



Exercising Prediction Validation Methods for Analysis of Quiet Supersonic Overflight

Alexandra Loubeau, Kathryn Ballard, William Doebler, Nathan Cruze
NASA Langley Research Center

185th Meeting of the Acoustical Society of America
Sydney, NSW, Australia
December 5, 2023
Session 2pNSb, Paper 2pNSb3

Enable development of noise-based standard for commercial supersonic aircraft

➤ Two regulatory noise issues for civil supersonic flight

- Limiting airport noise during subsonic flight
- Limiting sonic boom noise during [en route supersonic](#) flight

➤ Development of international standard

- Noise-based certification standard for supersonic en route (sonic boom) noise
- Certification standard would include noise metric, [test procedures](#), and noise limits

➤ Required characteristics of test procedures

- Well-defined “recipe” for all noise certification applicants to follow
- Incorporate both measurements and predictions
- Robust to apply in real-world situations

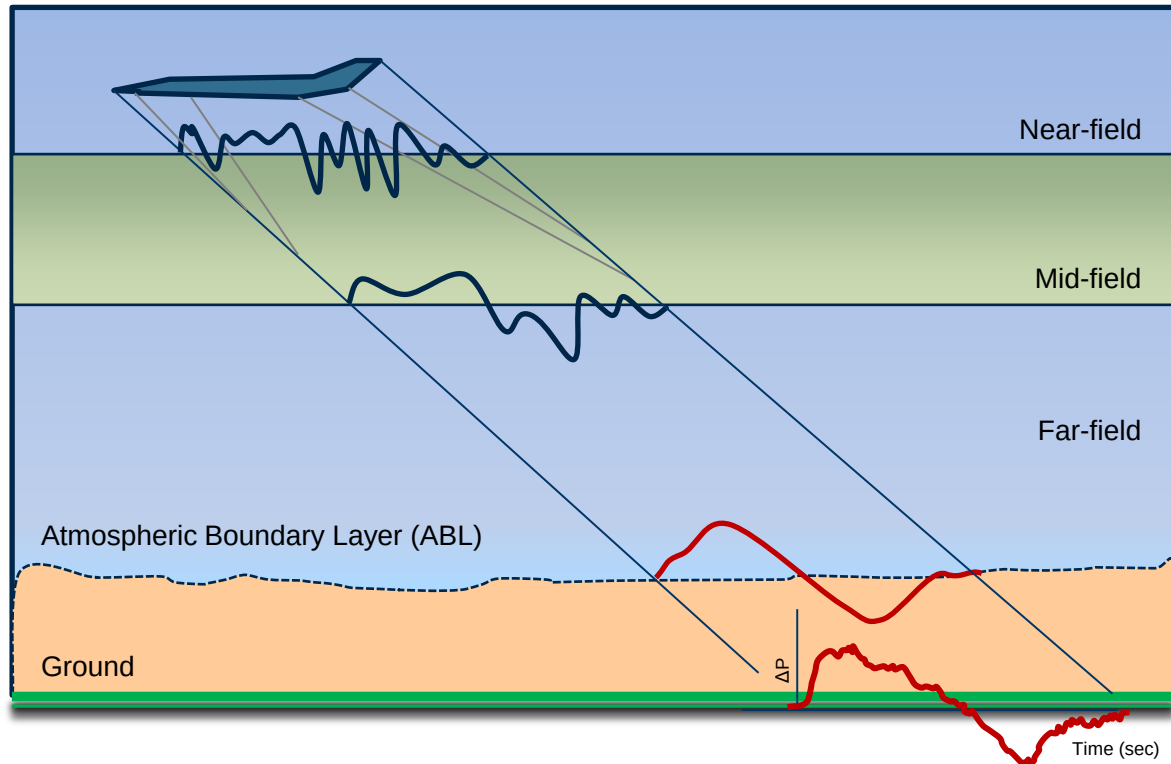


Notional Certification Procedure

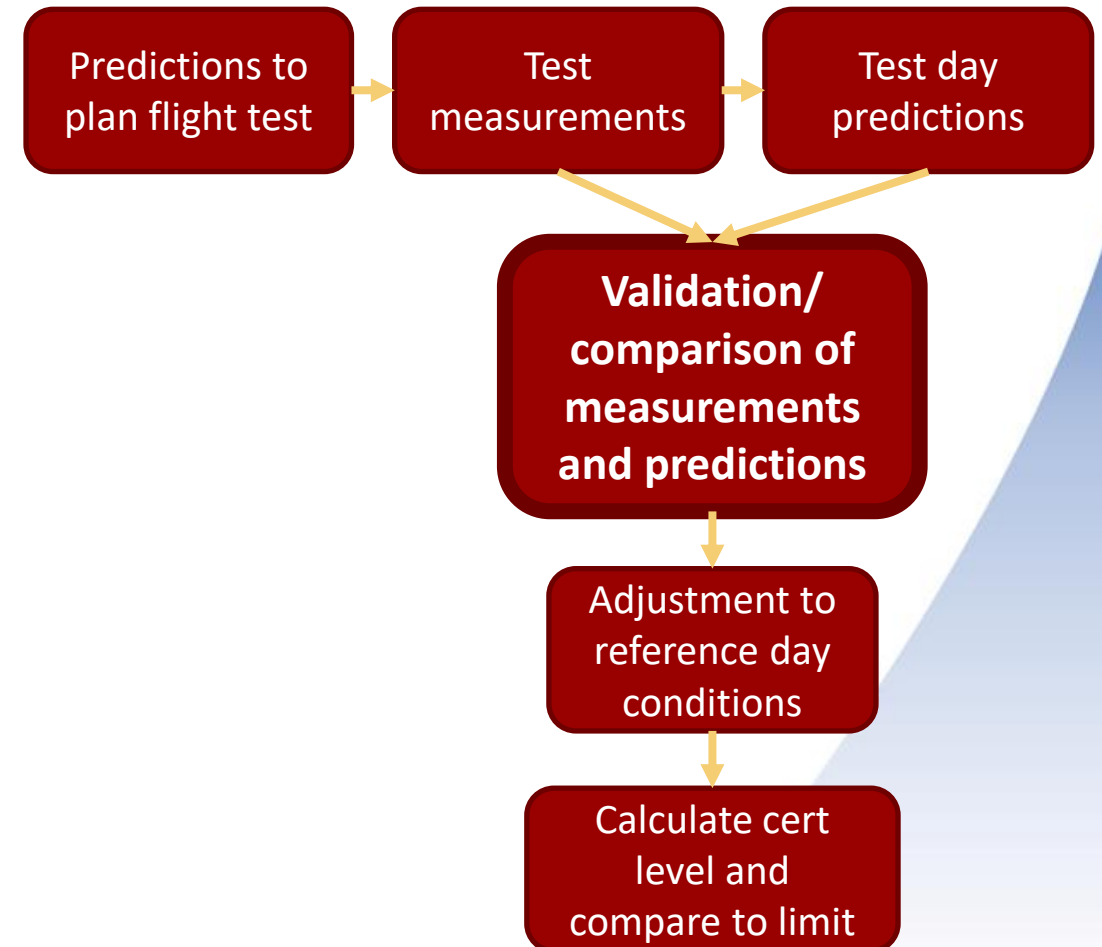


➤ Procedure must characterize noise performance at reference conditions

- Concept of fairness (“level playing field”)
- Procedure is not finalized yet



Notional Certification Procedure Steps



Procedure Development and Exercising



- **Need for a dataset to aid in developing and exercising procedures**
 - No existing measured shaped low-boom aircraft data
 - Procedures will be tested with X-59 data in the future
- **Procedure developed with simulation datasets in the past¹**
 - Results were limited because did not capture realistic variance of measured data
- **Extend analysis with measured N-wave data**
 - NASA SonicBAT measured data²
 - Previous exercise found that test day predictions without atmospheric turbulence effects show much less variation than measurements³
 - Perform new simulations of sonic boom propagation through atmospheric turbulence with SonicBAT conditions
 - Limitation: assesses N-wave data and not signatures from shaped aircraft, where atmospheric turbulence effects are expected to be reduced

NASA SonicBAT I Flight Test



➤ Flight test goal:

- Understand turbulence effects on ground measurements of sonic booms

➤ Flight test procedure:

- 20 flights (69 passes) at Edwards AFB, CA, USA (July 2016)
- Fly F-18 at Mach 1.39 at 10.4 km over microphone array through various levels of atmospheric turbulence
- Motor glider mic above turbulent ABL
- Meteorological measurements



SonicBAT Measured Data



- **Exercise certification procedure with data from 2 flights (6 passes total)**
- **Measured F-18 trajectory**
 - Altitude, heading, lat/lon, Mach, derivatives
- **Measured waveforms along microphone array**
 - Six sonic boom perception metrics computed for each windowed waveform⁴
- **Measured atmospheric profile from weather balloon**
 - Temperature, pressure, wind speed and direction, and relative humidity as a function of altitude
 - One profile per flight
- **Measured atmospheric turbulence parameters**
 - Weather balloon and SODAR data used to derive ABL height for each flight
 - Sonic anemometer used to collect temperature and 3D wind data to derive turbulence parameters

Turbulence Category	σ_{vector} (m/s)
Low	0.31
Medium	1.10
High	1.89

SonicBAT Predictions Through Atmospheric Turbulence



- 1. F-18 nearfield pressure from CFD, trajectory, meteorological data, array location**
- 2. PCBoom used to select ray closest to array midpoint**
- 3. PCBoom nonlinear acoustic propagation to top of ABL**
- 4. KZKFourier propagation through ABL to ground for 500 m virtual mic array**
 - 20 random realizations of turbulent ABL fields for each pass (~20,000 waveforms)
 - Sonic boom perception metrics computed
- 5. After prediction validation of method with measured data, predict reference day⁵ level with PCBoom**

Statistical Analysis for Prediction Model Validation

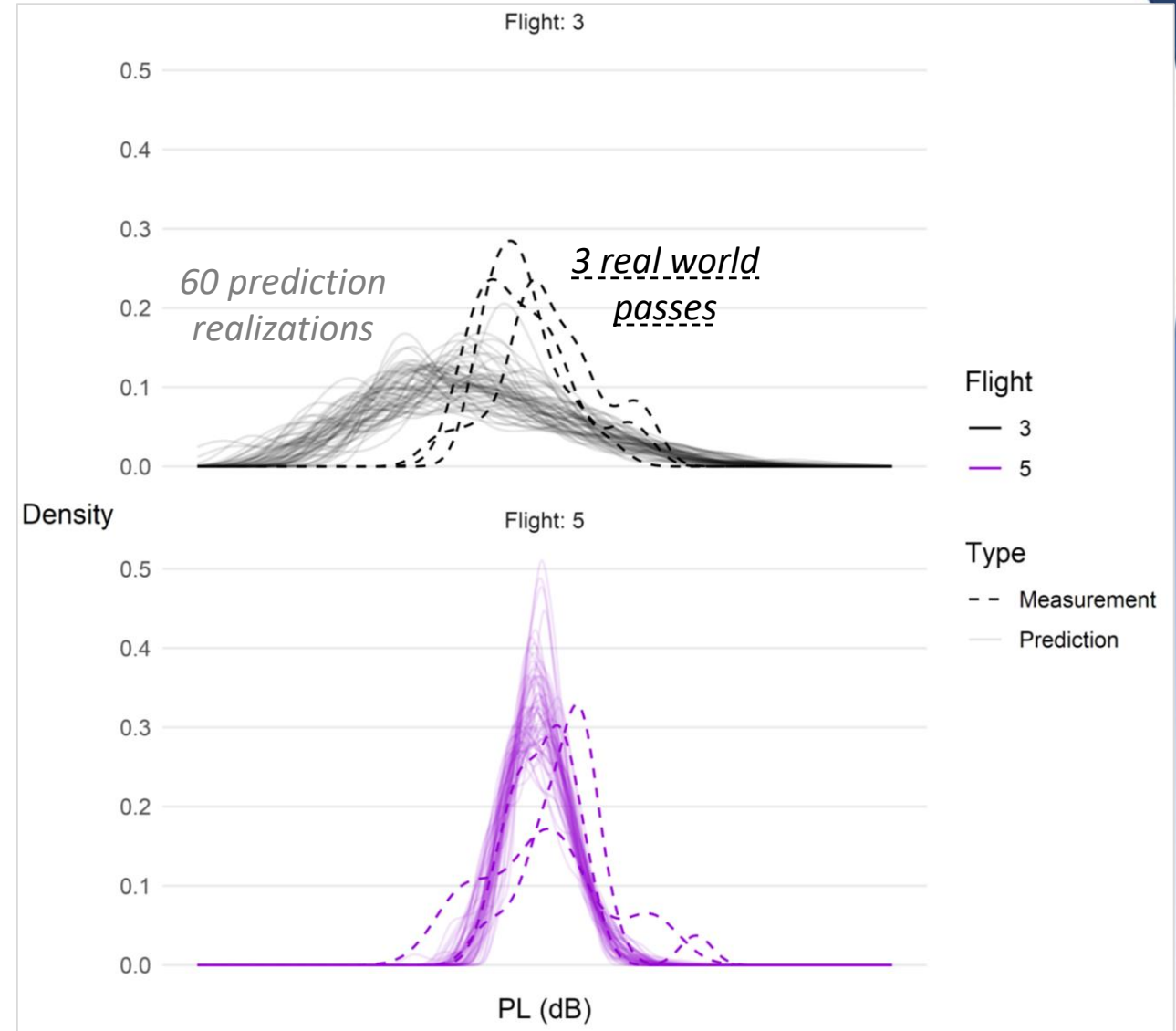


- Goal is prediction method validation
- Institute for Defense Analyses (IDA) statistical handbook⁶ suggests the following for validating models with real world data:

- Plot and summarize data
- Assess variability
- Hypothesis tests

- **Initial analysis with distributions**

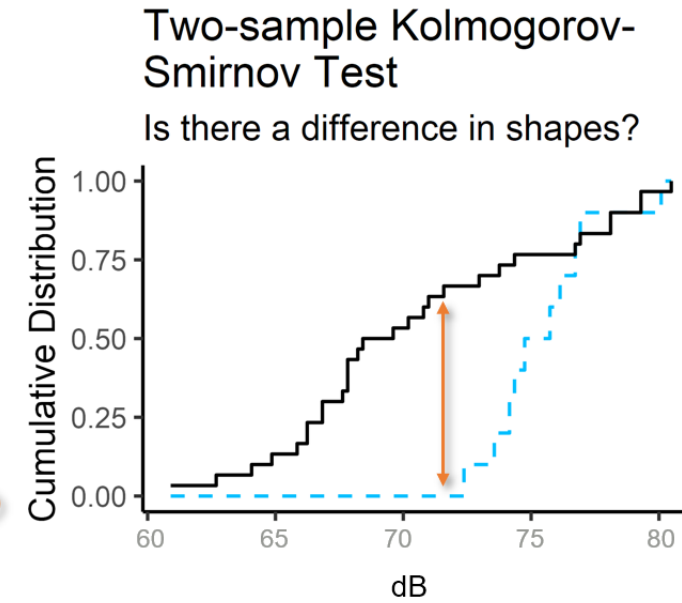
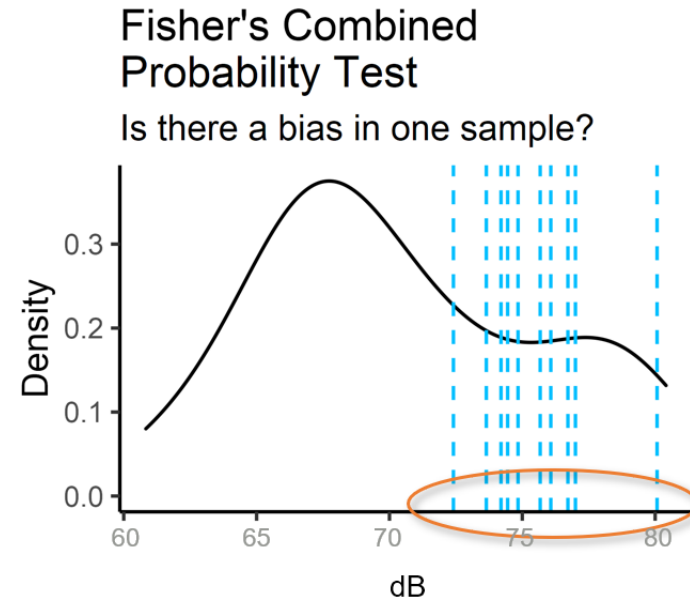
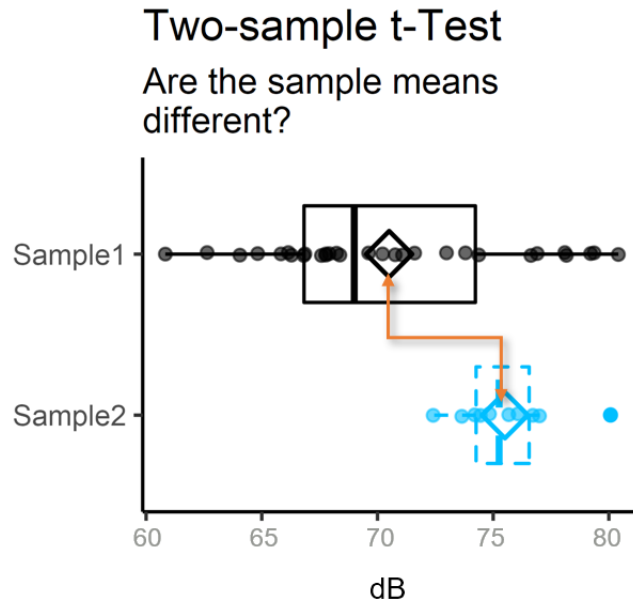
- Flights separated for further analyses because of differences in distributions
- Same trends for other metrics
- Considered subset of predicted data to match measured data resolution



IDA Statistical Analysis for Prediction Model Validation



- **Each test compares two groups of data in different ways**
 - Addresses differences in means and distributions
- **If any of three tests has p-value below threshold, conclude that two samples are different**
 - Significance threshold is adjusted to control for multiple testing



For SonicBAT subset, IDA tests show differences between measurement and prediction data and may be too stringent for this application

Equivalence Testing for Prediction Model Validation



➤ Equivalence testing is another method

- Look for evidence that two treatments are the same within some amount
- Typical example is comparison of a new medicine to the current standard: show that two treatments have same safety or effectiveness, within some margin

➤ For certification procedure, hypotheses of equivalence test are given below

$H_0: |\Delta_{mean}| > 1dB$ (Measurement and prediction means are different)

$H_A: |\Delta_{mean}| \leq 1dB$ (Measurement and prediction means are the same within 1dB)

- To validate the predictions, hope to reject the null hypothesis and conclude the alternative, that the means are the same within a small, practically negligible margin
- Need to define margin of equivalence beforehand

Equivalence Testing Results



- Flight 3 predictions tend to under-predict and results show differences at 1 dB
- Flight 5 results show predictions are the same as measurements, within 1 dB

Flight	Metric	Equal Variance	Pred Mean – Meas Mean	Equivalence Margin	TOST p.value	Equal Means
3	PL	TRUE	-2.260	1	0.999	FALSE
3	ASEL	TRUE	-2.563	1	0.999	FALSE
3	BSEL	TRUE	-1.263	1	0.881	FALSE
3	DSEL	TRUE	-1.523	1	0.993	FALSE
3	ESEL	TRUE	-1.913	1	0.999	FALSE
3	ISBAP	TRUE	-1.402	1	0.964	FALSE
5	PL	TRUE	-0.241	1	<0.001	TRUE
5	ASEL	TRUE	-0.250	1	0.001	TRUE
5	BSEL	TRUE	-0.379	1	<0.001	TRUE
5	DSEL	TRUE	-0.435	1	<0.001	TRUE
5	ESEL	TRUE	-0.343	1	<0.001	TRUE
5	ISBAP	TRUE	-0.316	1	<0.001	TRUE

Only with an equivalence margin of at least 3 dB do all tests show equal means

Medium turbulence

Low turbulence

- **Methods to validate predictions have been exercised for certification procedures**
 - While some methods compare mean values, others compare values across the distribution, and a combination of methods can be used
 - The methods were applied to a subset of an N-wave dataset to demonstrate the applicability and identify potential areas for improvement
 - Atmospheric turbulence variability was included as a factor in these analyses
 - Results were computed for six noise metrics (PL, ASEL, BSEL, DSEL, ESEL, and ISBAP)
- **Equivalence testing allows for practical validation of predictions with measurements**
 - For the two example N-wave flights evaluated here, predictions were only validated for both of the flights when the equivalence margin was set to 3 dB or greater
 - When only considering the lowest turbulence flight, an equivalence margin of 1 dB was sufficient for validating the predictions
 - Selecting and justifying the size of the equivalence margin requires further discussion
 - Only gives results for equivalence of sample means

Next Steps



- Develop extended equivalence test looking at means and variance
- Consider additional sample data
- Investigate other meteorological data
- Refine method for exercising with NASA X-59 data





1. Loubeau et al. Developing certification procedures for quiet supersonic aircraft using shaped sonic boom predictions through atmospheric turbulence. *J. Acoust. Soc. Am.* 149(4), 2021.
2. Bradley et al., Sonic booms in atmospheric turbulence (SonicBAT): The influence of turbulence on shaped sonic booms, NASA/CR-2020-220509, 2020.
3. Loubeau et al. Sonic boom prediction and measurement analysis methods for certification of quiet supersonic aircraft. *J. Acoust Soc. Am.*, 146: 2780, 2019.
4. Loubeau et al. Updated evaluation of sonic boom noise metrics. *J. Acoust Soc. Am.*, 144: 1706, 2018.
5. Loubeau et al. New reference day atmosphere for analyses of quiet supersonic overflight. *J. Acoust Soc. Am.*, 2023.
6. Wojton et al. Institute for Defense Analyses (IDA) handbook on statistical design & analysis techniques for modeling & simulation, IDA Document NS D-10455, 2019.

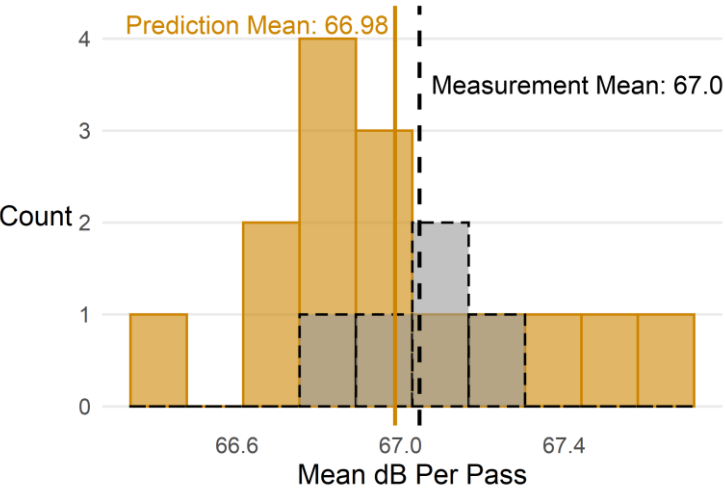
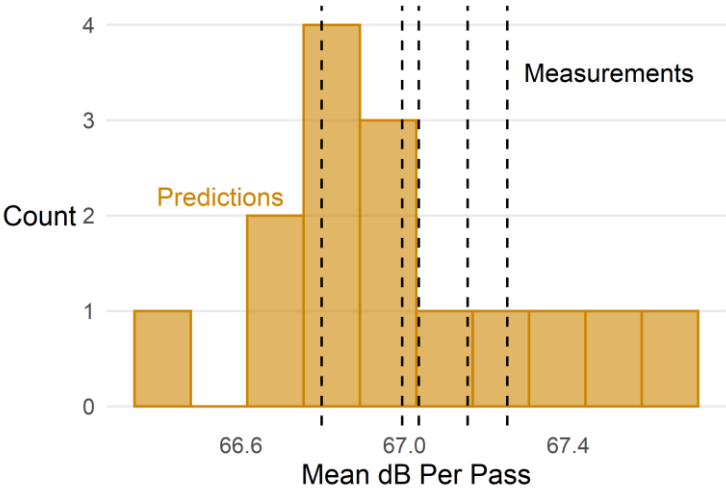
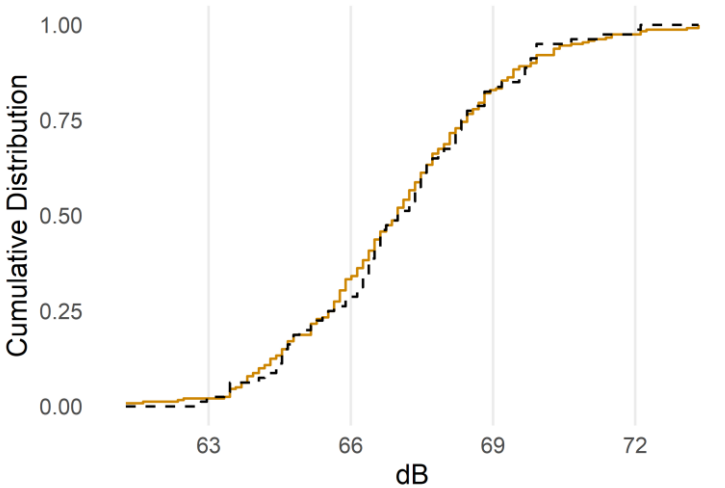


Backup Slides

- **En route noise certification procedures for quiet supersonic aircraft are in development under the International Civil Aviation Organization for a proposed standard that could enable civil supersonic overland flight. While these quiet supersonic aircraft do not exist yet, methods are being developed and evaluated using existing N-wave empirical datasets and low-boom simulation datasets. One potential aspect of future certification methods is the ability to use a validated prediction model to predict the certification noise level under reference day weather conditions. Methods to validate predictions, such as the two-sample t-test, Fisher's combined probability test, two-sample Kolmogorov-Smirnov test, and equivalence testing, have been exercised and modified for this application. While some methods compare mean values, others compare values across the distribution, and a combination of methods can be used. The methods were applied to a subset of an N-wave dataset to demonstrate the applicability and identify potential areas for improvement. Atmospheric turbulence variability was included as a factor in these analyses through KZKFourier propagation simulations with parameters matching actual flights and weather conditions. Results are presented for six noise metrics (PL, ASEL, BSEL, DSEL, ESEL, and ISBAP), and plans to expand the analyses are discussed.**

IDA Combination Test Example with Simulated Data



	T-test	Fisher's Combined Probability Test	Two-sample Kolmogorov-Smirnov Test
	Is there a difference in group means?	Is there a collective bias in the measurements?	Is there a difference in shape between the two groups?
Example Plot	Metric: PL ABL Height: 411.38m Prediction Mean: 66.98 Measurement Mean: 67.0 	Metric: PL ABL Height: 411.38m Predictions Measurements 	Metric: PL ABL Height: 411.38m 
Data	Mean of 16 microphones per pass - 5 measurement means - 15 prediction means	Mean of 16 microphones per pass - 5 measurement means - 15 prediction means	16 microphones from each pass - $5 \times 16 = 80$ measurement values - $15 \times 16 = 240$ prediction values
Assumptions	Samples are independent, approximately normally distributed	Set of independent tests	Distributions are continuous

IDA Combination Test Results for SonicBAT Data



- Statistically significant difference indicates that the code and measurements do not agree, and some remediation would become necessary before code could be used for certification

Flight	Metric	t-test p-value	KS test p-value	Fishers Combined p-value	Minimum p-value	Statistically Significant Difference?
3	PL	0.024	< .001	0.005	< .001	Yes
	ASEL	0.008	< .001	0.005	< .001	Yes
	BSEL	0.027	< .001	0.016	< .001	Yes
	DSEL	0.034	< .001	0.009	< .001	Yes
	ESEL	0.018	< .001	0.008	< .001	Yes
	ISBAP	0.050	< .001	0.010	< .001	Yes
5	PL	0.546	0.19	0.289	0.190	No
	ASEL	0.640	0.139	0.206	0.139	No
	BSEL	0.144	0.001	0.013	0.001	Yes
	DSEL	0.183	0.003	0.061	0.003	Yes
	ESEL	0.351	0.018	0.130	0.018	Yes
	ISBAP	0.215	0.014	0.086	0.014	Yes

90% confidence level

Medium turbulence

Low turbulence