and deployed operations
    - » Most datasets had very little metadata and lacked a good target position reconstruction

Test Design: Factors

- **Target Type: A / B**
  - Different general narrowband signatures
- **Target Noise: Quiet / Loud**
  - Newer generations of each target type getting quieter
- **Towed Array Used: A / B**
  - Different array acoustic gains
- **APB Version: APB-09 / APB-11 (Hard to Change)**
  - Feature changes in the new APB

Power to observe all main effects and interactions with 80% confidence are all well above 80%

Other Explored Factors

- **Operator Proficiency**
  - Measured using an in-house NUWC formula
  - Ranges between 1 (Junior) and 20 (Expert)
  - Based on several factors
    - » Number of years in Service
    - » Number of deployments
    - » Time since last deployment
  - Unable to attain a list of operator proficiencies prior to test
  - Did not plan against this factor due to inability to choose a particular operator during a run
  - We did use this in the analysis as a covariate

- **Clutter Level: Low / Medium / High**
  - Defined as the average number of broadband traces that are present on the system
  - An increased clutter level can complicate the search for a target due to the interfering contacts
  - Unable to use due to a lack of available metadata in the ONI tape database

Other Explored Factors

- **Operating Profile: Patrol / Transit**
  - Acoustic noise increases with higher speeds
  - Patrol speeds are typically low
  - Transit speeds can be much higher
  - Unable to use due to a lack of data in the ONI metadata

Test Design: Factors (Summary)

Target	System Under Test	Environment
Type: A / B	**Array:** A / B	**Clutter:** Low / Med / High
Noise: Quiet / Loud	**APB Version:** APB-09 / APB-11	
Operating Profile: Patrol / Transit	**Operator Proficiency:** Continuous [1,20]	

Test Design: Factors (Summary)

Used in Test Design (green) | **Used in Analysis** (blue) | **Not Used** (red)

Target	System Under Test	Environment
Type: A / B	**Array:** A / B	**Clutter:** Low / Med / High
Noise: Quiet / Loud	**APB Version:** APB-09 / APB-11	
Operating Profile: Patrol / Transit	**Operator Proficiency:** Continuous [1,20]	

Data Not Available During Design Process

Final Tape Library

		Array B	Array A
Type B	Quiet	2	1
	Loud	1	1
Type A	Quiet	2	2
	Loud	1	2

Approved for public release; distribution is unlimited.

Test Design Run Plan

Total Number of Runs

APB-11 (72 runs)

		Array B	Array A
Type B	Quiet	6	6
	Loud	6	12
Type A	Quiet	6	12
	Loud	6	12
	Null	6	

APB-09 (48 runs)

		Array B	Array A
Type B	Quiet	4	4
	Loud	4	8
Type A	Quiet	4	8
	Loud	4	8
	Null	4	

72

48

120

Executed Test Design

APB-11

		Array B	Array A
Type B	Quiet	6	5
	Loud	3	14
Type A	Quiet	10	16
	Loud	5	18
	Null	5	

APB-09

		Array B	Array A
Type B	Quiet	2	1
	Loud	0	4
Type A	Quiet	3	5
	Loud	2	6
	Null	2	

Total Number of Runs

82

25

107

Outline

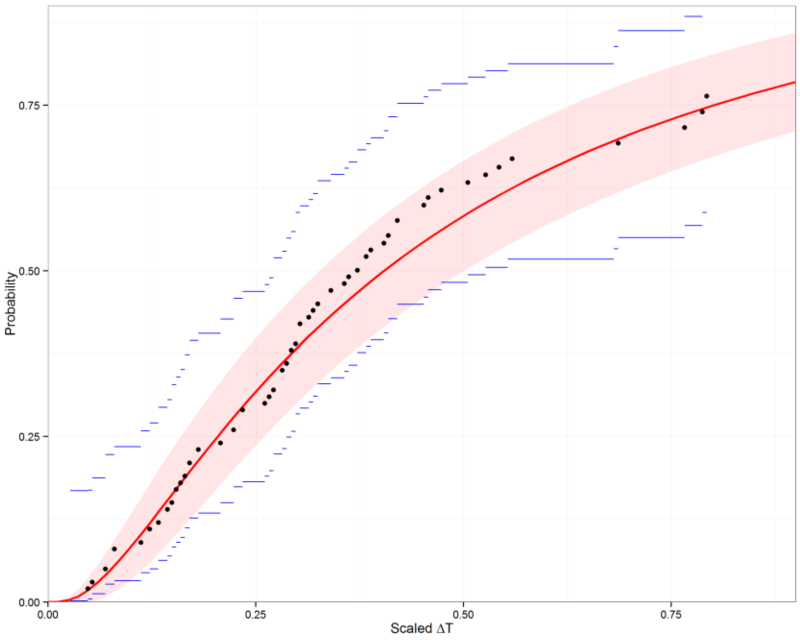
- Background
- Test Motivation
- Test Design
- **Test Results**
- Conclusions
- Recommendations

ΔT Analysis Method

- **Maximum Likelihood Fit (MLE)**
  - Explicitly handled non-detected targets as censored data
  - Assumed Lognormal Distribution
  - Distribution has two parameters

- **Used several methods to select the final model**

- **Validated the model by comparing the model predictions with the acquired data**



$$p(x|\boxed{\mu,\sigma}) = \frac{1}{2\sqrt{2\pi\sigma}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

Two Parameters	
μ:	Location Parameter
σ:	Scale Parameter

Approved for public release; distribution is unlimited.

ΔT Model Selection

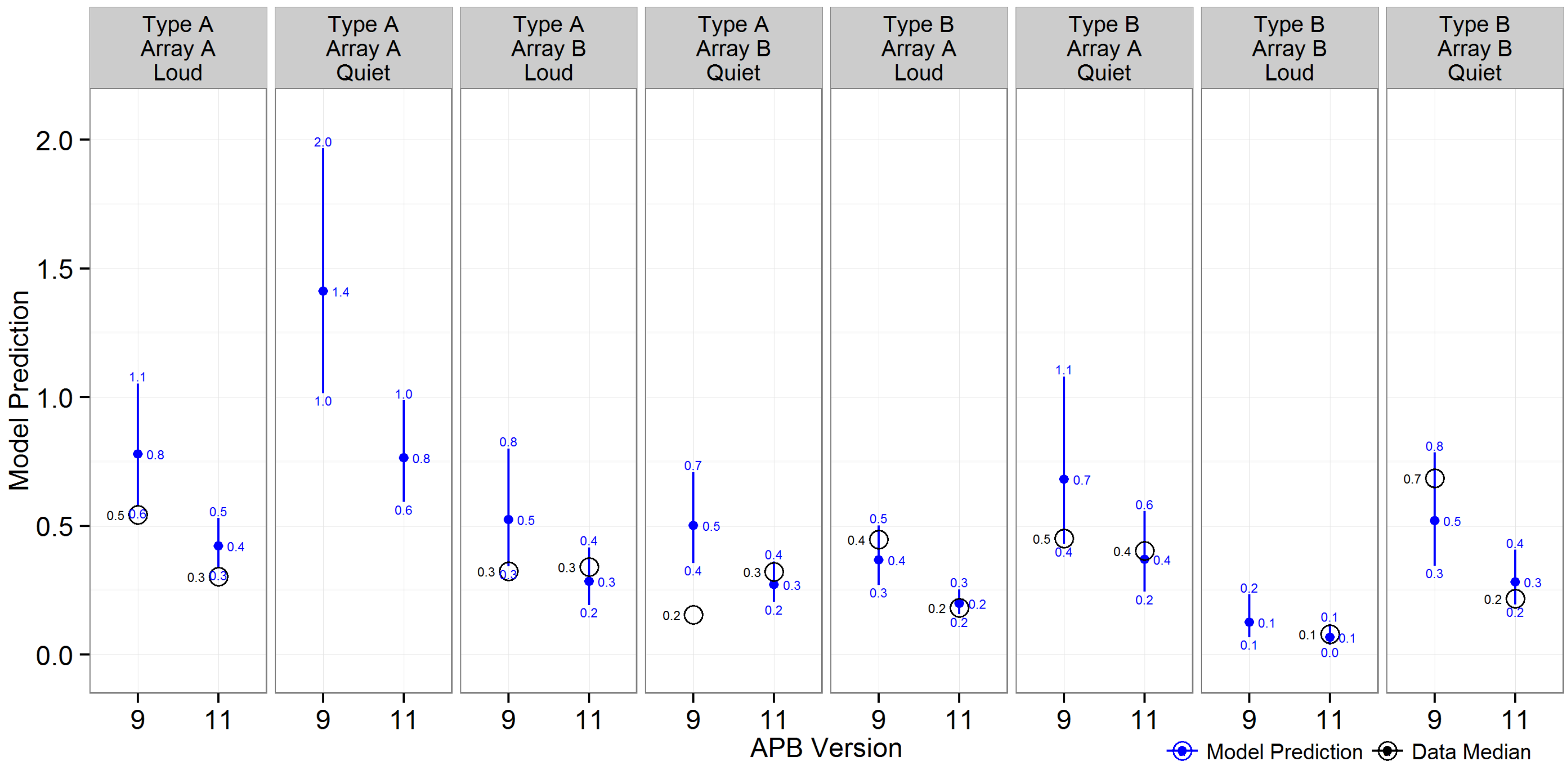
- **Model Complexity**
  - Explored up to three-way interactions for the location parameter
  - Explored main effects only model for the scale parameter
- **Used a variety of methods to select the final model**
  - Backwards model selection, using the metrics:
    - AIC
    - BIC
    - Total Likelihood
- **Final model**
  - Factors only affect the location parameter and the scale is considered constant
  - All factors studied significantly affected ASW performance

ΔT Final Model

$$\mu = \beta_0 + \beta_1\, RF + \beta_2\, APB + \beta_3\, Target + \beta_4\, Noise + \beta_5\, Array + \beta_6\, Target * Noise + \beta_7\, Target * Array + \beta_8\, Noise * Array + \beta_9\, Target * Noise * Array$$

Term	Description of the Effect	p-Value
β_1 (RF)	Increased recognition factors resulted in shorter detection times	0.0227
β_2 (APB)	Detection time is shorter for APB-11	0.0025
β_3 (Target)	Detection time is shorter for Type B targets	0.0004
β_4 (Noise)	Detection time is shorter for loud targets	0.0012
β_5 (Array)	Detection time is shorter for the Type B array	0.0006
β_6 (Target*Noise)	Additional model terms added to improve predictions. Third order interaction is marginally significant. Therefore, all second order interactions nested within the third order interaction were retained to preserve model hierarchy.	0.0628
β_7 (Target*Array)		0.9091
β_8 (Noise*Array)		0.8292
β_9 (Target*Noise*Array)		0.0675

ΔT Model Validation



Data are adequate to support a comparison of APB performance

- Model prediction results are consistent with the non-parametric median estimates
- Results are largely driven by the APB-11 data since it was the majority of the data

Note: Non-parametric medians calculated from the Kaplain-Meier estimator per bin and model estimates are the median with an 80% confidence interval

Approved for public release; distribution is unlimited.

Detection Method Analysis

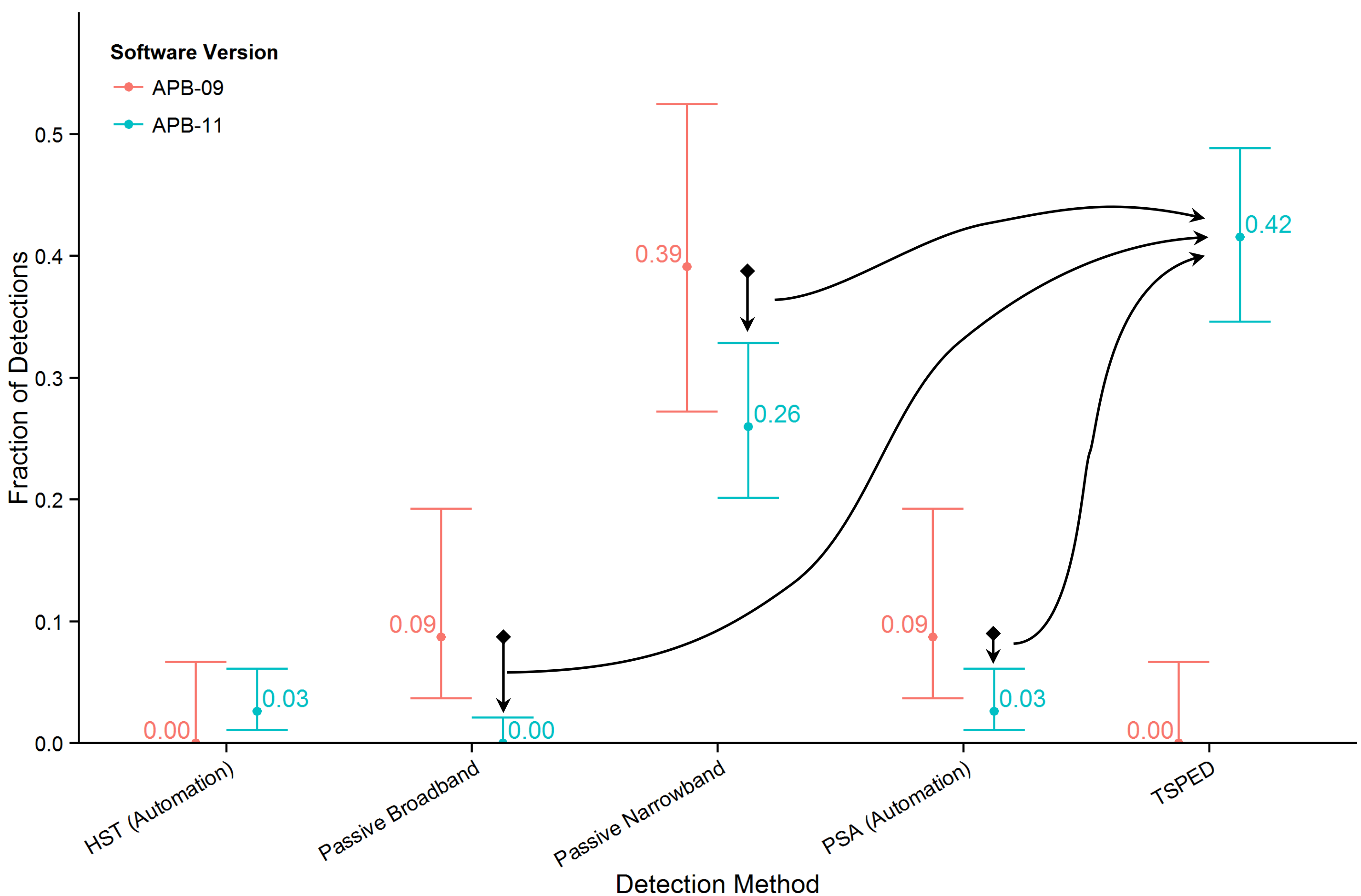
- **In addition to evaluating the ΔT as a function of the covariates, the method of detection was also investigated**
- **Detection Methods**

Harmonic Set Tracker:	Automated method to detect a repetitive signal in frequency space
Passive Broadband:	Operator detected target manually using broadband signal processing
Passive Narrowband:	Operator based detection method using narrowband signal processing
Priority Search Agent:	Automated method based on various acoustic properties of the signal
TSPED:	Operator based detection method using a hybrid broadband / narrowband processing (Only available in APB-11)

Detection Method Analysis

Detection Method	APB-09	APB-11	Total
HST (Automation)	0	2	2
Passive Broadband	2	0	2
Passive Narrowband	9	20	29
PSA (Automation)	2	2	4
TSPED	N/A	32	32
Total Detections	13	56	69
Missed Detections	10	21	31
Total	23	77	100

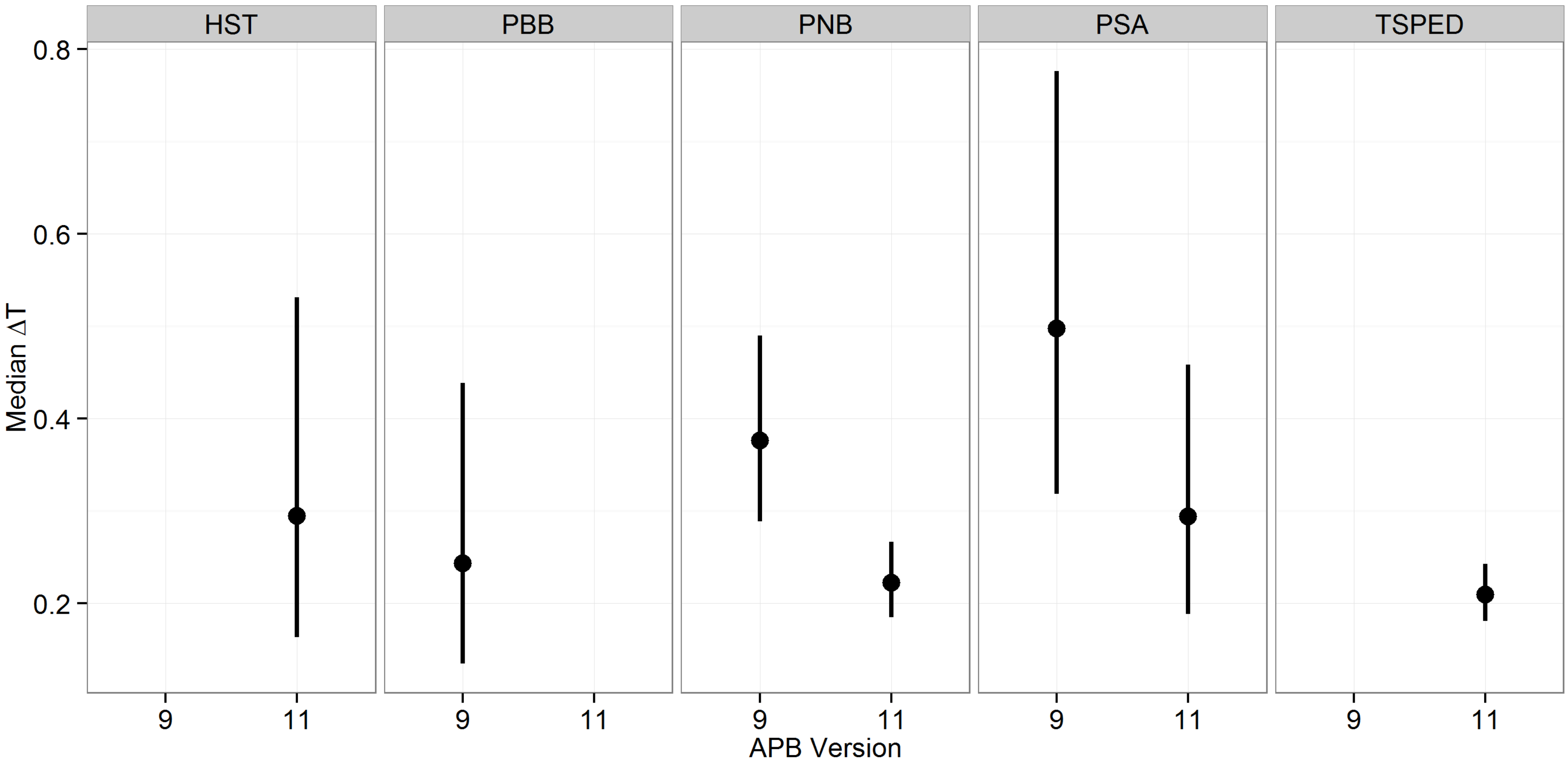
Detection Method Analysis



Note: Uncertainties determined with an 80% confidence Wilson Score Interval

Approved for public release; distribution is unlimited.

Detection Method Analysis



- **TSPED was the primary detection method on APB-11 and was competitive with PNB for the ΔT metric**

Approved for public release; distribution is unlimited.

Initial Detection Range

- **A limited set of data (4) had either one-sided or two-sided reconstructions**
  - Reconstruction determines the bearing and range between two contacts in an interaction
  - A full two-sided reconstruction requires that both platforms in the interaction record and report their position as a function of time
  - A one-sided reconstruction is the estimation of the targets position based only on the information recorded by the searcher
    - This reconstruction is only approximate due to the lack of information
- **Based on the four datasets with a reconstruction, there was no noticeable performance difference between the two studied APB versions**
- **No conclusions can be drawn because of the limited data**
- **The Navy should increase the number of datasets with ranging information to improve this analysis**

Approved for public release; distribution is unlimited.

Comparison With At-Sea Results

- **Results from the In-Lab testing do not reflect the measured values from the at-sea test phase**
  - The at-sea evaluation had an acoustic intelligence rider, with advanced training to detect threat targets, which affected their results
  - The ACINT is represented in the In-Lab evaluation as an operator with an Recognition Factor of ~16
  - The at-sea evaluations are limited since there was only a single target and the crew developed a string understanding of the target signature
- **This indicated that the In-Lab testing can be used for comparisons between APB, but not for an absolute assessment of ASW performance**

Outline

- Background
- Test Motivation
- Test Design
- Test Results
- **Conclusions**
- Recommendations

Conclusions

- **OIL testing is an effective method to evaluate ASW detection performance across different software versions**
  - Allows for a comparison of different APB software versions across controlled target and environmental conditions
- **Due to limitations in the testing method, OIL testing must supplement at-sea evaluations of ASW performance**

This is the first time that an A-RCI OT report has claimed a statistically significant improvement between different versions of the software

Approved for public release; distribution is unlimited.

Outline

- Background
- Test Motivation
- Test Design
- Test Results
- Conclusions
- **Recommendations**

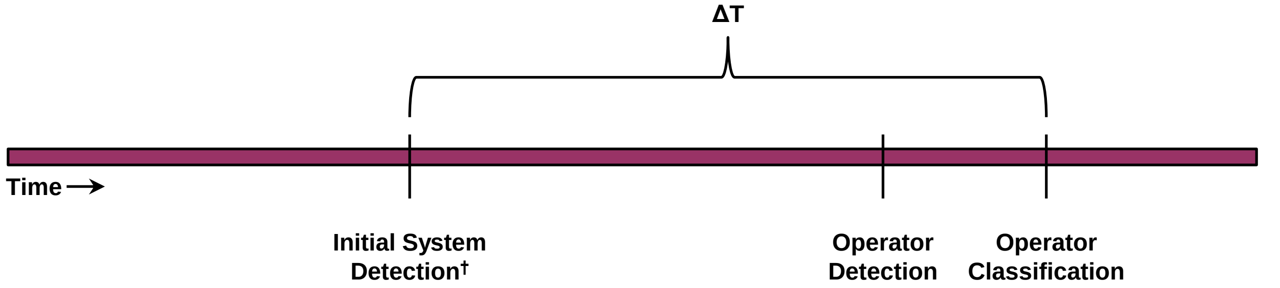
Recommendations

- **Incorporate OIL testing into future A-RCI test events**
- **Maintain in-water testing to evaluate other aspects of performance such as localization**
- **Dataset Improvements**
  - More datasets to choose from
  - Improve available metadata (clutter level, operating profile)
  - Add more emphasis to getting datasets that have at least one-sided reconstruction abilities
  - Have an all hands meeting prior to the commencement of test that identifies all the datasets that will be used in the testing
- **Account for operator proficiency in a controlled fashion**

Approved for public release; distribution is unlimited.

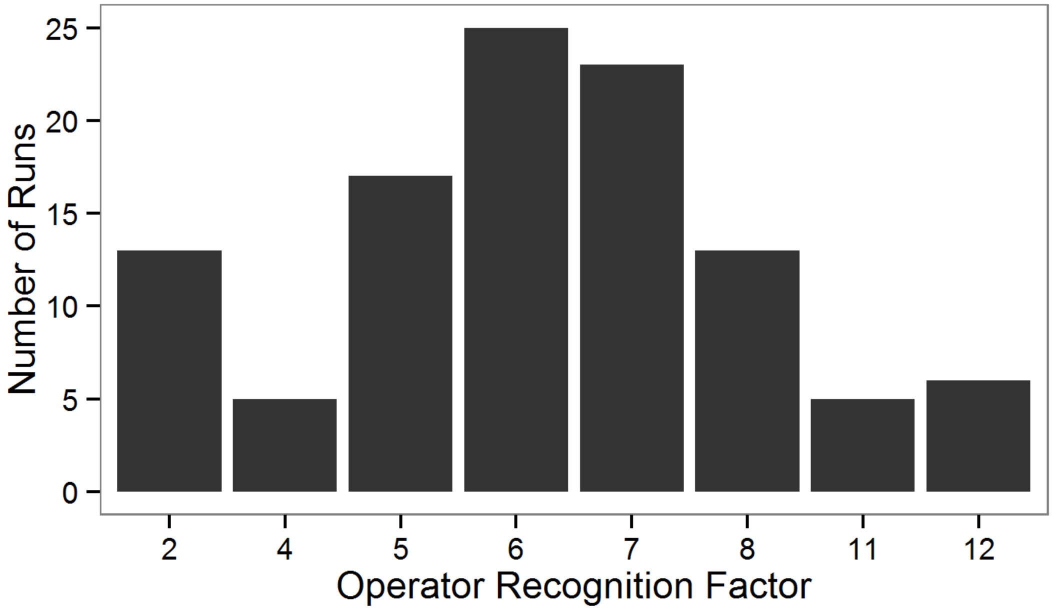
BACKUP SLIDES

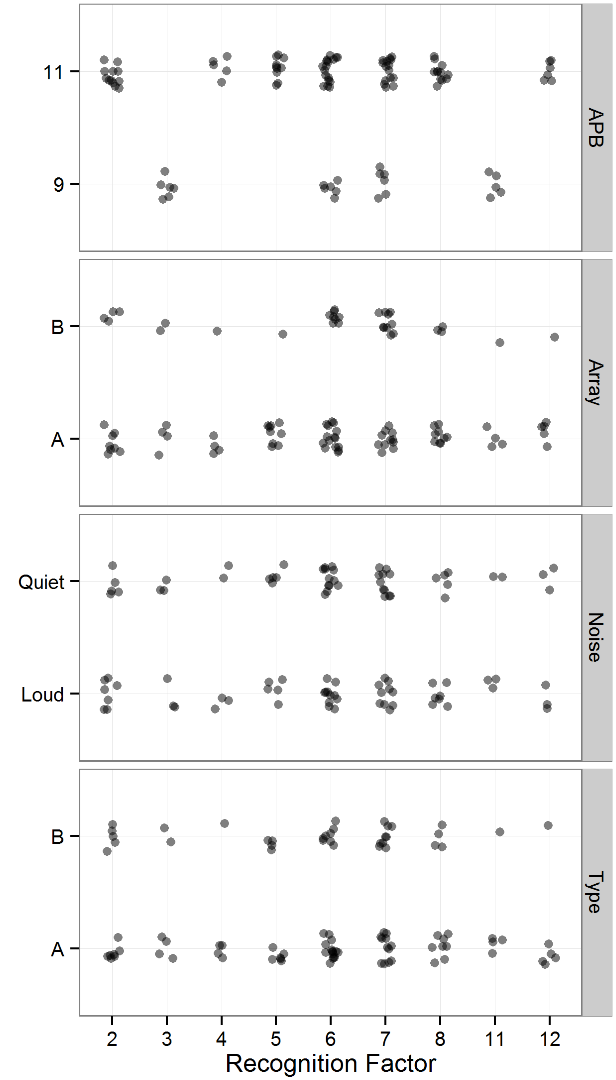
ΔT Metric Overview



Approved for public release; distribution is unlimited.

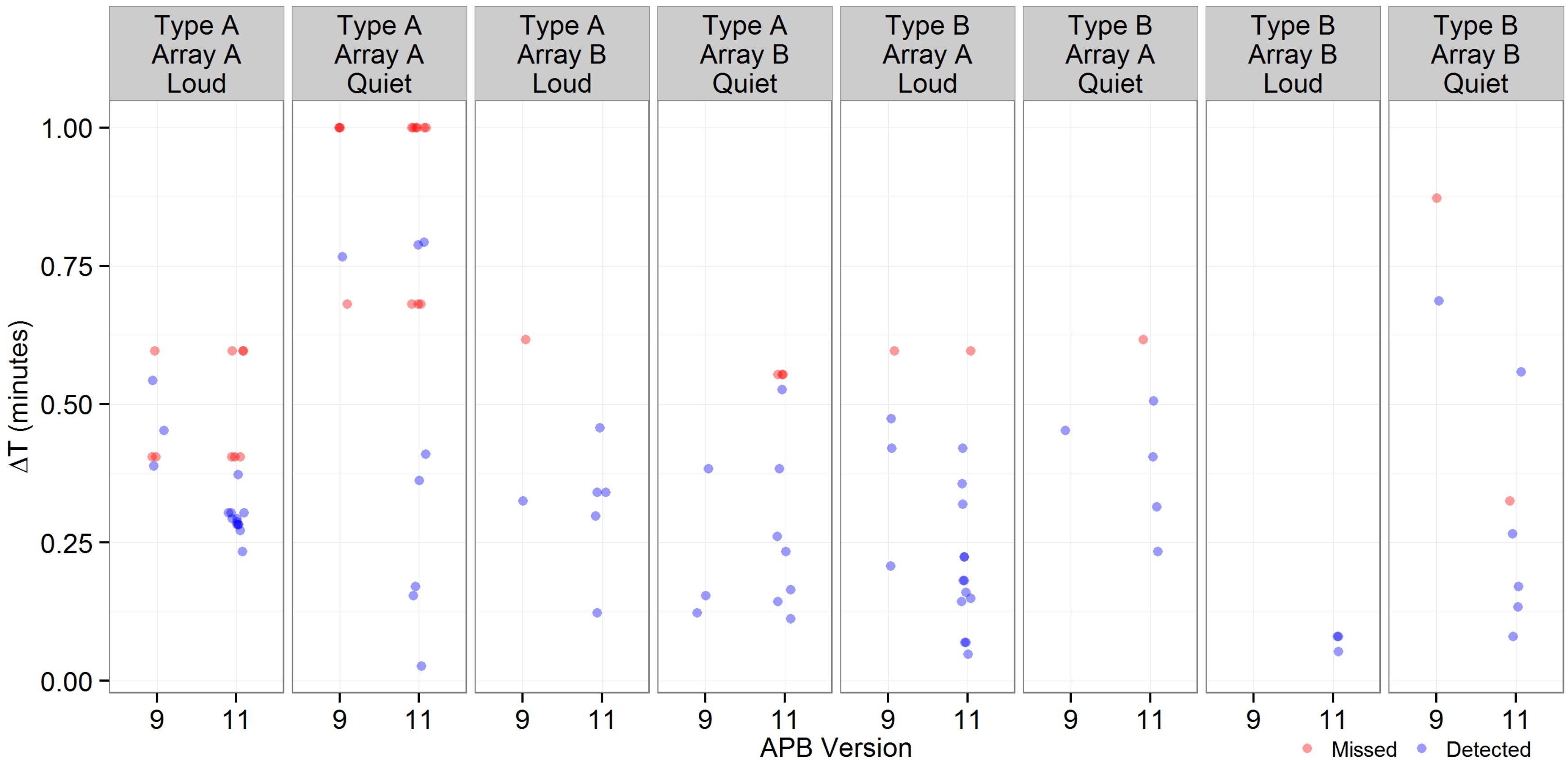
Recognition Factor Distribution





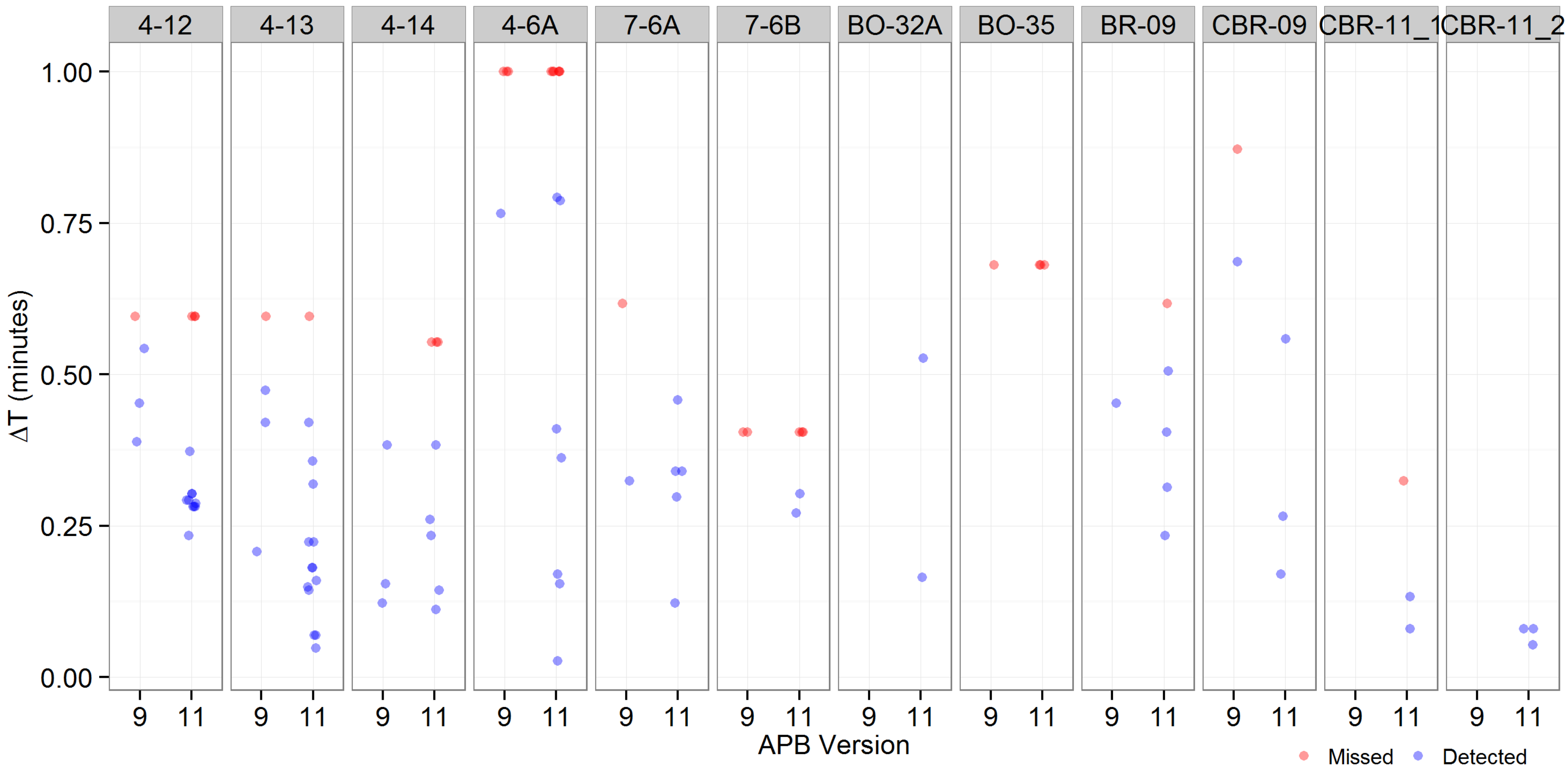
- **Testing had a good distribution of operator proficiency levels**
- **These different proficiency levels were distributed evenly throughout the test design space**

Raw Results from Test (By DOE Bin)



Approved for public release; distribution is unlimited.

Raw Results from Test (By Data Cut)



Approved for public release; distribution is unlimited.

ΔT Final Model

$$\mu = \beta_0 + \beta_1\, RF + \beta_2\, APB + \beta_3\, Target + \beta_4\, Noise + \beta_5\, Array + \beta_6\, Target * Noise + \beta_7\, Target * Array + \beta_8\, Noise * Array + \beta_9\, Target * Noise * Array$$

Term	Point Estimate	Std Error	p-Value
β_0 (Intercept)	-0.534	0.232	N/A
β_1 (RF)	-0.074	0.032	0.0227
β_2 (APB)	0.307	0.100	0.0025
β_3 (Target)	0.359	0.098	0.0004
β_4 (Noise)	-0.324	0.098	0.0012
β_5 (Array)	0.347	0.098	0.0006
β_6 (Target*Noise)	0.186	0.098	0.0628
β_7 (Target*Array)	0.011	0.098	0.9091
β_8 (Noise*Array)	0.021	0.098	0.8292
β_9 (Target*Noise*Array)	-0.180	0.098	0.0675
σ	0.763	0.069	N/A

Note: Fit assumes that low values are -1 and high values are +1 (a.k.a. JMP style)