NASA/CR–2020-220589/Volume I

Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas Risk Reduction for Future Community Testing with a Low-Boom Flight Demonstration Vehicle

Juliet A. Page
Volpe National Transportation Systems Center, Cambridge, Massachusetts
Kathleen K. Hodgdon
Penn State University, State College, Pennslyvania
Robert P. Hunte and Dwight E. Davis
Applied Physical Sciences Corporation, Groton, Connecticutt
Trent A. Gaugler
Gaugler Consulting, Easton, Pennslyvania
Robert Downs
Volpe National Transportation Systems Center, Cambridge, Massachusetts
Robert A. Cowart
Gulfstream Aerospace, Savannah, Georgia
Domenic J. Maglieri
Eagle Aeronautics, Newport News, Virginia
Christopher Hobbs
KBRWyle, Arlington, Virginia
Gary Baker
Volpe National Transportation Systems Center, Cambridge, Massachusetts
Matthew Collmar
Gulfstream Aerospace, Savannah, Georgia
Kevin A. Bradley
KBRWyle, Arlington, Virginia
Brian Sonak and Diana Crom
Penn State University, State College, Pennslyvania
Christopher Cutler
Volpe National Transportation Systems Center, Cambridge, Massachusetts

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov

- E-mail your question to help@sti.nasa.gov

- Phone the NASA STI Information Desk at 757-864-9658

- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/CR–2020-220589/Volume I

Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas Risk Reduction for Future Community Testing with a Low-Boom Flight Demonstration Vehicle

Juliet A. Page
Volpe National Transportation Systems Center, Cambridge, Massachusetts
Kathleen K. Hodgdon
Penn State University, State College, Pennslyvania

Robert P. Hunte and Dwight E. Davis
Applied Physical Sciences Corporation, Groton, Connecticutt

Trent A. Gaugler
Gaugler Consulting, Easton, Pennslyvania

Robert Downs
Volpe National Transportation Systems Center, Cambridge, Massachusetts

Robert A. Cowart
Gulfstream Aerospace, Savannah, Georgia

Domenic J. Maglieri
Eagle Aeronautics, Newport News, Virginia

Christopher Hobbs
KBRWyle, Arlington, Virginia

Gary Baker
Volpe National Transportation Systems Center, Cambridge, Massachusetts

Matthew Collmar
Gulfstream Aerospace, Savannah, Georgia

Kevin A. Bradley
KBRWyle, Arlington, Virginia

Brian Sonak and Diana Crom
Penn State University, State College, Pennslyvania

Christopher Cutler
Volpe National Transportation Systems Center, Cambridge, Massachusetts

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

Prepared for Langley Research Center
under Contract NNL15AA00C

May 2020

Acknowledgments

Technical Monitor: Jonathan Rathsam, NASA Langley Research Center

Point of Contact: Larry J. Cliatt, II, NASA Armstrong Flight Research Center

Available from:

NASA STI Program/Mail Stop 148
NASA Langley Research Center
Hampton, Virginia 23681-2199
Fax: 757-864-6500

Contents

List of Figures

List of Tables

List of Appendices

Acronyms

Abbreviation	Term
ABS	Address Based Sample
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
API	Application programming interface
ARMD	Aeronautics Research Mission Directorate
ATM	Automatic Transaction Machine
CAEP	Committee on Aviation Environmental Protection
CDMA	Code division multiple access
BDNL	B-weighted Day-Night Noise Exposure Level
CDNL	C-weighted Day-Night Noise Exposure Level
DDNL	D-weighted Day-Night Noise Exposure Level
EDNL	E-weighted Day-Night Noise Exposure Level
FDNL	Unweighted Day-Night Noise Exposure Level
DailyLLZd	Day-Night Noise Exposure Level of Zwicker loudness levels in phons, for diffuse incidence, calculated using a time constant of 70 msec and averaging across the two peaks.
DailyLLZf	Day-Night Noise Exposure Level of Zwicker loudness levels in phons, for frontal incidence, calculated using a time constant of 70 msec and averaging across the two peaks.
DailyPNL	Day-Night Noise Exposure Level of Perceived Noise Level (dB)
DailyISBAP	Day-Night Noise Exposure Level of Indoor Sonicboom Annoyance Prediction Level
CISBoomDA	Cockpit Interactive Sonic Boom Display Avionics, a version of PCBoom that allows pilot to see sonic boom footprint on ground while using flight simulator
COAMPS	Coupled Oceanographic Atmospheric Mesoscale Prediction System
COAMPS OS	Coupled Oceanographic Atmospheric Mesoscale Prediction System On Scene
ASEL	A-weighted Sound Exposure Level (dBA)
BSEL	B-weighted Sound Exposure Level (dB)
BYU	Brigham Young University
CSEL	C-weighted Sound Exposure Level (dBC)
DSEL	D-weighted Sound Exposure Level (dB)
ESEL	E-weighted Sound Exposure Level (dB)
FSEL	Unweighted Sound Exposure Level (dB)

Abbreviation	Term
dBA	A-weighted Decibels
dBC	C-weighted Decibels
CST	Commercial Supersonic Technology
DNL	Day-Night Noise Exposure Level
DS	Daily Summary, participant subjective data report
EAFB	Edwards Air Force Base
EDA	Exploratory Data Analysis
FAA	Federal Aviation Administration
GAC	Gulfstream Aerospace Corporation
GSM	Global System for Mobile Communications
HA	Highly Annoyed
ICAO	International Civil Aviation Organization
IFR	Instrument Flight Rules
IRB	Institutional Review Board
ISBAP	Indoor Sonicboom Annoyance Prediction Level
LBDM	Low Boom Dive Maneuver
LBFD	Low Boom Flight Demonstrator
LLZd	Zwicker loudness levels in phons, for diffuse incidence, calculated using a time constant of 70 msec and averaging across the two peaks.
LLZf	Zwicker loudness levels in phons, for frontal incidence, calculated using a time constant of 70 msec and averaging across the two peaks.
LOESS	Locally estimated scatterplot smoothing
MAC	Media Access Control
METARS	World Meteorological Organization surface weather reporting format
Npeak	Minimum (negative) pressure of recording (psf)
NQ	FAA Noise Quest website hosted by Pennsylvania State University (http://www.noisequest.psu.edu/)
OMB	Office of Management and Budget
PAO	Public Affairs Office
Peak	Peak (maximum) overpressure of recording (psf)
PL	Stevens' Mark VII Perceived Level (dB)
PLDN	Average Day-Night Perceived Noise Level (dB)
PNL	Perceived Noise Level (dB)
PrSG	Procedures Subgroup

Abbreviation	Term
PSEL	Perceived Sound Exposure Level (dB)
QA/QC	Quality Assurance/Quality Control
SE	Single Event, participant subjective data report
SERDP	Strategic Environmental Research and Development Program
SIGMETS	Significant Meteorological Event reporting format
SIM	Subscriber Identity Module
SMM	Social Media Monitoring
SonicBAT	Sonic Booms in Atmospheric Turbulence project
SPIKE	Sonic Pressure Integrated Kit Electronics
SRC	Penn State Survey Research Center
SSTG	Supersonic Task Group
TAF	Terminal Aerodrome Forecast
TCPA	Telephone Consumer Protection Act
USPS	United States Postal Service
Volpe-CARS	Volpe Center Acoustic Recording System
VPN	Virtual Private Network
WSPR	Waveforms and Sonic Boom Perception and Response
WSPRRR	WSPR Risk Reduction
ZDNL	Unweighted Day-Night Sound Level
%HA	Percentage Highly Annoyed

Foreword/Preface

For over 70 years, man has flown faster than the speed of sound. Concorde, the only successful commercial supersonic operation, provided the public with travel at Mach 2.0 but only over water due to its objectionable sonic boom. Retired in 2003, Concorde went down in history as a technological marvel well ahead of its time. Unfortunately, no civilian operational replacement has emerged. But that might be changing, thanks to the relentless pursuit of industry, NASA, and a small group of sonic boom experts.

NASA's Aeronautics Research Mission Directorate (ARMD) has identified a near term (2015-25) strategic goal of enabling the establishment of a standard for acceptable overland supersonic flight, in cooperation with international standards organizations [NASA, 2015]. Since then, ARMD has been developing and validating analytical design tools and technologies intended to enable the development of supersonic aircraft with low sonic boom. In the longer term (2025-35), ARMD seeks to advance its research to meet the desired sonic boom level in larger aircraft, as well as other challenge areas related to successful supersonic transports. This research will include the development and validation of technologies and tools to reduce propulsion emissions and noise affecting the airport community.

In 2015 NASA awarded a two part Community Response NASA Research Announcement (NRA) to the Applied Physical Sciences (APS) team for conceptualizing a sonic boom community response test in anticipation of a low-boom flight research program and executing the QSF18 flight experiment in Galveston, Texas. In early 2018 NASA took another step toward re-introducing supersonic flight with a $247.5 million contract to Lockheed Martin for the design, manufacture, and flight testing of a supersonic research aircraft, now known as the X-59, that reduces a sonic boom to a gentle thump.

NASA's actions are providing core leadership that will make it possible to realize quiet civilian supersonic flight over land. Such flight is currently banned in the United States and elsewhere. Changing the current regulations in the US and abroad will require extensive measurements showing that the advancements in sonic boom signature shaping technology are sufficient to find community acceptance. If these NASA programs are successfully executed, data acquired from the flight program will be used to guide policy on international standards for sonic boom.

Imagine a future where you could board a quiet supersonic transport aircraft and make a day trip across the Atlantic or Pacific oceans. With today's computational horsepower and analytical software tools coupled with the ingenuity of the human mind and passion to solve the world's most challenging problems, engineers and researchers are about to 'crack-the-code' to enable civil operations with the next generation of aviation technology.

Executive Summary

The world, including the FAA and ICAO, has taken note of the confluence of advanced aircraft design/modeling capabilities and low-boom technology maturation that will eventually allow for civil supersonic overland operations.

In 2011 NASA, in combination with some of the current team members, conducted a proof-of-principle pilot test using an F-18 low boom dive maneuver (LBDM) over Edwards Air Force Base (EAFB) to demonstrate techniques for gathering human subjective response to low booms. The Waveform and Sonicboom Perception and Response (WSPR) research project, [Page *et al*., 2014], was a practice session for future low boom testing using a purpose-built low boom research aircraft. Key WSPR outcomes included: confidence in survey instrumentation, modes of delivery, data acquisition and dose-response correlation, and subsequent statistical analyses procedures. However, this test covered a limited geographic area of approximately one square mile with participants recruited in cooperation with EAFB authorities who were at home most of the day and accustomed to hearing sonic booms from Air Force operations.

Following WSPR 2011, NASA solicited team proposals to address risk reduction for future community testing with a NASA Low-Boom Flight Demonstration (LBFD) vehicle. The team, led by Applied Physical Sciences Corp. (APS) and consisting of Penn State University Applied Research Laboratory, Penn State Survey Research Center, Volpe National Transportation Systems Center (U.S. Department of Transportation), Gulfstream Aerospace Corporation, KBRwyle, Eagle Aeronautics and Gaugler Consulting, was selected by NASA and tasked with executing the Waveforms and Sonicboom Perception and Response Risk Reduction (WSPRRR) project, a two phase, multi-year effort. The major accomplishments of the WSPRRR project are outlined below and described further in this report and its appendices.

Accomplishments:

WSPRRR Phase 1 (2015-2016)

- Developed a detailed conceptual plan for community dose-response testing with the NASA Low-Boom Flight Demonstrator (LBFD) within the contiguous United States which could support establishment of a new noise-based sonic boom standard for supersonic aircraft certification;
- Identified key risks and development areas associated with the planning, execution, and data analyses of such testing; and
- Proposed and obtained NASA approval to conduct Phase 2 risk reduction activities in priority research areas that would require further understanding prior to executing the proposed test including: additional sonic boom propagation meteorological analysis, weather hardening of acoustic instrumentation, development of protocols and procedures for engaging communities to obtain subjective response data and conducting a low-amplitude sonic boom dose-response test in a community not used to hearing sonic booms using the F-18 Low Boom Dive Maneuver (LBDM, a surrogate noise source for the LBFD).

WSPRRR Phase 2 (2016-2019)

- Developed a detailed test plan for conducting a low-amplitude sonic boom dose-response test in a community not used to hearing sonic booms using the F-18 LBDM, i.e. Quiet Supersonic Flights 2018 (QSF18);
- Planned and conducted a pre-test at AFRC of updated acoustic instrumentation, community surveys and geolocation methods prior to QSF18;
- Prepared Office of Management and Budget (OMB) and Institutional Review Board (IRB) applications and received approvals to conduct the QSF18 test;
- Executed the QSF18 test in Galveston Texas in November 2018;
- Prepared and delivered an objective and subjective measurement data archive to NASA; and
- Analyzed the QSF18 test data and inferred dose-response relationships, assessed survey methods, compared findings to previous community noise studies and identified lessons learned applicable to future LBFD testing.

Design and Refinement of a Sonic Boom Dose-Response Test:

In Phase 1 WSPRRR embarked on a spiral design process for a conceptual LBFD test plan. Testing in a community not used to hearing sonic booms introduces many challenges not encountered at EAFB, e.g. off-range focus/climb signature placement, participant mobility, wide area acoustic measurements, recruitment challenges and diverse community dynamics. In developing this plan, the need to test and measure expected future conditions for gathering pertinent dose-response databases for the FAA and international regulators was used as a guiding principle. Studies project future sonic boom noise exposure in the US to be no more than 10 events daily over certain regions of the country [Rachami & Page, 2010; Salamone, 2009].

One key element of a future LBFD community test design is ensuring adequate representation of the US population including climate, housing types and demographics. Techniques were created to identify communities for recruitment while considering flight planning and logistics, aircraft performance and range/endurance and seasonal meteorological effects on anticipated noise levels. A flexible, "balanced days[†]" noise exposure and statistical research design was created to quantify the necessary test participants and data points, acoustic measurement needs, operational tempo, community outreach/public engagement and finally recruitment strategies for establishing a dose-response relationship of low amplitude sonic boom noise in communities not used to hearing sonic booms. Throughout this process, risks associated with the LBFD conceptual dose-response test plan were identified and ranked. Several necessary design elements associated with these risks are now described.

First, participants of the WSPR 2011 test were residents living in EAFB housing and acclimated to noise associated with supersonic flight operations. While this test opportunity afforded a familiar environment with minimal risks, it deliberately did not broach challenges that communities not used to hearing sonic

---

[†] Balanced days refers to two days with similar CDNL values which are derived from a smaller number of loud events and a large number of quieter events. Additional description of the test design is contained in Appendix D.

booms present including: the absence of a predisposition to aircraft noise; potential unwillingness to participate in the experiment; safety and security of noise monitors and staff, and a host of other issues that present risks to the success of the eventual LBFD experiment and the attainment of certification for supersonic overland flight.

During WSPR 2011 only individuals who were at or near home during most of the day were recruited. Recognizing this would not be viable for future LBFD testing, a reliable non-intrusive method to determine participant location at the time of each event, yet sensitive to privacy concerns, would be necessary even if they didn't hear the noise or fill out a single event survey. QSF18 survey design leveraged location services in the web-based mobile enabled surveys and provided for reported home and work times from participants to ascertain their location. Additionally, the need for location identification factors into the OMB and IRB review processes. Participant location determination and evaluation of their single event and cumulative noise exposure were identified as high priority risk items.

Next, the geographic test area, including noise measurements and number of recruited participants, needed to be scaled up. WSPR 2011 covered approximately one square mile, included 100 test participants, and the sonic boom noise exposure, or "footprint", was relatively uniform over this area. The AFRC Pre-test increased to about 12 square miles, whereas QSF18 covered approximately 60 square miles with 500 participants with considerable sonic boom noise variation over the LBDM footprint.

Looking ahead to LBFD, each community deployment will likely include recruitment areas on the order of 2500 square miles with potentially tens of thousands of participants experiencing potential lateral noise variation over the footprint[‡]. Collective risk assessment and analyses suggested that more extensive use of a predicted noise footprint should be considered, anchored by fewer acoustic measurements and an interpolation scheme based on modeling. Such a move would lower both acquisition and field labor costs for deploying, operating, and recovering a large instrumentation burden. Noise measurements over increasingly larger areas requires adaptation of monitoring equipment, transition to cellular connectivity and refinement of recording techniques and protocols to ensure success. Equipment placement and monitor density must take into account the participant locations, dose determination methodology and the projected sonic boom footprint contours.

However, part of establishing the dose-response relationship is knowing what noise levels participants experienced during the course of their day. If their location is known and the sonic boom levels over the test area are quantified, one can estimate the noise level the participants experienced. This requires sufficient monitoring density or a method that augments measurement with model predictions. As the geographic area increases, monitoring equipment, installation and operation for a dense measurement area can become cost-prohibitive; further spaced apart empirical data node interpolation can be augmented with analytical predictions from a trusted tool such as PCBoom [Page *et al.,* 2010]. This is the

---

[‡] The LBFD cruise condition design likely includes some lateral variation across the footprint. Since the LBFD test incorporates constant Mach and altitude flight over the test area, any downtrack variation is expected to be due to stochastic effects from meteorological variability in the region during the flight.

method that the WSPRRR team developed and refined for the LBFD and adapted for QSF18.

Execution of the QSF18 Sonic Boom Dose-Response Test:

In November 2018, the QSF18 test was executed in Galveston Texas. Highlights of the test included:

- 9 flight days over a 2 week period
- 4 to 8 sonic thumps delivered daily with levels gradually increased during the test
- 500 members of the public recruited to participate in single event and daily summary surveys
- 15 noise monitors deployed to measure sound levels across the survey area
- 52 sonic thumps delivered over the test period
- 11869 single event surveys completed by the participants

What was Learned:

Outreach and recruitment were successful, but required significant advance NASA outreach to communities and helpdesk support during survey enrollment.

The AFRC Pre-Test and QSF18 test validated instrumentation upgrades and field checks of acoustic data, which supported on site decisions for noise dose and operational waypoints.

Logistics and site preparation went well, however urban ambient noise is a significant challenge for sonic boom metrics, indicating a need for more advance site scouting and low noise monitor siting. Manual waveform event identification techniques were effective, however are not feasible for larger scale testing with a significantly larger number of monitors. Further optimization of windowing and spectral subtraction of background noise is needed prior to the X-59 overflights.

The F-18 low boom dive maneuver introduced sounds beyond lateral cutoff that elicited participant response. The different acoustic character of these sounds requires different metric analysis procedures. Similar issues should be expected for future X-59 testing, thus development of appropriate analysis algorithms is necessary.

The F-18 low boom dive maneuver was useful for QSF18, but did not always deliver the intended PL metric values. Shortfalls in the propagation algorithms, as well as the complexity of the upper air meteorological profile in a humid coastal region, were factors.

Modeling to project subject noise dose was rigorously tested due to the nature of the F-18 low boom dive footprint. Steady level flight for X-59 should greatly simplify footprint contouring, but will likely require supplemental empirical stochastic analysis techniques for meteorological variability and turbulence-induced uncertainty.

Survey geolocation worked, facilitated by a combination of single event and daily summary survey location data. Background survey locations (home and work) should be validated during recruitment, prior to dose response testing, and survey fields should force entry of a georeferencable address.

More effective techniques are required to ensure participant compliance with the daily summary reports.

Further testing is required of the software for subjective data gathering. The dynamic nature of technology (internet browser, mobile device capabilities and location services) and the evolving topic of personal privacy (and the use of geo-tracking technology) will ensure this topic remains a high risk challenge.

Post test statistical analysis of the data gathered during QSF18 may be summarized as follows:

- The single event dose-response relationships for the metrics considered were established and showed a positive correlation between noise level and percentage highly annoyed (%HA) response.
- The correlation between cumulative daily dose and percent highly annoyed response was not statistically significant for QSF18. This finding is presumably driven by the lack of highly annoyed (HA) reports from which to estimate such a relationship in addition to the low levels of the sonic thumps.
- Reminders to participants resulted in significantly higher single event response rates among that group, however the opposite was true for the daily summaries, in that response rate was higher for those who did not receive single event reminders.

Past studies have suggested that the 75 dB Perceived Level (PL) target for X-59 at cruise will find community acceptance. If not, lower PL test levels will be needed. Levels may or may not decrease with lateral distance. If not, then lower PL's would require flights at higher altitudes which could have environmental or performance implications. It is vital to establish X-59 lateral patterns at cruise for each of the selected metrics.

What comes next?

The overarching driver affecting all aspects of future X-59 testing is the geometric breadth and scale of the dose-response testing and the operational tempo at which it needs to occur. This necessitates development of procedures and protocols for automated testing of data quality assurance and quality control during recruitment, subjective and objective data gathering resulting in near-real-time data analyses.

Additional consideration should be given to the desired aggregated dose-response database for X-59 community testing as a whole. What survey sampling techniques are plausible and what criteria (meteorological, seasonal, geographic, demographic) are needed to ensure a suitable dataset and what implications do those criteria have on recruitment? The overall site selection, prominent community process[§] and eventually the recruitment stencil[**] should be refined with an integrated process to account for survey sampling techniques, recruitment goals and database criteria.

While decades of objective data gathering have refined testing protocols that include near-real-time quality assurance processes, the corresponding subjective survey instruments are the result of a much more dynamic technology environment, hence near-real-time quality assurance processes tuned to the desired survey instruments require further development.

As the end of WSPRRR Phase 2 approaches, X-59 fabrication is now underway and future LBFD dose response test design is ongoing. Many lessons are still being learned as demonstrated with QSF18. As was suggested in Phase 1, test planning continues to be an iterative process and the team stands ready to support NASA's ongoing efforts.

[§] The prominent community process involves identifying and quantifying communities based on census information, local and regional boundary identifications and other parameters as described in Appendices A and B.

[**] The recruitment stencil is a repeatable process with deliberately randomized elements that can applied to each prominent community.

1. Introduction

1.1 Background

Sonic boom community overflight experiments conducted over St. Louis, Oklahoma, and Edwards Air Force Base during the 1960s as part of the U.S. Supersonic Transport (SST) Program and laboratory experiments led by NASA under the 1980s High Speed Civil Transport (HSCT) and the 1990's High Speed Research (HSR) programs, provide the foundation for the existing knowledgebase of the response of individuals and communities to sonic booms. The SST program demonstrated that the anticipated 1.5 lbs/ft^2 to 2.0 lbs/ft^2 SST booms would not be acceptable to the population at large and led to the current ban of supersonic overland civil flight. Research under HSCT and HSR developed and refined state-of-the-art sonic boom modeling and design tools and demonstrated in laboratory settings, that aircraft that exhibit low-amplitude shaped sonic booms in the range of 0.3 lbs/ft^2 to 0.5 lbs/ft^2 might possibly be acceptable to the general public. Technology explorations progressed on multiple fronts and on August 27, 2003, DARPA and NASA demonstrated with the F-5 Shaped Sonic Boom Demonstrator Aircraft (SSBD) that low-amplitude sonic boom flight was in fact possible via aircraft shaping.

After many decades of sonic boom progress [Maglieri *et al*., 2014], NASA's Aeronautics Research Mission Directorate (ARMD) established a series of projects, currently entitled the Commercial Supersonic Technology (CST) project, aimed at providing the research and leadership to enable the development of this new generation of supersonic civil aircraft. The project's near term objective was to develop tools and methods that will enable demonstration of overland supersonic flight with an acceptable sonic boom, and collect a large dataset of responses from a cross section of the population in the most natural of settings, i.e. the communities where people live and work. A low boom flight demonstration mission was to follow with two goals: 1) design and build a piloted, large-scale supersonic X-plane with technology that reduces the loudness of a sonic boom to that of a gentle thump; and 2) fly the X-plane over select U.S. communities to gather data on human responses to the low-boom flights and deliver that data set to U.S. and international regulators. This experimental aircraft, known as the X-59 Quiet SuperSonic Technology Demonstrator (QueSST) will be flown over U.S. communities starting in 2023.

QSF18 was the latest step in a series of experiments and analyses that are investigating human responses to aircraft-generated sonic booms. Figure 1-1 provides a comparison of the geographic and participant scales of the experiments including WSPR in 2011, the AFRC pre-test in 2017, the QSF18 test in 2018 and the anticipated LBFD testing in 2023 and beyond.

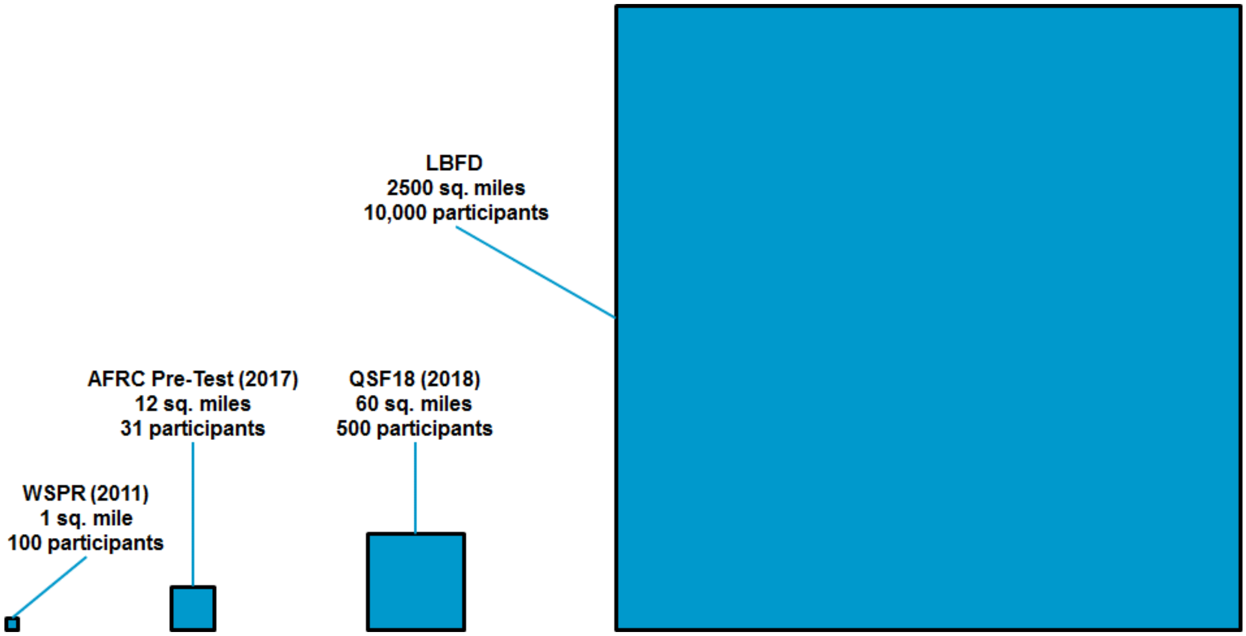


Figure 1-1 Comparison of sonic boom test areas and participant quantities for WSPR 2011, AFRC Pre-Test, QSF18, and the envisioned LBFD 2023

In 2011 NASA funded the Waveforms and Sonic Boom Perception and Response (WSPR): Low-Boom Community Response Program Pilot Test [Page *et al*., 2014]. This test was conducted over Edwards Air Force Base in California in 2011 and was designed to test and demonstrate techniques to gather data relating human subjective response to multiple low amplitude sonic booms using NASA's unique F-18 low boom dive maneuver (LBDM) [Haering *et al*., 2005]. It was in essence a practice session for further wider scale testing on communities not used to hearing sonic booms using a purpose built low-boom demonstration aircraft. Communities not used to hearing sonic booms present additional challenges beyond those overcome during the WSPR experiment. These include: the absence of a predisposition to aircraft noise; willingness to participate in the experiment; safety and security, and a host of other issues that present risks to the success of the experiment and the attainment of certification for supersonic overland flight.

The 2011 WSPR test was designed by members of this current project team and was executed in conjunction with NASA. The WSPR program addressed the following: design and development of an experiment to expose people to low-amplitude sonic booms, development and implementation of methods for collecting acoustical measures of the sonic booms in the neighborhoods where people live, design and administration of social surveys to measure people's reactions to sonic booms, and an assessment of the effectiveness of various elements of the experimental design and execution to inform future wider-scale testing. Key outcomes from that test were the confidence in the survey instrumentation, modes of delivery, data acquisition and dose-response correlation and subsequent statistical analyses procedures.

Building on the success of WSPR, the current effort, known as WSPR Risk Reduction (WSPRRR), consists of two phases. WSPRRR Phase 1, which was completed in 2016, described a conceptual dose-response test plan to address the following activities: recruitment, outreach, subject survey collection, correlation to noise, and statistical analyses. The Phase 1 report, “NASA Low Boom Flight Demonstrator Conceptual Test Plan for Community Response Testing, Risk Identification and Proposed Risk Mitigation Activities”, is attached to this report as Appendix A. Testing in communities not used to hearing sonic booms introduces many challenges not encountered in WSPR 2011 at EAFB, e.g. off-range focus/climb signature placement, participants whose locations change throughout the day, wide area objective measurement, and diverse community dynamics. The Phase 1 results provided the basis for a low-amplitude sonic boom subjective noise test in six different climate regions in the United States that will ultimately allow international regulatory agencies to draft a noise-based standard for the certification of civilian supersonic overland flight.

The Phase 1 report provides a rigorous risk assessment of all aspects of that conceptual dose-response plan. In response to those risks, follow-on risk-reduction activities were defined. The Phase 2 effort was comprised of two major additional risk reduction activities including 1) a “Pre-test” at Edwards Air Force Base in May 2017, which provided an opportunity to test the objective acoustic sensor system and subjective survey instrumentation in advance of 2) Quiet Supersonic Flight 2018 (QSF18) dose-response test over a larger geographic area with a community not used to hearing sonic booms utilizing the F-18 LBDM to create sonic thumps.

This document presents the end results of WSPRRR Phases 1 and 2. These results include development and documentation of a conceptual test plan (Appendix A) and a detailed test plan (Appendix D) for risk reduction testing of the NASA Low-Boom Flight Demonstrator (LBFD) within the contiguous United States.

1.2 Objectives

The objective of QSF18 was to conceptualize a future sonic boom community response test, use it to identify key risks and development requirements associated with the envisioned test; then to propose and conduct risk reduction activities in priority areas, and reassess the current status of testing in advance of the anticipated X-59 dose-response testing in 2023. Specific tasks included:

- Preparation of OMB and IRB applications to conduct the QSF18 testing
- Detailed test design including extensive analysis in support of QSF18 test site selection
- Execution of the QSF18 risk reduction field test
- Preparation and delivery of the QSF18 measurement data archive, noise exposure and community response databases
- Analyses of QSF18 test data to infer dose-response relationships, assess survey methods, and identify lessons for future LBFD tests
- Comparison of findings to previous community noise studies

Descriptions of these tasks are provided in the Overview of Accomplishments.

1.3 Overview of Accomplishments

The principal activity during WSPRRR Phase 2 was design, execution and analysis of the QSF18 community test in an area not used to hearing sonic booms. This risk reduction opportunity was conceived out of Phase 1 research and was conducted in Galveston Texas in November 2018. Extensive analyses, test design and preparation was conducted in order to ensure a successful test. The results of WSPRRR Phase 2 tasking are documented in this report, and include the following elements:

- *Preparation of OMB and IRB applications to conduct the QSF18 testing.* This milestone was critical from several perspectives. First, approvals by IRB and OMB are federally mandated prior to any public survey administration or human subject testing. Second, approvals support a scientifically valid and credible test design, ensuring that research methods and surveys were appropriate for the specified analyses. Third, nuances of the applications and informed consent, specifically regarding participant data handling, the use of mobile devices and respondent geo-referencing are important to understand prior to X-59 dose-response testing.
- *Planning and conduct of a pre-test of acoustic instrumentation and community surveys.* The AFRC pre-test gave confidence in hardware and survey software and deployment and survey reminder techniques. It allowed refinement of procedures prior to the QSF18 test in Galveston.
- *Detailed test design including extensive analyses in support of QSF18 test site selection.* Selection of a suitable test site in a community not used to hearing sonic booms had to be performed very carefully anticipating as many aspects as possible, so as not to create a problematic situation for future X-59 testing. The nature of the F-18 LBDM noise exposure introduced additional risk to the QSF18 test, so significant sonic boom analyses considering historical meteorological data were conducted to ensure full understanding of the range of potential footprints and the probability of delivering the desired noise dose to the community. This also led to the identification of community areas and geographic boundaries which guided subsequent recruitment and instrumentation deployment plans.
- *Execution of QSF18 risk reduction field test.* This test utilized the existing NASA F-18 LBDM to correlate human annoyance response with low level sonic thump noise in a coastal community setting. Past sonic boom research evaluated full scale N-wave booms, with levels that were approximately 1 psf or greater. The low level sonic thump is a new noise source of approximately 0.13 psf to 0.53 psf. This effort tested methods for remote aircraft basing and operations, community engagement, acoustic noise measurements, and conduct of community noise dose annoyance response surveys. QSF18 developed and evaluated research methods for future community response testing using the X-59 research aircraft. Figure 1-2 contains a summary of test data.
- *Preparation and delivery of the QSF18 measurement data archive and noise exposure and community response databases.* This comprehensive database was prepared and delivered electronically to NASA and includes both subjective and objective data as well as as-flown aircraft flight trajectories and forecasted and measured meteorological information.
- *Analysis of the LBDM test data to infer dose-response relationships, assess survey methods, and identify lessons for future LBFD tests.* A key accomplishment is the development of a dose-response relationship based on a logistic regression analysis for both single event and cumulative response data. Figure 1-3 presents the single event dose-response relationship, and

Figure 1-4 presents the corresponding daily summary relationship. Assessment of the performance of the geolocation aspects of the survey instruments was an important risk reduction activity (Figure 1-5).

- *Comparison of findings to previous community noise studies.* Cumulative dose-response data from QSF18 was found to be consistent with prior data including that at higher amplitudes during the 1970s and with lower sonic booms during WSPR 2011 (Figure 1-6).
- *Recommendations for future risk reduction activities.* During the QSF18 analysis phase and with the opportunity to revisit the Phase 1 efforts after some passage of time, several potential activities were identified for future consideration. These are described in Section 7.3 and fall into several areas including:
    - Influence of X-59 design, flight and low-boom performance capabilities on the community test plan
    - Site selection interdependencies including Demographics, Meteorological, Seasonal and Geographic (considerations for focus placement and avoidance)
    - Geographic nexus between prominent communities and the combination of flight operation patterns and lateral sonic boom noise distribution
    - Procedures to conduct near real-time Quality Assurance/Quality Control (QA/QC) of incoming data streams and data analyses
    - Objective measures refinement (e.g. ambient noise, metric analysis, expanded geographic coverage, automatic signal identification techniques, hardware reliability, security and weather handling improvements, meteorological stochastic quantification methods, low cell coverage)
    - Subjective tools and methods development (e.g. improved compliance, participant training, automation)
    - Relational subjective-objective database structure and associated IRB protocol development
    - Considerations for lateral cutoff sounds on test design, data acquisition and analysis

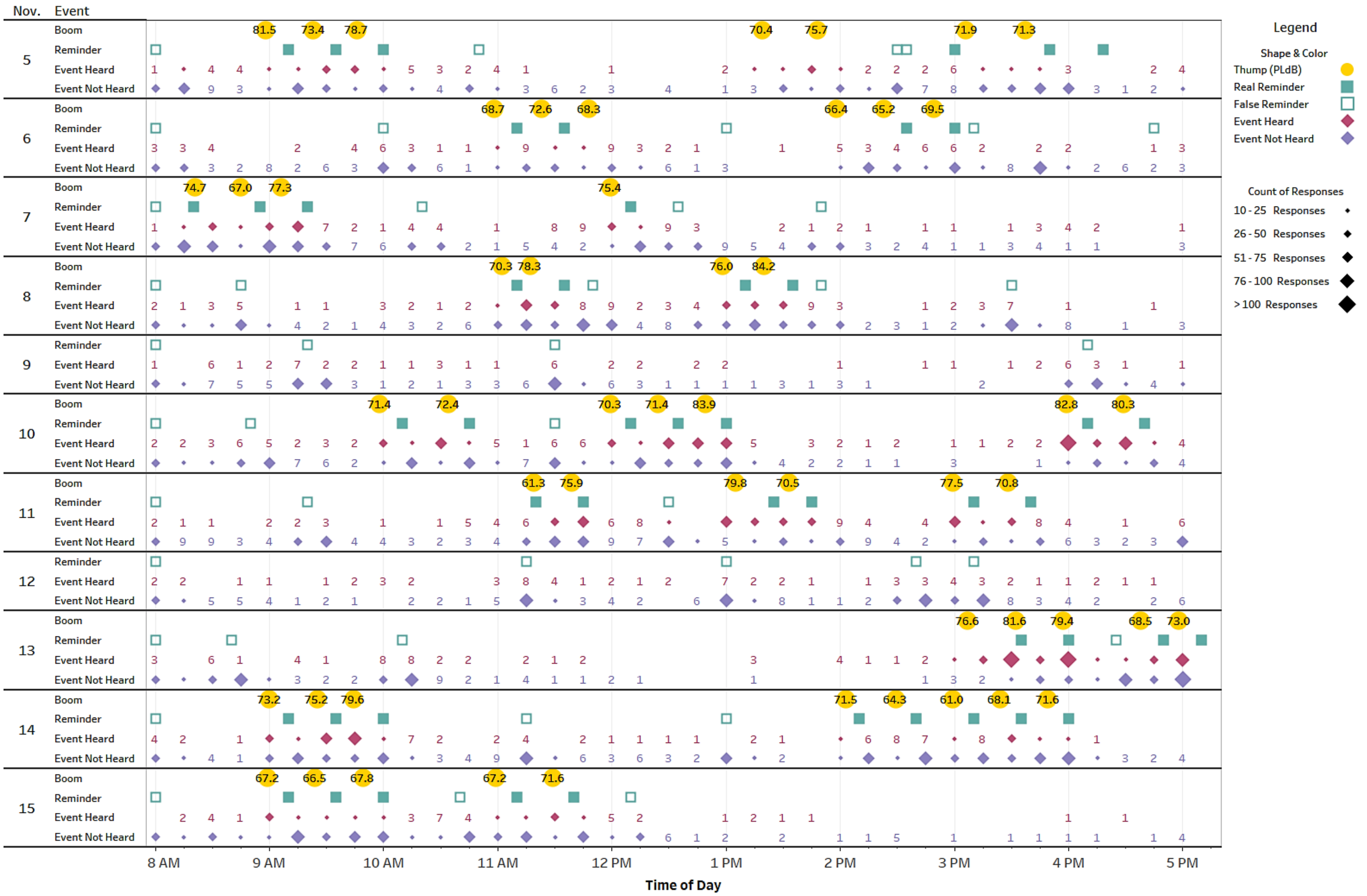


Figure 1-2 QSF18 Data snapshot: sonic thump events and average PLs on Galveston Island (yellow), real and false reminders (squares) and participant responses (diamonds)

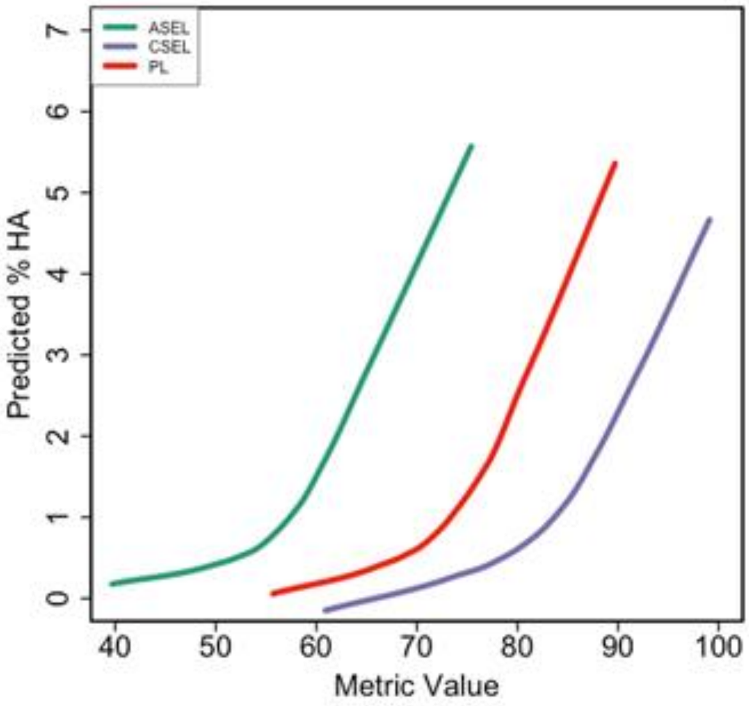


Figure 1-3 QSF18 dose-response relationships for single event response data

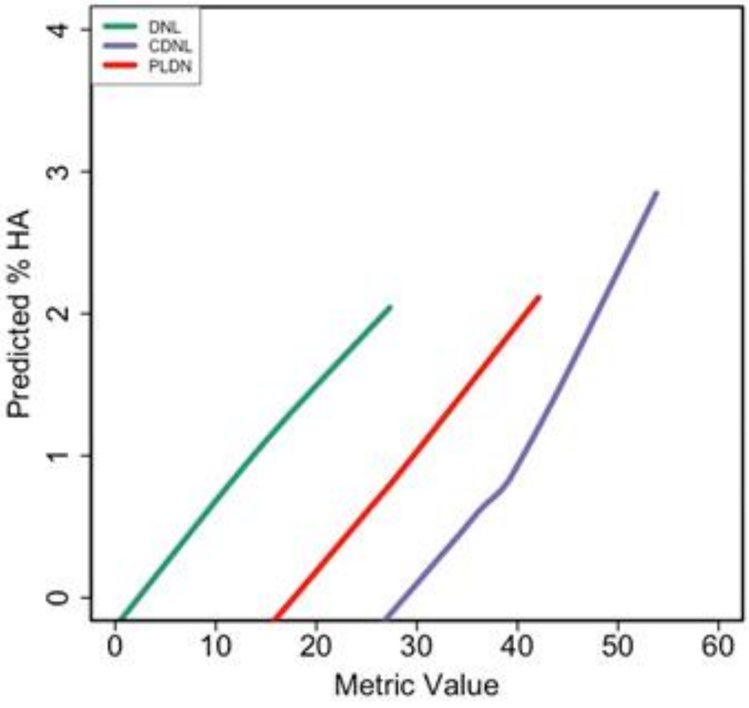
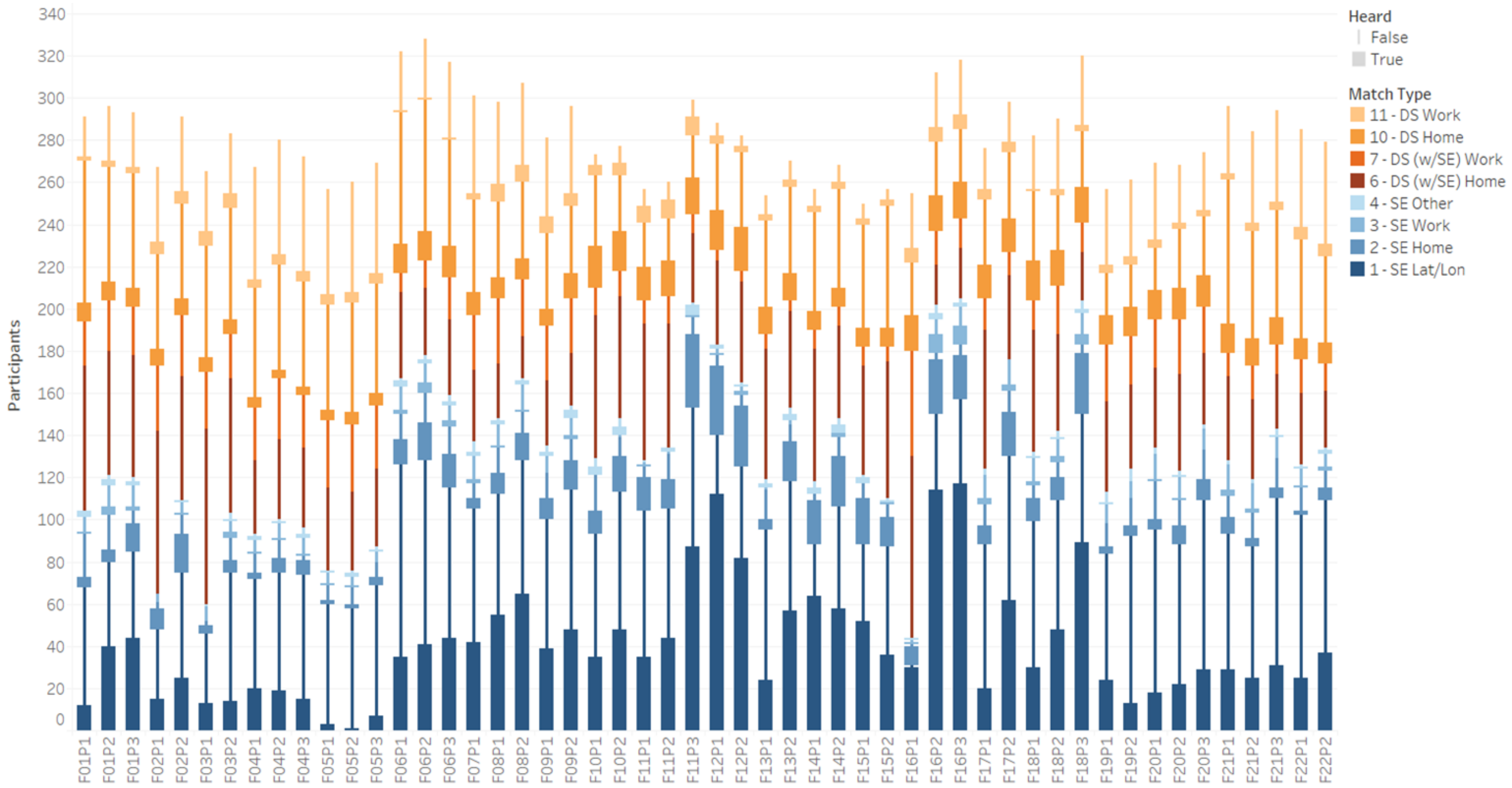


Figure 1-4 QSF18 dose-response relationships for cumulative daily summary response data



Does not include records for which a lat/lon couldn't be determined.

Figure 1-5 Geolocation data sources and response data summary

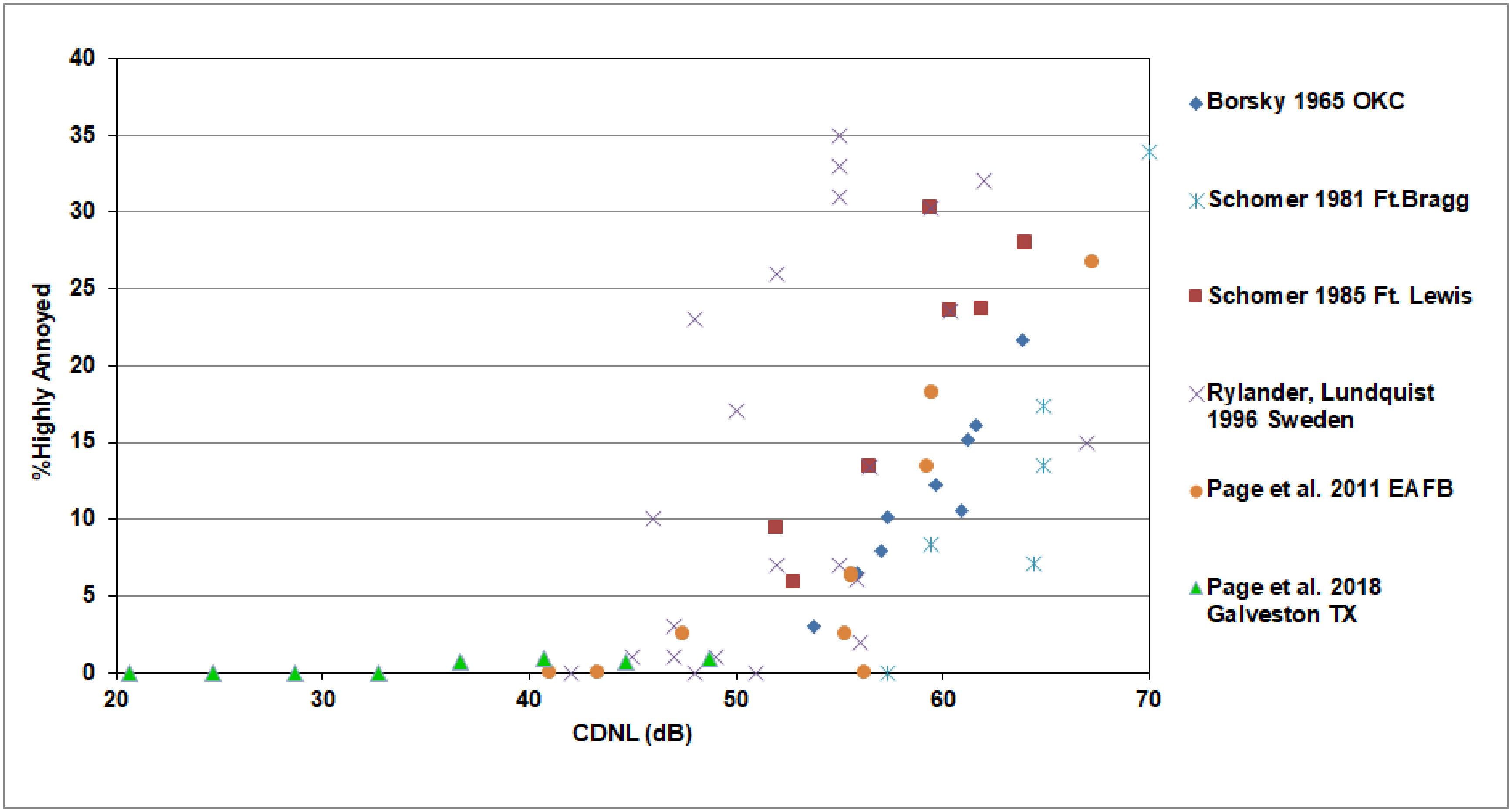


Figure 1-6 Comparison of QSF18 cumulative dose-response data with prior studies

2. Low Boom Flight Demonstrator Conceptual Test Plan and Risk Identification and Mitigation

In Phase 1 of this effort, the team executed several key actions required as part of the development of the low boom evaluation program. Specifically, these actions were:

(1) Developing a conceptual plan for testing of the NASA Low-Boom Flight Demonstrator (LBFD) within the contiguous United States;
(2) Identifying key risks and development areas associated with the planning, execution, and data analyses of such testing; and
(3) Proposing risk reduction activities in priority research areas that require further understanding prior to executing this test.

A detailed report describing the results of Phase 1 is provided in "NASA Low Boom Flight Demonstrator Conceptual Test Plan for Community Response Testing Risk Identification and Proposed Risk Mitigation Activities" (the Phase 1 report). The Phase 1 report is attached as Appendix A of this document.

The Phase 1 report was built on the team's WSPR 2011 experience, and describes a conceptual dose-response test plan to address the following activities: region and site selection; recruitment; community outreach; subject survey collection; noise dose design; acoustic measurements and collection of other objective data; correlation to noise; and statistical analyses. Communities not used to hearing sonic booms introduce many challenges not encountered at Edwards Air Force Base (EAFB), the site of the WSPR 2011 testing. These challenges include a wider range of climates, off-range focus/climb signature placement, participants whose locations change throughout the day, wide area objective measurement, and diverse community dynamics.

The conceptual test plan in the Phase 1 report provides the basis for a low-amplitude sonic boom subjective noise test in six different regions in the United States that will ultimately allow international regulatory agencies to support international policy allowing for the certification of civilian supersonic overland flight. The plan addresses community response to single event booms as well as daily cumulative response to multiple booms. The Phase 1 report presents the team's perspective on risk identification, prioritization and mitigation. A risk reduction plan was developed and key risk mitigation activities and outcomes were identified, along with proposed Phase 2 activities for further exploration and mitigation of high priority risks.

One key outcome of the Phase 1 activities was the identification and ranking of specific risks and recommendation that a test be conducted in advance of community testing with the low boom flight demonstrator. The Quiet Supersonic Flight 2018 (QSF18) test was thusly proposed. The 8 risks identified in Phase 1 (Figure 2-1), that require community participation include in priority order:

- #27 Participant Location Determination
- #23 No Subjective Response
- #33 Determination of noise at a participant’s Location
- #25 Participant Motivation
- #22 Low Boom Signature is a new noise source
- #21 Cross Community Comparison
- #26 Participant Recruitment Challenges
- #17 Media Response

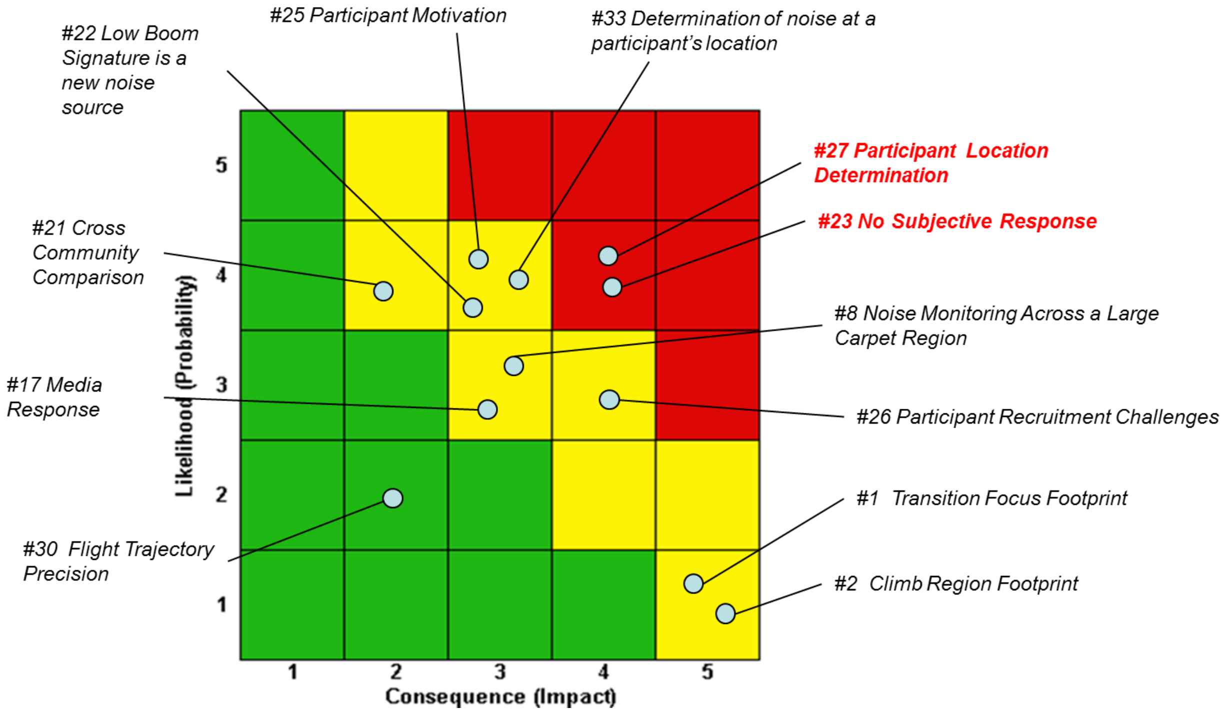


Figure 2-1 LBFD community response testing risks identified in Phase 1

As a result, the QSF18 test was recommended. This test provided an opportunity to engage the public on matters related to future testing using the X-59, including interfacing with public officials, emergency responders, local media and the public at-large. It also offered NASA AFRC Flight Operations Test Planning team the chance to further interface with regional air traffic management services for supersonic flights in the national airspace. Finally it provided an opportunity to coordinate logistical needs for remote aircraft basing, community recruitment and engagement and deployment of testing instrumentation for objective and subjective data gathering. The QSF18 site selection, test planning, actions and findings conducted as part of the Phase 2 efforts are detailed below in this report.

3. QSF18 AFRC Pre-Test

Prior to the QSF18 test in Galveston, Texas, a pre-test was designed and executed at Armstrong Flight Research Center (AFRC), in order to conduct risk reduction for acoustic and survey instruments and methods which were planned to be used in QSF18. The high priority methods tested at AFRC included: participant geolocation and survey web-based technology; acoustic instrumentation cellular integration; and sonic boom metric analysis and interpolation methodology. Overall, the 2017 Pre-Test provided execution details similar to those planned for QSF18 and in essence a "dry-run" in advance of going to an uncontrolled community not used to hearing sonic booms, including noise dose, flight operations/schedule and boom placement, participant recruitment, survey methods, SBUDAS acoustic measurement instrumentation and networking, and IRB approvals. A primary objective was to scale the acoustic array area from WSPR 2011 to a much larger area as shown in Figure 1-1. Specific risk reduction was required with regard to understanding the accuracy with which it would be possible to determine the location of a subjective response from a participant at the time of a sonic boom event, determining the effectiveness of the subjective survey methods, and determining the effectiveness of the cellular networking of acoustic data collection equipment across the full extent of the sonic thump footprint. Additionally, the AFRC pre-test was also designed to provide for Lessons Learned regarding the control and placement of the boom footprint from the F-18 LBDM within the test control area containing the ground acoustic array and test subjects. The pre-test also afforded an opportunity to test execution of communications, instrumentation setup and operation, and evaluate instrumentation set up time.

The test was conducted from 8-12 May 2017. Over the course of the three days 9 flights were executed with 21 booms delivered in the vicinity of 41 potential participants resulting in 252 boom recordings collected and the opportunity for collection of 861 responses (if every recruit participated and every participant responded to every boom). Participants received random text messages during the course of the day to remind them to be attentive for Sonic Booms. There were in fact 145 Single Event survey responses: 79 responses from WSPRRR Team Members, two responses where the ID was unknown, and 64 responses from AFRC participants. Figure 3-1 presents the noise monitor deployment positions. Six SBUDAS were deployed daily and operated by the team. The colored dots in the left portion represent WSPR 2011 sensor placement.

Figure 3-1 Noise monitor deployment positions

A detailed report describing the plan for the pre-test conducted at Armstrong Flight Research Center is provided in "Armstrong Flight Research Center Waveforms and Sonic boom Perception and Response Risk Reduction (WSPRRR) Test Plan". The AFRC pre-test test plan is included as Appendix F of this report. A detailed report describing the results of the pre-test conducted at Armstrong Flight Research Center is provided in "NASA Low Boom Flight Demonstrator Community Response Pre-Test Armstrong Flight Research Center May 8-12, 2017". The AFRC pre-test report is included as Appendix G of this report.

3.1 Lessons Learned From AFRC Pre-Test

The AFRC pre-test successfully provided lessons that were applied to the QSF18 Galveston testing, and which inform future LBFD test planning. Key findings include recommendations in the following areas:

- SBUDAS (acoustic data collection system) calibration, deployment, operation, and networking
- Operations, including base station and communications
- Metric calculations
- Survey techniques
- Geolocation
- Recruitment

The following subsections provide an overview of the lessons learned from the AFRC Pre-Test. A more detailed discussion of the lessons is provided in Appendix G.

3.1.1 SBUDAS Instrumentation

1. Multi-channel recordings having two channels provides valuable backup.
2. A dedicated LMR Radio Operator minimizes distractions and facilitates communications.
3. Limit calibration times to morning and at the close of each day – process is arduous and drift is small.
4. To avoid rain damage or shut down, need to determine method to weatherproof the noise monitors.
5. Stock kits with cones and reflective tape for night time and early morning operations.
6. Use smaller batteries and solar panels – simplify deployment.
7. Define at least two locations for each monitor - allows re-location in case of high ambient noise or poor modem connectivity.
8. Cellular networking problems – avoid cellular repeaters and disable non-vital network resources.
9. Cellular Modem VPN Configuration
10. Configure all modems with IPSec tunnels to all others - allow swapping modems between components.

3.1.2 Base Station

1. Recommend two base stations with individual operators - prevent overloading of operator and provide redundancy in the event of a base station failure.
2. Reliability of base station connection to VPN may be improved by direct connection to the internet rather than through cellular modem. Noise monitors and base station should be installed and field tested for validation in sufficient time prior to any flights.
3. Since the Command Center utilizes CISBoomDA, a version of PCBoom that allows a pilot to see the sonic boom footprint on the ground while using a flight simulator, for real time boom feedback, colocation of the base station is not required providing more flexibility in system deployment.

3.1.3 Window Length for Calculation of Metrics

1. For WSPR 2011 the window length for calculation of metrics was limited to 650ms (encompassing only the initial boom). Given that participants will be unacquainted with sonic booms it is likely that their response will be to the full event. The 650ms window as used during WSPR 2011 will continue to be utilized.

3.1.4 Communications

1. Minimize communications. Do not require acknowledgement unless a specific station is called.
2. Need direct link between all key responsible roles: relay induced delay.
3. Text messaging should be for information purposes only (as a log) – originate all decisions via radio transmission on the PI circuit.
4. LMR radios extremely useful for field coordination (far better than cell phone or text) and provided closed circuits. For the Community Response Test six months advance notice required.

3.1.5 Operations

1. Troubleshoot only between flights – can cause cascading problems that delay or nullify a flight.
2. Have at least one person not fixed to a location – assists problem mitigation.
3. Insert non-flight days for data assessment - necessary changes can be identified and introduced.
4. Common simple lexicon for characterizing audible booms – increases field note value.
5. For testing at secure facility, ensure unescorted access for all test team members.

3.1.6 Subjective Data Collection

1. Prompt participants to complete all survey protocols, this improves response rate and can be structured to minimize introduction of bias.
2. Use participant input to verify location: this reduces risk of inaccuracy of automated geolocation.
3. Survey Protocols:
    a. Initial emails from SRC going to receivers' spam folders, consider PSU outgoing address to ensure delivery.
    b. Clarify and manually enter location if uncertain whether automated location in GPS map is correct.
    c. Provide text prompts to encourage completion of background survey.
    d. Daily text prompt to remind participants to complete Daily Survey at end of day.
    e. Text just after each boom, and at random times: "A boom may have occurred. Did you hear a boom?"
    f. Include a link to the survey embedded within text messages.
    g. Evaluate option to go back within the individual survey when providing responses.
    h. Implement dates in selectable format rather than editable field.
    i. Investigate options for creating short cut to Qualtrics survey for iPhone and Androids.

4. QSF18 Site Selection and Detailed Test Plan

4.1 QSF18 Test Objectives

QSF18 provided the first flight test of a “low-boom” noise source over a community not used to hearing sonic booms, and an opportunity to gather data demonstrating methodology to correlate human annoyance with low level sonic boom noise. The assessment of community noise impact from civilian supersonic flight over land using a low boom dive maneuver included the investigation of relevant objective and subjective variables that affect the given noise environment.

Objectively, it was designed to adequately characterize the noise environment and identify appropriate metrics to represent it from empirically and analytically-derived measures. Subjectively, the test was designed to assess aspects of community impact including annoyance, attitudes, and the extent to which the noise interferes with daily activities. Correlations between objective and subjective variables can identify methods and metrics that relate to the subjective perception. Measurements of the single event and estimates of daily cumulative noise levels and associated survey responses are gathered to provide a comprehensive dose response data set.

Additionally, the test provided an opportunity to engage the public on matters related to this and future testing using LBFD, including interface with public officials, emergency responders, local media, and the public at-large. Finally, conducting F-18 research flight operations from yet another remote location offers NASA’s AFRC Flight Operations Test Planning Team the chance to build upon the Sonic Booms in Atmospheric Turbulence project (SonicBAT) off-range experience [Bradley *et al.* 2018]. The findings of this effort will provide lessons learned and further improve research methods for future community-scale response testing using the purpose-built LBFD.

Success criteria were defined in advance of the test design, and are itemized in Table 4-1.

Table 4-1 Phase 2 QSF18 success criteria

- No significant negative impact to either the Galveston community or the NASA X-59 / Commercial Supersonic Technology (CST) program
    - Mandates near real-time monitoring of data collection to provide input and guidance to an Adaptive Noise Dose Design that can respond to events almost as they occur
- Successful collection of operational lessons learned
    - Community Engagement and Participant Recruitment
    - Subjective Data Collection
    - Noise Monitor Deployment and Operations
    - Remote Basing of Aircraft Operations/Field Crew Logistics
- Assemble a scientific database sufficient for validating dose-response collection and analysis methods for future X-59 Community Testing

The following test parameters were established in cooperation with NASA to meet the objectives listed in Table 4-1:

- Community signatures, i.e. "thumps" (minimum of 32)
    - Plan includes 8 flight days, with an average of 3 flights daily
    - 2 to 3 "thumps" per flight, separated by a minimum of 20 minutes
    - The potential for up to 52 "thumps" over the test period
- Single monitor at Alpha (Scholes airport) must be in operational condition for Go decision
- Recruitment –Minimum of 400 participants
    - Planning 8000 recruitment letters sent
- Response Rates –Minimum of 7% responding regularly to noise
    - Based on WSPR2011 and AFRC Pre-Test 2017
- Subject locations positively identified 95% of the time
    - Not relying solely on automated methods to report locations

The determination of the minimum number of participants was based on a power analysis conducted during the development of OMB materials (see Appendix H). The response rates were based on past low boom tests and the geolocation rates were based on WSPRRR team based testing of the geolocation system. The targeted recruitment (500 participants) met the requirement for the number of participants (400 to 500) necessary to detect a dose response relationship with a slope of 0.015 for a power of 80%, as described in the OMB material (Appendix H). The recruitment yielded 500 potential participants prior to the test. This number dropped to 496 participants by the start of the test as described in Table 5-4. We did not identify a specific number of cumulative daily dose responses necessary in advance.

As will be described later, 5796 single event responses associated with a sonic thump and with a measurable noise dose were ultimately acquired. This is the result of an 8.5% participant enrollment rate, 51 thumps, and an overall 22.7% (5796/25500) single event dose determination success rate. For cumulative daily dose, 2585 combinations of participant - test days, out of a possible 4500 (500 participants * 9 flight days) were successfully computed, resulting in a 57.4% cumulative daily dose determination success rate.

4.2 QSF18 Site Selection

Following the Phase 1 LBFD Test design and Risk Reduction examination, NASA opted to consider a potential community test using the F-18 LBDM [Haering *et al.,* 2005]. The following criteria were identified to help guide the community selection:

- Coastal location where the loud focus boom from dive maneuver can be placed over water, away from residences
- Community should not be accustomed to hearing sonic booms
- Nearby NASA or military airfield for F-18 basing and operations
- Sufficient population density to recruit people and gather hundreds of survey responses to each boom event

- Coordination with local air traffic control feasible to facilitate execution of the F-18 low boom dive maneuver without interfering with commercial flight operations

After an assessment of continental United States regions, the following communities (Figure 4-1) were identified as candidate QSF18 test locations (listed in order of preference):

1. Galveston, TX
2. Melbourne, FL
3. Panama City, FL
4. Gulf Shores/Orange Beach, AL
5. Cape Cod, MA

Figure 4-1 Candidate communities under consideration for the Phase 2 risk reduction test

This process is consistent with the recommendations from the “Phase 1 Low Boom Flight Demonstrator Conceptual Test Plan” (Appendix A). In Phase 1 an extensive due diligence process was conducted to identify five candidate communities: Cape Cod Massachusetts, Melbourne Florida, Panama City Florida, Orange Beach Alabama, and Galveston Texas. In Phase 2, a detailed review was conducted of each of these communities relative to a multitude of variables and the risk that each of these variables introduces to each of the design aspects for the event. The method employed to identify risks was the same as executed for the overall design of the event in Phase 1.

Each of the variables considered in the site selection weighting process is based on a series of analyses that included the following elements:

- Storms / Seasonal Hurricane Storm Assessment (within 100 nmi of the potential site) obtained from the NOAA historical hurricane track database[††]
  - Number of named storms
  - Storm strength summation
  - Number of storm days
- Meteorology and atmospheric effects on the F-18 Low Boom Dive Maneuver Boom delivery repeatability using 10 years of historical upper air data over the potential test period months based on mean – most prevalent value, psf, (See Appendix B and Appendix C).
  - Extent and placement of low boom footprint area, nmi^2 for 60% vs. 80% probability
  - Extent of high psf areas, nmi^2 for 50% probability of $p>0.75$ psf
  - Extent of focal zone area, nmi^2 for 50% probability $p>1.5$ psf
  - Placement of focal zone onshore, nmi^2 $p>0.75$ psf
  - Month-to-month variability, standard deviation of monthly (mean – most prevalent level, psf)
- Boating prevalence including numbers of marinas and marine businesses within 50 and 100 miles and the number of USCG registered boats within 50 and 100 miles.
- Upper Airspace Use (based on FAA historical data during August to November 2016 and reflective of the number of flight miles aircraft traveled above 25 kft in the identified region).
- Demographics of communities utilizing the prominent community process identified in Phase 1 and reflective of aggregated community deviation from the overall USA distribution based on census data and including the following parameters: male/female, age, ethnicity, race, education, income and unemployment rate by county, (Appendix B).
- Social, Community & Cultural Capital factors have been patterned after biofuel refinery siting studies and augmented with additional parameters. This category includes: HS Diploma and College graduates by age 25, income, voter turnout, census return rate, cooperation/ collaboration with local government (assessed by number of POCs identified), local media outlets, numbers of museums and outreach venues, anthropogenic noise levels (L50 dBA levels), local transportation noise (aviation and non-aviation source), recent media reports (count of articles about noise in local media).
- Housing – Occupation Rate and Housing Types based on US Census Bureau, American Housing Survey, 2015[⊗]: total housing units, occupied units, owner vs. rented units, vacant housing total counts, occasional use housing and housing construction type (evaluated as a weighted difference in construction characteristics from US overall).
- Cellular Coverage based on crowd sourced Verizon cellular coverage in the region of interest including average download Mbps, average upload Mbps, latency ms and signal bars
- Public Works facility count including police and fire stations, town halls, post offices, libraries, medical facilities, public schools.

[††] Data retrieved from https://coast.noaa.gov/hurricanes/ 05 Sept. 2017. Storm strength based on NOAA categorization including named hurricanes of category H1 to H5, Tropical Storm, Tropical Depression and Extratropical events.

[⊗] US Census Bureau, American Housing Survey, 2015 nearest available survey region utilized included Boston, Houston, New Orleans and Orlando for the five potential sites under consideration.

The relative importance of each site selection category was linked together based on a relative ranking assignment using a Z-score risk assessment methodology (subtracting the mean and dividing by the standard deviation). Z-scores were evaluated in a spreadsheet. NASA provided the final decision.

Ultimately Galveston, Texas was selected for the QSF18 test. Its island geography allows boom placement over high population density while minimizing exposure to high amplitude sonic booms. The community is not accustomed to sonic booms. Nearby NASA facilities at Ellington field provides opportunity for aircraft basing and NASA Public Affairs Office (PAO) support is possible in the area. The backup location selected was Melbourne, Florida. It has the highest population of communities considered, however residents have some familiarity with sonic booms because of SpaceX and past booms from the NASA Space Shuttle. There is a higher probability of exposure to loud focus booms along the barrier islands and the area does have a higher concentration of commercial upper airspace use. The region does offer NASA facilities (Kennedy/SLF) for basing and NASA Public Affairs Office (PAO) support is available. This basing facility was utilized during SonicBAT deployment. Since community engagement is a key risk area that needs to be explored, the familiarity of Melbourne from the recent Sonic Booms in Atmospheric Turbulence (SonicBAT) experiments might reduce effective risk investigation.

4.2.1 Lessons on Site Selection

The QSF18 Test had specific coastal requirements necessitated by the F-18 LBDM. This simplified the site selection process, however it was still an intensive effort to perform basic analysis on multiple locations, which were eventually down-selected.

Future site selection activities should identify requirements in terms of general parameters which are amenable to automated processes. For example, geographic areas for placement of focused booms could be identified in terms of area and distance from other features (e.g. test area or airport). GIS tools could then be used to search for potential locations meeting such criteria. Supplemental requirements could then be applied (e.g. runway length and population density) to refine potential sites. Future work should also consider the sequence of site criteria selection, as this will likely affect computational effort.

Meteorological considerations must always be considered when assessing sites, since this has a direct effect on desired boom delivery suitability and success rates. The importance of the effects are a consequence of the details of the operational flight trajectories relative to the upper air profiles. One needs to consider seasonal effects, but also monthly and daily variations. During QSF18 it was observed that the diurnal meteorological variation had an impact on waypoint planning[††] (see Section 4.3), in particular the component of onshore and offshore wind components. Site selection analysis considered historical twice daily upper air data at geographically available monitoring sites, which were sometimes hundreds of miles away. Any future test planning should examine finer temporal resolution meteorological data to provide information relevant to test design event timing (morning, afternoon,

---

[††] Waypoint planning in this context refers to the process of using atmospheric profiles and aircraft trajectory information to determine a geospatial location where the maneuver should be executed to deliver a target overpressure at a specific ground location.

evening and nighttime booms). Such data could be empirical or the output from suitable modeling. The location of the meteorological data should be close enough to the potential sites that there are not significant differences in wind speed, direction, temperature, and relative humidity. In particular, differences in wind speed or prevailing wind direction will result in differences between the modeled and realized footprints, and differences in relative humidity will affect ground loudness levels. The acceptable distance between meteorological data and potential sites is likely to vary with geographic location – for inland locations with no significant terrain variation an acceptable distance is likely larger than it is for potential sites near a coastal region.

Given the increase in geographic area for X-59 testing, it might also be prudent to consider multiple meteorological assessments across the area of interest. Furthermore, the local weather patterns and upper air variability trends across the test area should be considered. The goal of such an analysis should be to understand how the expected distribution of boom levels across the test area might change by the hour, day, week or month.

The social capital investigation is a laborious process. It requires researching site specific social norms and preferences regarding aviation noise and community tolerances and adaptability. Reliance on federal data sets (census, housing, maritime, transportation statistics) which have varied refresh periods (annual to decadal), can impact confidence in the selection data. One also needs to consider special analyses such as the off shore oil rig and boating prevalence studies for the LBDM. These kinds of analyses could be tied to specific geographic considerations (avoid areas, airspace constraints) or could be impacted by local events. For example, the annual Galveston motorcycle rally (Lone Star Rally 2018) was held during the weekend, 1-4 November, just preceding the start of QSF18 testing on 5 November.

4.3 Test Plan

A full description of the experimental design and detailed test plan for Quiet Supersonic Flights 2018 Test (QSF18) was documented in “QSF18 Detailed Test Plan for Community Response Testing in Galveston Texas” as part of Phase 2. This test plan is included as Appendix D of this report. The test plan provides full detail of the test execution, including:

- Test objectives
- Success criteria
- Participant recruitment plan
- Survey design and methods
- Sonic boom analysis for test dose design
- Noise dose design
- Objective measurements, including instrumentation suite and laydown, operations and staffing
- Noise metrics
- Analysis plan
- Go/no-go criteria

The proposed test plan included matched cumulative daily noise doses which were distributed across the two week long test window. The noise exposure design provided the test day, time of the flight, number of sonic thumps per flight, noise level for each flight, daily sonic thump noise exposure, number of flights per day, cumulative daily noise dose and the test day in the design that most closely matched that daily dose. The intent was to afford the ability to compare responses across different days and cumulative noise doses, in addition to comparing across single event dose responses.

Each QSF18 cumulative daily noise exposure represents a sum of the single event exposures for each test day. The noise exposure design for the anticipated daily range at Galveston was from 42 to 52 CDNL, corresponding to a range of 32 to 48 average day-night perceived level (PLDN). The noise dose range further inland was lower because the locations are further from the dive maneuver. The design afforded paired comparisons at Galveston across test days and the potential for comparisons to other communities on different test days.

The proposed noise dose plan considered acclimation to a new noise source in the community. For the first few days, the noise dose was planned to have lower cumulative daily doses, either due to level or number of thumps, to afford an introduction of the noise to the community. Previous research has shown that the net effect of habituation and sensitization is dependent on the interaction between stimulus (noise) level and number of stimuli (sonic thumps) [Petrinovich, 1984]. That is, the level and number of sonic thumps per day may affect the ability of a community to acclimate to the noise or the rate of the acclimation. This is in keeping with anecdotal recommendations that a new noise source should be introduced gradually to communities in order to afford the community the opportunity to adjust and acclimate to the noise. A short introductory period was planned, and the intent was to present the highest number of sonic thumps on test days that occurred later in the field test.

In consideration of the test community, an additional “if-then” layer was added to the proposed plan, with defined incremental steps in noise dose.

- Plan exposure not to exceed 80 PLdB per event for days 1 – 4 (or 1 – 3).
- If the survey results and other feedback indicate the community would tolerate higher levels, plan exposure not to exceed 85 PLdB per event for days 5 – 6 (or 4 – 5). Include sonic thumps at lower levels as well.
- If the survey results and other feedback indicate the community would tolerate higher levels, plan exposure not to exceed 90 PLdB per event for day 7 (or 6 – 7). Include sonic thumps at lower levels as well.
- If the survey results and other feedback indicate the community would tolerate higher levels, plan exposure not to exceed 95 PLdB per event for day 8. Include sonic thumps at lower levels as well.

QSF18 assessed test methodologies which are being proposed for use during future X-59 community tests. Noise exposure for NASA's X-59 aircraft is anticipated to be approximately 75 PLdB directly under track, with noise levels laterally off-track of the flight path on order of 70-75 PLdB. The low loudness level sonic thumps developed for QSF18 were not anticipated to elicit a large number of responses for assessing the % Highly Annoyed as defined in the Test Plan. As such, the level of the sonic thumps in the planned noise dose were not anticipated to have a highly notable impact on the test community.

Because sonic thumps are a new noise source, a NASA decision was made, just prior to conducting the test, to limit the field noise dose even more than was proposed in the original test plan. Field noise doses were also affected by weather conditions that impacted both the schedule of the test flights and the noise propagation across the sonic thump footprint. The field dose descriptions and the number of thumps at each level over the duration of the field test are provided in Table 4-2 and Table 4-3. "Waypoint planning" is the selection of daily test points relative to the local weather conditions of the test day to ensure the proper noise dose is delivered to test community. There were fewer Low and Medium level thumps, with just over half of the thumps presented at the Quiet level. Appendix T provides more detailed summary tables with metrics calculated from measured data at each sensor, and a comparison to the design levels.

Table 4-2 Sonic thump levels and descriptions

ID	Waypoint	Description	Based on Projected LBFD Undertrack Metrics		Notes
			PSF max	PLdB max	
U	5	Ultra quiet	n/a	n/a	Evanescent waves only on Galveston Island
Q	4	Quiet	0.13	73.7	
L	3	Low	0.20	79.7	
M	2	Medium	0.28	84.0	
MH	6	Medium-High	n/a	n/a	Added to provide waypoint between M & H[‡‡]
H	1	High	0.53	93.3	

Table 4-3 Number of sonic thumps in field test

Actual Level (based on median)	# of booms	
	Actual	Design (11/1/18)
Quiet - Q	28	14
Low - L	15	19
Medium - M	5	17
High - H	2	2
No data:	2	
Totals:	52	52

During the test planning phase, detailed analyses were conducted to address two topics: (1) findings for work done to assess the effect of meteorology on the selection of dive location for the F-18 LBDM; and (2) refinement of flight go/no go criteria. The results were documented in a technical memorandum that

---

[‡‡] The requirement to add MH, a waypoint between M and H, arose during test execution, and the waypoint planning process was modified accordingly to facilitate sonic thump placement.

was delivered to NASA. The boom analyses conducted for these topics overlap significantly, thus it was decided to present them in a combined document. The document, entitled “QSF18: Supplemental Meteorological Analysis and Go/No-Go criteria” is included as Appendix E of this report.

4.4 Lessons Learned on Experiment Design

4.4.1 Lessons on Subjective Design

In the Single Event Survey, an open ended field was provided for input from respondents. It was intended to allow respondents to add a qualitative descriptive input on the listening experience in addition to the defined response scales. Input received was often about test logistics rather than the listening environment or experience. Some comments were received that indicated that respondents did not clearly understand that they could be prompted for a single event response up to 10 times per day, and that not all reminder messages were associated with a sonic thump. Future instructions should include specific examples for submission formats. If funds allow, the Penn State Survey Research Center (SRC), or survey provider, can download the open ended data fields at the end of each day. This real time data cleaning will increase cost, but may help identify confusions on the part of respondents in sufficient time to provide feedback to respondents during the course of test execution.

The QSF18 research team provided datasets to NASA at the completion of the field test. The approach protects the respondents’ name, and uses a unique ID number for each participant, to protect the confidentiality of their responses. This approach borders on providing fully identifiable data, because it includes the home and work locations provided in the survey responses. As such, a data sharing agreement is recommended that defines conditions on which the data can be accessed, and a similar agreement should be developed as a requirement recommended for X-59 test data.

It is recommended that the X-59 data be archived in a defined database that is included in the IRB and OMB documentation. Over the conduct of multiple tests, there is the potential to establish a repository of data. If institutions other than NASA have access to the data, NASA may want to add requirements for use of the data. This could be in the form of a data agreement, or a process requiring IRB approval in order to gain access to the data. The terms and requirements for use of a data repository would be up to NASA.

Potential data delivery options include the following types of datasets:

- Fully Identifiable data: provide all the information that was gathered except respondents’ identity.
- Partially de-identified data: include the lat/long location data but remove the home addresses to protect household identify. This might affect the ability for researchers to fully use the dataset.
- Fully de-identified data: this would include noise dose and response data, but without the location associated with the dose.

If a data repository is established the informed consent language should be modified to accommodate potential future use of the data by other researchers. The NASA IRB would provide input on the potential data repository and the terms of access to the dataset. The language required by the IRB is function of risk in the research. Risk has two parts, the probability of harm and the potential magnitude of harm to participants. The QSF18 team took steps to protect the participant's ID for confidentiality purposes. For this dataset, if an ID was revealed, there is a low probability of harm and presumably no magnitude of harm if someone learned of a participant's location or even their noise ratings. As such, the QSF18 test was considered minimal risk research.

With a minimal risk design, the research can present such data as graphs with indicators representing participation locations across the quadrants, but not a respondent's name. While someone could make an effort to use the GIS data to reveal the home or work location indicated by the lat/long data point, they won't know which member of the household or work location participated, and there is minimal risk to the participant if the home/work location is revealed. The responses are associated with the unique ID, not a name. In the future, clear language should be added to provide for consent for future use of the data. The language should state: "We may use your research information for other research studies or may share your information here or at other institutions for future research efforts without additional informed consent."

4.4.2 Lessons on Objective Design

More time should be allocated in the schedule for pretest[§§] and posttest data cleaning. The data cleaning and dose response calculations are complex with various inputs. The surveys requested location input on both the background survey, the Single Event and the Daily Summary to provide a level of redundancy in the location response in the event that one of the fields was not fully completed. This requires the combining of multiple data sets for location information. The redundancy on data gathering should be maintained but the compilation of locations should be summed in one file to facilitate dose calculation for dose response models.

The home addresses were identified by address based sampling and should be sufficiently formatted to be identifiable by latitude and longitude. The work addresses should also be associated with lat/long coordinates prior to the test. Clarifications for address location can be made by email or text to facilitate dose response calculation after the test. This real time data cleaning will increase cost, but may help improve noise dose calculations.

[§§] Pretest data refers to geopositional information reported by the participants, including their home and work addresses. Pretest data cleaning is the necessary action of verification to ensure geolocation / position identification during the participant recruitment and acceptance process.

5. QSF18 Execution

5.1 OMB and IRB Applications

Appropriate approvals were obtained to ensure that the QSF18 field test was in compliance with regulatory guidelines. Both Office of Management and Budget (OMB) and Institutional Review Board (IRB) approval was obtained. Further information is contained in Appendix H and Appendix I.

IRB approval was required because the research involved the use of human participants. An IRB Authorization Agreement was signed to indicate that Pennsylvania State University would rely on the NASA Langley Research Center IRB for review and continuing oversight of the research. Approval was granted by the NASA IRB before the field test on August 31, 2018. Both PSU and NASA participate in the Collaborative Institutional Training Initiative (CITI) IRB web-based training and certification that is shared across academic institutions, government agencies, and organizations in the U.S. and around the world. All WSPRRR team members that participated in the conduct of the research completed the CITI training. By completing this training, all team members complied with both the PSU and NASA IRB training requirements.

The Paperwork Reduction Act (PRA) of 1995 requires that the US Federal Office of Management and Budget (OMB) approve each collection of information by a Federal agency before it can be implemented. The information requested is intended to ensure that agencies employ effective survey and statistical methodologies that are appropriate for the type of information that is to be collected. The OMB approval included development and submission of the OMB Information Collection documents and adherence to the process. This included submitting the required Paperwork Reduction Act supporting statement, a notice published in the Federal Register on March 05, 2018 providing a chance for any interested individuals to comment on the proposed information collection within 60 days, and submission of the final Paperwork Reduction Act clearance request, including any public comments received to OMB in order to obtain approval. OMB approval was provided in August 2018, with no requests from the public during the Federal Register notices, and no requirement for clarifications was issued by OMB. All survey documents presented the following OMB statement: "This information collection meets the requirements of 44 U.S.C § 3507 as amended by section 2 of the Paperwork Reduction Act of 1995. The OMB control number for this collection is 2700-0167, which expires on 8/31/2021."

The QSF18 field test was successfully executed with the sonic thump noise source over a community not used to hearing sonic booms to test all aspects of the community response effort, consistent with the test design approval by IRB and OMB.

5.2 Pre-Test Activities

AFRC initiated weekly planning sessions as of 28 February 2018 through the conclusion of QSF18. A total of 35 meetings of 1.5 hour duration with at least 16 participants were conducted leading up the event.

A comprehensive review of the Galveston area was conducted employing geographic information system (GIS) technologies by Applied Physical Sciences and Volpe National Transportation Center. Google Earth™

proved enlightening as an overview early on and then again during the execution of the event. Google Earth allowed for the early identification of open areas, large swaths of industrial areas, cemeteries and bayou communities. GIS data for the Galveston community was mined expanding the detailed dataset beyond what is available in Google Earth; this allowed the compilation of a matrix of all public's works and open spaces for each of the four quadrants selected for recruitment and noise monitor placement. Appendix B provides a detailed discussion of the site selection and gridding process, including the use of four quadrants.

A review of available climatological data and the relationship of humidity and winds aloft to PCBoom predictions was further explored by Volpe. Extensive PCBoom predictions were generated for multiple waypoints offshore from the community to enable adjustment of boom intensity based upon community response. These analysis activities are described in Section 6, Appendix B, and Appendix C.

NASA arranged a site visit to Ellington Air Force Base and the Galveston community 16-18 April 2018. During this visit AFRC focused on Ellington Air Force Base and flight operations while NASA personnel and members of the APS team focused on Scholes Airport in Galveston as a potential base of operations and investigated candidate noise monitor locations throughout the community. The team was divided into pairs with each pair assigned to one or more quadrants and provided a list of candidate noise monitor locations. For each potential monitoring site location several parameters were investigated.

- A data sheet was completed which noted the following details:
    - Building terrain and landscape
    - Area activity level/noise assessment
    - Nearby restroom facilities for person staffing noise monitor
- Cellular connectivity as measured using a cellular phone app "Speedtest™" by Ookla. This evaluated smartphone upload and download speeds in Mbps.
- Push to talk handheld radio performance was checked at each site

All candidate sites were evaluated over the course of two days; the data sheets were collected each day, evaluated and rated on a scale of A (best), B, and C (least suitable). On the third day one individual was located at Scholes airport with a laptop computer connected to the internet using its designated cellular modem. A second individual then travelled to one or two of the A graded sites in each quadrant with a second computer connected to the internet with a second cellular modem intended for use with the Sonic Boom Unattended Data Acquisition System (SBUDAS). A TCP/IP connection over a cellular virtual private network (VPN) was established at each of the sites and throughput was further confirmed.

Noise monitors were assembled and tested at Gulfstream Aerospace well in advance of the QSF18 test with cellular throughput over the VPN evaluated between Gulfstream in Georgia and Applied Physical Sciences in Connecticut. Gulfstream hosted SBUDAS familiarization training for NASA and field personnel on the contractor team on 5 September 2018. Field personnel were instructed concerning noise monitor assembly, troubleshooting, and set up. Ultimately as a final test prior to shipment, each team member transported two noise monitors to designated sites around Savannah, Georgia,and noise measurements were collected over the cellular VPN on each of the Noise Monitor Base station computers located at the Gulfstream facility.

5.3 Recruitment

5.3.1 Recruitment Strategy

The detailed site selection process in Appendix B and test plan in Appendix D provide the rationale for the recruitment, sample size justification and methodology. The following presents an overview of the strategy, details of the process, and the execution of the recruitment for the QSF18 field test.

QSF18 recruitment consisted of dividing the test area into four quadrants under the sonic thump footprint and then randomly selecting households from the general population using targeted Address Based Sampling (ABS). ABS is sampling from address lists that are updated via the United States Postal Service (USPS). The USPS maintains the Address Management System (AMS) for sorting and sequencing of mail, in Computerized Delivery Sequence (CDS) files [AAPOR Report, 2016].

The USPS CDS has six address groups:

- City Carrier Residence Only
- City Carrier Business
- City Carrier Combination Residence and Business
- Post Office Box
- Rural Route and Contract Delivery Service Route [U.S. Postal Service 2013b]
- Combined Delivery Type

Researchers can purchase survey samples from vendors who sell ABS samples. Vendors differ in the source of their addresses, the services they provide, and their geographic coverage. It is essential to have a reputable vendor because the researcher will not have access to the sample frame or the sampling process. Primary vendors hold a Delivery Sequence File Second Generation (DSF2) or CDS license with the USPS. Vendors can enhance the USPS lists with addresses from additional sources such as local tax records, phone directories, or credit card databases. Some vendors may also provide related data such as geocodes, phone numbers, and demographic information as available.

P.O. Box addresses are often excluded from sampling frames to minimize duplication of addresses. Households with both a P.O. Box and a city-style address can receive mail at either address, and may have no linkage between the two addresses in the sample frame. A housing unit may have a physical address and multiple P.O. Boxes for multiple persons living in the same housing unit. The risk of duplication between mailing addresses and P.O. Boxes is high, and the chance of locating a housing unit on the basis of the P.O. Box is low, which is why P.O. Box are often excluded. Some P.O. Boxes constitute their own route in the CDS. These P.O. Box addresses typically have no corresponding city style address and are not duplicates of other housing unit addresses on the frame. Some vendors label the P.O. Box group in the CDs as Only Way to Get Mail (OWGM). The OWGM P.O. Box addresses can be retained in the frame for mail surveys when other P.O. Boxes are removed.

The effect of not having P.O. boxes for recruitment during this test was minimal since the desired sample size was to contact 8000 prospective individuals across a well populated area. The sample was randomly generated from the sample frame. Enrollment was done on a voluntary basis, from a randomly selected sample of addresses.

5.3.2 Recruitment Execution

A list of households within the footprint area was compiled by Survey Sampling International. From this list, a systematic random sample of all qualifying households was selected using a random starting point and a sampling interval, in order to reach the sample size of 8000 households for recruitment letters. The qualifying file is sorted using zip codes. A complete, nine-digit ZIP Code (zip + 4) consists of two parts. The first five digits indicate the destination post office or delivery area. The last 4 digits represent a specific delivery route within that overall delivery area. The sample is sorted by the five digit zip code, then by zip +4 before every nth address is randomly sampled. This process is known as “nth-ing”. The test region was divided into quadrants, as shown in Figure 5-1. Quadrant B had the largest potential number of households from which to select the ABS sample. Quadrant C had fewer potential households in the ABS sample. The team had also discussed recruiting fewer respondents from Quadrant C since the population was less dense in that Quad and since there was a limited number of noise monitors to distribute across the boom footprint. Table 5-1 summarizes the ABS samples and final recruitment quantities by quadrant. The respondents were recruited on a first come, first enrolled basis. The enrollment was closed once a sufficient number of individuals expressed an interest in participating. The distribution of respondents across the quadrants within the anticipated test region is indicated in Figure 5-1.

Table 5-1 Recrutiment by quadrant

Area	Quadrant	ABS Sample	Final Recruitment
SW Galveston	A	1914	148
La Marque, Bayou Vista, Tiki Island	B	3615	212
Texas City	C	559	20
NE Galveston	D	1914	116
	Total	8002	496

Figure 5-1 Final distribution of respondents across quadrants

Recruitment letters with a $2 pre-incentive were sent out to the 8000 addresses with a household unique ID to be used for enrollment. The letters were addressed to the head of household, although any eligible member of the household could enroll. Any member of the household that was qualified could enroll into the study by submitting a background survey. Enrollment was contingent on being over 18 years of age, willing to provide at least an email address for communications, and living and working within the anticipated boom footprint. The consent and background survey required the respondents to provide an email contact, and also requested a mobile phone number.

The home address was confirmed and work address requested (but not confirmed, in order to help reduce respondent burden) to ensure respondents both lived and worked within the footprint. The recruitment letters were prepared and mailed beginning on Thursday 10/04/18 and ending on Tuesday 10/09/18. Reminder post cards were sent to households approximately 10 days after the initial mailings. A second set of reminder post cards was planned, but not sent because sufficient enrollment was reached after the first set of reminder post cards were sent.

The first respondents to complete the enrollment and background survey on line were enrolled, with a target of enrolling 500 respondents. The respondents began enrolling on 10/9/18 and sufficient potential enrollment was almost reached by 10/17/18. On-line enrollment remained active for 3 more days to ensure sufficient sample size. The actions in the recruitment process are listed in Table 5-2.

Table 5-2 QSF18 recruitment process

Recruitment Action	Date
Letters stuffed/Start mailing	10/4/18 – 10/9/18
Columbus Day; No mail	10/8/18
Enrollment period start	10/9/18
Reminder post cards printed	10/12/18 – 10/13/18
Reminder post cards mailed	10/16/18
Enrollment approaching 500	10/17/18
Email confirmation sent	10/19/18
Enrollment period end	10/20/18
1st email confirmation reminders sent	10/23/18
Cell phone confirmation sent	10/26/18
2nd email confirmation reminders sent	11/2/18

Of the 8000 invitations to participate that were sent out, 1348 were returned as undelivered due to no forwarding address or vacant address, resulting in an undeliverable rate of 16.85%. We anticipated only 10% undeliverable during the planning process. Currently, undeliverable rates are approximately 18%[***]. Of the 8000, delivery was successful to 6652 homes, resulting in 83% delivery rate. The enrollment closed with 568 initially enrolled into the study, and an enrollment rate of 8.5%. It is speculated that the rapid enrollment was due to NASA Outreach efforts with community and news outlets prior to and during the recruitment period.

The enrollment by date is provided in Table 5-3. A text/email confirmation request was sent and time allowed for prospective respondents to respond online that they acknowledge their enrollment in the QSF18 NASA study. Repeated requests were sent as necessary to prompt a confirmation response. There were 544 requests for enrollment confirmation emails and text messages sent. The confirmation requests exceeded the 500 sample target as some attrition was anticipated during the enrollment process. Confirmations were received from 500 of the 544 requests that were sent. Of those, 341 respondents confirmed by both email and text, 64 were email only confirmation and 95 were text only confirmation. The time frame for the test date was not announced prior to the start of the recruitment and enrollment.

[***] *Undeliverable-as-Addressed (UAA) Statistics by Mailing Industry Quarterly Report (Q4 FY18).* ACS Nixie stats by industry for FY18 Q4. December 03, 2018 https://postalpro.usps.com/undeliverable-addressed-uaa-mail/FY18QTR4_INDNIXCNT. Currently, the undeliverable rates as experienced by the PSU Survey Research Center are approximately 18%. For the fourth quarter of 2018, the USPS listed the undeliverable rate across all industries at 28.6%. The observed lower rate is an indication of the integrity of the ABS data obtained by PSU SRC.

A number of participants dropped out once the test dates were announced due to lack of availability during the test period. Some were also rejected if they were under 18 years of age or if they worked outside of the sonic thump footprint. Other participants were eliminated because they lacked internet access. Table 5-4 shows the attrition of the number of participants from the completion of recruitment through the end of the test.

Table 5-3 QSF18 enrollment date

Enrollment Date	Total submitted	Dropped: Not willing	Dropped: Under 18	Dropped: Work out of Area	Net
10/9/2018	**87**		**2**	**4**	**81**
10/10/2018	**119**		**1**	**7**	**111**
10/11/2018	**79**		**1**	**8**	**70**
10/12/2018	**35**			**3**	**32**
10/13/2018	**27**				**27**
10/14/2018	**28**			**2**	**26**
10/15/2018	**47**			**4**	**43**
10/16/2018	**44**	**1**		**3**	**40**
10/17/2018	**31**	**1**		**1**	**29**
10/18/2018	**54**	**1**		**2**	**51**
10/19/2018	**34**	**1**			**33**
10/20/2018	**1**				**1**
Total	**586**	**4**	**4**	**34**	**544**

Table 5-4 QSF18 participant attrition

Total number of recruits	544	Result of recruitment
Total number of recruits who responded to confirmation requests	500	Result of confirmation requests
Participants at commencement of testing	496	Attrition due to unavailability
Participants at completion of testing	476	Attrition due to some participants never submitting reports

Once respondents had confirmed their willingness to participate, they were assigned to a response group (email/text) and reminder type within each group (with reminder/no reminder). The groups were email reminder, email no reminder, text reminder, text no reminder. Those participants who only responded to the confirmation request via email were assigned to the email group (n=64). Those that responded to the confirmation request by text only were placed in the text group (94). The rest responded to the confirmation message by both methods and were then randomly placed in either the email or text group. Once response groups were assigned, random assignment for reminder/no reminder was made within each group. The target was to have 125 respondents in each reminder type/group. Participants were not overtly told to which groups they were assigned. At the start of the test there were 496 respondents, but only 476 were still participating by the end of the test period.

5.3.3 Participant Recruitment Lessons Learned

The QSF18 recruitment was successfully executed in a short period of time. There are some modifications that would facilitate recruitment in the next test.

- The scheduled start of enrollment was delayed due to an unanticipated extension in the testing of the GPS application for participant location. This did not impact the effectiveness of the enrollment, but it did affect the timing of the shipment of the recruitment letters. The delay in schedule resulted in the recruitment letters being sent out over a holiday weekend. The majority of the letters (6500) went out in the first shipments on 10/4/18 and 10/5/18 using the US Postal System. Approximately 1500 mailers were sent on 10/8/18 but were delayed in processing due to the Columbus Day Federal holiday falling on 10/8/18. For future tests, all potential households should receive the mailers within a few days of one another.
- We initially planned on a one month recruitment period, with initial invitation letters followed by 2 sets of reminder post cards. Recruitment was successful with a 10 day, rather than a 1 month recruitment period. Future efforts should plan for at least a 14 to 21 day recruitment period in the event that enrollment is not as rapid for the next test.
- NASA received a number of email and phone questions that should have been directed to the PSU Survey Research Center. The recruitment letter should have listed a phone number at the Survey Research Center for questions.
- The effect of P.O. boxes being omitted in the ABS of the USPS may become an issue in a less populated area where the prevalence of P.O. box addresses can be higher.
- The undeliverable rate was 16.85%, due to no forwarding address or vacant address. Methods for weeding out undeliverable letters should be investigated.
- It is speculated that the rapid enrollment was due to NASA Outreach efforts with community and news outlets prior to and during the recruitment period – these efforts should be continued.
- Over-recruit by a certain percentage, to account for participants that drop out due to lack of availability during the test period, or who are found to work outside of the sonic thump footprint.

5.4 Flights

The F-18 was operated out of Ellington Field and the control room established there provided the test director oversight and control of the testing operations. The acoustic field crew, led by the field crew chief, had a base of operations at Scholes Airport, in the test community and close to the acoustic and meteorological instrumentation. Radio communications and protocols were established between the control room and the field crew. Prior to each day of operations the approximate flight and sonic thump times and desired noise dose levels were established based on the noise dose test design and the anticipated weather. During the course of daily operations the design was adapted between flights as allowed for in the noise dose plan.

Flight Waypoint Planning

As described in earlier sections, daily noise dose was varied by changing the number of events per day and varying the individual flight levels. The single event noise dose was varied by shifting the position of the dive waypoint to effectively move the footprint relative to the study area. Because meteorological conditions are known to have significant effects on how the footprint from a low-boom dive maneuver is formed, dive waypoints were calculated individually for each flight using forecast upper air profiles. In practice, the forecast model was updated at six-hour intervals, and waypoint planning was conducted using the latest possible forecast that would allow the waypoint package to be delivered to the ground crew and PI at least 90 minutes prior to takeoff.

Waypoint planning utilized PCBoom and followed the procedure developed by NASA AFRC. Additional waypoints were incorporated into the planning process mid-test as will be described. A waypoint package was comprised of a Garmin .gpx file containing 4-6 waypoint positions, forecast data used in modeling to plan waypoints, screenshots and a .kml file of modeled footprints for each dive waypoint, and a text summary listing:

1. Latitude/longitude coordinates of dive waypoints and aircraft heading,
2. Propagation times to noise dose design sites,
3. PL (dB) predictions using thin shock and Burgers at noise dose design sites,
4. Maximum overpressure modeled using thin shock and Burgers at noise dose design sites.

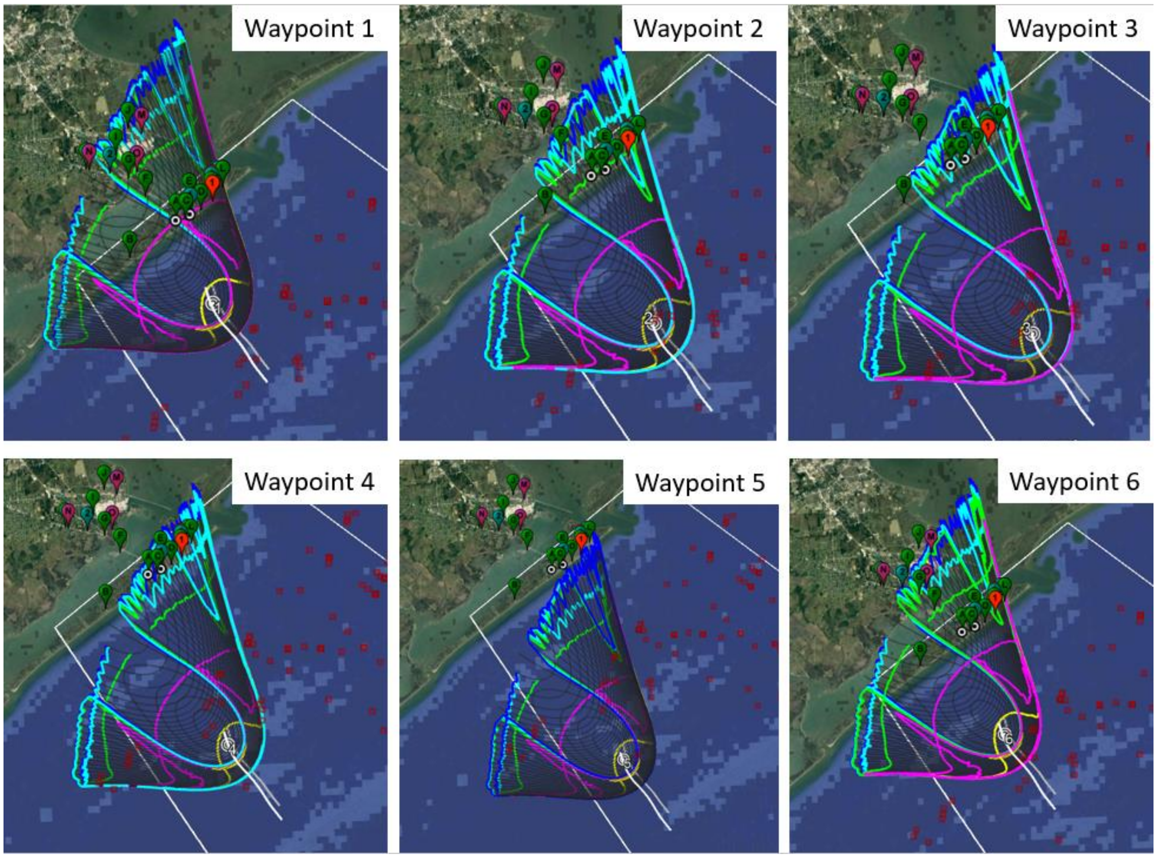


Figure 5-2 A collection of modeled footprints using dive waypoints 1-6

Dive waypoints as utilized in the execution of flights are summarized in Table 4-2, and graphics illustrating waypoints 1-6 are shown in Figure 5-2. The first four dive waypoints are based on targeting modeled overpressure levels at the primary noise dose design site (Scholes airport). Dive waypoint #5 was added prior to execution of the first flight, and its position was determined by placing the downtrack edge of the footprint at the coastline such that the entire footprint was offshore. The goal was to expose participants to evanescent waves only. Dive waypoint #6 was requested by the PI partway through the test, and was added to the waypoint planning procedure for subsequent flights. Placement of dive waypoint #6 was at the midpoint of dive waypoints #1 and #2. The goal in executing a dive using waypoint #6 was to give more options for mid-flight changes to planned waypoints if reports from noise monitors differed from expected levels. Across the executed flights, all of dive waypoints 1 – 6 were utilized at least once.

Go / no-go

In the flight test planning stage, a set of go/no-go criteria were developed including considerations of aircraft readiness, meteorological conditions, instrumentation readiness, etc. As part of flight execution and objective data collection, pre-test procedures called for the field control center to report to the PI on noise monitor status. The “go” criterion for instrumentation was to have at least one monitor operational – this condition was met for every event. No-go conditions resulting in flight delays, cancelations, or in-flight termination of dives were small in number and were typically related to weather conditions. These situations are described specifically in the bulleted list below.

Flights as executed

A total of 52 thumps over 22 flights on 9 test days were executed. The number of thumps and flights per day followed noise dose design subject to constraints of test-day weather. A summary of executed flights and thumps is given in Table 5-5 along with the sequence number of the boom. Noteworthy aspects of flight execution are as follows:

- Field reports indicated that the first thump was not heard except evanescent waves at one location. Post-flight modeling for flight 1, pass 1 showed that the footprint was shifted relative to the planned position to the southwest, and only one monitor (Bravo) recorded a thump.
- Flight 4 was delayed by one hour due to fog in the area; updated waypoints were calculated and provided to the team.
- For flight 5, pass 3 the pilot reported having to fly around a thunderstorm to line up for the dive.
- For flight 6, pass 3 was executed four minutes earlier than planned to due to a lower than expected fuel level.
- Flight 7 was originally planned to include three passes. This was reduced to 2 passes due to building clouds in the area, and the pilot terminated the second dive at roll-in due to clouds at the waypoint.
- Due to low-level clouds and fog, additional fuel reserves were carried on flights 8 and 9 in case of an Instrument Flight Rules (IFR) divert. As a result, only two passes were planned for flights 8 and 9 since the NASA F-18 research aircraft were not allowed to fly into visible clouds or fog during QSF18.
- During flight 11, a new dive waypoint (#6) was added at the midpoint between waypoints 1 and 2 to allow great fidelity in targeted overpressure level. Dive waypoint 6 was subsequently added to preflight waypoint planning.
- The first dive on flight 13 was terminated due to loss of radio contact between the aircraft and mission control center. The MCC radio was replaced and the first dive executed on a delay.
- The first planned flight on 20181113 was canceled due to high winds on the ground, icing conditions at altitude, and clouds at dive altitude. Weather conditions changed enough to allow two afternoon flights.

Table 5-5 Summary of flights and thumps as executed

Date	Flight	Number of passes	Sequence No.
20181105 (7 thumps)	1	3	QSF001,QSF002,QSF003
	2	2	QSF004,QSF005
	3	2	QSF006,QSF007
20181106 (6 thumps)	4	3	QSF008,QSF009,QSF010
	5	3	QSF011,QSF012,QSF013
20181107 (4 thumps)	6	3	QSF014,QSF015,QSF016
	7	1	QSF017
20181108 (4 thumps)	8	2	QSF018,QSF019
	9	2	QSF020,QSF021
20181110 (7 thumps)	10	2	QSF022,QSF023
	11	3	QSF024,QSF025,QSF016
	12	2	QSF027,QSF028
20181111 (6 thumps)	13	2	QSF029,QSF030
	14	2	QSF031,QSF032
	15	2	QSF033,QSF034
20181113 (5 thumps)	16	3	QSF035,QSF036,QSF037
	17	2	QSF038,QSF039
20181114 (8 thumps)	18	3	QSF040,QSF041,QSF042
	19	2	QSF043,QSF044
	20	3	QSF045,QSF046,QSF047
20181115 (5 thumps)	21	3	QSF048,QSF049,QSF050
	22	2	QSF051,QSF052
	Total	**52**	

Dive waypoints were placed based on modeled overpressure at Scholes airport. Measurements of PL at that location (monitor Alpha) are plotted in Figure 5-3, together with corresponding ambient levels. Note that flight 1, pass 1 is excluded as no thump was recorded at Scholes airport for that event. The highest PL recorded at Scholes airport was 85 dB; recall that this site is farthest uptrack site of the three noise dose design sites and thus expected to receive the highest level. An indication of the ability to deliver the desired overpressure levels at Scholes Airport is provided in Figure 5-4. This graphic shows that while the F-18 was able to successfully provide the target Quiet (.13 psf), Low (.20 psf) and Medium (.28 psf), obtaining the High (.53 psf) booms was not as successful. Figure 5-5 illustrates the ability to deliver the desired PL. As shown, the PL metric delivery was consistently lower than planned, especially for High booms. There are several reasons that the measurements are different from predictions. Propagation modeling did not include the effects of clouds on ground signatures. This likely resulted in the as-flown metric values being lower than desired. This is described in more detail in Section 6.1.4. Another reason is that the predicted waypoints were based on the pre-flight early morning atmospheric soundings. Another factor is that the location of the predicted footprints and boom levels was based on a fixed aircraft weight. The aircraft weight is reduced for each pass due to fuel consumption. Reduced weight means a lower boom level, although not a big change in itself. However, when the aircraft weight changes, the dive profile and Mach time history also change, which in turn causes the boom footprint to move. It is important to note that the number of booms for these four categories is Quiet (n=10), Low (n=23, not including Flight 1 Pass 1, for which the boom was only detected at sensor BRAVO), Medium (n=15) and High (n=3) so there were not as many attempts at delivering the High booms.

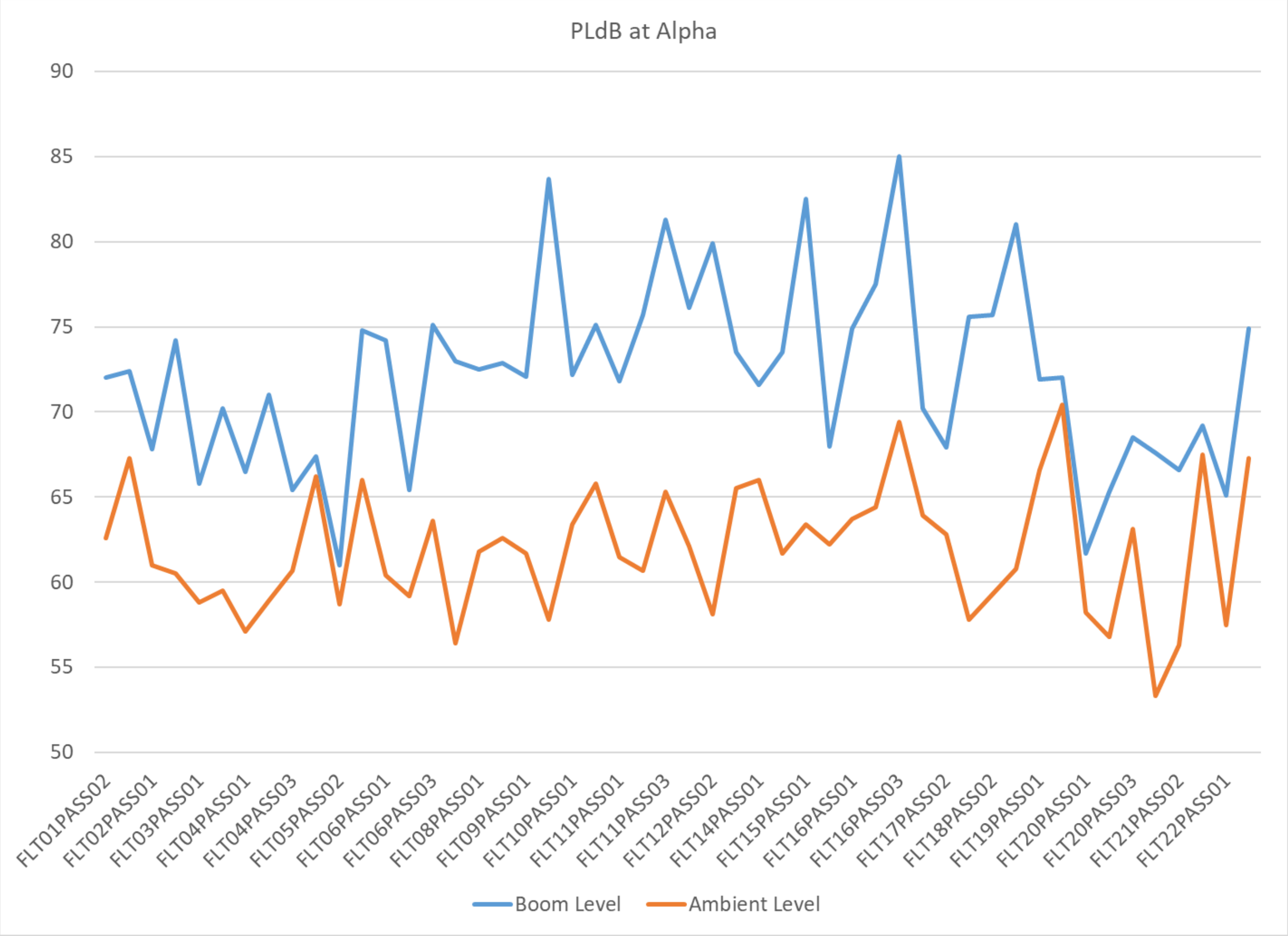


Figure 5-3 Measured levels at primary noise dose design site (monitor Alpha at Scholes Airport)

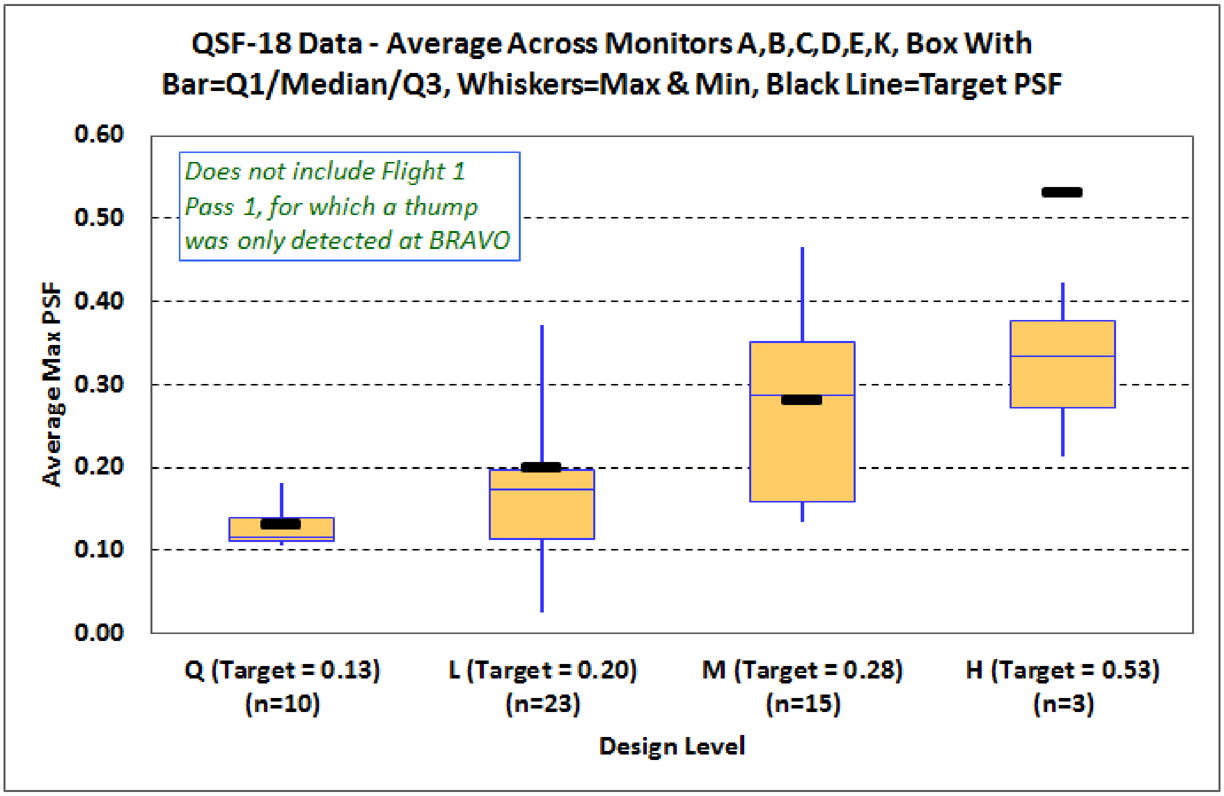


Figure 5-4 Summary of overpressure levels

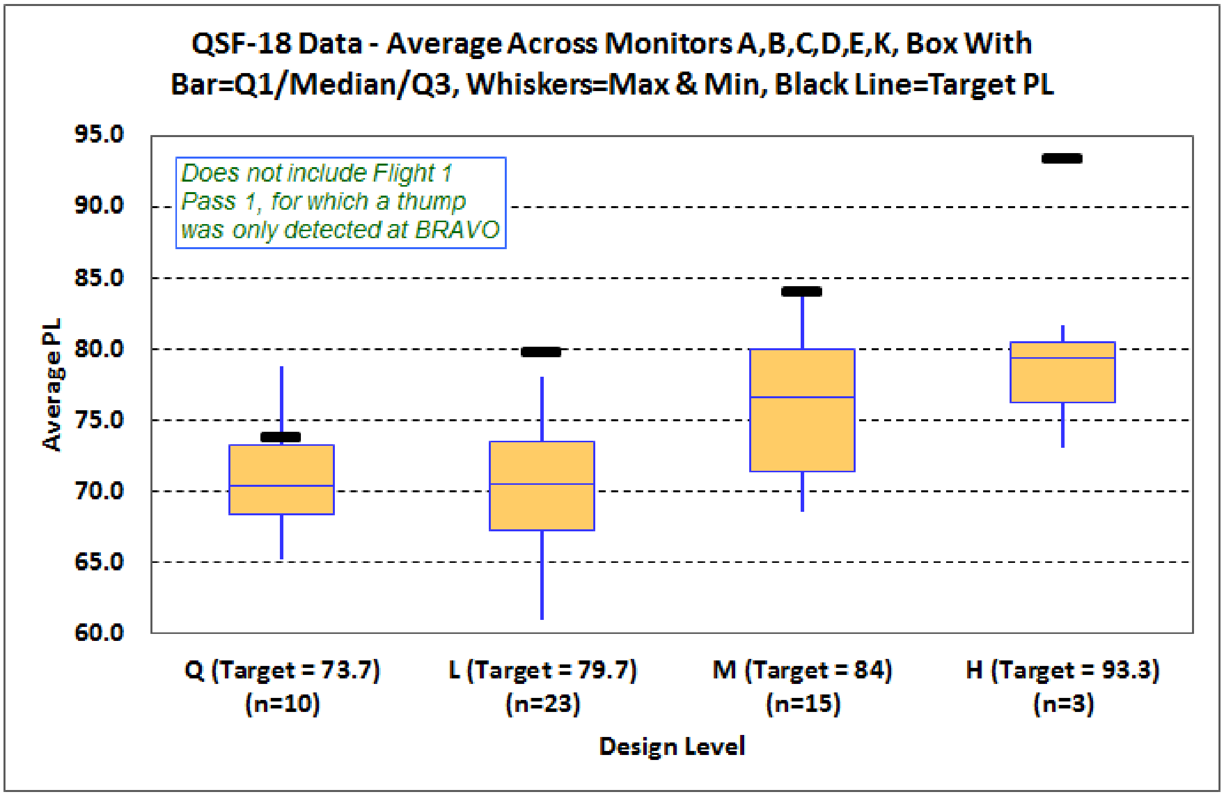


Figure 5-5 Summary of PL levels

5.4.1 Flight Lessons Learned

As mentioned in Section 5.4, the two F-18 research aircraft were based out of Ellington Field near Houston, Texas and flown to the test off the coast of Galveston. All aspects of QSF18 flight operations and logistics, including local airspace coordination, were managed by NASA personnel which were also based at Ellington during the test period.

WSPRRR team tasking included close coordination of waypoint planning, noise modeling and test point delivery with the NASA team to develop the planned sonic thump noise exposure within desired locations around the Galveston and surrounding communities. As such, topics documented in this section are limited to this scope only, however there are likely additional lessons learned from the crew and flight operations test team.

1. Timely delivery of “as-flown” aircraft tracking data would have helped identify aircraft navigation system issues earlier which resulted in insufficient delivery on noise exposure on the test area during the early test period.
2. Misdiagnosis of low noise levels due to aircraft navigation system issues created a false concern among the test team that waypoint planning and noise propagation modeling were erroneous. This issue resulted in rework of early waypoint planning and uncertainty in test point decision making to correct the low noise exposure within the community.
3. X-59 operational flight tempo of up to 5 flights per day may be overly ambitious given that NASA intends to only have one research aircraft. QSF18 utilized two aircraft and three pilots which gave some leeway in making flight times and crew availability. Even so, QSF18 only accomplished a maximum 3 flight per day operational tempo.
4. Recalling the cloud cover “knock-off” call on one inbound test point (Flight 7 Pass 2), recovery of the aircraft at the bottom of the LBDM would have put the pilot in the cloud deck so the test point was called off. X-59 supersonic flight passes will not involve the complex dive maneuvers, however test area meteorological conditions will need close monitoring due to the sometimes rapidly changing weather conditions of new, unfamiliar test areas.
5. Additionally, cloud cover modeling during the pre-test noise propagation analyses was insufficient for test planning in the humid coastal environment. During X-59 test planning this should be considered as well.
6. At some point during QSF18 testing, the waypoint planning team discussed that having a bigger “box” with more freedom for inbound heading angles for the dive would have been advantageous. This may not have been looked at closely enough ahead of test deployment when flight clearance was coordinated with FAA to define the supersonic flight box. More effort should be placed on operational trajectory flexibility during X-59 community test planning to allow for responding to changing weather conditions and community response feedback during the test window.
7. After a few days of testing, the waypoint planning team began posting each day’s planned flight schedule with takeoff, “Mark”, and anticipated propagation times as well as planned thump

loudness levels at measurement point Alpha at Scholes airport on the wall in the field crew control room at the beginning of the day. The field crew found this practice quite useful prior to entering the field to deploy their equipment. Although radio communications were available and utilized throughout the day, the advanced test point knowledge provided for improved situational awareness especially around expected measurement times. It should also be noted that the team found it difficult to get changes out to the field crew if they occurred within the test day. For X-59, this practice should be kept and further refined to allow for near real-time text or email communication of test point changes or flight time adjustments throughout planned test periods in addition to radio communications.

5.5 Objective Data Collection

The Gulfstream Aerospace Corporation (GAC) Sonic Boom Unattended Data Acquisition System (SBUDAS) is an environmentally protected and remote sonic boom recording system. The SBUDAS, or more commonly called “field kit”, is a purpose built data acquisition system developed for community noise measurements. Each SBUDAS is fitted with a Verizon Wireless powered cellular modem which provides remote connectivity. Therefore, the main responsibility for field crew members is to deploy, calibrate and retrieve the field kits each day. Interaction with the system is accomplished remotely by a central “host station” operator. During the QSF18 campaign twelve (12) SBUDAS were planned for deployment and operated by two host station operators; one SBUDAS was damaged in shipping so ultimately 11 were deployed during QSF18 by 6 Field Personnel and operated reliably throughout the test.

Each SBUDAS contains the components shown in Table 5-6 and depicted in Figure 5-6.

Table 5-6 Instrument list

Component	Description	Quantity
NI cRIO-9023	CompactRIO Controller	1
NI 9234	DSA Module	1
NI 9870	Serial Interface	1
NI 9381	Multifunction I/O Module	1
G.R.A.S. 40AN	Low Freq., Free-field Microphone	1&1*
G.R.A.S. 26AJ	Preamplifier	1&1*
G.R.A.S 41AO	Microphone Environmental Enclosure	1
G.R.A.S. 12AQ	Power Module	1
DIGI WR21	LTE Cellular Modem	1
20A Solar Controller	Solar Charge Controller	1
Bioenno Power P/N BLF-1220AS LiFePO4 Battery	12V 20Ah (240 Watt-hr) Lithium Iron Phosphate Rechargeable Battery	1
Solar Panel	40W Solar Panel	1
SanDisk Flash	64GB USB 3.0 Flash Drive	1
G.R.A.S AA0008	LEMO Cable	1
Garmin 16xHVS	GPS Receiver	1
	Lock for the Box	1
Energizer AA Batteries	Batteries for Calibrator	2
1&1* - 1 as primary and 1 as backup		

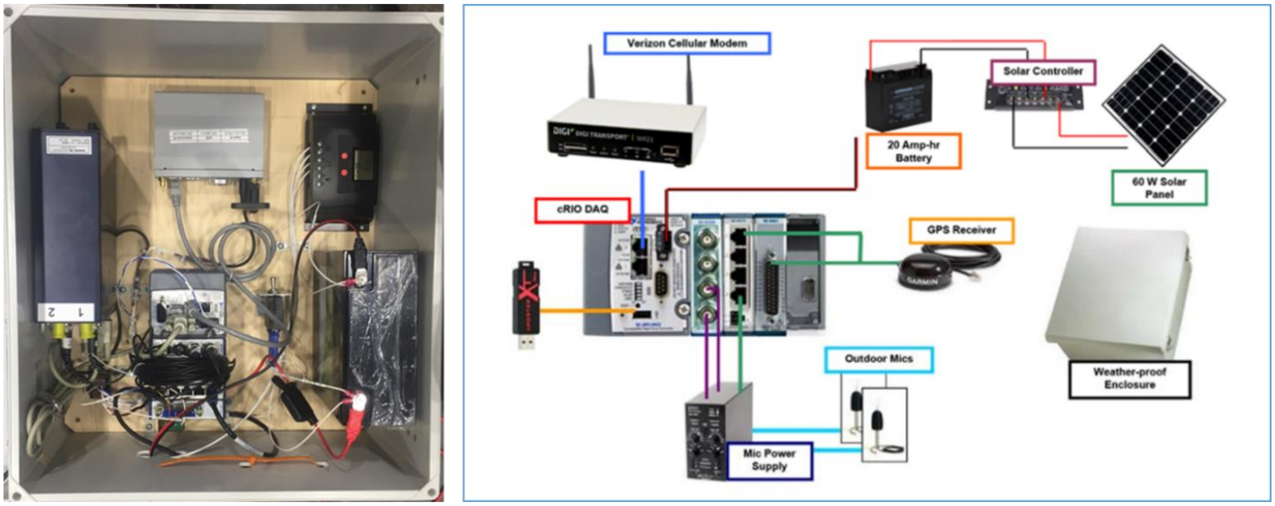


Figure 5-6 SBUDAS electronic data acquisition components and interfaces

Insrument calibration used a B&K 4231 acoustic calibrator. It develops a 1000Hz tone, and the 94 dB setting was used when calibrating. Instrumentation was typically calibrated once per day. Table 5-7 provides the instrument specifications.

Table 5-7 Instrument specifications

G.R.A.S. 40AN 1/2" Ext. Polarized Free-field Microphone, Low Frequency	Freq range: 0.5 Hz to 20 kHz	Dyn range: 14 dBA to 149 dB	Sensitivity: 50 mV/Pa
G.R.A.S. 26AJ 1/2" SysCheck Preamplifier with integrated connector	Freq range: 2.5 Hz - 200 kHz	Noise: 1.8 µV	Gain: -0.35 dB
G.R.A.S. Power Module Type 12AQ	Frequency response:	For gain - 20 dB to 50 dB: 10 Hz to 100 kHz ± 0.1 dB	2 Hz to 200 kHz ± 0.2 dB
National Instruments NI 9234 Dynamic Signal Acquisition (DSA) Module	24-bit resolution	Anti-aliasing filters	102 dB dynamic range

The deployment of the eleven SBUDAS and four Sonic Pressure Integrated Kit Electronics (SPIKE) noise monitors[†††] for QSF18 is shown in Figure 5-7.

[†††] The NASA SPIKE units were deployed for QSF18 as part of the planned instrumentation, placed in lower priority locations due to the a) lack of network connectivity and near real time data analysis capability and b) manned triggering and operational requirements. Due to instrumentation issues no data was acquired from the SPIKE units.

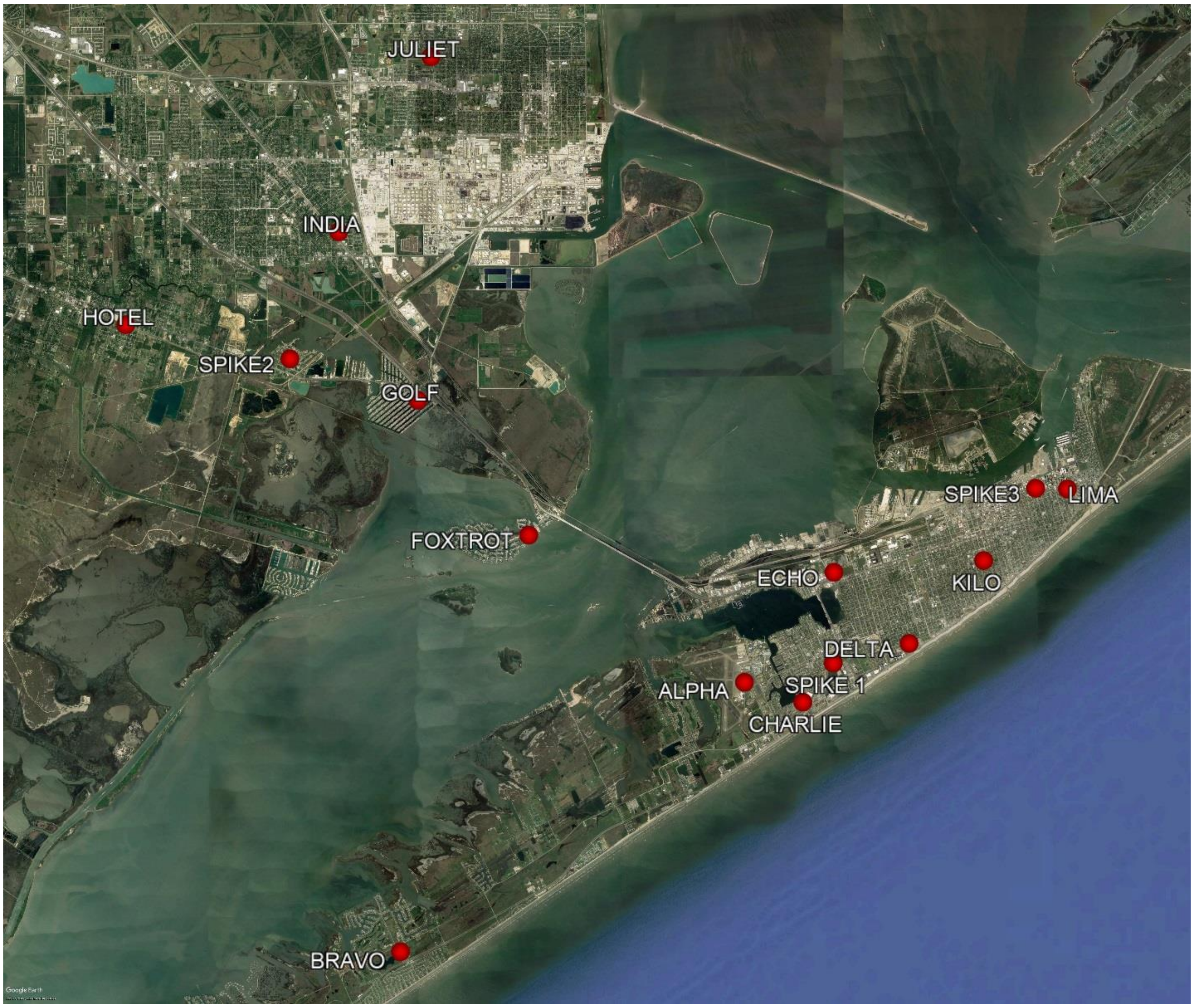
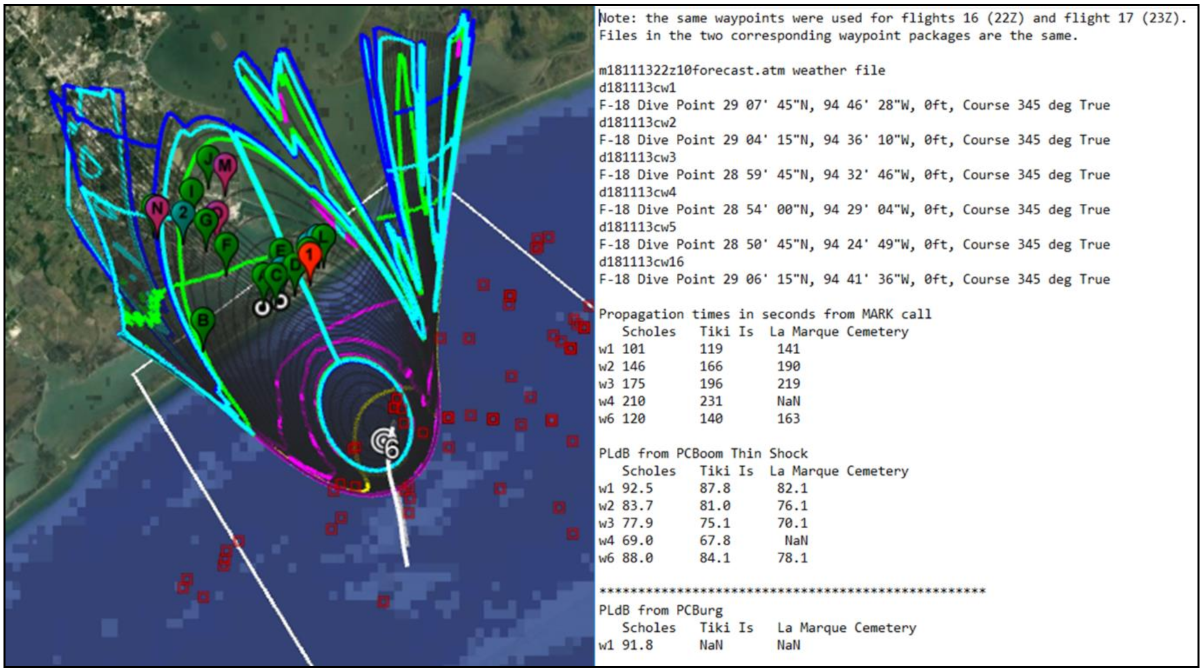


Figure 5-7 SBUDAS Noise Monitor placements during QSF18 are denoted using the phonetic alphabet

Two SBUDAS were deployed by a single Field Operator two hours before the commencement of flight operations on a daily basis. Deployment and calibration of all noise monitors was typically accomplished in less than one hour. Field personnel would then stand by their second SBUDAS leaving the first unattended throughout the course of the Flight Day.

Prior to each flight a “Waypoint Planning” email denoting anticipated waypoint and accompanying PCBoom predictions was distributed to all personnel via email. Figure 5-8 provides an example.



```
Note: the same waypoints were used for flights 16 (22Z) and flight 17 (23Z).
Files in the two corresponding waypoint packages are the same.

m18111322z10forecast.atm weather file
d181113cw1
F-18 Dive Point 29 07' 45"N, 94 46' 28"W, 0ft, Course 345 deg True
d181113cw2
F-18 Dive Point 29 04' 15"N, 94 36' 10"W, 0ft, Course 345 deg True
d181113cw3
F-18 Dive Point 28 59' 45"N, 94 32' 46"W, 0ft, Course 345 deg True
d181113cw4
F-18 Dive Point 28 54' 00"N, 94 29' 04"W, 0ft, Course 345 deg True
d181113cw5
F-18 Dive Point 28 50' 45"N, 94 24' 49"W, 0ft, Course 345 deg True
d181113cw16
F-18 Dive Point 29 06' 15"N, 94 41' 36"W, 0ft, Course 345 deg True

Propagation times in seconds from MARK call
   Scholes   Tiki Is  La Marque Cemetery
w1 101       119       141
w2 146       166       190
w3 175       196       219
w4 210       231       NaN
w6 120       140       163

PLdB from PCBoom Thin Shock
   Scholes   Tiki Is  La Marque Cemetery
w1 92.5      87.8      82.1
w2 83.7      81.0      76.1
w3 77.9      75.1      70.1
w4 69.0      67.8       NaN
w6 88.0      84.1      78.1

**************************************************
PLdB from PCBurg
   Scholes   Tiki Is   La Marque Cemetery
w1 91.8      NaN       NaN
```

Figure 5-8 Waypoint Planning Prediction distributed via email prior to Flight 17 of QSF18. PCBoom Prediction overlay (left) Waypoint Summary (right)

Ground control at Scholes Airport would inform all field personnel via LMRS radio of the status of each flight: Take off, turn in, and commencement of the low boom dive maneuver. "Turn in" defines when the F-18 aircraft began the final turn toward the test community, signifying the start of the test point. "Turn in" was generally XX minutes before the "Mark" call. "Mark" signified the start of the low boom dive maneuver. Field personnel would note the time from commencement of the low boom dive maneuver to the time that they heard any sonic thump. After a suitable time period, the NASA Field Crew Lead would request auditory reports from SBUDAS and SPIKE field personnel in sequence.

Portability and Power Management Earlier versions of the SBUDAS utilized Marine Lead Acid Batteries and large solar panels which made deployment a laborious process. The current SBUDAS data acquisition electronics fit within an 18x24x10 inch environmental enclosure with enough space left to accommodate the 12 Volt 20 Amp-hour rechargeable battery. A small solar panel provided charging through the day and additionally contributed to weather proofing of the electronics enclosure. Figure 5-9 depicts the battery and charging hardware. All SBUDAS were left on a trickle charge overnight when they were returned at the end of each flight day. The SBUDAS enclosure weighed approximately 15lbs, plus another 5 lbs for the solar panel. The entire system could be deployed and calibrated in less than 10 minutes.

Figure 5-9 Lithium Iron Phosphate rechargeable battery (top left), trickle charger (bottom left), SBUDAS with solar panels prior to deployment at GAC Savannah GA (right)

Weather proofing During the AFRC Pre-Test there was one day when a brief but intense rain shower passed over the test area which resulted in the cancellation of data collection and a hasty effort to secure all of the equipment before any microphones were damaged. For QSF18 it was decided that all microphones would be mounted in their environmental enclosure with hydrophobic wind screens to protect them in the event of rain (Figure 5-10). This was key as there were several days which had brief but intensive rain showers. During the rainfall the SBUDAS were left in the field and the wind screen was changed with a dry spare once the weather had passed.

Figure 5-10 Left: GRAS Microphone with environmental enclosure and wind screen Right: SBUDAS with microphone mounted in environmental enclosure with wind screen on tripod as deployed during QSF18.

Unattended Operation/Cellular VPN Eleven SBUDAS and four SPIKE noise monitors were deployed in support of QSF18. The SBUDAS were all controlled and monitored by two host-stations and operators located at Scholes Airport. Although all eleven could have been operated by a single individual, the second host station was included to minimize risk and manage workload. Field personnel placed each SBUDAS in the morning and assisted the host station operator with calibration in the morning and prior to securing the noise monitors on a daily basis; otherwise there was no further action required in the field with respect to the SBUDAS. Half of the SBUDAS were truly unattended all day with some in public places; the other half had field personnel on hand but only for subjective feedback to corroborate objective data and not for SBUDAS operation. All performed with excellent reliability collecting measurements for every boom event at all deployed stations.

Improved Near Real Time Feedback The SBUDAS Host Station Operator had the ability through a National Instruments Labview™ Interface to control, monitor the health and review data collected on the SBUDAS on a near-real-time basis. Preliminary pressure and PLdB measurements for all SBUDAS were plotted and distributed via email to critical team members within minutes of each boom event as shown in Figure 5-11.

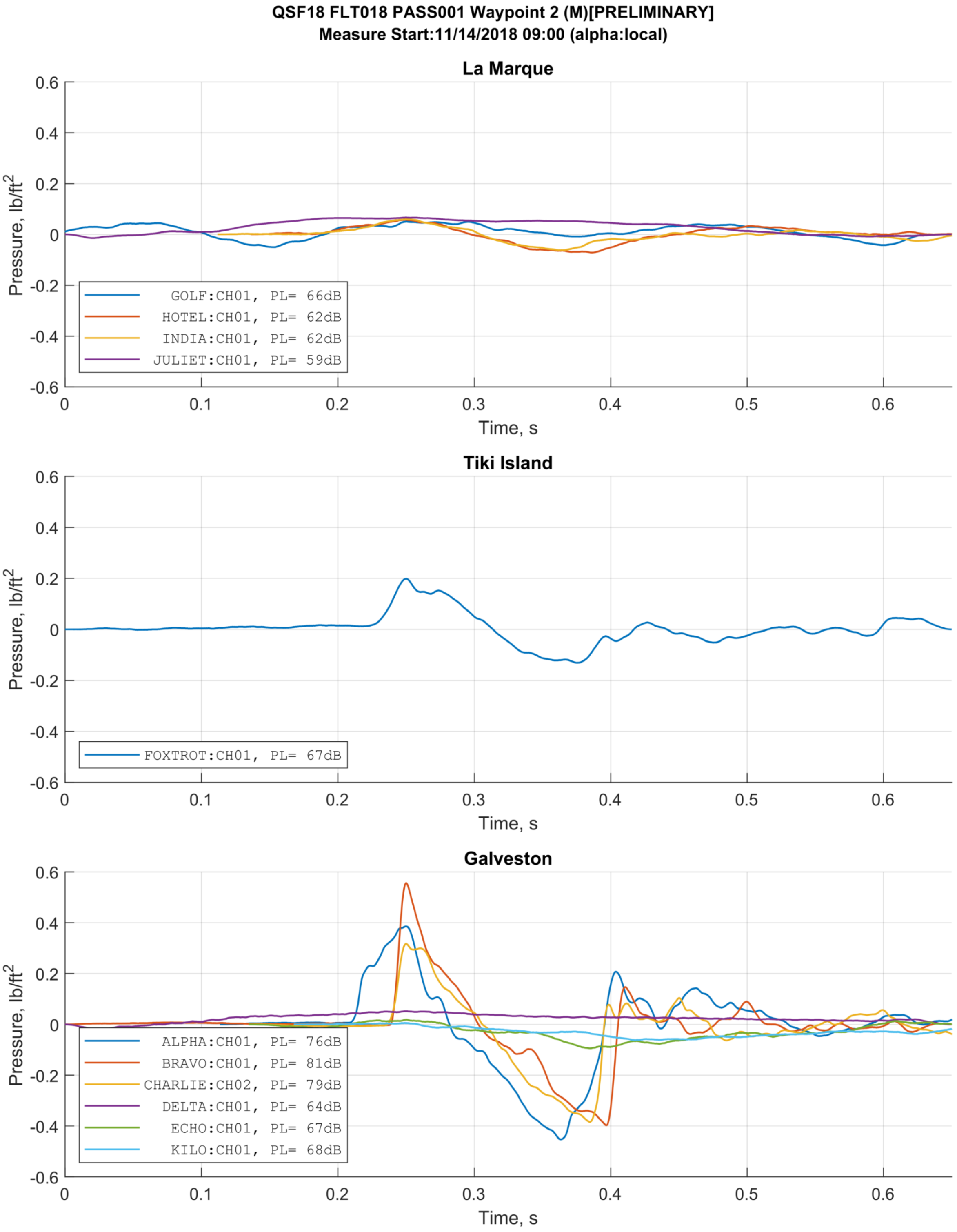


Figure 5-11 SBUDAS near real time feedback following each boom event.

This confirmed successful data collection and provided valuable feedback to event coordinators for adjustment to waypoint selection to ensure that satisfactory levels were maintained over the community.

5.5.1 Objective Data Collection Lessons Learned

Valuable lessons learned during the Armstrong Flight Research Center (AFRC) Pre-Test conducted in May 2017 led to several significant improvements to the SBUDAS. These improvements included weather proofing, improved batteries and solar power, true unattended operation and near-real-time acoustic reports distributed via email, which served as critical situational awareness for the team throughout the test execution.

5.6 Subjective Data Collection

The subjective data collection was executed via participant surveys. These surveys were executed in conjunction with geolocation techniques, to enable dose response analysis. Detailed discussion of the surveys and geolocation are as follows.

The team used a custom java script written by the Penn State Survey Research Center (SRC) that was applied within Qualtrics[‡‡‡]. Survey data management and geolocation were implemented through:

- Qualtrics survey on a GPS enabled device
- PSU SRC application implemented to identify respondent's location
- Application identifies latitude and longitude of respondent
- Phone presents graphical map of location with query "Is your location correct?"
    - If yes, app proceeds to sonic thump questions
    - If no, application prompts respondent to enter current address

In compliance with IRB requirements, the respondents provided informed consent to have location services enabled on their device and to allow their location to be retrieved and sent through the mobile survey. The SRC custom java script provided the latitude and longitude position of a participant responding through the Qualtrics survey on a GPS-enabled device using any web browser. However, when the location services was turned off, the respondent was asked to input their location via an explicit survey location question.

The survey instruments used for QSF18 are summarized in searchable outline form in Appendix J. Screen shots of the surveys as seen by participants are provided in Appendix K. The survey questionnaires were formatted in a mobile enabled web platform that is https, using Qualtrics, a web based survey software

---

[‡‡‡] Qualtrics website: https://www.qualtrics.com/support/survey-platform/getting-started/survey-platform-overview/

tool. The src.survey.psu.edu page was a "hosting" page for information that included the project information and the enrollment survey. The survey instruments included background survey, single event, daily summary and final feedback survey to provide a comprehensive dose response data set, designed to assess annoyance due to sonic thumps and responses on a set of features such as demographic variables, respondent attitudes towards noise in general, their perceived noise sensitivity, and their perceived ability to habituate. The surveys were designed to gather sufficient data to develop a dose-response model to assess the percent of respondents that are highly annoyed by sonic thumps. The surveys were also designed to gather sufficient data to support statistical analysis to identify underlying relationships and contributing factors. Participants were required to provide a home and work address (as appropriate), and an email address. Participants were asked to provide a cell phone number, but were not required to provide this information. Answers to other questions were optional. Multiple methods were incorporated to identify respondent locations, including an automated geolocation feature to ease burden of use, and responses to other survey questions.

Respondents' name, email, cell phone number and address were used as needed for test communications. Each household was given a unique enrollment code, and each respondent was provided with a unique link to access the surveys which could be completed multiple times. Each link was associated with an access code (or "ID code"). The participants entered their unique ID code which corresponded to their unique responses. This system allows for the tracking of respondents, and it does not put any personally identifying information (such as name, email, etc.) into the shared data file. Only the respondent code is included with the response data. The surveys cannot be completed without the code. The home and work addresses provided were used for determination of the noise dose only. The noise dose was associated with the respondent's ID code, and not the respondent's identity.

The communications with respondents were conducted via email across all of the various survey instruments and the respondent groups. Reminder group respondents received additional communications. All potential respondents were sent an email reminder in the AM to participate, with the single event link. An end of day email reminded them to complete the daily summary and included the daily summary link. The respondents in the non-reminder group only received the AM and PM reminders. They did not receive reminders throughout the day. Only respondents in reminder groups received reminders throughout the day. Reminder group respondents were informed to look for texts or emails throughout the day, depending on their group assignment.

In addition to the reminder with the single event link that all respondents received in the AM, the reminder group were also sent text messages or emails with the survey link throughout day reminding them to listen. Some of the texts were sent just after a thump occurred and some were sent as random "false reminders". There was a maximum of 10 reminders per day. The flight schedule with the associated pre-thump and false reminders is summarized in Figure 1-2. The invitation to complete the post test feedback, with a code and a link to the feedback "portal" page, was sent to all respondents on November 16, 2018. Access to the final feedback survey was closed on November 19, 2018. As detailed in Table 5-8, they were sent text messages or emails with the survey link throughout the day reminding them to listen. The electronic communication schedule is detailed in Table 5-9.

Table 5-8 Electronic communications

Electronic Communications	
Pre Field Test Communications Test	
Test Communications were sent to test communications, the geolocation map, and confirm contact.	
Text Testing (no subject line):	This is a communications test for the NASA survey. Please respond at:${l://SurveyURL}
Email testing: (subject line: NASA Survey)	This is a communications test for the NASA survey. Please respond at:${l://SurveyURL}
Daily Morning Reminders	
Morning reminder(subject line: NASA Survey)	Thank you in advance for participating in the NASA survey today. You can use the link below to report a sonic thump any time you hear one throughout the day. You can use the link as many times as necessary. You will also receive an email at the end of today for your daily summary survey. If you have any technical issues, please email the PSU Survey Research Center at: srcwebsurvey@psu.edu Use this link to access the NASA Survey: ${l://SurveyLink?d=Take the Survey} Or copy and paste the URL below into your internet browser: ${l://SurveyURL} Follow this link to opt out of future emails: ${l://OptOutLink?d=Click here to unsubscribe}
Single Event Reminders	
Both email and text have the same message; email has a subject line, the text message does not have a subject line.	
Subject line: NASA Survey Reminder	Did you just hear a sonic thump? Please click on the NASA survey link, indicate your location, and answer yes or no. Please complete the additional survey questions. If you are driving, please wait until you have stopped to complete the survey. ${l://SurveyURL}
Daily Evening Reminder for Daily Summary	
Subject line: NASA Daily Summary Survey	Thank you for your participation today in the NASA survey. Please complete your NASA daily summary survey via the link below. If you have any technical issues, please email the PSU Survey Research Center at: srcwebsurvey@psu.edu. Follow this link to the Survey: ${l://SurveyLink?d=Take the Survey} Or copy and paste the URL below into your internet browser: ${l://SurveyURL} Follow the link to opt out of future emails: ${l://OptOutLink?d=Click here to unsubscribe}
Final Feedback	
Subject line: NASA Final Feedback	We appreciate your participating in this NASA research study! At this time, please complete the final feedback survey. You will receive compensation of $25 per week for the two weeks of the survey for a total amount of $50 as an expression of appreciation. If the survey is terminated before the end of the first week, participants who completed the survey until its termination will receive $25. If the survey is terminated after the first week, but before the end of the second week, participants who completed the survey until its termination will receive $50. Please allow 1 month for the compensation to be processed. If you have any technical issues, please email the Penn State Survey Research Center at survey srcwebsurvey@psu.edu. Follow this link to the Survey: ${l://SurveyLink?d=Take the Survey} Or copy and paste the URL below into your internet browser: ${l://SurveyURL} Thank you for your participation and feedback!

Table 5-9 Electronic communications schedule

Electronics Communications Schedule				
Times Sent	Text Message Group		Email Group	
	No Reminder	Reminder	No Reminder	Reminder
Sent Pre-Test to confirm communications contact	Text Test	Text Test	Email Test	Email Test
Sent Each Morning	AM Reminder Email	AM Reminder Email	AM Reminder Email	AM Reminder Email
Sent After Each Event		Text Reminder		Email Reminder
Sent Each Evening	PM Reminder Email	PM Reminder Email	PM Reminder Email	PM Reminder Email
Post Test Feedback	Feedback Reminder Email	Feedback Reminder Email	Feedback Reminder Email	Feedback Reminder Email

The response data were accessible on-line so that the team could monitor the annoyance response to each sonic thump. The annoyance data included the date and time of the sonic thump response and the single event annoyance rating. The daily summary data were also monitored to track the cumulative annoyance response. Because respondents were permitted to enter daily summary data the following morning, the review and closure of daily survey was conducted at noon on the day after the test day. The daily summaries gathered the date, cumulative annoyance rating, and time of survey submission for respondents.

During the flight test period the SRC Qualtrics survey did not experience any outages, and the SRC personnel were able to successfully monitor the communications throughout the test. There were no disruptions in on-line access to the surveys during the field test.

Some of the planned flights had to be rescheduled due to weather. An additional flight day was added on Test Day 11 to capture some of these flights. The survey instruments did not include this date, so this revision required the survey to be edited on-line to add the additional date to the response bubbles. For the single event survey, as designed, if the respondent reports that they did not hear the thump, then the survey inquiries stop and they are not asked about annoyance. However, an oversight occurred in the Day 11 edit when the additional date response bubble was added. The survey was saved with the coding allowing progression to the next question, as this was the default save mode. The survey editor neglected to add the code that prohibited advancement to the annoyance question if the thump was not heard. This oversight allowed a respondent that did not hear the thump to report on their annoyance. The statistical analysis used these annoyance ratings. Note there were 3 total people who reported HA and not heard on Day 11 (out of the 752 who reported not heard on that day), where 1 was very annoyed and 2 were extremely annoyed. Thus, this issue had a negligible impact on results.

5.6.1 Subjective Data Collection Lessons Learned

The survey implementation provided sufficient data for analysis of both single event and cumulative daily annoyance response data. On the single event surveys, respondents were asked if they first heard the event, and were then asked to rate their annoyance. As such, there were respondents that indicated that they did not hear the event. This group of respondents were presumably receiving reminders, since they responded for something they did not hear (the analysis did not include an item by item review to verify that all responses that an event was not heard were from respondents that received reminders, however such a review could be conducted). Respondents were informed that the design included several "false reminders" throughout the day, so it would also be appropriate to not hear non-events. The sonic thump levels were quiet enough that the non-reminder respondents did not respond as frequently to single events. It is recommended that all respondents receive reminders for all single events to improve response rates. The inclusion of false reminders embedded within the design offsets the risk of respondent responding to the reminder rather than the event.

The surveys were reviewed by both the PSU and NASA IRB's, and contained clear and specific language for respondents to follow. Based on questions to the PSU SRC, and the lack of full submissions on the Daily Summary, there were still points of confusion. For instance, respondents were asked to manually enter their location if the automated location provided in the GPS map was not correct. The respondents did not provide street addresses as expected. This concept needs to be reinforced with an example address, because the entries were not as detailed as was anticipated.

Text and email prompts were successful in encouraging responses to single events, with text prompts being the most effective. Emails were sent every morning as a reminder to listen and every evening to remind respondents to complete the Daily Summary with the link to the survey embedded within the messages. The design should reiterate that a Daily Summary submission is required every day from every respondent even if they didn't hear any thumps that day. It is recommended that the incentive be made contingent on submitting "X" daily summaries. It isn't pragmatic to require 100% participation, but it should be at least 50% or more for the Daily Summary.

Based on lessons learned from the AFRC test, the survey included the potential for respondents to go back within individual sections of the survey when providing responses. This was done to make it easier to complete the surveys. The dates were implemented in selectable format rather than an editable field to make survey completion easier.

Surveys were accessed by individual on-line links. There were some issues with respondents starting a survey and leaving it open without completion. The surveys were left "open" so that respondents could partially complete a survey and, if interrupted, return to finish where they left off in the survey. Some individuals returned and started new surveys, resulting in multiple submissions with different time stamps. Other multiple submissions were observed with the same time stamp, indicating that they hit submit multiple times at the time of submission. Some open surveys resulted in more than one daily summary being submitted for the same individual. The surveys were closed by the SRC before the field test began each day to ensure that there was a new set of data collection. It is recommended that the surveys still be left open to provide a full range of options to the respondents. Instructions should provide examples of the potential for multiple submissions. The submissions should be dealt with in the data cleaning process. Section 6.2.1.2 describes how multiple submissions were handled in the analysis. Protocols for data handling should be developed and refined before the X-59 test, in particular the potential for multiple survey submissions from a unique ID for a single event. Flexibility should be maintained in the survey design to support the respondents, but the research should have clearly defined data cleaning protocols.

The following actions are recommended to modify survey instructions and implementation to encourage respondent participation.

- Develop Frequently Asked Questions located on both NASA and PSU SRC pages.
- Reiterate contact PSU SRC for enrollment questions and NASA for research background.
- Reiterate instructions to respondents for all survey completion actions.
- Add language to end-of-day reminder to always complete the daily survey, even if they didn’t hear anything.
- Provide “address example” at open field to enter address if auto-geo location is not correct and require address conformity checks in survey software to improve geolocation success rate on user provided address data.
- Consider a trade study on recommended procedure for respondents. The success of the auto geolocation is dependent on whether the respondent keeps geo-location on, or if it is turned on at the time of responding. If the procedure is to keep it turned on, the battery will run down more rapidly.

5.7 Measurement Data Archive

A full measurement data archive for QSF18 was assembled and delivered to NASA. This section provides an overview description of the structure and contents of this archive. Full details, with description of structure and contents, and file nomenclature, are provided in the Description of Data included in the archive. Figure 5-12 and Figure 5-13 show the structure, with representative details. Large numbers of directories and files exist for each flight, and the structure and directory and naming details are similar for each. To assist using the data, certain subdirectories contain readme files that provide specific information for interpreting the data.

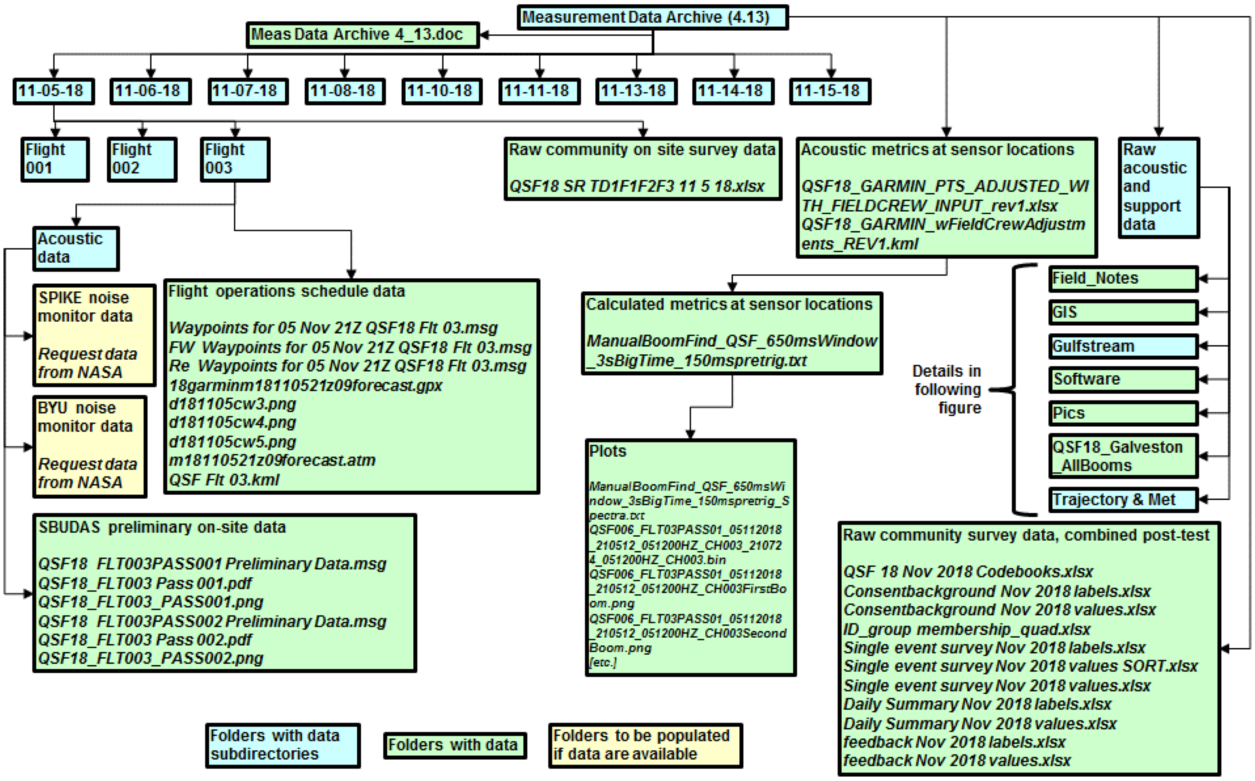


Figure 5-12 QSF18 Measurement data archive structure

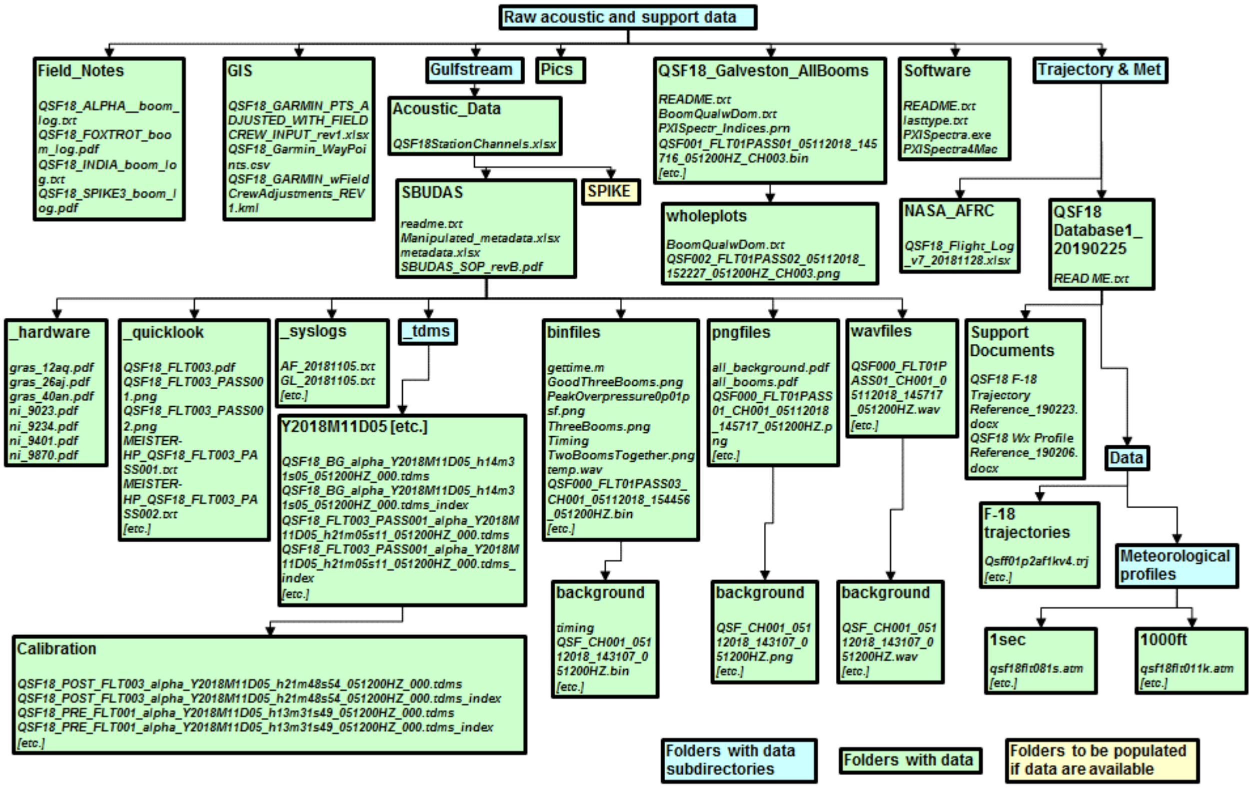


Figure 5-13 QSF18 measurement data archive structure – raw acoustic data

As shown, the set of folders underlying the top level consists of:

- A folder for each flight day with preliminary data collected and distributed on site for each flight
- Raw community survey data combined post test
- Acoustic metrics at sensor locations
- Raw acoustic data

Raw acoustic and support data

Seven folders are provided underneath the top level raw acoustic data folder. These are:

- Field notes
- GIS
- Gulfstream
- Pics
- QSF18_Galveston_AllBooms
- Software
- Trajectory and Met

6. QSF18 Experimental Results

The primary result of the QSF18 test is the community dose-response relationship to low level sonic booms. This relationship was built upon a number of analyses that included both subjective and objective analyses, as well as hybrid elements that leveraged and combined both empirical datasets. The relationships between these data streams are presented in Figure 6-1 (the data flow diagram), which summarizes the data and analysis flow for the QSF18 Galveston test. The following sections describe the various analyses and results from the test.

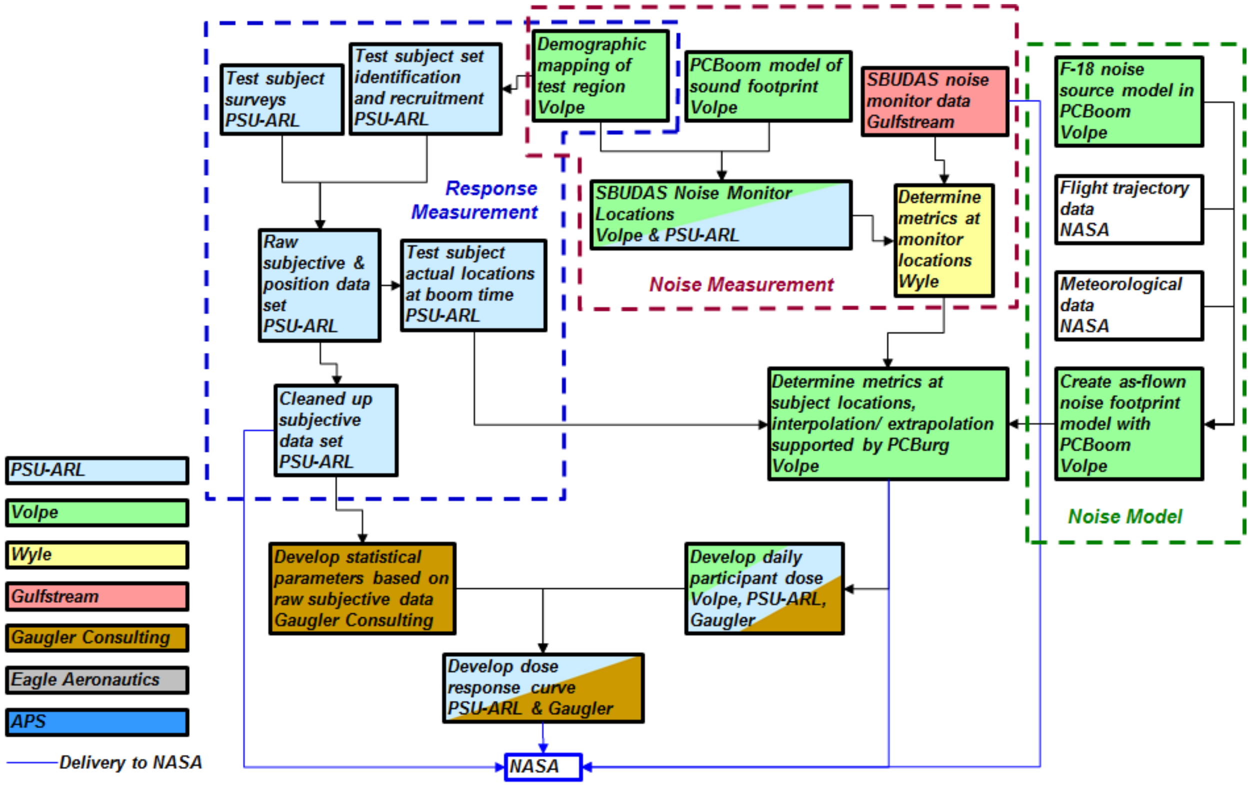


Figure 6-1 QSF18 data flow diagram

6.1 Objective Data Analysis

Analysis of the objective data comprised a multistep process. The acoustic data was processed and metrics determined at sensor locations (Section 6.1.1). The participant locations were determined at the time of each event based on a combination of the single event, daily summary and known event times (Section 6.1.2). Sonic boom footprints were calculated using the as-flown tracking and meteorological data, and were fused with the georeferencing data via an analytically guided interpolation process of the empirical data to ascertain each participant's noise exposure for each event (Section 6.1.3). The cumulative daily dose was then computed.

6.1.1 Determination of Metrics at Sensor Locations

The recordings in the measurement archive represent a sampling of the sonic thump footprint created by the sound emissions from the aircraft when it is traveling supersonic. Additionally, Appendix T provides summary tables showing thump metrics calculated from measured data, in comparison to the design levels. The low boom dive maneuver can create many complex signatures, including two sonic thumps at a location on the ground. Each thump is generated from a different point along the supersonic portion of the trajectory. An example recording from the Bravo monitor is shown in Figure 6-2. The recording clearly shows two sonic thumps separated by approximately 1.4 s. The figure shows the first shock of each thump marked with a red x.

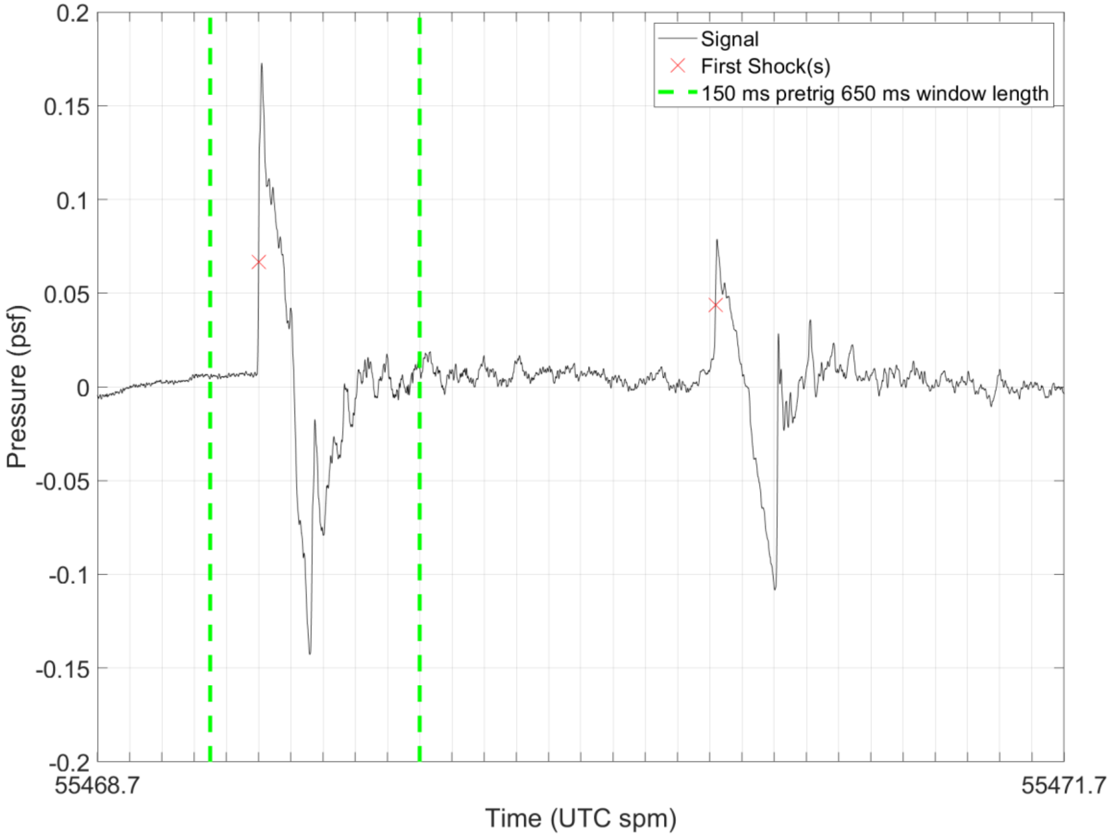


Figure 6-2 Recorded signature at Bravo during the second pass of the first flight (QSF002) on 05 Nov 2018

The Test Plan detailed the use of the Auto Boom Finder program [Hobbs, 2012] for identifying the sonic thumps in the recordings; however, due to the complexity of some of the signals, the Auto Boom Finder program was unable to identify portions of the records that were clearly associated with emission from the aircraft when it was traveling supersonic. While the Auto Boom Finder program was able to identify both the sonic thumps in Figure 6-2 because they are well separated, the program failed to trigger (identify) the first shock of the boom pictured in Figure 6-3. The trigger marker in the figure shows where the program estimated the first sonic thump (boom) to begin. As can be seen in the figure, the program was unable to find the first shock. This is an example of the two sonic thumps arriving at nearly the same time. This can occur at the edges of the footprint.

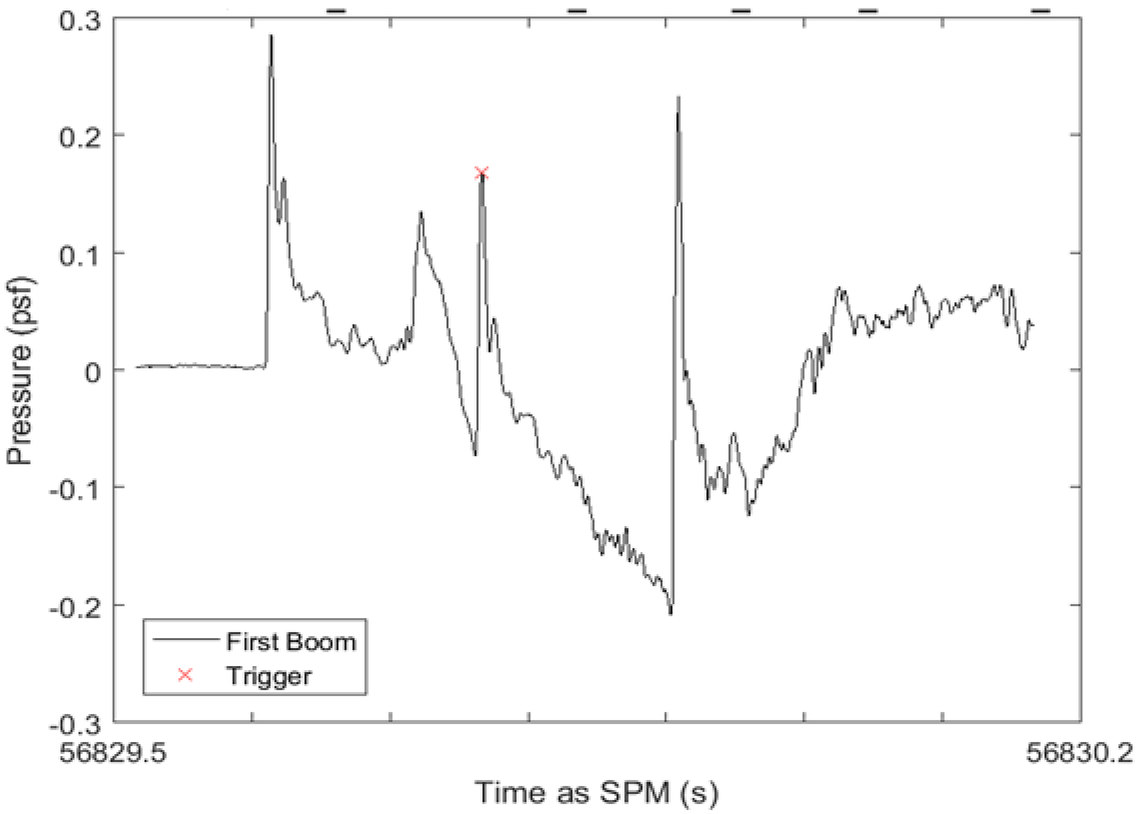


Figure 6-3 Auto Boom Finder result looking for events in recording at Kilo from flight 1 pass 2 (QSF003)

Another type of signature that the Auto Boom Finder program was unable to identify was often recorded when the monitor's location was beyond the edge of the footprint predicted by PCBoom Version 6.70b. An example is shown in Figure 6-4. While this type of waveform is a result of the supersonic portion of the aircraft's trajectory, it typically does not have shock structure characteristic of conventional sonic booms. It can have a different period, oscillate more than one cycle, and be heard as a rumble. The Auto Boom Finder program was not able to identify (trigger on) the waveform in Figure 6-4. The algorithm for finding sonic booms that is used by the Auto Boom Finder program is dependent upon frequency content and shape specific to N waves with a specific period [Hobbs, 2012]. Events that deviate from this shape, including overlapping thumps, very low amplitudes, and off-the-carpet events, may elude the program's algorithm.

An example of a calculated footprint's peak overpressure contours overlaid on the Galveston area showing the locations of the monitors Alpha (A) through Kilo (K) along with the location of participants on the single event survey can be seen in Figure 6-5. The figure shows a three second clip of the recording at the indicated monitors with the peak overpressure noted in the time trace. All graphs on the footprint have the same scale. This test point was the 28th supersonic pass of the aircraft. The measurement campaign had a total of 52 passes of the aircraft traveling faster than the speed of sound. There was a practice flight where the aircraft did a pass without going supersonic. The sequential numbering of the recordings is related to the flight and pass number of the aircraft in Table 5-5. All footprints and overpressure traces at monitors similar to these figures can be found in Appendix Q.

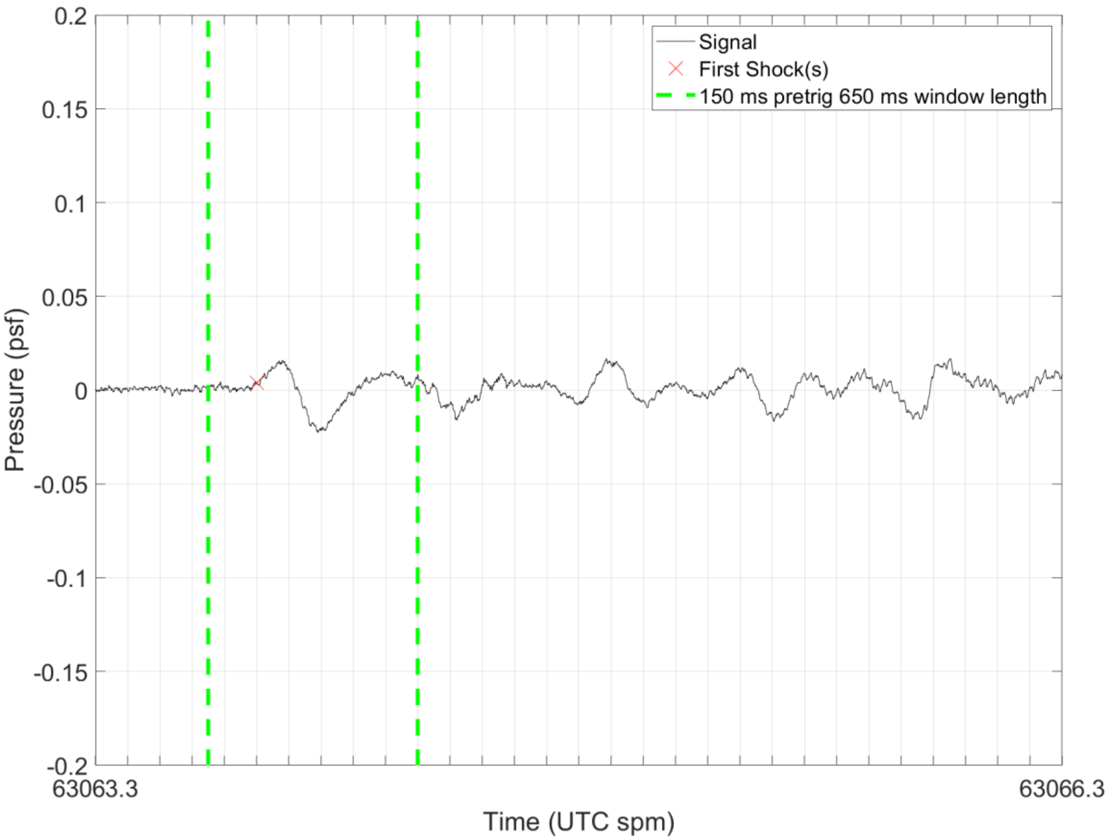


Figure 6-4 Waveform recorded at Hotel during QSF052 with beginning of the waveform manually identified

An important note regarding the analysis of the monitor data: the field crew at the monitoring sites did not report hearing any aspect of the flight (sonic thump, rumble, etc.) during the first pass of the first flight (QSF001) on 5 November 2018. The intitial review of the monitor data did not find any noise from the aircraft that was generated when it broke the sound barrier. Reports from the Field Lead indicated the aircraft was off course when it began the low boom dive maneuver. A subsequent review of the data did show that the calculated footprint was well to the south of the study area; however, the monitor at Bravo did record an event that was clearly from the aircraft while traveling supersonically. The calculated footprint is shown with the recording at Bravo in Figure 6-6. It was reported as being heard by five study participants near the Bravo monitor.

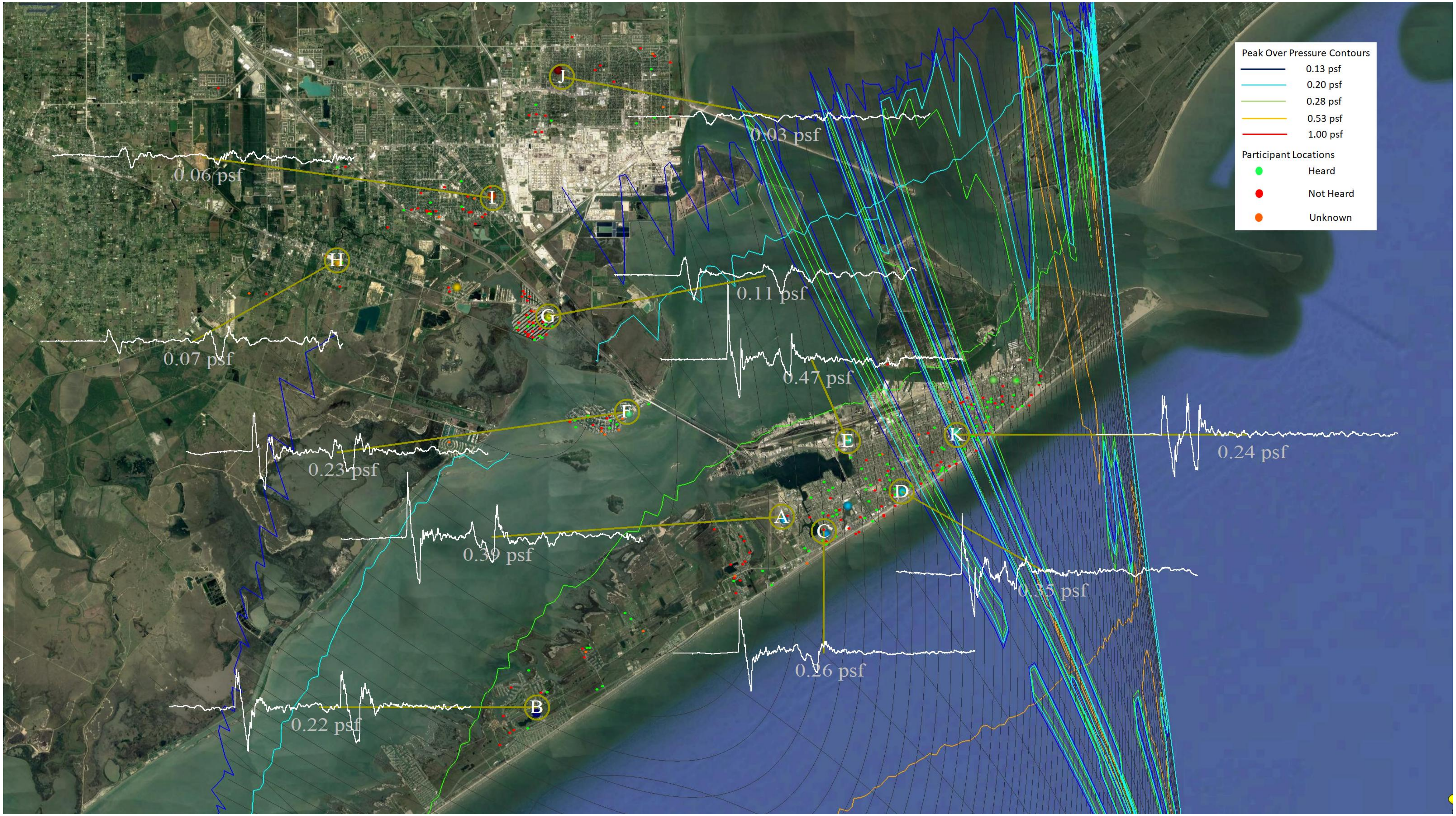


Figure 6-5 Calculated footprint showing peak overpressure contours from QSF028 with recordings and participant locations

Figure 6-6 Calculated footprint showing peak overpressure contours from QSF001 with recordings and participant locations

The Auto Boom Finder program was unable to identify recorded signatures that lacked the conventional N-wave shape, either from monitors off the footprint or near the edge when the two sonic thumps overlapped; therefore, the beginning of the sonic thumps were identified manually by visual inspection for all SBUDAS data. The time of the beginning of the first sonic thump and second thump (if observed) were tabulated for use in the metric analysis described below.

Once the start of a noise event was manually identified, the analysis followed the same methodology as the first WSPR program [Page *et al.*, 2014]. A thump such as the one shown in Figure 6-7 was manually identified by its first shock. The length of time from the recording was identified as starting a Pretrig amount of time, 250 ms in this figure, before the first shock, and would extend to a duration of 650 ms as shown in this figure. The ends of the record would be smoothly transitioned to zero with a half cosine that is Taper ms long (100 ms in Figure 6-7). The acoustic metrics were then calculated from the resulting waveform.

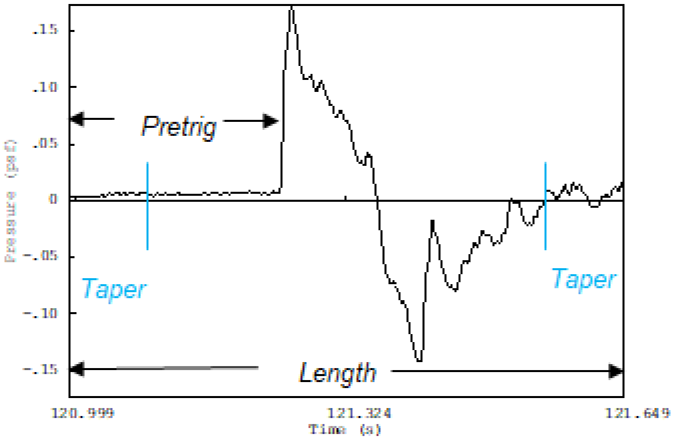


Figure 6-7 Window parameters used in calculating metrics

Two different window lengths were used in the analysis: 650 ms and 3 s. The first was to match the window length used for WSPR with the focus being on the first thump, and the second was to capture both thumps (if two exist) along with any rumble after the thumps. The metrics calculated for analysis are shown in Table 6-1.

Table 6-1 Metrics calculated from recordings

Metric	Unit	Description
PL	dB	Stevens' Mark VII Perceived Level
CSEL	dB	C-weighted Sound Exposure Level
ASEL	dB	A-weighted Sound Exposure Level
FSEL	dB	Unweighted Sound Exposure Level
LLZf	Phons	Zwicker[§§§] loudness for frontal incidence
LLZd	Phons	Zwicker loudness for diffuse incidence
PNL	dB	Kryter's Perceived Noise Level
BSEL	dB	B-weighted Sound Exposure Level
DSEL	dB	D-weighted Sound Exposure Level
ESEL	dB	E-weighted Sound Exposure Level
ISBAP	dB	Indoor Sonic Boom Annoyance Prediction Level
Peak	psf	Maximum value of record
Npeak	psf	Minimum value of record

In the event there were two thumps, a set of metrics was calculated for each thump with a 650 ms window length, a set of metrics for a 3 s window starting just before the first thump, and a set of metrics for the 650 ms just preceeding the first thump's window. This was the ambient noise defined as the section of the recording immediately before the first thump. Only the ambient, first thump with 650 ms window length and first thump with 3 s window length were considered for analysis. In Figure 6-8 the portions of the recording at a monitor that were used for analysis are marked. In this example the first shock of the first thump (found manually) was used to define three portions of the recording: a 650 ms window containing the first thump (using a 150 ms pretrig in this example and marked with green vertical lines in the figure); the ambient which is the 650 ms of the recording just before the first green line; and the 3 s window which begins at the first green line. These three portions of the recording were prepared for analysis using a 100 ms half-cosine taper to smooth the ends to zero.

[§§§] Zwicker loudness may be of use to future researchers, possibly for sound beyond the cutoff.

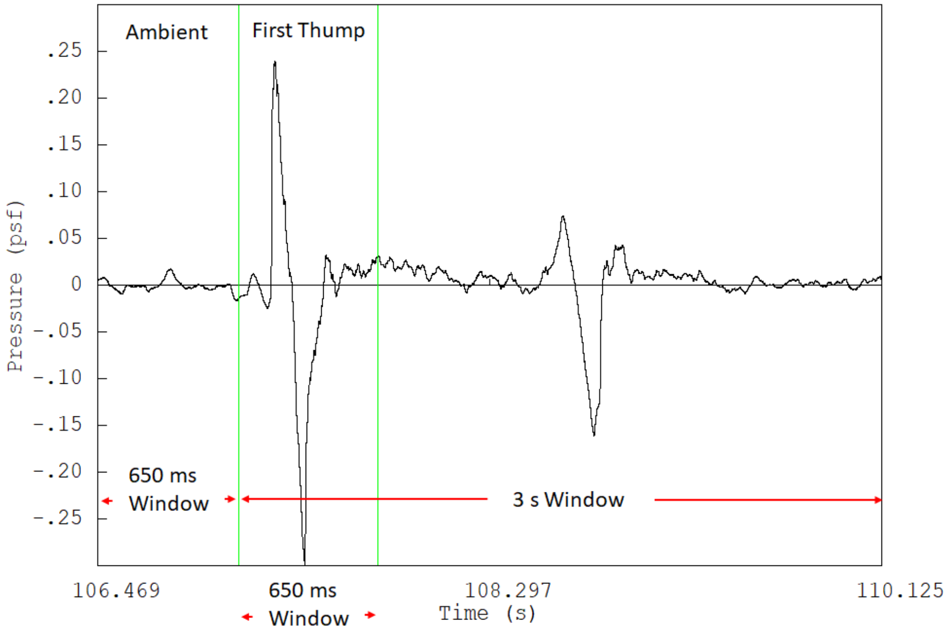


Figure 6-8 Example of portions of recording used for analysis

For WSPR, a pretrigger duration of 250 ms was used to isolate the first thump for analysis. A comparison of the PL values of the QSF18 measurements in 2018 with those made at Edwards Air Force Base during the WSPR project in 2011 is shown in Figure 6-9. The figure shows that much higher amplitude booms were recorded at Edwards in comparison with those at Galveston. Furthermore, the relationship between the level of loudness and the peak amplitude is similar for events with a peak amplitude of 0.1 psf or higher. The lower amplitude thumps from Galveston show a different relationship between the Mark VII level of loudness and peak amplitude. The apparent rate of change (slope) of the PL metric versus the peak overpressure decreases for lower amplitude thumps. This may be the relative influence of ambient noise in the computation of the metric as the thump's contribution diminishes. A change in the slope of the Mark VII level of loudness versus peak amplitude would be expected as the energy in the thump diminishes relative to the constant energy of the ambient; thus, the calculated level of loudness for the smallest amplitude thumps would be based on the constant energy of the ambient and never go below a certain value.

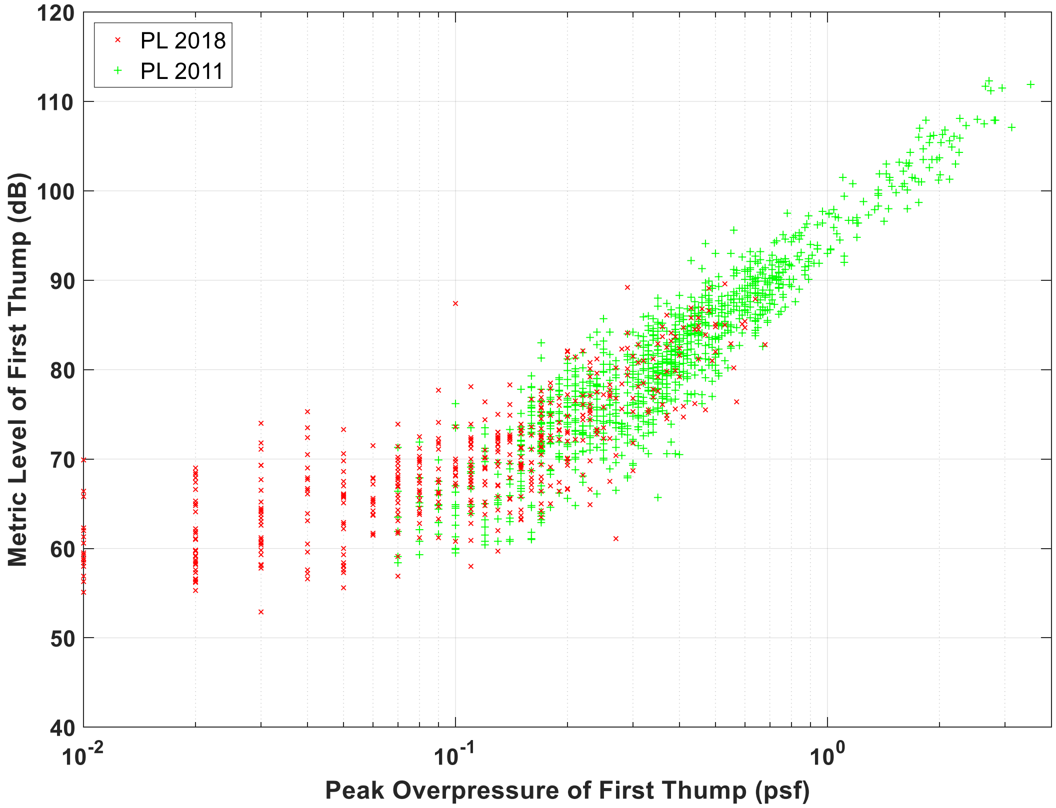


Figure 6-9 Comparison of Stevens' Mark VII level of loudness measured during WSPR (2011) and QSF18 (2018)

The pretrigger duration for computing the metrics for the final analysis of the Galveston data was 150 ms. This was chosen based on visual inspection of the events recorded. While the resulting level of loudness did not change for the analysis of the first thump as shown in Figure 6-10, decreasing the pretrig duration does allow for longer duration sonic booms expected for future aircraft which are expected to be much longer than the F-18s used for the current measurements campaigns; thus, while the duration of the thumps generated by the 17 m long F-18 for the low boom dive maneuver is approximately 160 ms when they reach the ground, a civilian supersonic aircraft like the Concorde at 62 m fuselage length would generate a thump on the ground that is much longer in duration.

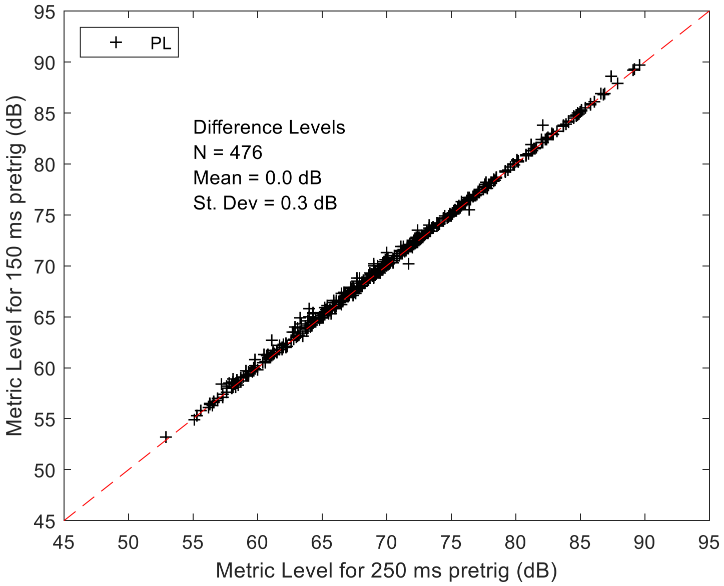


Figure 6-10 Comparison of decreasing pretrig duration when calculating Stevens' Mark VII level of loudness

In order to determine which metrics were contaminated by ambient noise, either because the ambient was abnormally high or because the thump was very low in amplitude, a comparison was made of the metric of the first thump with the metric of the ambient. In many instances, the low levels of the recorded events both inside and outside the footprint were too close to the ambient to be usable for analysis. The difference between the first thump and ambient metric values was used as a surrogate for the signal to noise ratio (SNR). Difference levels of 1, 3, and 5 dB were used as test criteria for determining whether or not a thump recording's metric value could be used in the dose determination analysis (see Section 6.1.3 which describes calculation of noise dose at participant locations).

An example of how many of the measured first thumps have a Stevens' Mark VII Perceived Level (PL) greater than the ambient and by how much is shown in Figure 6-11. The PLs of 143 of the 476 events identified as first thumps (30%) were 10 dB or greater than the ambient, and 314 of the 476 events identified as first thumps for the analysis (66%) had PLs more than 3 dB above the ambient. The first bin in the figure ranged from -5 to 1.01 dB, and the following bins are 2 dB wide. Because the metrics were tabulated to the nearest tenth of a decibel, any metric with a difference level of greater than 1 dB was reported in the bin to the right of the first bin. Any recording with an ambient level of loudness greater than the first thump can be explained by the thump having an insignificant contribution and the ambient varying as the first thump is recorded. The number of recorded events that are above the ambient by a certain level was different for different metrics. The metrics calculated from the recordings for the entire measurement campaign are available in the electronic archive. A sample of the metrics for the portions of the record discussed above are shown in Table 6-2. The actual file in the Measurement Data Archive is located at:

Meas data archive 4.13\Acoust metrics sensor locations\Calculated_Metrics\ManualBoomFind_QSF_650msWindow_3sBigTime_150mspretrig.txt

The Measurement Data Archive contains the ASCII output of the computations. It also contains fields not used for this analysis.

Summary statistics of all the recordings with identified events can be found in Table 6-3. There were a total of 575 recordings. Of the 575 recordings, there were 476 that were determined to contain noise from the aircraft emanated during the supersonic portion of the trajectory. This was done by visual inspection. All analyses used the monitor recordings with identified events from QSF002 through QSF052. The LIMA monitor operated for only four passes.

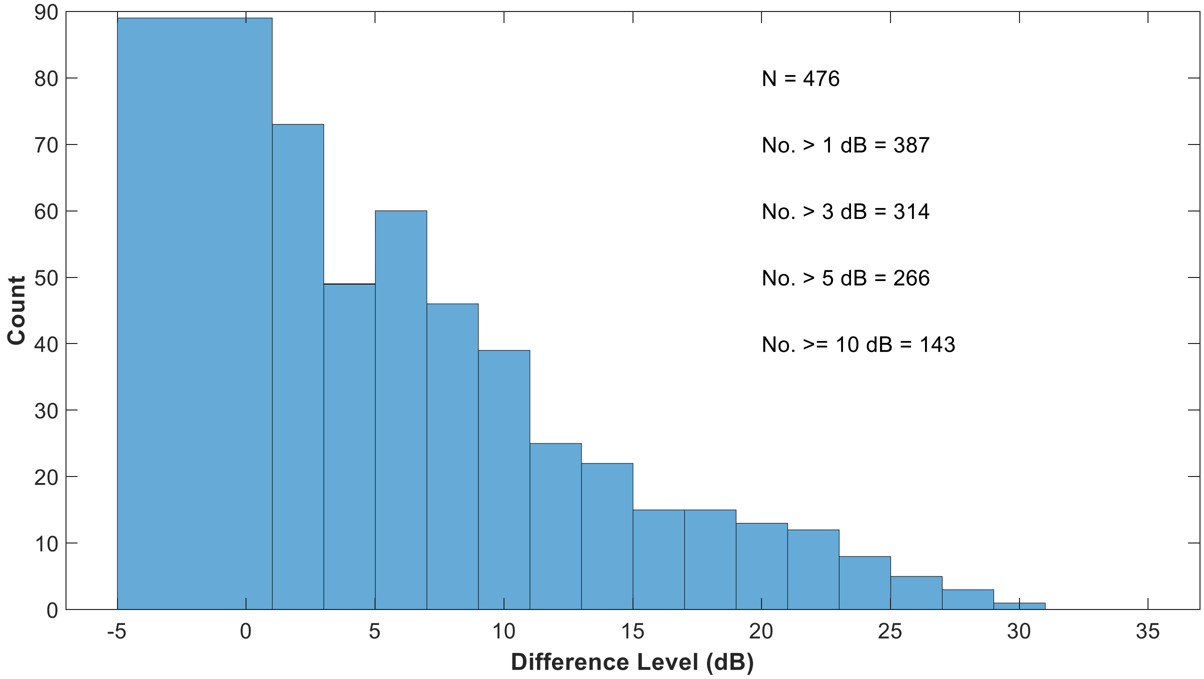


Figure 6-11 Histogram of differences in Perceived Level between first thump and ambient

Table 6-2 Example metrics

Ambient													1st Thump													2nd Thump													3 s Window													Seq.	Flt.	Pass	Mon.
PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF	PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF	PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF	PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF				
69	65	56	88	83	84	76	60	62	60	73	0.05	-0.01	85	97	69	108	98	98	93	83	84	78	97	0.50	-0.38	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	87	97	70	108	100	100	95	83	85	79	98	0.50	-0.38	QSF037	16	03	Alpha
68	61	53	81	82	82	75	56	60	57	72	0.02	0.01	77	87	60	101	90	90	86	75	75	70	88	0.18	-0.14	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	80	88	63	102	93	93	88	75	76	71	90	0.18	-0.14	QSF037	16	03	Bravo
68	60	52	78	81	81	75	53	60	58	72	0.01	-0.02	77	88	59	104	92	92	86	75	76	70	89	0.17	-0.26	75	85	57	97	89	90	83	72	73	67	87	0.12	-0.11	82	90	64	106	96	97	91	77	79	73	93	0.17	-0.26	QSF037	16	03	Charlie
69	68	55	84	82	83	76	62	62	60	75	0.03	-0.03	76	88	60	103	89	90	85	74	76	69	88	0.28	-0.21	72	84	56	97	85	86	80	69	72	65	84	0.10	-0.10	79	90	63	105	93	93	89	76	78	72	91	0.28	-0.21	QSF037	16	03	Delta
69	72	53	92	83	83	75	62	62	59	77	0.06	-0.03	82	95	65	108	94	95	91	81	82	76	95	0.50	-0.33	75	88	59	102	90	90	83	73	75	68	87	0.09	-0.22	84	96	67	110	97	97	93	82	84	77	96	0.50	-0.33	QSF037	16	03	Echo
66	60	50	75	79	79	71	53	56	54	70	0.01	-0.01	78	90	61	106	90	91	87	76	78	71	90	0.35	-0.28	75	85	57	99	88	89	83	72	74	67	87	0.14	-0.11	81	91	63	106	94	94	90	77	79	73	92	0.35	-0.28	QSF037	16	03	Foxtrot
64	62	49	76	78	79	69	53	55	53	70	0.01	0.00	69	84	52	100	87	87	78	69	71	63	82	0.16	-0.11	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	74	85	58	101	91	91	83	71	73	66	86	0.16	-0.11	QSF037	16	03	Golf
60	59	43	69	73	74	65	51	52	49	66	0.01	0.00	83	96	67	109	96	96	93	84	84	78	95	0.68	-0.27	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	84	97	67	109	97	97	94	84	85	78	96	0.68	-0.27	QSF037	16	03	Hotel
64	64	48	79	77	78	69	55	56	53	70	0.02	-0.01	70	85	53	100	88	88	77	70	73	65	83	0.15	-0.14	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	74	86	58	101	91	91	81	71	73	66	86	0.15	-0.14	QSF037	16	03	India
62	61	46	76	75	76	67	52	53	50	69	0.01	-0.01	67	83	50	100	83	84	75	68	71	62	81	0.11	-0.15	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	73	84	55	102	87	87	79	69	72	64	85	0.11	-0.15	QSF037	16	03	Juliet
71	65	55	88	84	85	78	58	62	60	76	0.05	0.03	80	90	63	107	93	94	89	77	79	72	92	0.37	-0.35	76	85	58	98	90	91	84	71	73	67	87	0.09	-0.15	83	92	66	108	97	97	93	78	80	75	94	0.37	-0.35	QSF037	16	03	Kilo

Table 6-3 Statistics of recordings at monitors

Monitor	Number 1st Thumps[1]	Number 2nd Thumps	Section: Statistic\Metric	Ambient													1st Thump													2nd Thump													3 s Window												
				PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF	PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF	PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF	PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP	MxPSF	MnPSF
Alpha	50	30	Max	76	74	75	93	90	91	84	66	69	67	78	0.09	0.01	89	97	74	109	99	99	98	87	86	83	98	0.53	-0.01	81	93	65	108	94	94	90	80	81	76	93	0.29	-0.01	90	98	74	111	101	101	99	87	87	83	100	0.53	-0.01
			Min	56	51	39	62	68	69	60	45	46	44	61	0.00	-0.03	57	60	40	77	70	71	62	50	51	47	65	0.01	-0.49	58	60	41	74	71	72	63	51	52	49	66	0.01	-0.33	64	65	47	79	77	78	69	56	56	53	72	0.01	-0.49
			Average	63	61	47	73	76	76	68	55	55	53	69	0.01	-0.01	70	81	54	97	84	84	78	68	70	63	81	0.16	-0.14	66	76	50	93	80	81	73	63	65	59	77	0.07	-0.08	75	83	59	99	88	89	82	71	73	67	85	0.16	-0.14
Bravo	48	23	Max	78	74	67	91	91	92	86	68	70	69	79	0.04	0.00	90	96	75	107	100	101	98	88	87	84	99	0.53	-0.01	77	92	61	108	92	92	86	78	80	73	90	0.31	-0.01	90	97	76	111	101	101	99	88	88	85	99	0.53	-0.01
			Min	53	51	36	59	66	67	57	43	44	42	60	0.00	-0.08	58	64	42	81	74	75	64	56	57	53	71	0.01	-0.42	59	69	42	84	73	74	65	56	57	52	70	0.02	-0.33	65	70	48	85	81	82	72	63	64	59	77	0.02	-0.42
			Average	63	63	47	74	76	77	69	55	56	53	70	0.01	-0.01	71	81	55	96	84	85	79	69	70	64	82	0.15	-0.12	69	80	53	96	83	84	77	68	69	63	81	0.10	-0.10	76	84	60	98	89	90	84	72	73	68	86	0.15	-0.12
Charlie	51	25	Max	75	76	60	93	88	88	85	71	70	68	82	0.06	0.00	81	93	64	109	93	93	91	80	81	75	92	0.39	-0.01	78	87	61	103	91	91	87	74	75	70	88	0.18	-0.03	84	94	67	109	97	97	94	81	82	76	95	0.39	-0.01
			Min	56	51	39	62	69	69	60	44	47	44	61	-0.01	-0.06	57	57	40	77	70	71	61	49	49	46	64	0.01	-0.46	62	66	45	88	75	76	67	55	57	52	71	0.02	-0.18	65	70	48	85	78	79	71	59	59	55	74	0.01	-0.46
			Average	62	60	45	73	74	75	67	53	54	51	68	0.01	-0.01	69	80	53	97	83	83	77	67	69	63	81	0.14	-0.14	67	78	51	96	81	82	75	65	67	61	79	0.07	-0.09	74	83	58	100	88	88	82	71	72	67	85	0.14	-0.14
Delta	47	27	Max	89	76	76	92	100	100	99	76	85	82	89	0.05	0.01	89	99	75	109	100	101	99	88	87	82	100	0.64	-0.01	74	88	59	100	86	86	83	74	75	69	85	0.14	-0.01	95	99	82	109	106	106	105	88	90	87	100	0.64	-0.01
			Min	54	50	38	59	67	68	58	43	44	43	60	0.00	-0.04	57	56	40	74	71	72	62	49	50	47	64	0.01	-0.42	58	61	41	80	72	73	63	51	54	49	68	0.01	-0.15	66	66	48	77	79	80	72	58	58	55	73	0.01	-0.42
			Average	63	62	47	74	76	77	69	55	55	53	69	0.01	-0.01	69	79	53	97	83	83	76	66	69	62	80	0.15	-0.14	67	75	50	93	80	80	73	63	65	59	77	0.06	-0.08	74	82	58	99	88	88	82	70	72	66	84	0.15	-0.14
Echo	47	28	Max	70	70	55	94	83	84	78	63	63	61	76	0.03	0.00	85	94	70	109	95	96	93	83	83	78	95	0.47	-0.02	81	93	64	107	93	93	90	80	81	75	93	0.35	-0.01	86	95	70	110	97	98	95	84	84	79	97	0.47	-0.02
			Min	52	52	36	60	66	67	56	44	45	42	60	-0.01	-0.11	56	61	40	84	70	71	62	50	53	48	66	0.02	-0.46	55	60	38	77	68	68	59	48	51	46	65	0.01	-0.30	63	66	45	86	76	77	68	57	59	54	72	0.02	-0.46
			Average	61	60	45	72	74	75	67	53	54	51	68	0.01	-0.01	70	81	54	97	83	84	78	68	70	63	81	0.15	-0.14	67	77	51	93	81	82	75	64	66	60	78	0.08	-0.09	74	83	58	99	88	88	82	71	72	67	85	0.15	-0.14
Foxtrot	48	28	Max	74	71	60	92	87	88	83	68	68	67	79	0.05	0.00	85	98	70	108	99	100	94	85	86	80	98	0.59	-0.01	77	87	61	102	90	91	86	76	76	71	88	0.15	-0.01	87	98	71	109	100	101	95	85	86	80	99	0.59	-0.01
			Min	54	52	37	63	67	68	58	44	45	43	60	0.00	-0.06	56	64	39	80	70	70	61	51	55	48	68	0.02	-0.40	54	62	37	78	68	68	59	51	51	48	67	0.01	-0.17	62	68	45	83	76	77	68	58	59	54	73	0.02	-0.40
			Average	62	62	46	73	75	76	68	54	54	52	69	0.01	-0.01	70	82	54	98	84	84	78	69	71	64	82	0.18	-0.15	67	78	50	94	81	81	74	64	67	60	78	0.07	-0.09	75	84	58	100	88	89	83	72	73	67	86	0.18	-0.15
Golf	42	24	Max	79	74	66	95	90	90	90	66	76	74	80	0.06	0.01	85	96	70	111	97	97	94	84	84	79	96	0.57	-0.01	86	96	71	109	97	97	95	85	85	81	97	0.37	-0.01	89	99	73	112	100	100	97	87	88	83	100	0.57	-0.01
			Min	53	53	37	62	67	67	58	46	49	46	62	0.00	-0.07	55	61	39	78	69	70	61	50	52	47	65	0.01	-0.49	62	67	47	78	76	77	69	56	57	53	72	0.01	-0.38	65	69	49	84	79	80	72	58	60	57	74	0.02	-0.49
			Average	63	62	47	73	76	77	69	55	56	53	69	0.01	-0.01	71	81	55	97	84	85	79	69	70	64	82	0.17	-0.15	70	80	55	95	84	84	78	68	69	64	81	0.10	-0.12	76	84	60	99	89	90	84	72	73	68	86	0.17	-0.15
Hotel	35	17	Max	73	71	62	90	87	88	81	65	66	65	77	0.07	0.02	86	96	69	109	97	97	95	85	84	80	97	0.68	-0.01	77	90	61	101	90	91	86	77	77	71	89	0.20	-0.02	87	97	70	110	99	99	96	85	85	80	98	0.68	-0.01
			Min	54	55	36	61	67	67	58	47	47	44	62	0.00	-0.05	59	59	42	74	72	73	63	52	53	49	66	0.01	-0.47	60	60	43	83	73	74	65	53	54	50	68	0.01	-0.14	67	67	50	81	80	81	72	59	60	57	74	0.01	-0.47
			Average	64	61	47	73	76	77	69	54	55	53	69	0.01	0.00	71	80	55	96	84	85	79	68	69	64	82	0.15	-0.14	68	77	52	95	82	82	76	65	67	61	78	0.08	-0.09	76	82	59	98	89	89	84	71	72	67	85	0.15	-0.14
India	28	16	Max	70	72	56	92	83	84	78	63	63	61	77	0.06	0.00	86	99	72	109	101	101	95	86	87	81	98	0.60	-0.01	75	89	59	102	91	91	83	74	76	69	87	0.16	-0.01	87	100	72	110	102	102	96	87	87	81	99	0.60	-0.01
			Min	53	52	36	61	66	67	57	46	46	43	60	0.00	-0.05	53	61	37	78	69	70	59	53	53	48	67	0.01	-0.48	59	66	42	78	73	73	64	53	59	49	71	0.01	-0.22	63	69	45	83	79	80	68	60	61	57	75	0.01	-0.48
			Average	62	61	46	75	75	76	68	54	54	52	69	0.01	-0.01	70	81	54	98	84	84	78	68	71	64	82	0.18	-0.15	66	77	49	93	80	80	73	63	66	59	77	0.06	-0.09	74	84	58	100	88	89	82	71	73	67	85	0.18	-0.15
Juliet	30	15	Max	74	74	60	88	87	87	84	70	69	68	79	0.04	0.00	85	96	69	109	97	97	94	83	84	78	96	0.60	0.00	74	89	58	101	90	90	85	75	76	69	87	0.18	-0.02	86	97	70	110	98	99	95	84	85	79	98	0.60	0.00
			Min	53	54	37	61	68	68	59	46	46	44	61	0.00	-0.05	55	58	38	72	69	69	61	49	49	46	63	0.01	-0.38	60	63	44	85	76	76	67	55	56	52	70	0.02	-0.16	63	65	46	78	76	77	69	57	56	54	71	0.01	-0.38
			Average	64	63	49	74	77	78	70	56	57	55	70	0.01	-0.01	71	81	55	98	84	85	79	68	70	64	82	0.15	-0.13	68	77	52	95	82	83	76	66	67	61	79	0.07	-0.09	75	83	59	99	89	89	83	72	73	68	85	0.15	-0.13
Kilo	46	25	Max	71	69	55	88	84	85	78	63	63	61	77	0.05	0.03	84	94	68	107	94	95	92	83	82	78	95	0.38	0.00	78	91	62	106	92	92	87	78	79	73	91	0.27	-0.01	85	95	69	108	97	97	94	84	84	79	97	0.38	-0.01
			Min	52	51	35	59	65	66	56	42	44	41	59	0.00	-0.02	56	60	39	71	69	70	61	50	54	47	67	0.01	-0.36	55	63	38	79	68	69	60	50	53	46	66	0.01	-0.24	63	67	46	81	76	77	69	57	59	54	73	0.01	-0.36
			Average	62	60	46	72	75	76	67	53	54	51	68	0.01	0.00	70	80	53	96	83	83	77	67	69	63	81	0.14	-0.13	67	76	50	93	80	81	74	64	66	60	77	0.06	-0.09	74	83	58	99	88	88	82	71	72	67	85	0.14	-0.13
Lima	4	2	Max	65	67	48	85	78	78	70	56	57	54	72	0.02	0.00	83	92	68	108	94	94	92	81	80	77	93	0.41	-0.16	72	80	57	98	85	85	81	70	70	66	82	0.09	-0.08	83	93	68	108	95	95	93	81	81	77	93	0.41	-0.16
			Min	60	56	44	70	73	74	64	50	50	48	65	0.01	-0.03	67	84	51	100	83	84	75	68	72	63	81	0.18	-0.39	68	76	52	94	82	82	76	66	66	61	78	0.04	-0.11	71	85	54	101	86	87	78	69	73	64	84	0.18	-0.39
			Average	63	62	46	78	75	76	67	53	54	51	69	0.01	-0.02	77	89	61	103	90	91	86	77	78	72	89	0.27	-0.24	70	78	54	96	83	83	78	68	68	64	80	0.07	-0.10	79	90	62	104	92	92	88	77	79	72	90	0.27	-0.24

6.1.1.1 Development of Acoustic Metrics at Sensor Locations Lessons Learned

A result of this test showed the impact of ambient noise when recording thumps. This occurs because the spectrum of the ambient is significantly broader than the spectrum of a thump and has higher levels in portions of the spectrum that influence the calculation of the various metrics. Of the 476 events, there were 387 that had a PL difference of 1dB above the ambient, 314 that were 3 dB above the ambient, 266 that were 5 dB above the ambient, and 143 that were 10 dB or more above the ambient. Higher ambient levels can be expected in urban and commercial/manufacturing areas in the communities that are to be flown over by the X-59. Future work should be done to adequately address how to separate the ambient from the recorded thump levels. Doing so will also assist in the interpolation of data measured across a larger area where it is expected that the ambient noise will be significantly different at monitor locations, and will more accurately determine the recorded levels attributable to the thump alone.

Another lesson learned, in terms of the data processing of the monitor recordings, is that a summary statistics program should have been run on the data set in the field. A simple report of the peak overpressure as part of a summary statistic program for all recordings would have shown that there was an event to consider at BRAVO based on the fact that a 0.2 psf peak overpressure had been recorded. Because a record of the initial look at the recordings was not consulted for the final analysis, this event was missed. Missing the thump recorded at Bravo during the first supersonic pass of the campaign occurred because the data review and field crew reported there should not have been any events to find in the records. Subsequently, the first pass's recordings were not analyzed. A simple report of the peak overpressure for all recordings would have shown that there was an event to consider at BRAVO.

6.1.2 Georeferencing Participant Locations

This section describes the data processing and GIS analysis used to obtain each of the participant locations at the time of each sonic thump event. Due to the variation in metric levels across the footprint from the F18 Low Boom Dive maneuver utilized at the QSF18 tests, and because the intent of this analysis was to test procedures and protocols, it was important to locate participants at the time of the thump events as accurately as possible. During the development and execution of the analysis procedures using the QSF18 analysis, notes and lessons learned were documented and are listed in section 6.1.2.1.

Successful automatic geolocation of survey respondents was achieved for 8462 out of 11869 (71%) single-event surveys, which corresponds specifically to the number of single-event responses in which respondent indicated that the map showed correctly, and the data included a valid set of latitude / longitude coordinates. There were two success criteria for automatic geolocation. First, the respondent indicated the map displayed their location correctly. Second, a valid set of latitude and longitude coordinates was recorded in the survey. Note that the 71% success rate is does not include responses for which locations were subsequently determined through manual geocoding of addresses or locations inferred from daily summary responses.

Participants' home and work addresses had to be converted to latitude and longitude. There were a number of inconsistencies in the addresses that needed to be cleaned up prior to geocoding. Specific examples include:

- The wrong zip code for the address.
- The state included in the city/town box, such as “Galveston TX”, and then the state included again in the state box.
- Spaces omitted in the town name, such as “Texascity”.
- Everything included in the address box, such as “123 maple st galvestion tx 77550” and then a separate city, state, and zip field also filled in.

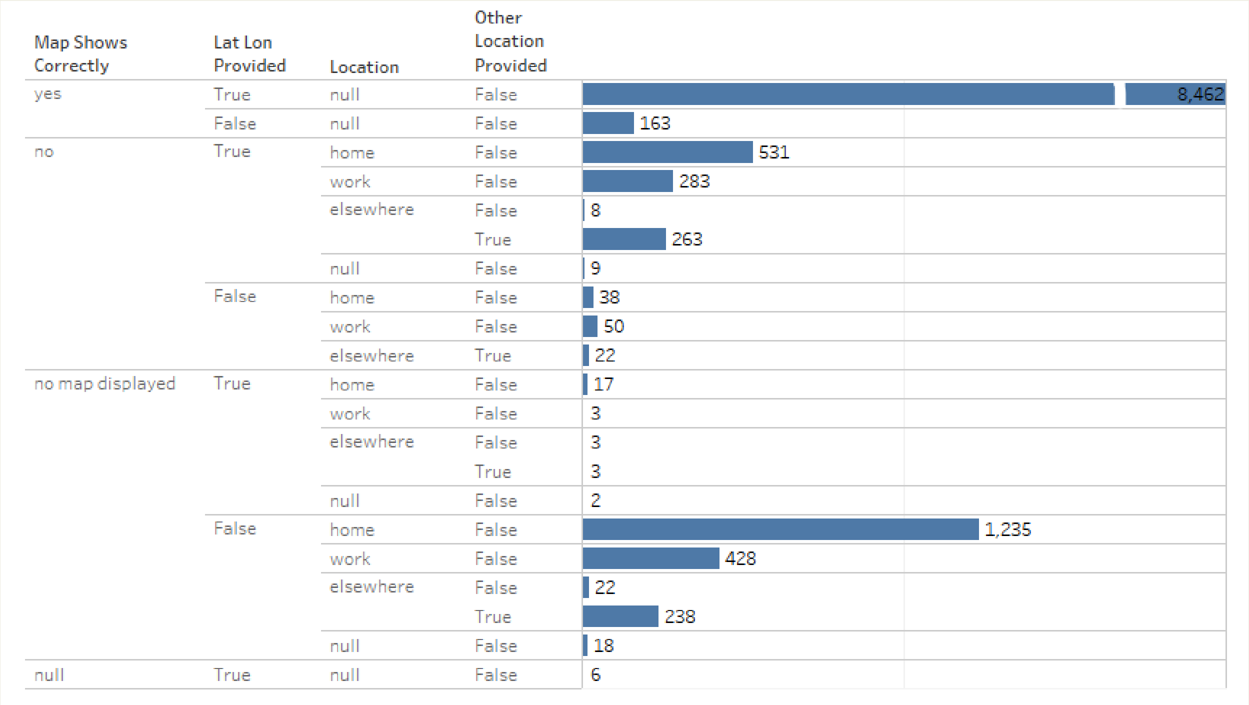
More automated checking of addresses at the time of input is recommended. A possible approach leverages websites that conduct some sort of address verification. If a survey signup page was built by a web developer they could use an application programming interface (API) to verify addresses. One option is “SmartyStreets.com”, but UPS also has one. Address verification process details depend on how participants apply (e.g. over phone, online, both). A less automated way would be to test addresses of new participants on a daily basis using something like SmartyStreets “List” based checking. If any problems with addresses come up the participant could be contacted or deleted prior to the beginning.

In a number of cases Google Maps was used to search for an address to confirm the change. Updates to the original addresses were only made when there was reasonable confidence that the change to the address was correct. Out of the 500 participants, updates to both home and work addresses were made for 34 participants, updates to only the home address were made for 22 participants, and updates to only the work address were made for 25 participants. There were 13 home addresses and 8 work addresses that were not usable (e.g. no address provided). Once the address data had been cleaned the ArcGIS World Geocoding Service was used through Esri’s ArcGIS Pro Software to geocode the addresses.

In addition to home and work addresses, other locations were provided for many single event surveys when the participant wasn’t at or near home. There were 396 such unique locations in the dataset. Unfortunately these addresses are almost completely freeform and many were difficult or impossible to locate. Of the 396 unique locations, 299 could be located while 97 could not be located. A significant number of the locations that ultimately could be located required research and cleaning, for example “l.a morgan school 36th street” was updated to the official address of the LA Morgan Elementary School which is “1410 37th st”. Of the 97 that couldn’t be located, some could be roughly located but not with enough precision to assign a reasonable latitude and longitude (e.g. midtown Galveston) while others were completely unusable (e.g. driving to work).

Table 6-4 illustrates the occurrences of combinations of four data elements related to locating the participants for the single event records. The fields “lat/lon provided” and “Other Location Provided” contain derived Boolean values based on the existence of latitude/longitude coordinates or “somewhere else” respectively. Over 71 percent of the records fall into the “Map Shows Correctly” = True and “lat/lon provided” = True category. Over 15 percent of the records indicate the location to be home and over 6 percent of the records indicate the location to be work.

Table 6-4 Combinations of elements related to locating participants for single event records



Map Shows Correctly	Lat Lon Provided	Location	Other Location Provided	
yes	True	null	False	8,462
	False	null	False	163
no	True	home	False	531
		work	False	283
		elsewhere	False	8
			True	263
		null	False	9
	False	home	False	38
		work	False	50
		elsewhere	True	22
no map displayed	True	home	False	17
		work	False	3
		elsewhere	False	3
			True	3
		null	False	2
	False	home	False	1,235
		work	False	428
		elsewhere	False	22
			True	238
		null	False	18
null	True	null	False	6

Once all addresses were manually cleaned and located where possible, the next step was to bring all of the data elements together for further processing and analysis. A Python script was developed to do this. The script reads all necessary source data from Excel and text files (i.e. geocoded addresses) into in-memory data structures. This takes the data from Excel, which in some cases is organized in a way that makes it very difficult to work with, and puts it in a format that makes it easy to examine, debug, and analyze.

The next step was to examine each participant's location at the time of each sonic thump event. For each of the 26,000 unique thump/participant combinations, the following information was passed into the "process participant for thump" function:

- The information about the particular thump event (date, time).
- The information about the participant, including the participant's geocoded home and work addresses.
- Any single event survey records for that participant on the day of the thump event being processed, including "other locations".
- The daily summary, if it exists, for that participant on the day of the thump event being processed.
- Geocoded "other locations" for lookup if necessary.

With this data, the following logic was used to locate a participant at the time of the thump. A unique "location assignment code" was assigned to each location determination method.

1. First, determine if the participant recorded any single events for the day.

If there were single events recorded, then the closest single event within 20 minutes of the thump time was associated with the thump (see lessons learned #6 in section 6.1.2.1) and the location was determined using the following logic:

- If that single event indicated that the map showed correctly and the latitude and longitude were non-zero, then the recorded latitude and longitude was used (location assignment code 1).
- Otherwise, if the location indicated was home or work then the coordinates for the respective geocoded address were used (location assignment codes 2 and 3).
- Otherwise, if the location indicated was “other” and the address was successfully geocoded then the coordinate for the geocoded “other” address were used (location assignment code 4).
- Otherwise no location could be determined (location assignment code 5).

If there were single events but none within the 40 minute time window, 20 minutes on either side of the thump, then:

- If the daily summary indicated the participant was at home or at work at the time of the thump then the corresponding location was used (location assignment codes 6 and 7).
- If the daily summary didn’t indicate they were at work or home at the time of the thump then no location could be assigned (location assignment code 8).
- If there was no daily summary then no location could be determined (location assignment code 9).

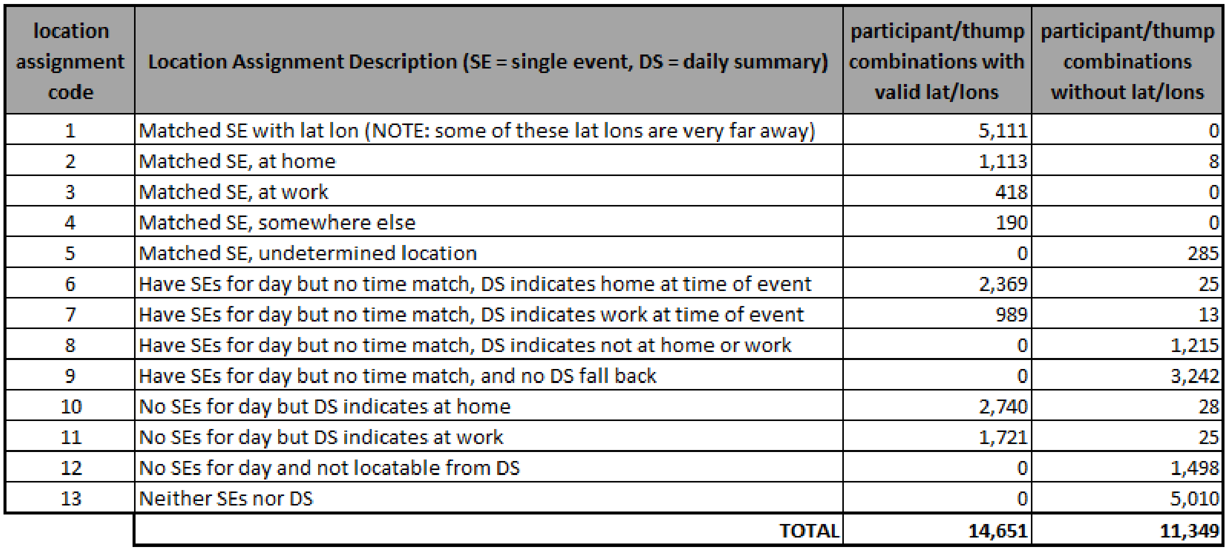
2. If the participant had no single events on the day of the thump and there was a daily summary then the logic used is similar to that when there are single events in the day but none can be associated with the thump event:
    - If the daily summary indicated the participant was at home or at work at the time of the thump then the corresponding location was used (location assignment codes 10 and 11).
    - If the daily summary didn’t indicate they were at work or home at the time of the thump then no location could be assigned (location assignment code 12).

3. If there were no single events and no daily summary on the day of the thump event then no location could be assigned for the particular thump/participant combination (location assignment code 13)

This processing script created a file named participant_locations.txt with 26,000 rows, one for each thump/participant combination. The main elements of the file include the thump_id, the participant id, the latitude/longitude coordinates of the person at the time of the thump (when possible), a determination for whether it was heard or not, and the location assignment code.

Table 6-5 describes the different location assignment codes and shows the occurrences of each along with information on results with and without latitude/longitude coordinates. It should be noted that each location assignment code either does or does not have latitude/longitude coordinates with relatively few exceptions. For example, location assignment code 2 mostly results in valid latitude/longitude coordinates with the exception of 8 participant/thump combinations, something that occurs due to one or more addresses which couldn’t be located.

Table 6-5 Location types and frequencies



location assignment code	Location Assignment Description (SE = single event, DS = daily summary)	participant/thump combinations with valid lat/lons	participant/thump combinations without lat/lons
1	Matched SE with lat lon (NOTE: some of these lat lons are very far away)	5,111	0
2	Matched SE, at home	1,113	8
3	Matched SE, at work	418	0
4	Matched SE, somewhere else	190	0
5	Matched SE, undetermined location	0	285
6	Have SEs for day but no time match, DS indicates home at time of event	2,369	25
7	Have SEs for day but no time match, DS indicates work at time of event	989	13
8	Have SEs for day but no time match, DS indicates not at home or work	0	1,215
9	Have SEs for day but no time match, and no DS fall back	0	3,242
10	No SEs for day but DS indicates at home	2,740	28
11	No SEs for day but DS indicates at work	1,721	25
12	No SEs for day and not locatable from DS	0	1,498
13	Neither SEs nor DS	0	5,010
	TOTAL	**14,651**	**11,349**

6.1.2.1 Georeferencing Participant Locations Lessons Learned

1. The same data shouldn’t appear in different places when handing off study results to multiple teams and researchers. It would be clearer to have one file with the final information for the participants.
2. Cleaning addresses is manually intensive and potentially error-prone. More automated checking of addresses at the time of input is highly suggested, especially if significantly more participants are to be used in future studies.
3. All work and home addresses should be geocoded prior to acceptance into study. Presumably no one should be missing a home address. If a work address is missing there should be some indicator as to why (e.g. retired, unemployed) to differentiate cases where the data was not provided.
4. Odd combinations of location-related variables sometimes seem to contradict each other and make it difficult to decide how to handle them. An example of this would be when the value for “map shows correctly” = ‘no’ but latitude/longitude coordinates and an address were included. In such a case the latitude/longitude coordinates cannot be considered valid, and the data should not be included in the data set.
5. Organizing the data in a database using a standard relational structure[****] would make it much easier to work with and make analyses less error-prone. Additionally, a relational data approach

---

[****] Relational structure refers to the way data is typically stored in databases, for example, there is a one to many relationship between survey participants and single events recorded (where single events recorded could be 0).

would remove the difficulty associated with having a code book with separate labels and variables sheets.

6. Improvements to the structure of the data entry method should also be investigated. An example is replacement of the text box with a movable icon on a map that would show address or latituge/longitude, possibly mitigating the effort to manually clean data.
7. Each thump is processed individually for each participant. If thump events are close enough to each other then the same single event response could possibly be associated with two different thump events.

6.1.3 Determination of Metrics at Participant Locations

This section provides a concise description of the methods used to estimate metrics at participant locations, based primarily on three sources of input data:

1. measured metrics at monitor locations,
2. participant locations at boom times, based on single event (SE) / daily summary (DS) response data, and
3. results of PCBoom / PCBurg footprint modeling using version 3 (received 29 November 2018) as-flown trajectory data and measured atmospheric profiles.[††††]

The measured metrics dataset contains levels calculated using 650 ms durations as well as 3000 ms durations along with corresponding 650 ms ambient levels. For this analysis, both sets of metrics were used. Furthermore, minimum thresholds were placed on metrics relative to ambient levels: measured levels must have been at least 1 dB, 3 dB, or 5 dB above local ambient to be included in the analysis. Together, duration and ambient threshold criteria resulted in six complete sets of metrics at participant locations. For each event, the set of monitors whose metrics were above the ambient thresholds were considered the “usable” monitors for that event/metric combination. At least one usable monitor was required to determine metrics at participant locations. Note, however, that ambient levels from 650 ms windows are not directly comparable with the longer duration 3000 ms metrics. The metrics from the 3000 ms windowed data are relatively insensitive to the ambient criterion. This is because ambient noise is relatively steady over time frames on the order of seconds, while sonic booms and sonic thumps are short duration transients, thus a longer time window incorporates an increased amount of ambient signal roughly proportional to the window length, while a longer time window does not incorporate a significantly increased amount of sonic boom or sonic thump signal. For example, in the 1 dB / 3000 ms dataset, all monitor signals passed the ambient check except one monitor on one event.

Survey response data in the form of single event response and daily summaries were used to determine participant locations and correlate those locations with specific booms. Geolocation analysis procedures are summarized in Table 6-5 with the specific details explained in more detail in the preceding section. Due to the varied nature of these data, several possible location assignment codes were defined to track

---

[††††] Updated trajectory data and atmospheric profiles were received on 25 February 2019, after metric databases were distributed for dose-response analysis. A comparison of noise footprints modeled with version 4 input data is included at the end of this section.

the provenance of location data throughout the analysis. Metrics were determined for all locatable participant responses and further analysis could be conducted using, for example, only single event response data. Note that not all addresses were locatable using geocoding; some addresses were incomplete or otherwise indeterminate. In those cases no latitude / longitude coordinates were available to determine metrics at participant locations.

Metrics calculated from modeled ground signatures in PCBoom 6.70 / PCBurg 4 were used to supplement measured metrics. For each event, PCBoom programs FOBOOM 6.70, PCBFOOT 6.66, and WCON were run using v3 as-flown trajectory data and v3 measured atmospheric profiles to model dive footprints. PCBoom input files were assembled into a deliverable archive.

ASCII output files (in .pdx format) from WCON describing levels within modeled footprints were used to investigate how regions inside and outside the modeled footprint correlated with measured signatures at monitor locations[‡‡‡‡]. It was observed in some cases that measured signals outside the predicted footprint had characteristics like those of ground boom signatures and thus appeared to be within the actual realized footprint. To account for uncertainty in the actual locations of footprint edges, margins around the modeled footprints were developed through comparison with boom quality ratings. As part of the metric calculation process, each measured signature was examined and assigned a rating of 1-4 indicating that the signature:

1. Could be clearly attributed to the aircraft and originated during the supersonic portion of the trajectory ("good")
2. Appeared to have overlapping booms ("overlap")
3. Had characteristics of both a boom and other features ("nasty")
4. Appeared to be a rumble

Since quality rating 1 included signatures which did not strictly appear to be N-wave booms, an additional criterion was applied in footprint margin determination: maximum overpressure of at least 0.1 psf. Boom quality ratings for all measurements are shown in Figure 6-12 by event number. Considering quality and overpressure, inspection of the relationship between notional downtrack margins and the number of additional signals with rating 1 enclosed by notional margins indicated that a downtrack margin of 2.9 nmi provided a compromise between monitor inclusivity and margin size. Figure 6-13 shows that relationship quantitatively. For a margin of 2.9 nmi, 30 measured signals across the events considered are included in the footprint margins. The margins would need to be extended by approximately 1 nmi to add another measured signal rated 1, and doing so would incorporate "rumble" measurements. Booms 48-52 (flight day 9) were characterized by high overpressures at cutoff and appeared to have fundamentally different margins; in those cases, a downtrack margin of 4.2 nmi was used. A similar approach was used for lateral margins: monitors within 0.5 nmi of modeled footprint edges typically had quality ratings 1 and met the overpressure criterion. An example of the boundaries of a modeled footprint and its margins is shown in

---

[‡‡‡‡] Footprint cutoff margin determination was based on FOBoom propagation and not PCBurg results, because only the locus of ground boom locations / cutoff locations was relevant for this purpose, not signature amplitudes or waveforms.

Figure 6-14. For visual reference, an outline of the coastline is drawn in black together with an offshore area comprising oil rig locations and requested airspace (large trapezoid).

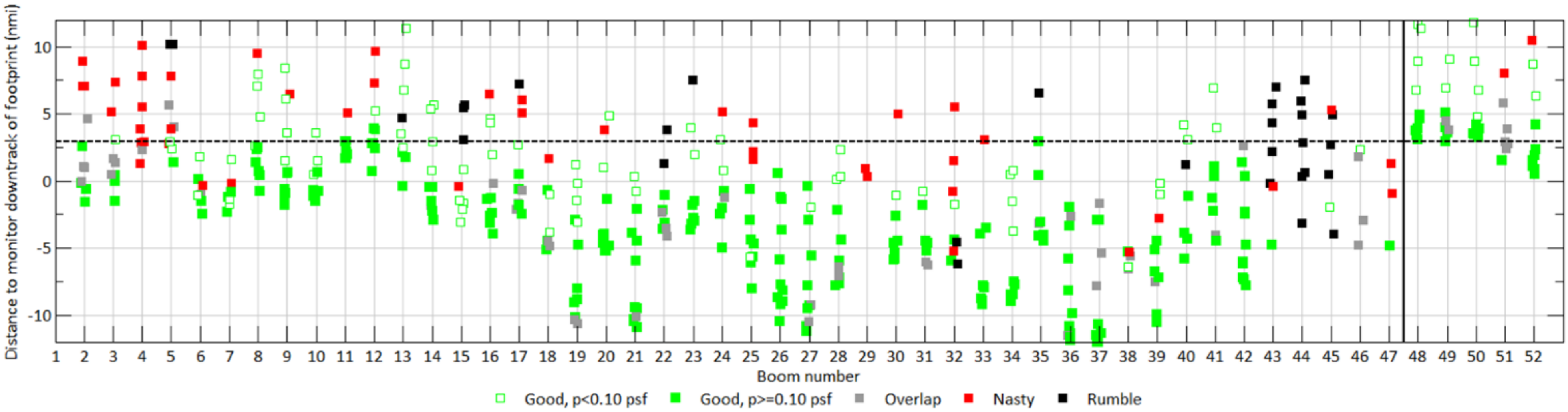


Figure 6-12 Comparison of boom quality rating with distance from footprint edge

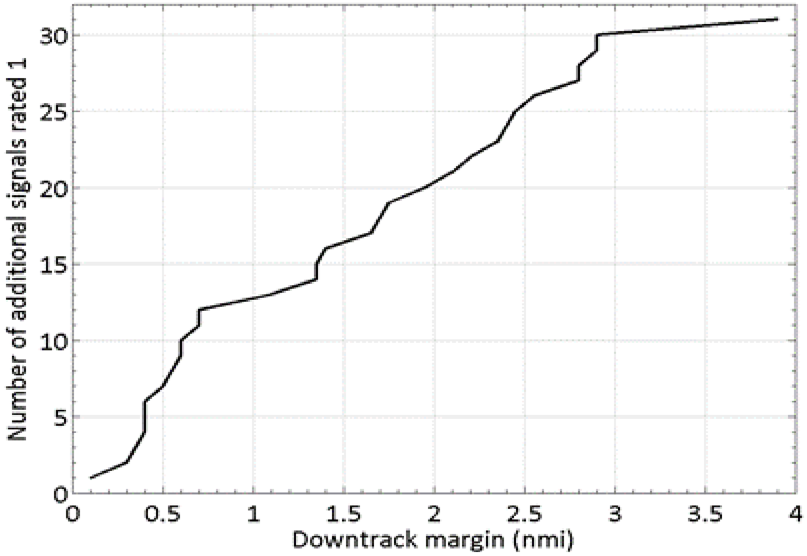


Figure 6-13 Relationship between monitor signals rated 1 and downtrack distances beyond modeled cutoff

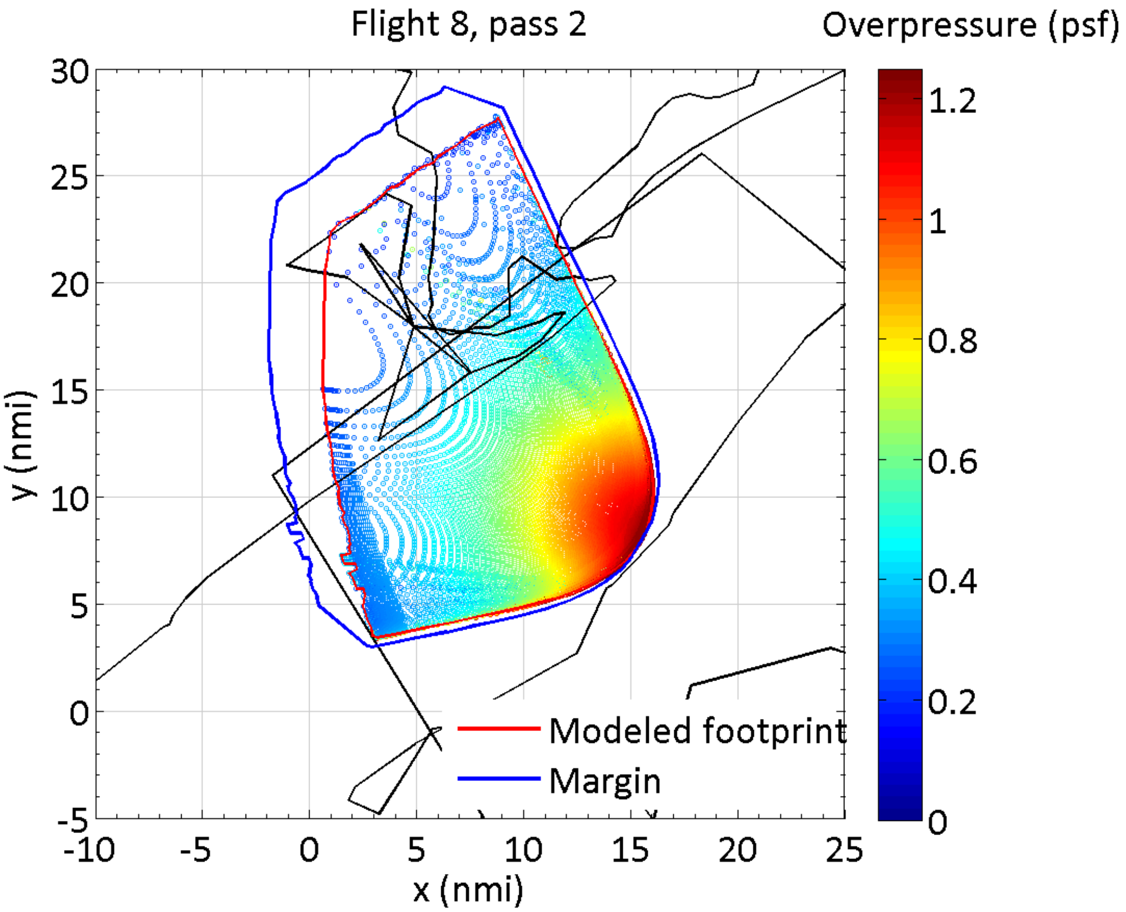


Figure 6-14 Example of borders of modeled footprint and its margins

For the purpose of determining metrics at subject locations, all survey participant locations within the recruitment area or near a monitor were included. Whether participant locations were within the footprint and also within its margins was tracked in the noise database using Boolean values to facilitate additional analyses.

The complete set of participant locations included several points in Houston, the Dallas area, and some as far away as Ecuador. Those points were not included in determination of metrics. Rather, the recruitment area (comprising four quadrilaterals A-D, Figure 6-15) was used as a criterion for deciding if metrics should be determined at a specific location. The recruitment area had four monitor locations near its borders: BRAVO at the southwestern edge on Galveston Island, and HOTEL, INDIA, and JULIET along the northwestern edge in Hitchcock, La Marque, and Texas City, respectively. To include participant locations that were outside the recruitment area but relatively close to one of these four monitor locations, metrics were also determined at participant locations within 2 nmi of these four monitors regardless of whether the locations were inside the recruitment area. That dimension was selected to include clusters of participants outside the northwest border of the recruitment area without exceeding typical monitor separation distances.

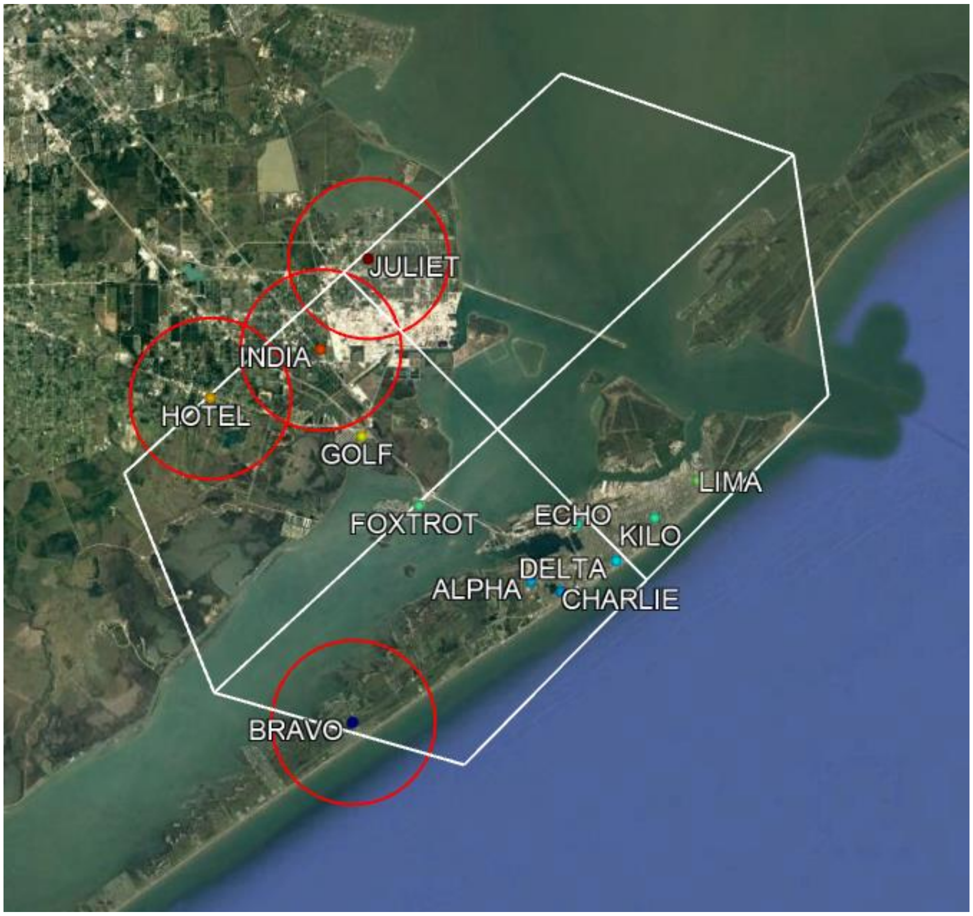


Figure 6-15 Recruitment area (white quadrilaterals) and monitor margins (red circles with radius 2 nmi) for monitors near borders

To account for the molecular vibrational relaxation effects on loudness metrics in propagation due to the molar concentration of water vapor in the atmosphere, the PCBoom Burgers' equation module PCBurg 4 was used to supplement FOBoom/PCBFoot footprint modeling. PCBurg is a much more computationally intensive tool than FOBoom 6.70, requiring up to several minutes to model propagation for a single ray[§§§§]. As such, relatively coarse meshes comprising a few hundred points per footprint were constructed, as illustrated in Figure 6-16. An outline of this process is as follows:

1. Run FOBoom with keyword BURGERS to generate the necessary inputs to PCBurg (.age and .ssg files)
2. Run PCBFoot with run option 7 to add a full summary to the ASCII output file (.asc file)
3. Use WCON to identify the points that enclose the low peak overpressure portion of the footprint (generally where undertrack overpressure is less than around 0.5 psf), plus a few nmi offshore.
4. Construct a square grid of points with 1 nmi spacing, and remove all points not in the region specified in the previous step.
5. Using PCBFoot .asc files to generate a list of rays containing PCBoom referenced (x, y) coordinates of ground intersection points, (T_{ac}, ϕ) for each ray, and the PCBoom identifier boom type (1 is carpet boom).
6. For each square grid point, find the closest ray intersection point which has PCBoom identifier boom type 1; add the PCBoom referenced (x, y) ground intersection coordinates to the list of Burgers mesh points and add a line to the PCBurg batch file for the corresponding T_{ac}, ϕ.
7. Run PCBurg for each grid point identified in step 6 to calculate ground metrics with molecular relaxation effects. In practice, these runs were distributed across many machines and run concurrently in batch mode at a sampling frequency of 51,200 Hz.
8. The output of PCBurg is a signature file for each grid point. Ground metrics are included in the header information for the second signature (indicated by header phrase "Refl = 1.9" which also indicates the free-field boom pressure was multipled by a ground reflection factor of 1.9). Parse ground metrics from all signature files and correlate with (x, y) locations from step 6.

[§§§§] See Lonzaga, J., "Recent Enhancements to NASA's PCBoom Sonic Boom Propagation Code", https://doi.org/10.2514/6.2019-3386, for a description of a faster implementation of Burgers equation propagation modeling, compared with PCBurg 4.

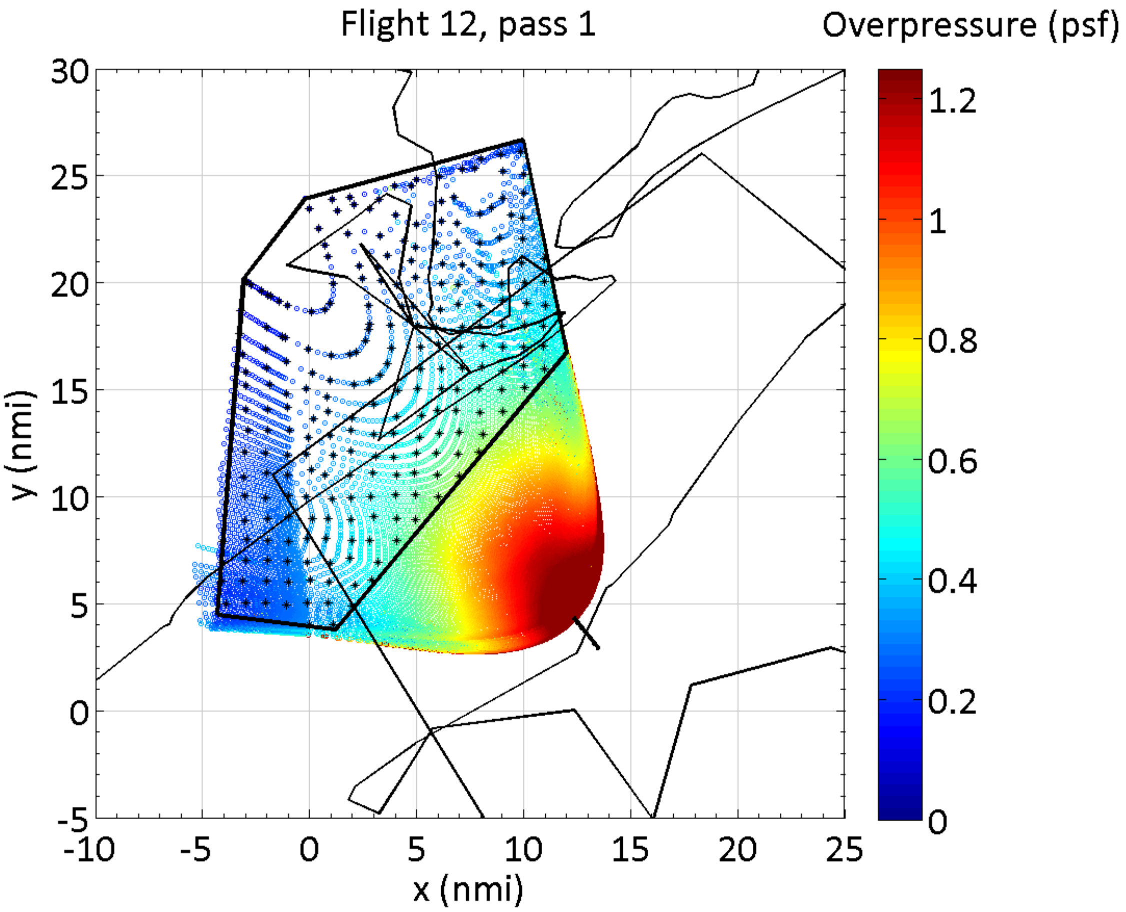


Figure 6-16 Example Burgers mesh (black * points) overlaid on footprint from WCON (colored circles)

Offsets between measured metrics and modeled metrics at monitor locations were used to account for observed overprediction of the model by anchoring the Burgers meshes to physically realized levels. Note that PCBurg only provides metrics for Pmax, PL, ASEL, CSEL and unweighted SEL[*****]. Time domain pressure signatures (in psf) are included in the data archive so a further analysis could include calculation of other metrics from PCBurg output files. Signature files in that archive have names such as "SIG_F12P1_195.TXT" indicating the flight number, pass number, and mesh point number.

Prior to analysis, latitude and longitude coordinates describing survey participant locations and monitor locations were converted to a local coordinate system using the Haversine formula. Based on a combination of participant location relative to footprint and recruitment area, usable monitor metrics available, proximity to monitors, and some interrelationships among these quantities, each combination of event/metric/location was assigned a "noise method type" that describes how metrics were determined in each instance. These noise method types are listed in Table 6-6 and the corresponding

---

[*****] Unweighted SEL is labeled as "ESEL" in PCBurg output. During analysis, those levels were incorrectly interpreted as E-weighted SEL, leading to some additional uncertainty in ESEL metrics at participant locations. An examination of the impact is included in section 6.1.3.2.

methods are described in detail in Appendix O. These types can be grouped into three categories:

1. Noise method types 1, 2, and 3 are normal scenarios in which the interpolation method depends on proximity to monitors and footprint margins, and whether PCBurg metrics were available.
2. Noise method types 0, 4, and 5 indicate scenarios where no metric determination could be made due to missing data or locations outside study area, and levels were set to zero.
3. Noise method types 6, 7, 8, and 9 are special cases dealing with locations that were both outside the Burgers mesh modeled footprint and in locations downtrack of the Bayou Vista area monitor (GOLF) when the GOLF monitor level was missing and/or when monitor levels at further downtrack locations HOTEL, INDIA, or JULIET were missing.

Table 6-6 Method types for determining metrics at participant locations

Noise method type	Location description	Metric level at participant location
0	No participant location data	0
1	Within 0.5 nmi of usable monitor	Monitor level
2	Inside the Burgers mesh modeled footprint	Burgers mesh interpolation anchored to measured levels
3	Inside study area and either outside Burgers mesh modeled footprint or Burgers metrics unavailable	Interpolation/extrapolation of measured metrics
4	Outside study area	0
5	All monitor levels below ambient threshold	0
6	More than 2 nmi downtrack from GOLF, outside Burgers mesh modeled footprint, with 1-2 monitors missing from set of HOTEL, INDIA, JULIET	Interpolation/extrapolation across HOTEL, INDIA, JULIET monitor(s) only
7	Downtrack from GOLF, outside Burgers mesh modeled footprint, with all monitors missing from set of HOTEL, INDIA, JULIET	Using level from Bayou Vista (GOLF)
8	Less than 4 nmi downtrack of Tiki Island (FOXTROT), outside Burgers mesh modeled footprint, with all monitors missing from set of GOLF, HOTEL, INDIA, JULIET	Using level from Tiki Island (FOXTROT)
9	More than 4 nmi downtrack of Tiki Island (FOXTROT), outside Burgers mesh modeled footprint, with all monitors missing from set of GOLF, HOTEL, INDIA, JULIET	Using level from Tiki Island (FOXTROT)

Table 6-7 provides a detailed listing of the parameters used and their numerical values for the metric determination process. A thorough discussion of this process is provided in Appendix O.

Table 6-7 Summary of numerical values used in metric determination process

Description	Value
Window length – used for determining which set of metrics to use	650 ms or 3000 ms
Level above ambient – used to select which measured values to exclude	1, 3, or 5 dB
Monitor margin – participants within this range of a monitor are assigned measured metric	0.5 nmi
Burgers mesh margin – monitors within this range of the Burgers mesh are used to anchor modeled metrics	0.75 nmi
Edge monitor radius – the study area is increased to include regions within this range of monitors near the edge (BRAVO, HOTEL, INDIA, JULIET)	2.0 nmi
p, exponent for inverse distance weighting interpolation	3
Lateral footprint margin	0.5 nmi
Downtrack footprint margin (events 1 – 47)	2.9 nmi
Downtrack footprint margin (events 48 – 52)	4.2 nmi

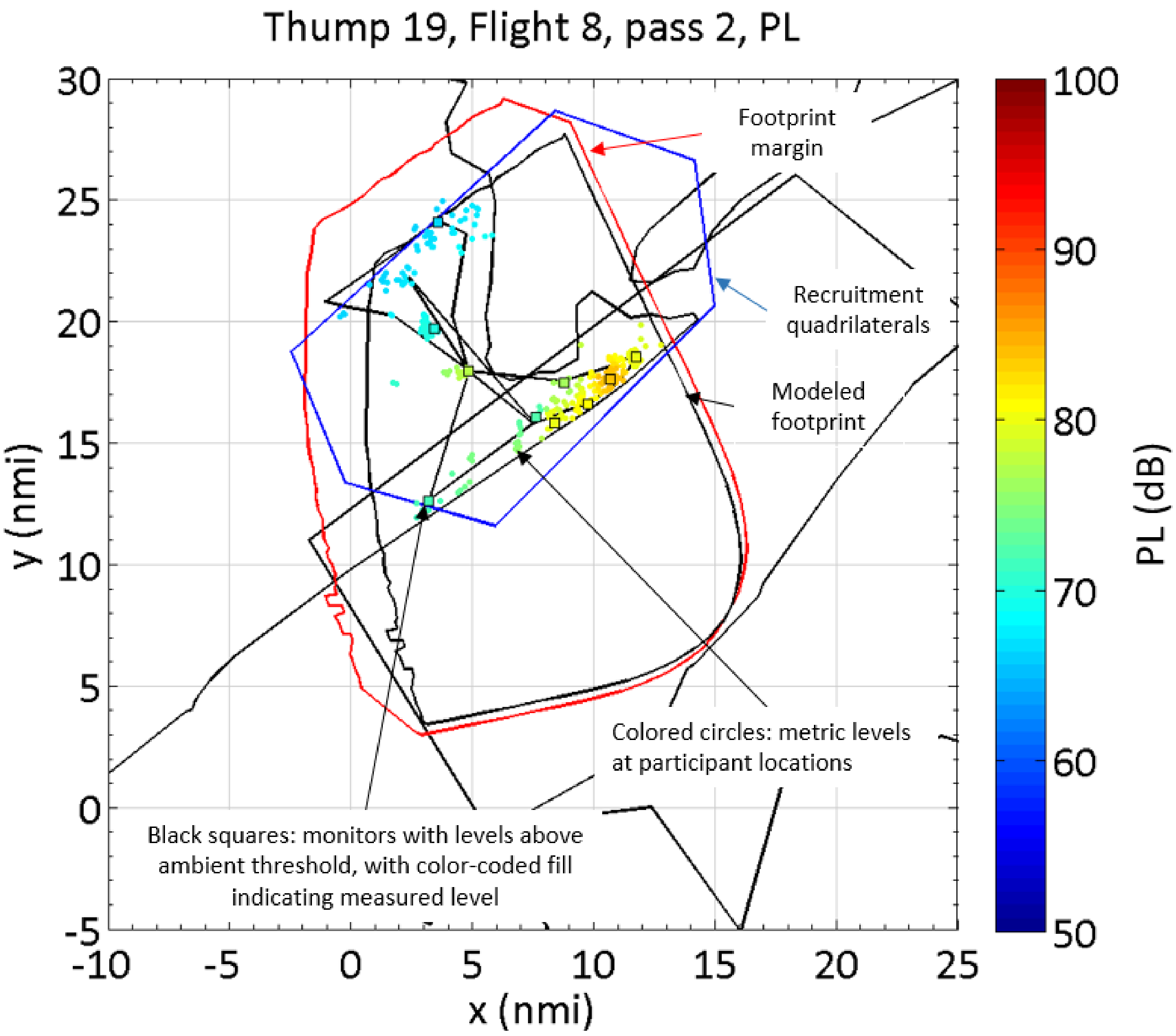


Figure 6-17 Example of metrics at participant locations

An example of metric levels at participant locations compared with measured levels is shown in Figure 6-17. Single event participant noise databases include participant location assignment codes, noise method types for each metric, and Boolean flags indicating if a specific location is inside the modeled footprint, inside footprint margins, and inside the study area. Using these fields, cumulative noise doses can be calculated using more or less restrictive requirements. For each set of noise metrics, daily noise doses were calculated using all participant locations inside the footprint margins. In other words, if a participant was outside the footprint at the time of a thump, the noise from that thump is not included in the DS dose calculation, but it would be used in the SE dose-response analysis. Moreover, even if a respondent was in the footprint, but was not locatable, these doses would also be omitted from the DS dose. It is certainly true that this could result in underestimates of the DS noise dose. However, there is no reliable way to include the dose from these events. One could assume that if a respondent could not respond with a SE report about a thump, then the noise was not noticeable to the respondent, and therefore this underestimate may in fact be quite negligible. Noise dose for a given flight day is specified by cumulative metrics calculated in a manner similar to Day Night Level (DNL) or Community Noise Equivalent Level (CNEL). Since no booms occurred outside the local time period 0700 – 1900, no penalties were added for evening or night hours. Cumulative levels in dB are calculated for each combination of participant ID, flight day, and N single event noise metrics (SE) as:

$$\text{Cumulative Level} = 10\log_{10}\left[\sum_{i=1}^{N} 10^{SE_i/10}\right] - 49.4$$

The cumulative level is essentially an energy sum of single event levels, with a standard factor of 49.4 dB removed to account for normalization to a 24 hour day[†††††]. For example, for one flight day a participant experiencing the five single event PLs tabulated below in Table 6-8 would receive a noise dose of 38.1 dB[‡‡‡‡‡].

Table 6-8 Example cumulative noise dose calculation

PL (dB)	10^(0.1*PL,i)
79.7	93325430.1
83.3	213796209.0
82.9	194984460.0
74	25118864.3
75.9	38904514.5
	sum(10^(0.1*PL,i)) = 566129477.8
	PLDN (dB) = 38.1

6.1.3.1 Comparison of Modeled and Measured Metrics

A detailed quantitative comparison was made between measured and modeled metrics using the Burgers meshes to determine metrics at monitor locations (see Appendix M for a discussion of PCBoom best practices). PCBurg output metrics were used without windowing or other modification. The scope of that comparison is limited to monitor recordings that were both within the modeled footprint and whose signals were sufficiently above the ambient level. Comparisons of the difference between modeled and measured PL are plotted in Figure 6-18, and show that levels for QSF18 flights are typically overpredicted by the model. Comparing levels for other metrics as in Table 6-9, the overprediction is consistent as shown in the mean differences. The number of samples available for these statistics is less than half of the boom recordings – the reason is that for many events, such as those using dive waypoints 4 or 5, several or all of the monitors were outside the modeled footprint and no comparison could be made.

[†††††] 10 log10(24 hours/day × 60 minutes/hour × 60 seconds/minute) ≈ 49.4. See, for example, “Calculation of Day-Night Levels Resulting from Civil Aircraft Operations”, EPA Report 550/9-77-450, Bishop *et al.*, March 1976.

[‡‡‡‡‡] Research into other impulsive noise sources often uses Zpeak or other metrics, whereas PLDN has typically been applied to sonic boom measurements.

Table 6-9 Differences between modeled and measured levels at monitors inside modeled footprints, with levels at least 1 dB above ambient

Metric	Mean difference	median	std	N
PL	6.1	6.6	5.8	233
CSEL	3.8	3.6	4.8	239
ASEL	8.1	8.4	6.3	222
ISBAP	4.3	4.7	5.2	239
MxPSF	0.02	0.05	0.11	239

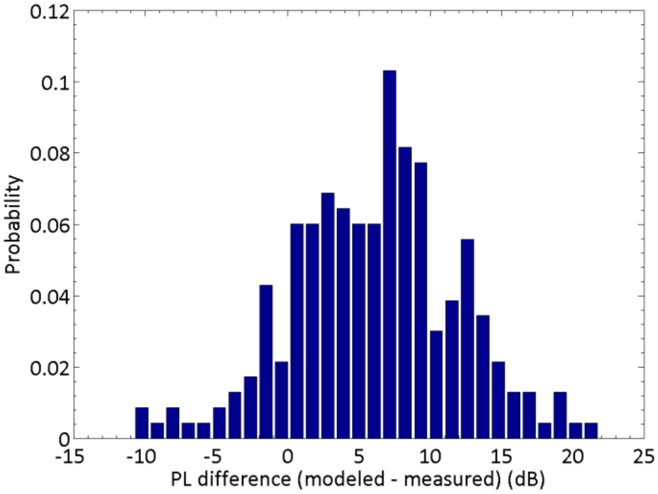


Figure 6-18 Difference between modeled and measured PL across all measurements inside modeled footprint with level at least 1 dB above ambient

To investigate some of the larger PL differences in Figure 6-18, boom 32 is considered. For that event, monitors CHARLIE and KILO recorded signals with PL differences of -21.9 dB and -19.4 dB compared to modeled values, respectively. The graphic comparison in Figure 6-19 shows that although the monitors were well inside the footprint (i.e. not near cutoff) the pressure signatures were not clean N-waves.

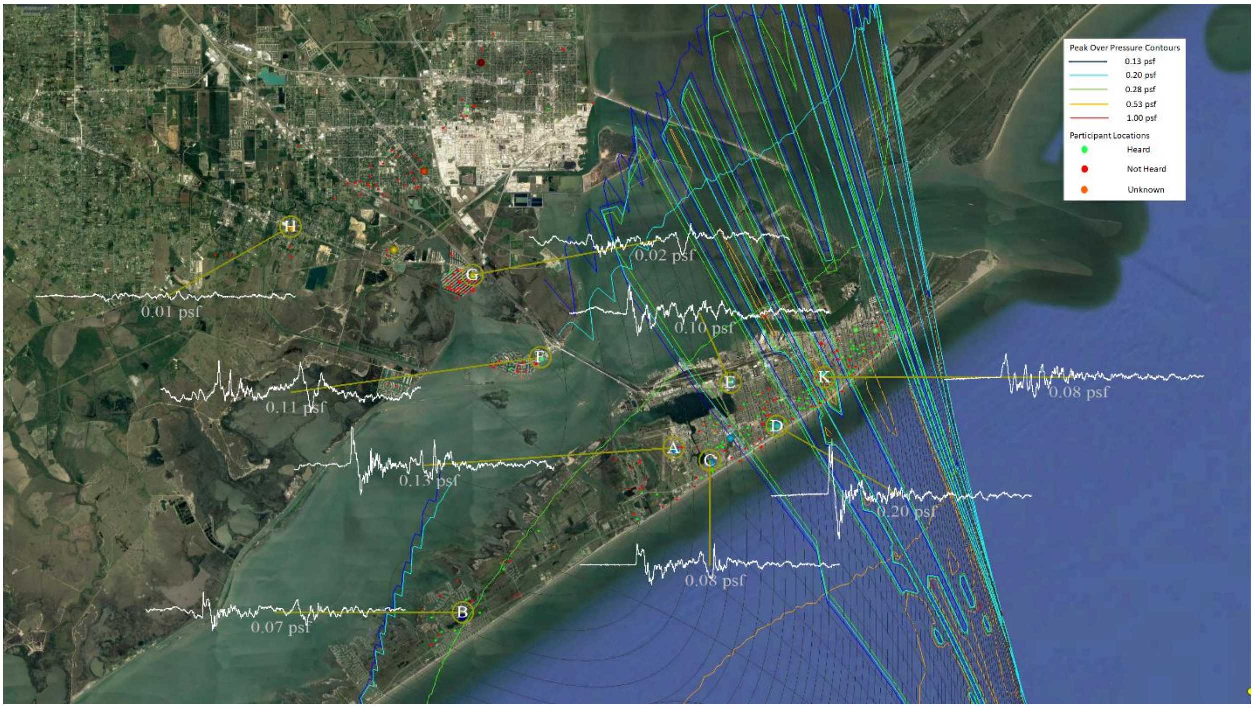
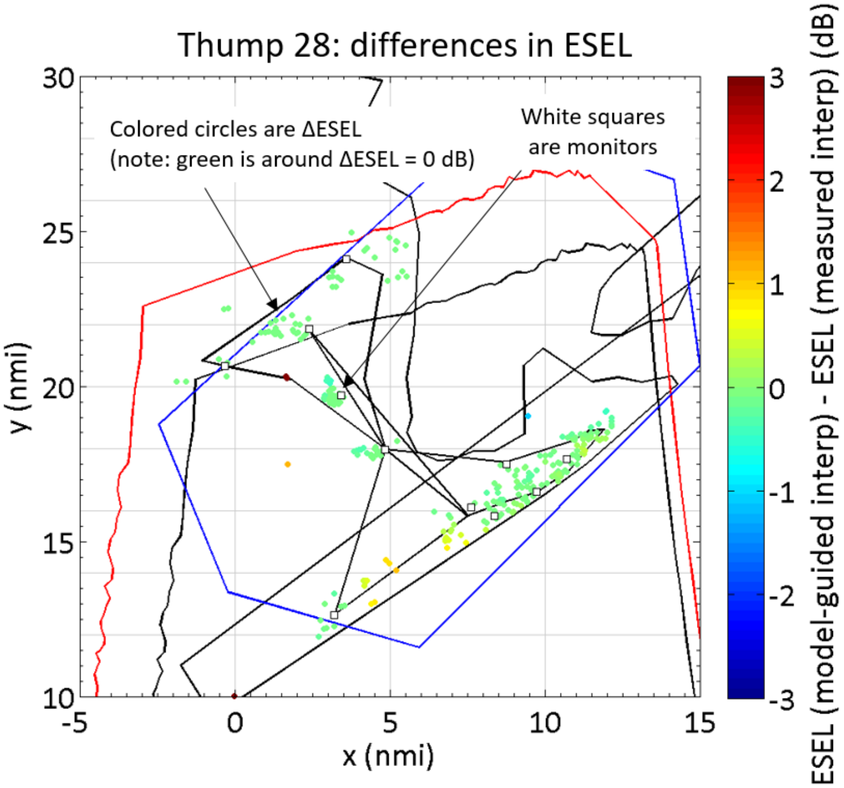
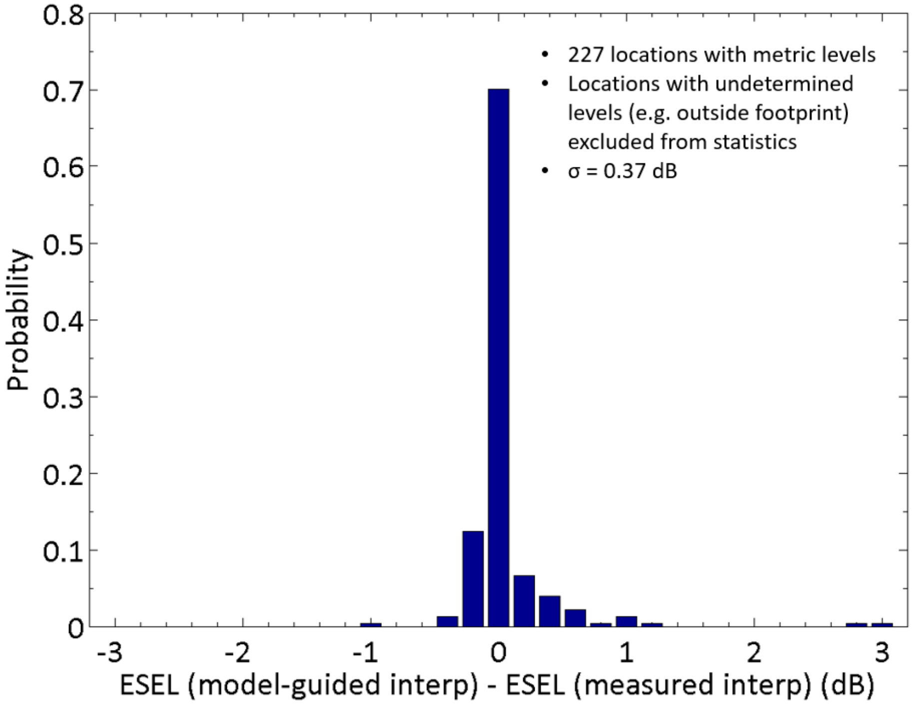


Figure 6-19 Comparison of modeled footprint (color contours) with measured pressure signatures (white insets)

6.1.3.2 Impact of ESEL Metric Interpretation

As mentioned above, following the determination of metrics at participant locations, it was discovered that the metrics labeled "ESEL" in PCBurg output actually represent the unweighted sound exposure level (sometimes abbreviated as FSEL or ZSEL). In the preceding analysis, those levels were interpreted as E-weighted SEL and used to make adjustments to interpolated ESEL values. Because the methods for determining metrics at participant locations were anchored by measured levels, the effect was not to introduce an offset such as that between FSEL and ESEL. Rather, interpolated metrics are affected by the difference in spatial gradients in ESEL and FSEL. To evaluate the impact of using FSEL to adjust interpolated ESEL, the metric determination process was re-run for an example case, but instead of using model-guided interpolation of measured ESEL, direct interpolation of measured ESEL was used. The differences in ESEL metric levels were calculated at each participant location and the results are shown in the upper portion of Figure 6-20. A distribution of level differences between the two methods is plotted in the lower portion of Figure 6-20. Taken together, it appears that the different methods do not introduce a consistently high or low offset and that for most locations the difference in interpolated ESEL is smaller than 0.5 dB due to the high correlation between ESEL and FSEL metrics. The impact is largest at locations far from monitors where the difference can be as large as 3 dB (overpredicted) though the total number of affected data points is small as shown in the histogram.





- 227 locations with metric levels
- Locations with undetermined levels (e.g. outside footprint) excluded from statistics
- σ = 0.37 dB

Figure 6-20 Difference in ESEL metric due to model-guided interpolation and direct interpolation of measured metrics

6.1.3.3 Uncertainty Quantification of the Metric Interpolation Methods

To evaluate the effectiveness of the interpolation methods and estimate the uncertainty, a monitor signal was dropped from the measurement data set and the interpolation methods described in Section 6.1.3 were used to determine levels at the location of the dropped monitor. Then, interpolated and measured values representing the same location could be directly compared. This procedure was repeated for all event/location combinations where a valid measurement was recorded. It should be noted that this method tends to overestimate the uncertainty, since it reduces the number of measurements used to anchor the interpolation scheme.

Comparing differences between interpolated and measured PL at monitor ALPHA, the interpolation scheme will tend to underestimate the level as indicated by a median value of -1.55 dB. Across all sites, however, the mean and median differences in PL are -0.03 dB and 0 dB respectively, indicating there is not a consistent under- or overestimation of PL across the footprint. A histogram of PL differences across all sites is shown in Figure 6-21. The model is overpredicting levels at measurement sites, so modeled levels are not used directly for assessment of metrics at participant locations. Interpolating measurements at monitor sites by dropping monitors shows that the interpolation does not not introduce a consistent bias. Figure 6-21 is comparing measurements with interpolated measurements (that is, no model bias). This is different from the comparison provided in Figure 6-18, which directly compares modeled with measured data, and thus does include model bias.

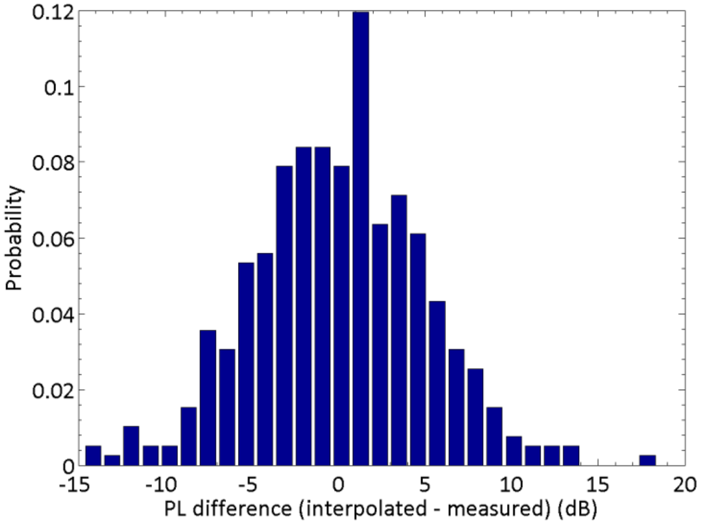


Figure 6-21 Distribution of differences in PL between interpolated and measured levels across all sites (N = 394)

Statistical quantities describing differences between interpolated and measured levels for all metrics are listed in Table 6-10 and Table 6-11. The mean interpolated levels are generally within 4 dB of the measured levels. While this is larger than the corresponding values from the WSPR 2011 data set (mean differences less than 2 dB), it should be noted that the study area and typical monitor separation distance are larger in the current work. In scenarios where large local (i.e. recorded by a single monitor) differences exist in measured data, this type of analysis is not able to reproduce the measured levels characterized as higher or lower than those measured at all neighboring monitor sites and will indicate a commensurately large difference between interpolated and measured levels. An example of this is shown in Figure 6-22, in which the monitor at DELTA recorded a level significantly higher than the three nearby monitors (PL = 87.9 dB at DELTA, compared with 74.7, 76.0, and 76.4 dB at ECHO, JULIET, and CHARLIE respectively). When the measurement from DELTA is dropped in the interpolation effectiveness assessment, the interpolated level at DELTA is 77.1 dB, or -10.8 dB compared to the measured level. The interpolation methods, either model-guided or via direct use of measured metrics, are not able to account for local spiking unless such phenomena are present in a measured signal.

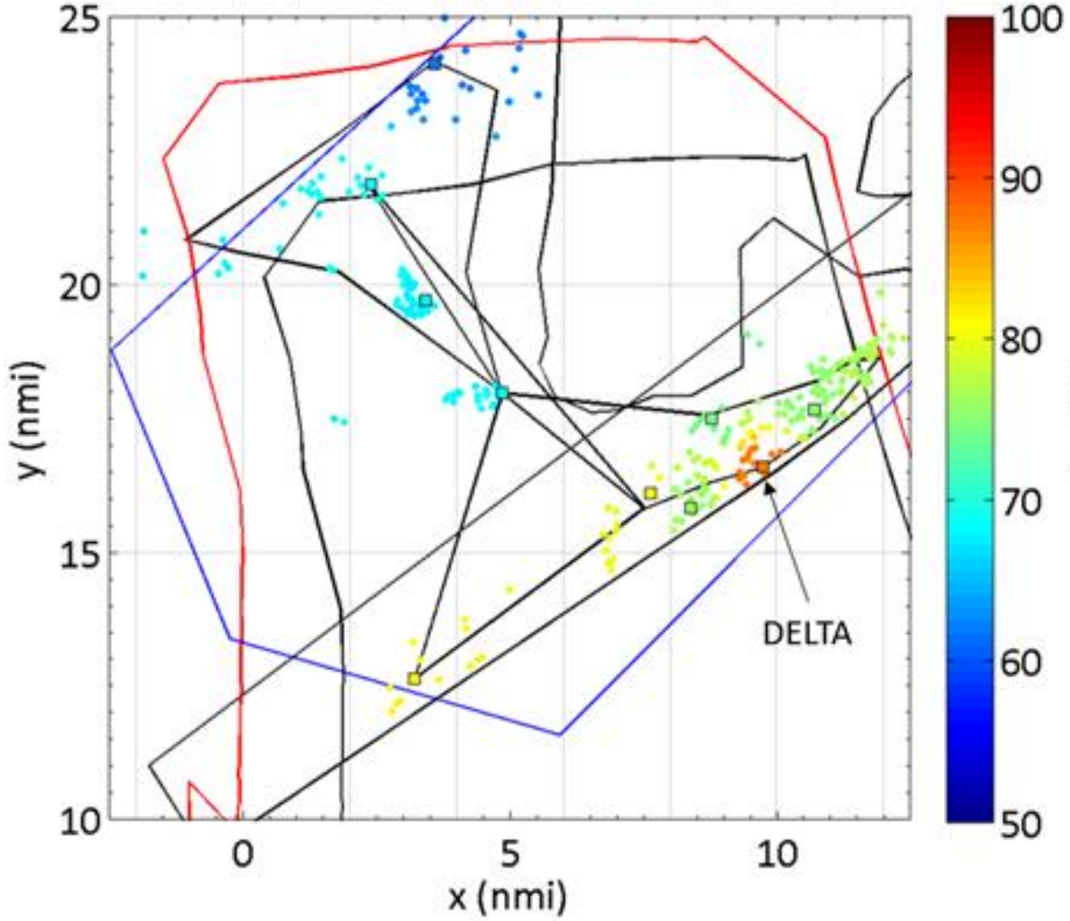


Figure 6-22 Example of locally high metric level recorded at one site (DELTA)

Table 6-10 Statistics for interpolated minus measured levels at each monitor site

	PL (dB)			CSEL (dB)			ASEL (dB)			FSEL (dB)			LLZf (dB)			LLZd (dB)			N
	mean	median	Std	mean	median	Std	mean	median	Std	mean	median	Std	mean	median	Std	mean	median	Std	
ALPHA	-1.2	-1.6	4.6	-1.4	-1.8	3.3	-1.5	-1.8	5.4	-1.3	-0.6	2.6	0.1	-0.2	3.9	0.1	-0.2	3.8	50
BRAVO	-1.6	-1.4	5.6	-2.5	-2.3	5.0	-1.6	-0.2	6.1	-3.8	-3.9	4.3	-2.5	-2.1	5.7	-2.4	-2.1	5.6	48
CHARLIE	1.2	1.5	4.3	1.3	1.5	3.3	1.3	1.5	4.5	0.7	0.1	2.3	-0.2	0.3	3.8	-0.2	0.3	3.8	51
DELTA	0.3	0.5	5.1	0.4	-0.2	4.6	0.3	-0.5	5.8	0.3	-0.3	2.5	0.1	-0.2	4.3	0.0	-0.3	4.2	47
ECHO	-1.0	-0.7	4.0	1.1	1.3	4.2	-0.6	0.4	4.5	1.2	1.1	2.6	-0.2	-0.5	3.3	-0.3	-0.4	3.3	47
FOXTROT	-1.2	-0.1	4.4	-2.0	-2.0	4.5	-0.5	0.6	5.1	-0.7	-1.1	2.8	-0.8	0.3	4.3	-0.7	0.6	4.2	48
GOLF	3.3	2.6	4.6	4.7	3.8	4.9	2.8	1.8	5.9	1.9	1.7	3.2	2.3	2.5	3.9	2.2	2.4	3.8	42
HOTEL	2.5	2.6	7.0	2.6	1.7	7.8	2.7	3.1	7.6	2.6	2.0	5.7	3.5	4.3	6.4	3.4	3.7	5.5	35
INDIA	-1.1	-1.5	3.6	-0.3	-1.2	4.4	-1.9	-2.0	3.1	-0.9	-1.9	4.2	-1.3	-1.3	4.9	-1.9	-1.5	4.2	28
JULIET	2.7	1.4	6.8	1.2	0.9	5.1	3.3	1.8	8.0	1.8	0.9	4.5	4.2	4.6	5.8	4.2	4.7	5.6	30
KILO	-0.3	-0.6	4.9	-0.5	-0.5	4.4	0.1	-1.0	4.8	-0.2	-0.4	2.8	-0.3	-0.5	4.4	-0.2	-0.3	4.4	46
LIMA	-0.9	0.7	4.6	0.5	1.0	2.2	-0.9	1.1	5.1	0.0	-0.3	1.5	0.6	1.3	3.1	0.6	1.2	3.2	4

Table 6-11 Statistics for interpolated minus measured levels at each monitor site, continued

	PNL (dB)			BSEL (dB)			DSEL (dB)			ESEL (dB)			ISBAP (dB)			MxPSF			N
	mean	median	Std	mean	median	Std	mean	median	Std	mean	median	Std	mean	median	Std	mean	median	Std	
ALPHA	-0.5	-0.9	5.2	-0.4	-1.1	4.0	-1.0	-1.7	3.3	-0.5	-0.7	4.3	-0.4	-0.9	3.6	-0.03	-0.04	0.09	50
BRAVO	-3.4	-2.2	6.7	-3.9	-2.9	6.0	-4.0	-3.5	5.2	-3.6	-2.7	6.1	-3.6	-2.8	5.3	-0.04	-0.03	0.11	48
CHARLIE	0.3	1.2	4.8	0.2	0.8	4.0	0.6	0.6	3.2	0.1	0.8	4.1	0.4	1.1	3.8	0.03	0.02	0.08	51
DELTA	0.1	-0.1	5.8	-0.3	-0.9	5.2	0.1	-0.7	4.3	-0.2	-0.8	5.3	0.0	-0.8	4.5	0.01	0.01	0.08	47
ECHO	0.1	0.1	4.6	1.1	1.2	4.3	1.2	1.7	3.5	0.8	1.6	4.4	0.2	0.5	3.5	0	0.02	0.08	47
FOXTROT	-1.0	-0.2	5.4	-1.1	-0.5	5.3	-1.4	-1.2	4.2	-0.7	0.1	5.1	-1.8	-1.1	4.5	-0.02	0.01	0.08	48
GOLF	3.6	3.6	5.8	4.0	3.4	4.5	3.5	2.6	4.3	3.2	3.0	4.5	3.9	2.9	4.3	0.05	0.03	0.07	42
HOTEL	3.0	4.1	8.8	3.7	4.1	6.7	2.5	1.7	6.1	3.2	3.8	6.7	2.5	2.8	5.5	0	0	0.1	35
INDIA	0.4	-1.5	6.5	-1.0	-0.2	3.9	-0.5	-1.1	3.5	-0.7	0.2	4.2	-0.7	-0.2	3.3	-0.01	-0.01	0.03	28
JULIET	3.0	4.0	5.5	2.0	1.4	6.7	2.1	0.9	6.1	2.4	1.4	6.5	1.3	0.2	4.2	0.01	0.01	0.03	30
KILO	-0.6	-1.2	5.4	-0.5	-0.7	5.1	-0.6	-0.8	4.0	-1.5	-1.1	9.3	-0.4	-0.5	4.1	0	0.01	0.08	46
LIMA	-0.2	1.3	4.5	0.0	1.6	4.0	0.4	1.5	3.2	0.0	1.8	4.4	0.4	1.2	3.3	0.04	0.04	0.06	4

Similar to data from Table 6-10 and Table 6-11, standard deviations across all monitors and all events can be calculated for each metric as in Table 6-12. These values show that for the range of conditions tested the standard deviations between single event dose based on interpolated measurements and measured metrics are in the range of 4 – 6 dB. Note, however, that this estimate is somewhat conservative as the measurement sample size was reduced to perform the analysis (dropping monitors) and that differences between measured and interpolated metrics are expected to be lower for locations close to monitors.

Comparison of monitor locations relative to footprint (i.e. undertrack vs offtrack, uptrack vs downtrack) does not show a clear trend between monitor location and mean difference across all metrics. There may be several factors including site-specific considerations leading to the distribution of measured vs interpolated differences observed. Table 6-12 combines measurements across all points to evaluate how differences between measured and interpolated values are distributed across all points. One could potentially look at confidence intervals, though since the 52 events are a mix of different waypoints (effectively placing monitors in different footprint locations), aircraft weights, and different atmospheres, a detailed analysis may conclude that approach is not valid. Another possible approach would be to look for cases with the same waypoints and similar atmospheric profiles to find subsets of events that are essentially repeats. Traditional statistics might be better suited to those data subsets. Statistical analysis of data from SonicBAT may be used to inform repeatability of ground measurements and enable comparison with QSF18 data, possibly giving insight into how much of variability is due to differences in aspects other than atmospheric profiles.

In general, instrument bias is a factor for measurements, however experience with analyzing the QSF18 data set suggests it may be small compared with other sources of variability. Response data specific to each transducer at low frequenices could be used to test that impression, and analysis of data from the co-located BYU microphones could also provide insights that might help to answer the question of measurement uncertainty.

Lessons learned on sources of uncertainty include:

- Doing quick looks at objective and subjective data and preliminary analysis / modeling during the test is valuable in both understanding data being collected and making corrections where possible.
- Observer reports from the field are a rapid and direct means of identifying differences.

Table 6-12 Standard deviations of measured metrics minus interpolated metrics at monitor locations across all events, all monitors

PL	CSEL	ASEL	FSEL	LLZf	LLZd	PNL	BSEL	DSEL	ESEL	ISBAP
4.9	4.9	5.5	3.7	4.5	4.4	5.8	5.3	4.5	5.8	4.5

6.1.3.4 Determination of Metrics at Participant Locations Lessons Learned

1. Determining metric levels near footprint edges requires additional consideration. The ability to predict lateral cutoff especially in the presence of winds at low grazing angles is difficult using

the current version of PCBoom. This introduces complexity to studies using the LBDM as the edge of the footprint often went through the study area. For X-59, if possible, avoid recruitment near the edge of the thump area, or consider moving the flight track to place edges of the thump area away from recruitment area.

2. Determining metric levels in footprint interior is more straightforward than near downtrack cutoff, as modeled levels can be used to guide interpolation of levels in between monitors.
3. More work is needed to better understand reason(s) for observed differences in measured and modeled levels for the F-18 low boom dive maneuver. Possible changes to the best practices could include use of Shulten's 3D curved earth ray path geometry in PCBoom[§§§§§].
4. Output from PCBurg labeled "ESEL" represents unweighted sound exposure level rather than E-weighted SEL. Since modeled metrics were used for adjustment of metric levels at participant locations, and all participant levels were anchored by measured levels, the impact to ESEL metrics in the participant noise database is minimal. In the future, for all metrics, the pure interpolation method should be employed *in addition to* the model guided interpolation to allow for evaluation of dose-response relationships using two techniques and to provide an indication of potential problems in either modeled or measured values.

6.1.3.5 Influence of PCBoom Input Data on Modeled Metrics

The preceding analysis was completed using version 3 as-flown trajectories and measured atmospheric profiles, which were distributed by NASA AFRC on 29 November 2018. An updated set (version 4) of as-flown trajectory files and post-processed atmospheric data was distributed by NASA AFRC on 25 February 2019. Due to a constraint on the analysis timeline, however, the metrics at participant locations were not recalculated using updated modeled levels from the v4 PCBoom input files. In this section, a comparison of modeled levels from each set of input data is made to evaluate the potential impact on metrics at participant locations. Ground boom modeling was repeated using PCBoom, with both v3 and v4 trajectory and atmosphere files for a selected event (boom 28: flight 12, pass 2). This example was chosen because it had a typical footprint shape considered to be representative of a nominal case. Comparing the footprint and overpressure contours in Figure 6-23, the results appear to agree closely. Some difference in contours can be observed at a fine enough scale, but the difference is qualitatively minimal.

[§§§§§] These features were not utilized because of lack of a systematic PCBoom validation analysis prior to executing the QSF18 field test. It was decided to retain consistency with the QSF18 flight waypoint planning process during the data analysis portion of Phase 2. In the future, use of the Shulten ellipsoidal earth algorithms should be considered.

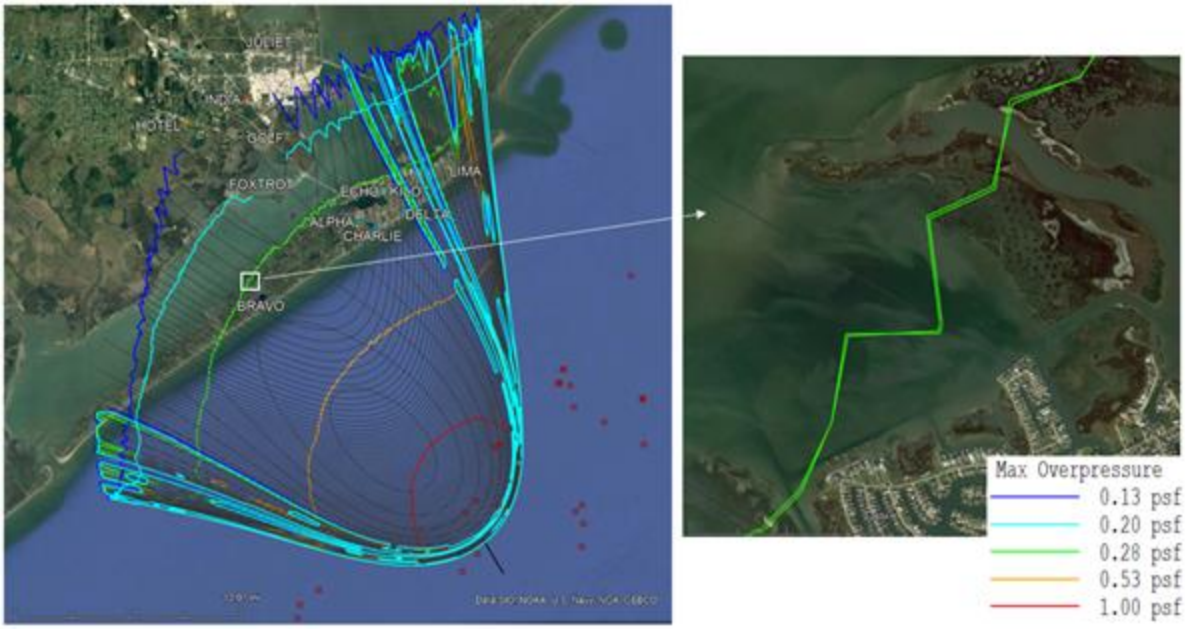


Figure 6-23 Footprint comparisons using version 3 and version 4 PCBoom input files for boom 28. Isopemps and overpressure contours appear coincident (left), though differences on the order of 100 ft are apparent on a fine scale (right)

To evaluate the difference quantitatively, PCBurg was used to calculate PL at ray ends landing near monitors ALPHA (Scholes Airport), FOXTROT (Tiki Island), and INDIA (La Marque). Results in Table 6-13 show that the difference in modeled levels is less than 0.1 dB. For this case, it appears that the update to v4 input files has no significant impact on modeled ground boom levels.

Table 6-13 Comparison of modeled metrics using version 3 and version 4 PCBoom input files, for event 28

	PL (dB)		MxPSF	
	v3 inputs	v4 inputs	v3 inputs	v4 inputs
ALPHA	83.72	83.74	0.29	0.29
FOXTROT	76.80	76.83	0.19	0.19
INDIA	70.91	70.99	0.12	0.12

6.1.4 Objective Data Analysis Lessons Learned Summary

Shortfalls in processes to adequately handle ambient noise were evident in the signature identification, metric calculations and subtraction of ambient spectra from the event. While some of these problems might be mitigated by more careful selection of quieter monitoring locations, this problem is not expected to be isolated to QSF18, and will likely be a recurring theme during future LBFD testing in urban areas.

More work needs to be done to the modeling and associated input data gathering to more reliably predict the edges of footprints and have suitable analysis procedures for points outside the footprint. It is necessary to evaluate the footprint extent (with a margin) and identify participant locations as inside or outside of the footprint. Though the margin was determined using empirical data, additional sensor granularity could have been useful, especially for the larger overpressure events. It has been suggested that other ray tracing modes in PCBoom can better evaluate cutoff position, however a methodical study and validation of modeling cutoff needs to be conducted well in advance of X-59.

Georeferencing, while an objective analysis activity, relied heavily on the location data reported by the participants. The lesson that subjective data should be scrutinized more closely as it arrives, especially to ensure valid address locations, is strongly reinforced by the subjective analysis.

Propagation modeling did not include the effects of clouds on ground signatures. This likely resulted in an overprediction of the metric values. Additional modeling capabilities that include propagation through clouds should be added to PCBoom and validated so it can be incorporated into the waypoint planning and metric evaluation process prior to X-59. Additionally the method that was employed in the field for estimating the cloud altitude levels (upon which improved modeling will rely) was based on a single balloon launch. Procedures for estimating cloud cover, density and altitude bands need to be determined for the purposes of sonic boom prediction with PCBoom.

Due to legacy mislabeling of metrics in PCBoom, the guided interpolation of the metric values at participant locations relied on an incorrect computational mesh. This introduced uncertainty near the footprint edges and was unfortunately promulgated into the dose-response analysis (for ESEL only). It was determined that overall the differences were of the order of 0.5 dB ESEL in the center of the footprint and 3 dB ESEL overprediction near the edges.

6.2 Subjective Data Analysis

All participants received test information via email communication. In order to test single event survey reminders, the respondents were divided into two groups with communication provided by either email or text message. Once response groups were assigned, random assignment for reminder/no reminder was made within each group. The target was to have 125 respondents in each reminder type/group, and the breakdown of participants by group is provided in Table 6-14. Participants were not overtly told to which groups they were assigned. All recruited participants completed the Background Survey at the time of their enrollment, but not everyone responded on all other surveys.

In the analysis of some of the data below, the 500 recruited participants will be referred to as “All Subjects” and the 476 people that submitted any number of single event (SE) and daily summary (DS) reports are referred to as “Responders”. Both groups are presented to better represent the demographics of the sample and to facilitate comparisons to similar groups in future studies.

Table 6-14 Quantity of respondents

Number of Responders by Group for both SE and DS Surveys (476 total)			
Email - No Reminder	**Email - Reminder**	**Text - No Reminder**	**Text - Reminder**
119	120	114	123

6.2.1 Summaries of Survey Responses Only

6.2.1.1 Single Event Report Summaries

The Single Event Survey was to be completed after each sonic thump event. The initial single event response dataset contained 11869 rows of data, which means that 11869 single event surveys were initiated. Before considering the noise dose, the rates at which the respondents submitted single event reports were characterized. The Single Event Response rates by reminder group and flight pass are presented in Figure 6-24. The process for assigning each response to a specific flight pass is described, and detailed results are provided, in Appendix R. The plot is for flight passes only and does not include false reminders or “no event” responses. These proportions per group use as denominators the group sizes reported in Table 6-14.

The data in the plot clearly indicates that the reminders worked in generating SE reports from respondents, and in general, text reminders (purple) were more effective than email reminders (green). The no reminder groups (red and blue) typically generated SE reports from no more than about 20% of the group members.

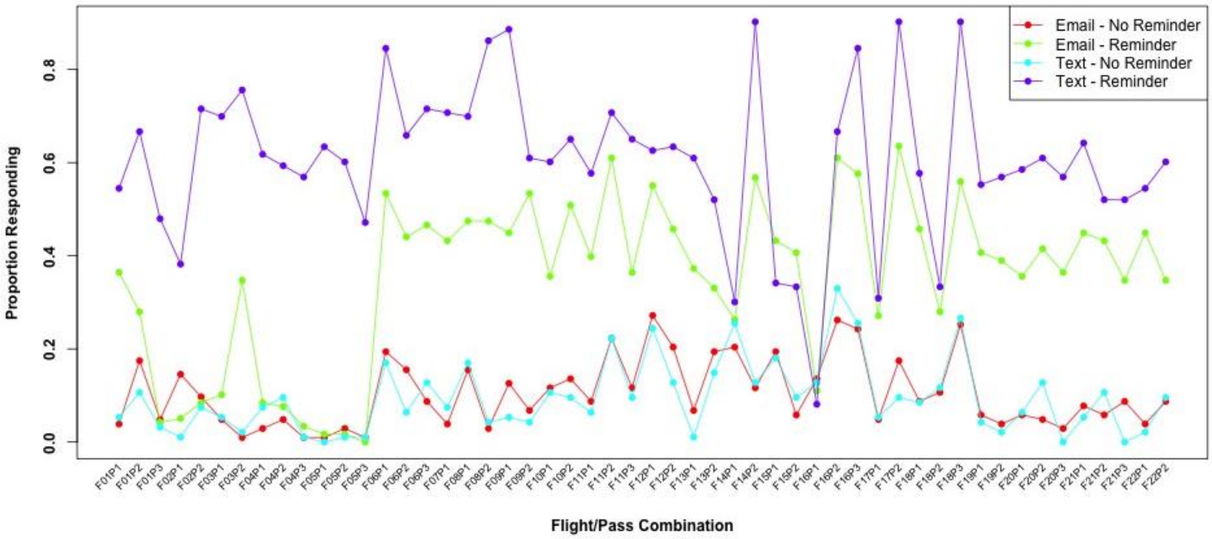


Figure 6-24 QSF18 single event response rate by reminder group

These data present the number of SE reports that were submitted. The respondents indicated whether or not they heard a sonic thump. Figure 6-25 shows the frequency with which individual thumps achieved a percentage of “heard” responses in 10% bins from 0 to 100. For example, only two thumps were reported being heard by less than 10% of the respondents to the event, while 7 thumps were reported being heard by 10-20% of the respondents to the event. It is interesting to note that the vast majority (40/52) of the thumps were heard by fewer than half of the people who submitted a report for that thump.

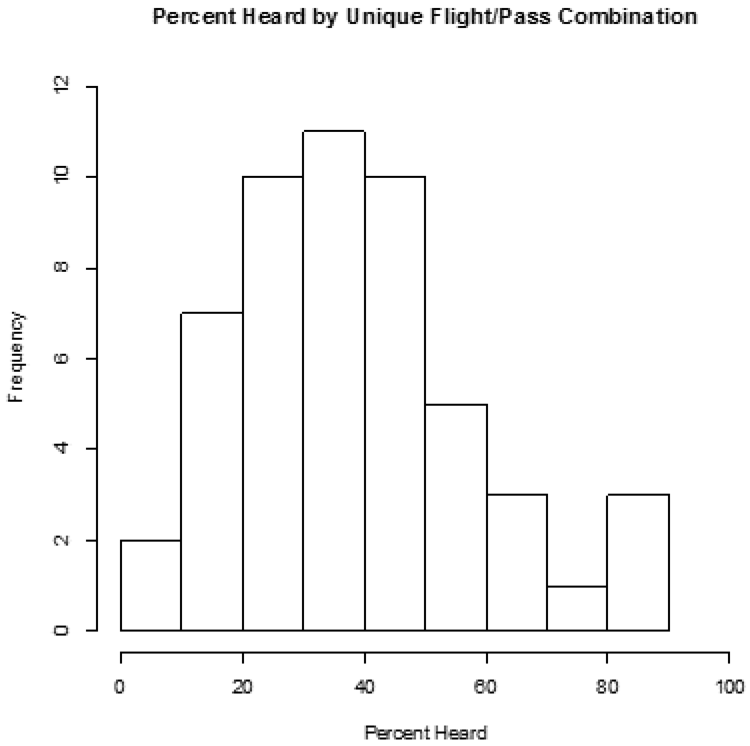


Figure 6-25 QSF18 histogram of single event percent heard among respondents to the event

Figure 6-26 shows the percentage of the reports from each reminder group that were associated with HA ratings, and now for easier visualization the two no reminder groups are combined. The data are presented for individual passes or sonic thumps and also include responses to false reminders and "no event"on the far right of the Figure 6-26. While some of these events show a sizable proportion of people giving HA ratings (for example, Pass 2 of Flight 9 has the no reminder group at 10% HA), note that the sample sizes for the number of reports within a group can be quite small. Figure 6-24 shows that approximately 10% of the no reminder groups (total 233 respondents) are responding to this thump. So if one observes just under 10% HA for 23 SE reports, then one is observing 2 HA reports total in that group.

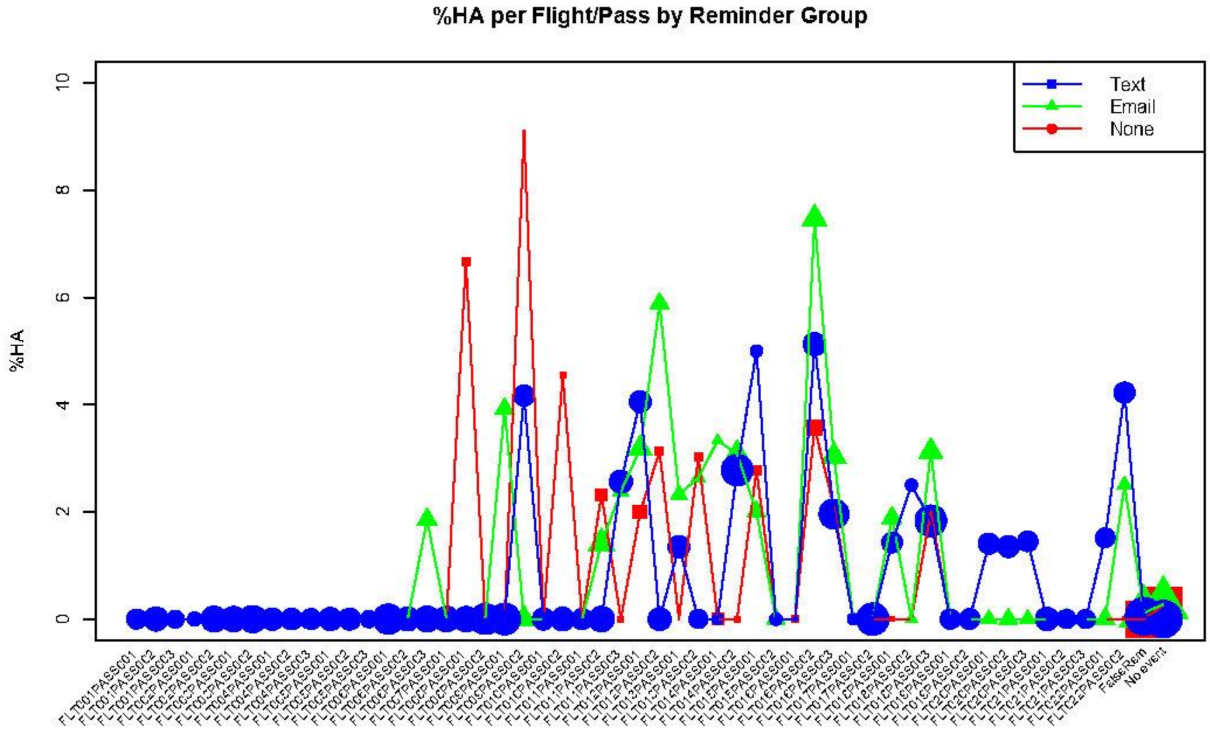


Figure 6-26 QSF18 single event %HA by reminder group and flight/pass, false reminder and no event, points are sized to represent total number of reports by subgroup

When these reports are not disaggregated by flight and pass, one can visualize an overall propensity to be highly annoyed by reminder group and event type. Figure 6-27 shows that across all thumps, 1.27% and 1.24% of the reports received by the no reminder and email reminder groups were recording HA responses, while for the text reminder group this was 0.8%. By contrast, the false reminders almost never prompted HA responses. Reports that were received which were not associated with a sonic thump or a false reminder registered HA responses for less than 0.5% of the reports. These are assumed to be instances where respondents mistook a different environmental noise for a sonic thump.

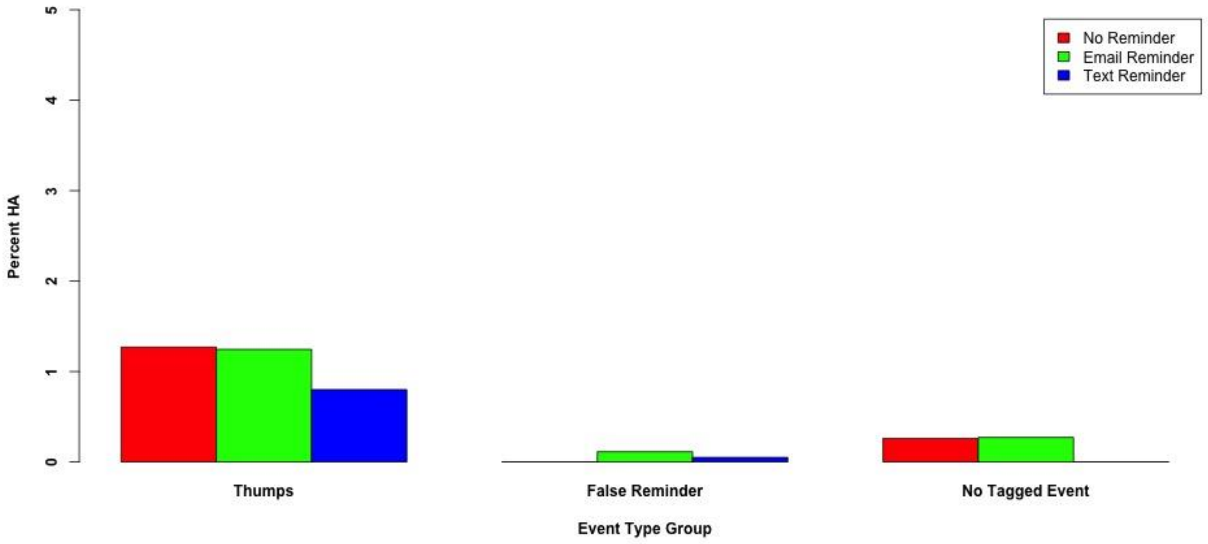


Figure 6-27 QSF18 single event %HA by reminder group and event type

6.2.1.2 Daily Summary Report Summaries

The initial daily summary response dataset contained 3411 rows of data, which means that 3411 daily summary surveys were initiated. Before considering the noise dose, the rates at which the respondents submitted DS reports were characterized. Figure 6-28 gives a histogram of the number of DS reports submitted per respondent. It shows that the mode of the distribution is 11 reports (one for each test day), but only 103 respondents submitted 11 reports. The next most common number of submissions is 10, but Figure 6-28 also shows that some respondents submitted more than 11 total DS reports. In the web-based survey system, a respondent could begin a report, leave it open, and then return to finish the report before submitting. All open reports were closed before the start of each test day. Some respondents began a second separate report, resulting in more than 1 Daily Summary per unique ID when the reports were closed. This topic will be addressed in more detail in Section 6.2.2.

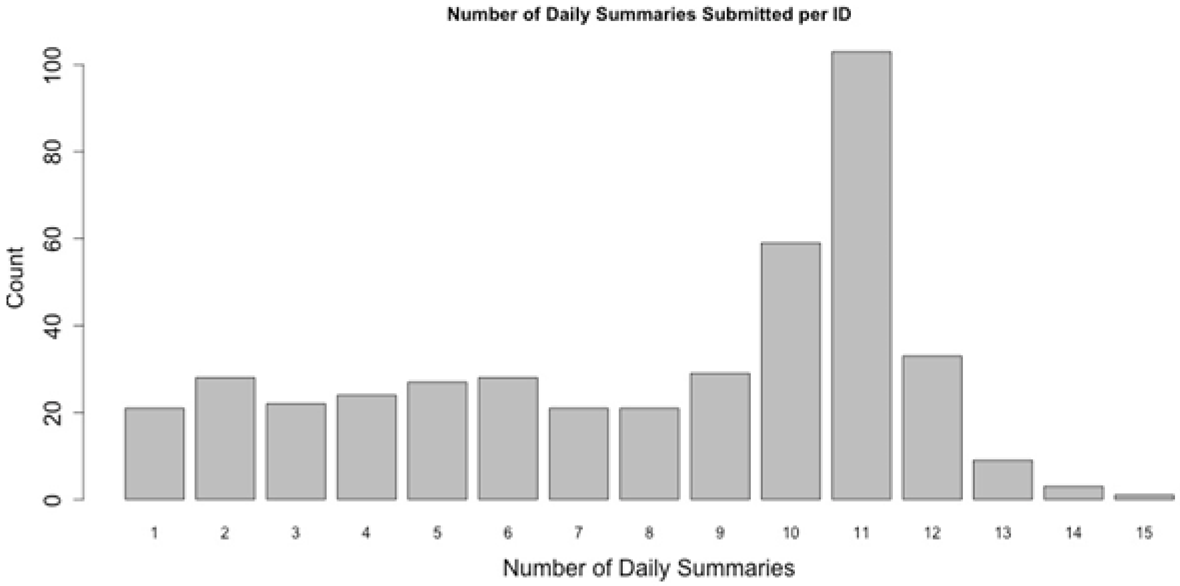


Figure 6-28 Number of daily summary reports submitted by respondents

Any analysis of this data would optimally include a single DS response per individual per day. Unfortunately, with the information gathered, it was impossible to accurately label some of the duplicate submissions as good/accurate and others as inaccurate. Two possible solutions are to (1) exclude all data per individual on a day for which they submitted more than one report and/or (2) include it all. As the model below already allows for correlation between responses from the same individual, option (2) will accurately account for this. Some model testing was done where the data from duplicates were excluded, and none of the results changed in any meaningful ways. As a result, all data are included in the analyses presented below.

In an effort to understand the effect, if any, of reminders on DS response rate, Figure 6-29 shows similar histograms for the email reminder, text reminder, and combined no reminder groups. The y-axis here shows the raw count of individual respondents submitting a specific number of DS reports, so it is not of specific interest. The main intent is to see the general shape of the plots, and note that they are all quite similar to one another, indicating that the reminders did not substantively change the way respondents handled the daily summary report submission.

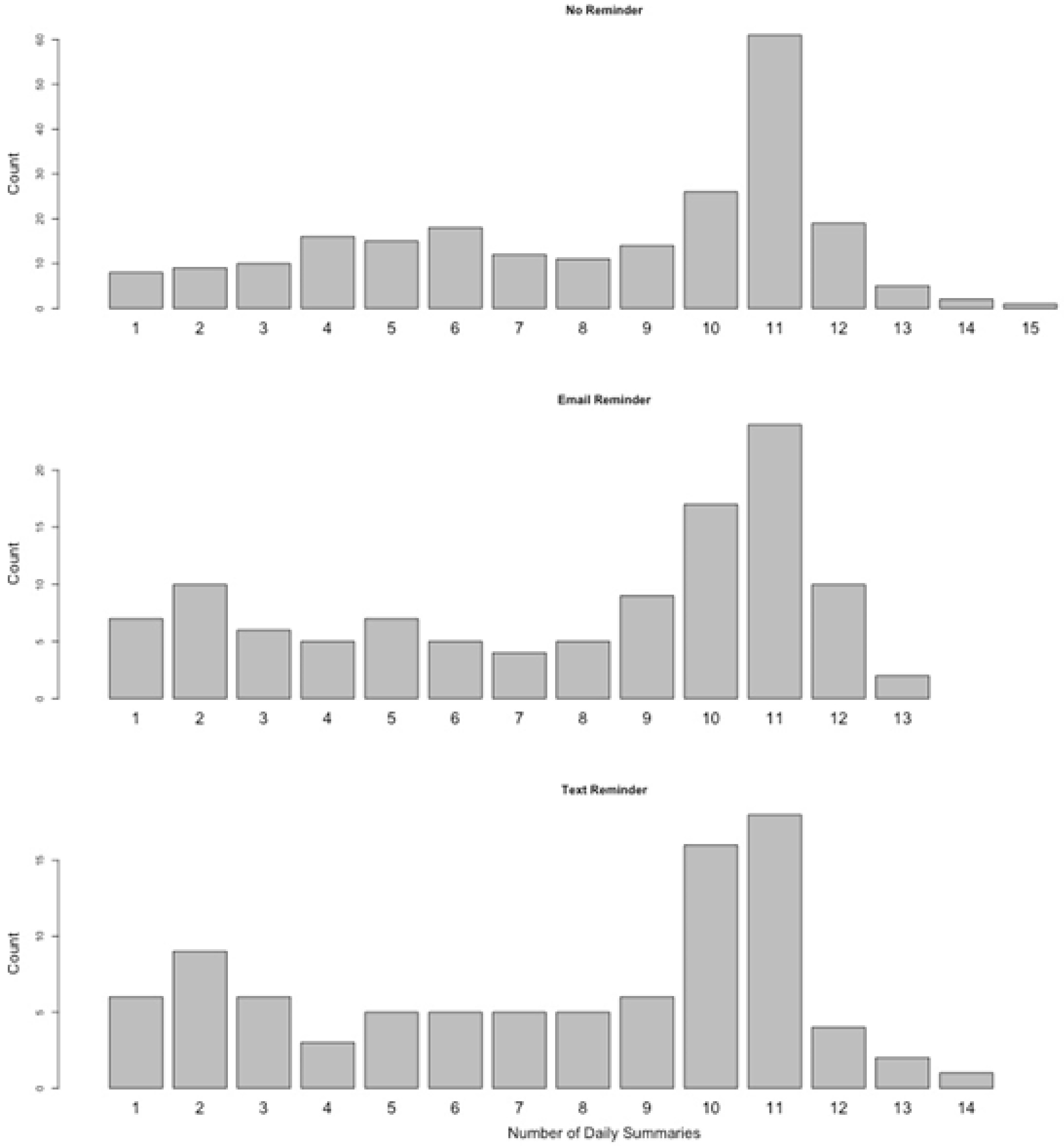


Figure 6-29 Number of daily summary reports submitted by respondents by reminder group

Figure 6-30 shows how the overall DS response rate varies by test day (while Figure 6-29 aggregated data across dates). It shows that with some variability, typically about 60% of the respondents submitted a DS report on each day of the study period.

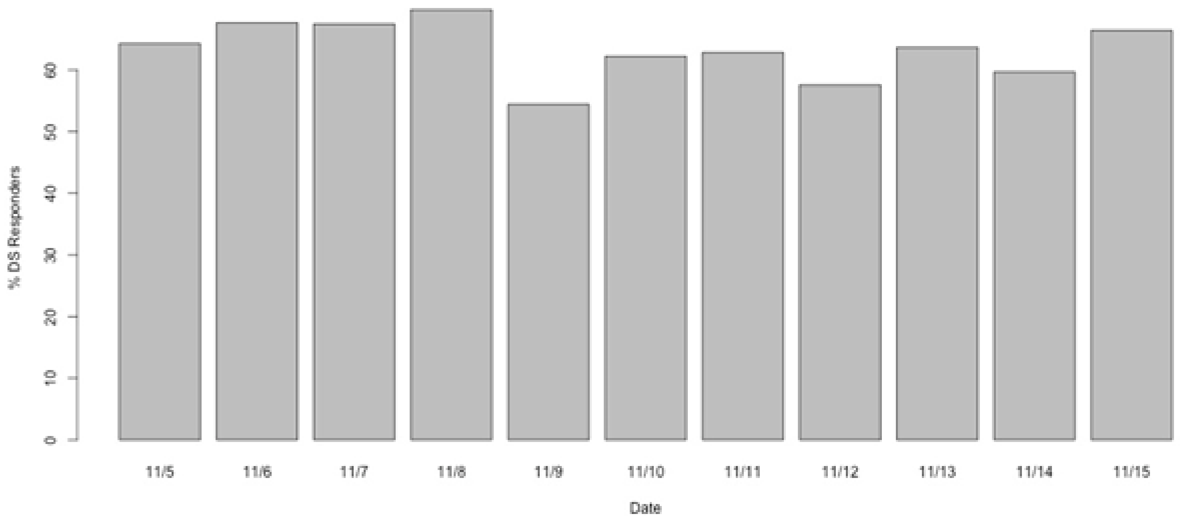


Figure 6-30 Daily summary response rates for each test day

In an effort to understand the impact of the event reminder group, the DS data (presented in Figure 6-31) has been broken out into the three reminder groups. The text reminder group is consistently responding at the lowest rate, followed by the email reminder group, while the no reminder group is most likely to respond on each day. Current information does not explain the difference in these response rates. Potentially this is a topic for future investigation.

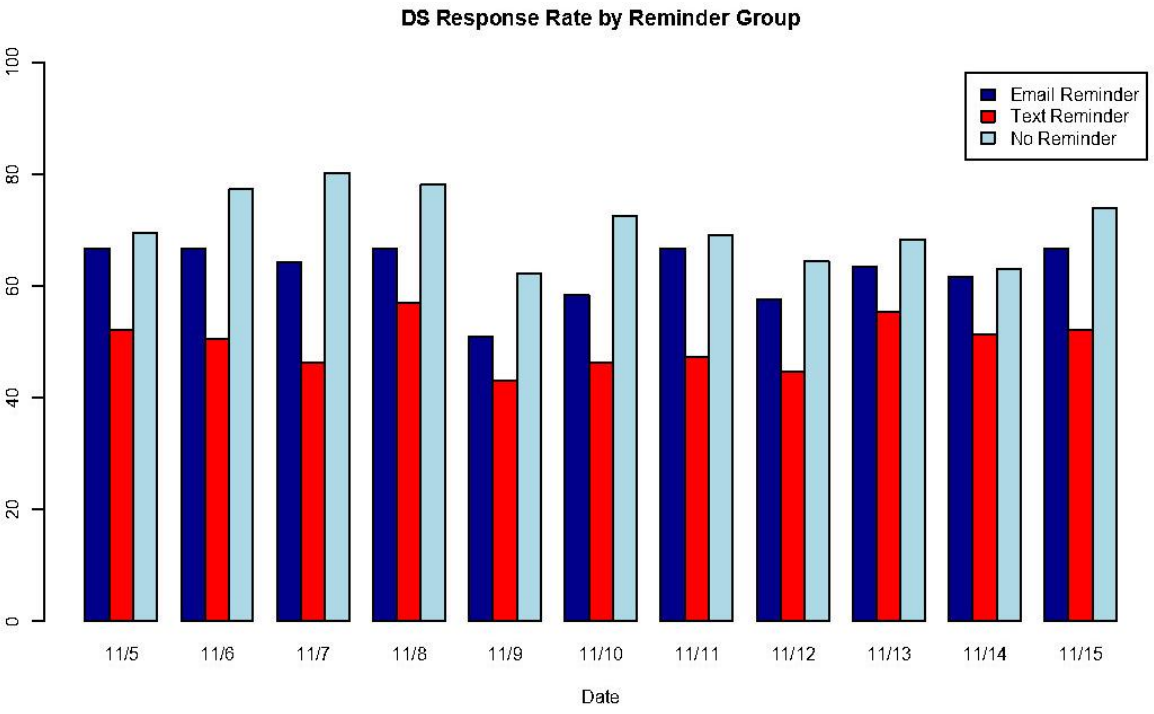


Figure 6-31 Daily summary response rates for each day by reminder group

6.2.1.3 Background Survey Summary

The background survey (Appendix J and Appendix K) gathered standard demographic characteristics (gender, age, education, etc.) and other types of characteristics specific to noise. The full set of summaries for participant demographic data obtained from surveys, plus derived noise habituation and sensitivity scales, is provided in Appendix L. These plots provide either histograms (for quantitative variables) or bar plots (for categorical variables) of the responses to the various background questions, some of which were incorporated into the dose response model. These plots are presented both for the full set of recruits, and for the respondents only, simply to confirm that those who chose to respond were not different in any way as measured by the background survey. A representative result of the standard demographics portion of the survey is presented in Figure 6-32.

The majority of the questions on the background survey are stand-alone demographic questions. However, two additional sets of questions were asked to provide more data for statistical analysis. The first set of additional questions (four in quantity) form a group meant to measure the latent construct of ability to habituate to noise. The full set of summaries for these questions is provided in Appendix L. These consist of bar plots corresponding to categorical variables, as shown in the example in Figure 6-33. Cronbach's alpha was used to conduct a reliability analyses on these four questions. Cronbach’s alpha is a measure that assesses how closely related a set of items are as a group. It is a measure from 0 to 1 indicating the extent to which the group of questions is interrelated, with 0 being unrelated and 1 being related. Cronbach’s alpha changes from .66 for the 4 question scale to .75 for the 3-question scale that omits the first reverse-coded item. This shift in value implies that the reverse-coded nature of the first question made the four question scale less reliable than the three question scale. A reverse-coded item is one in which a larger response indicates a lower scale value. The habituation scale included only three of the questions. The questions included on the QSF18 survey were previously utilized in a past NASA sponsored low boom research effort [Page, *et al*., 2014]. They were initially evaluated in an investigation of community attitudes towards blast noise that was sponsored by the DoD Strategic Environmental Research and Development Program SERDP WP-1546. The QSF18 design mirrored the blast research effort by utilizing a background survey, a daily summary survey and responses to single events. The noise sensitivity questions were used as a noise sensitivity index in the blast noise research [Nykaza, *et al*. 2014]. The development of the surveys was based on published recommendations from the International Commission on Biological Effects of Noise [Fields *et al*., 2001], and a review of noise sensitivity literature from as early as Weinstein [1978] to more recent considerations of noise sensitivity to impulsive military noise [Luz 2005]. The responses to the questions regarding sensitivity and habituation to noise are combined (via simple addition of the three questions used) to form a noise habituation scale, which will be used to explain annoyance in the dose-response models. The noise habituation scale is depicted in Figure 6-34.

The second set of additional questions (five in quantity) asked participants about their sensitivity to annoyance by common noise sources, including barking dogs, thunder, street traffic, commercial aircraft and military aircraft. A full set of summaries for these responses is provided in Appendix L. These consist of bar plots corresponding to categorical variables, as shown in the example in Figure 6-35. After conducting reliability analyses on these five questions, it was concluded that they form a cogent single scale measuring a respondent's sensitivity to noise (Cronbach's alpha of .76 for the 5-question scale). The responses to these questions regarding annoyance to common noise sources are combined (via simple addition) to form a noise sensitivity scale, which is incorporated into the dose-response annoyance models. The distribution of the characteristics is depicted in Figure 6-36.

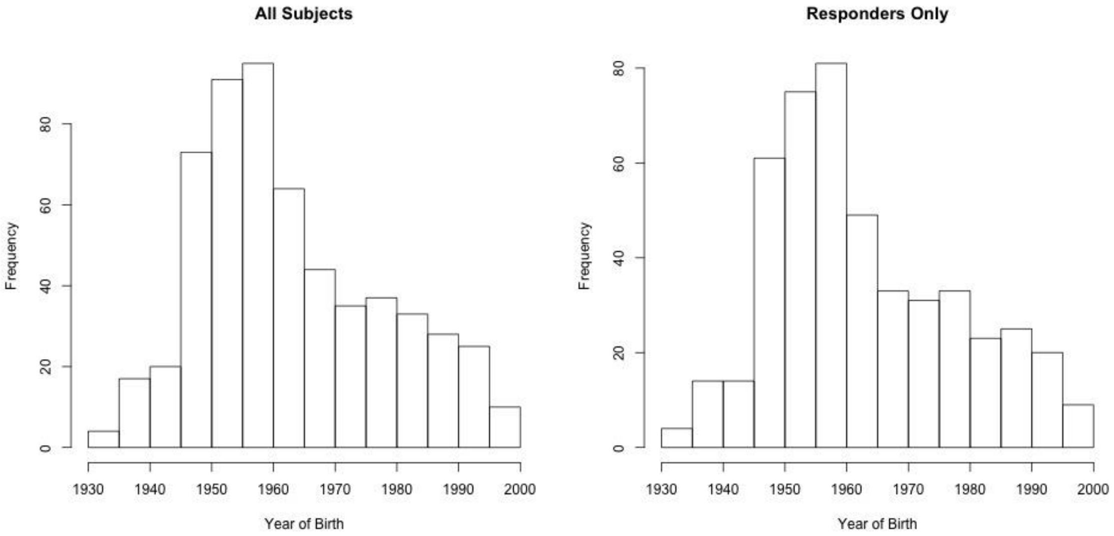


Figure 6-32 Distribution of birth year of recruited sample

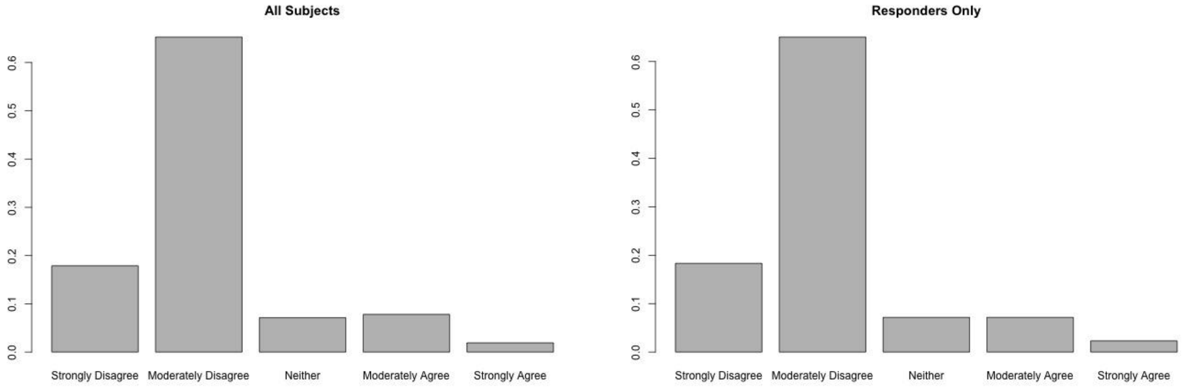


Figure 6-33 Distribution of responses to “with time most people adapt to noise” for recruited sample

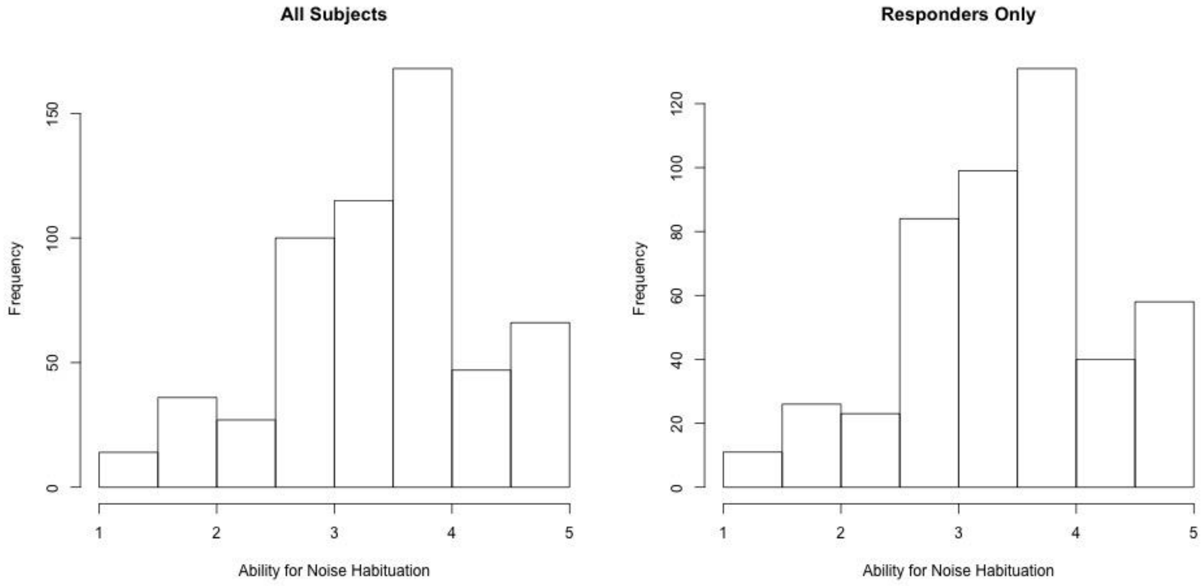


Figure 6-34 Distribution of calculated ability to habituate scale for recruited sample

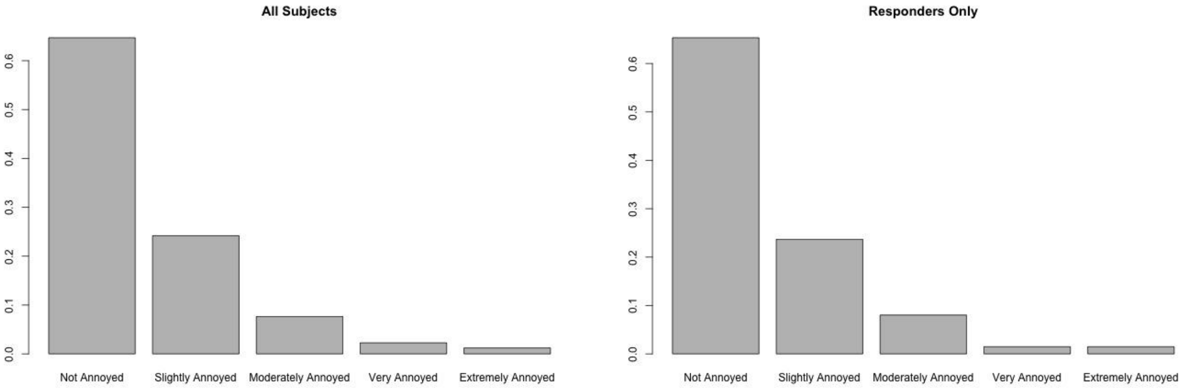


Figure 6-35 Distribution of responses to annoyance with military aircraft for recruited sample

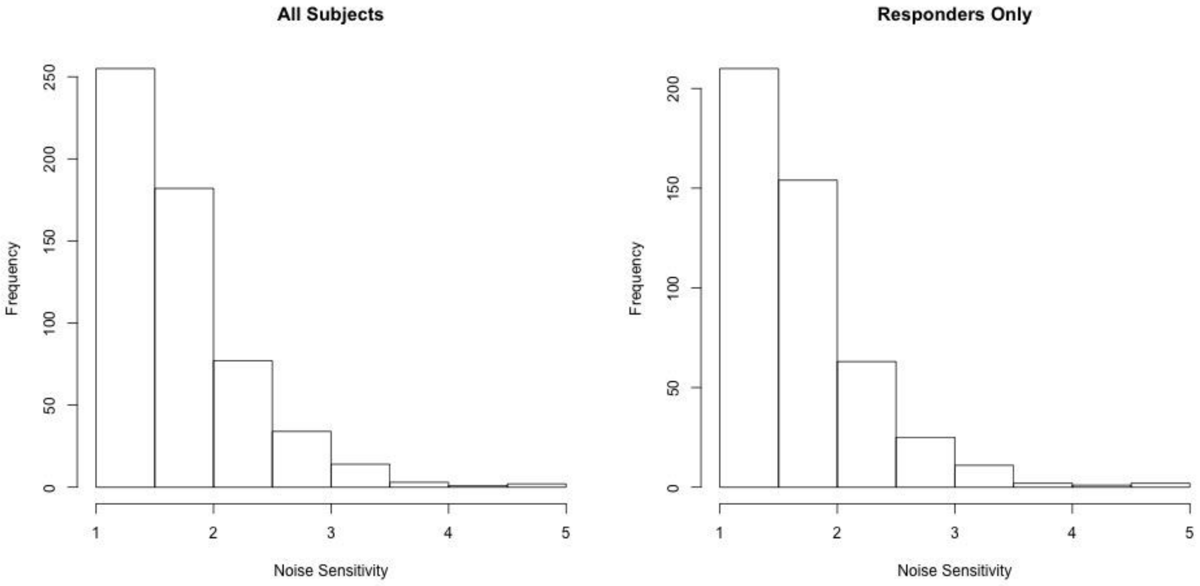


Figure 6-36 Distribution of calculated noise sensitivity scale for recruited sample

6.2.2 Subjective Data Analysis Lessons Learned

There is an ongoing need for more and better quality assurance/control with the subjective survey responses. Respondents submitting more than one daily summary, or multiple SE reports for the same event, are issues that need careful attention for future efforts. Some amount of participant error will be inevitable, but investigations should be conducted into safeguards that will reduce participant error.

The survey was designed to strike a balance between providing a format that was easy to complete with a format that controlled response fields. For example, the daily summary research design allowed respondents to leave and return to the survey in the event that they were interrupted during their response. This design feature was in effect throughout the test. A consequence of this was that some respondents inadvertently opened more than one daily summary survey for response. All surveys that were still open were closed by the Survey Research Center prior to 8 AM, before testing began each day. However, if a respondent had more than one form open, then each form was recorded for that date. The coding of the survey responses can be modified to limit the number of respondents submitting more than one daily summary, or multiple SE reports for the same event. Other potential changes can be made in the coding of the surveys. A deliberate decision was made to force respondents to select the time of the single event in 15 minute intervals. This decision was based on the noise plan that thumps would be at least 20 minutes apart, and the assumptions that grouping the time of the event in 15 minute intervals would support that rate of noise exposure. This approach can be modified if needed. The driving consideration is to provide at least, preferably more than 20 minutes between thumps, so that respondents can differentiate between thumps. How that data is captured can be modified for future field tests.

In addition, even though instructions were clear, it is possible that some respondents used daily summary forms incorrectly, judging by the fact that they would come in at all times of the day. This suggests there would be value in investigating development of improved guidance for respondents. Similarly, the ways in which respondents report event times should be investigated to identify ways to improve accuracy.

Another potential topic for future investigation is the seemingly counterintuitive finding in the QSF18 daily summary data that the text reminder group is consistently responding at the lowest rate, followed by the email reminder group, while the no reminder group is most likely to respond on each day.

6.3 Statistical Correlation and Dose Response Curve

One of the primary objectives of the analysis was to characterize the dose-response relationship, assessing the percent highy annoyed (%HA) in response to sonic thump sounds. Participants provided ordinal responses, (5 levels, which were mapped onto a binary (HA/non-HA) response) as described in Section 6.3.1.1. As a result, models that associate an individual's dichotomous (HA/non-HA) response primarily as a function of the objective measure of noise were fit. The experimentally assigned reminder group and the respondent's experimental quadrant were also accounted for in the model. In addition, several of the pieces of information measured on the background survey, as they may reasonably contribute to an individual's annoyance to noise, were incorporated; perhaps most importantly, an individual's score on the habituation scale and the noise sensitivity scale were included. Precisely because of these individual-level differences in perceptions of noise, the model also incorporates participant-specific intercept terms that allow for a respondent's annoyance ratings to be correlated with one another, but models them to be unrelated to responses from other participants.

6.3.1 Statistical Correlation and Dose Response Analysis

6.3.1.1 Single Event Analysis

As indicated above, the initial single event response dataset contained 11869 rows of data. Not all of these could be used in developing a dose response relationship. In order to be included, a respondent's position must be locatable by the methods described in Section 6.1.2, and must also have an associated single event noise dose at that location. The details about how reports were tagged to events are given in Appendix R. Note that event time for each participant is obtained via one of the survey questions (E2) that asks the respondent about the time at which the event/reminder occurred. Due to weather and other circumstances outside of the team's control, the thumps and reminders were not always spaced as planned. While the majority of responses were readily associated with specific events, there were some cases where reminder and events were closely spaced such that they were not as easily associated with specific response reports.

As an example of how this total of 11869 reports becomes a smaller usable dataset, for the PL metric at the 5 dB ambient threshold using the 650ms window, after including only those who were locatable and had a noise dose from a thump associated with them, the dataset included 5796 rows of data. Note that if every person had submitted a SE report for each of the 52 thumps, and all were locatable, there would have been 26000 reports; as such, these 5796 reports represent 22.3% of the possible data. Further, because the models also incorporated information from the background survey, a respondent must also have given full responses to these questions as well. For the 5796 reports referenced above, after the reduction from the background survey, this dataset has 5634 usable entries. The number of non-zero noise dose rows differs for different metrics and different ambient thresholds, but not by more than 200 observations in either direction. Table 6-15 summarizes the total usable sample size for each of the acoustic metrics.

Table 6-15 Total usable sample size for different metrics

Metric	Sample Size
PL	5634
ASEL	5634
BSEL	5727
CSEL	5834
DSEL	5829
ESEL	5732
FSEL	5829
LLZd	5629
LLZf	5629
ISBAP	5727
PNL	5629
MxPSF	5834

The most important variables in the development of this model are the noise dose (the calculation of which is described above), and the annoyance response. The annoyance response is defined as the dichotomized highly annoyed (HA) vs. not highly annoyed (not HA) rating of an event, which is derived from the ordinal rating provided by the respondent. A rating of HA corresponds to a respondent rating of "Annoyed" or "Extremely annoyed" on the original 5-point Likert annoyance scale. Respondent ratings of "Not at all annoyed", "Somewhat annoyed", and "Moderately annoyed" all correspond to a not HA response. Appendix J and Appendix K provide participant survey summaries and screen shots that illustrates these five ordinal annoyance choices. In addition to these variables, the respondent ID was accounted for. This variable served as a marker for a random intercept for every different individual in the dataset. This serves to induce a correlation structure in the data, such that all responses from the same individual will be modeled with a correlation, and therefore not treated as independent responses. The model also accounts for demographic factors and attitudes as measured in the background survey. The QSF18 model is of the form:

$Y = XB + B_M Met + ZA + E$, where:

Y is the binary annoyance response being modeled (HA or not HA), which is a function of:

Non-noise co-variables

X is a matrix of covariates that help to explain the annoyance response (see Table 6-21)

B is a px1 vector of coefficients to be estimated

Noise effects

B_M is a coefficient indicating the effect of the objective measure of noise

Met is a vector of the objective measures of noise

Random effects

Z is an nxk matrix of random effects (e.g. coding for participant specific intercepts, see Table 6-20 and Table 6-23 and discussion thereof below)

A is a kx1 vector of random variables (e.g. participant specific intercepts, see Table 6-20 and Table 6-23 and discussion thereof below)

E is an nx1 vector of estimation errors

In particular, since the responses were dichotomized, this model is a *random coefficients logistic regression model*. The model was fit in SAS PROC GLIMMIX. In the 'class' statement, the categorical variables that were declared are the respondent's ID, quadrant, reminder group, and gender. The model could have included whether or not they had children under 6, the education level, quality of hearing, and home and neighborhood noisiness, but it was difficult to achieve model convergence with these extra categorical covariates included. It was possible to achieve model convergence (after manual tuning[******] to convergence criteria on likelihood thresholds) with the inclusion of the quantitative covariates age, household size, and the two constructed scale variables for habituation and noise sensitivity. In general, it is strongly suspected that the relative dearth of HA responses made the effect of many extra covariates difficult to estimate. This lack of variability in the responses, coupled with the expectation that the vast majority of it is attributable to the noise dose, makes the tiny effect of the various covariates difficult to discern for this data set. In the absence of model convergence, the significance of the parameters can not be definitively determined, however these results do not contradict the possibility that they are not significant.

As these models were all run on individual responses, the model outputs all give insight into a single individual's propensity to rate a single event as highly annoying. There are some respondents who in general demonstrate a higher propensity to rate events as HA, even at lower noise dose levels, than other (presumably less noise sensitive) individuals. From a visualization perspective, this makes such individual level modeling misaligned with dose response pictures from prior studies (see Table 6-35), where the proportion of a group who were highly annoyed generally displayed a roughly monotonic increase as a function of dose.

In order to aggregate individual level non-monotonicity of the estimated dose-response function, the model outputs are smoothed with a locally estimated scatterplot smoothing (LOESS) procedure (in SAS

[******] The noise_650ms_5dB_SE_analysis.sas code delivered to NASA provides full detail of the manual tuning process. Wherever the code has "pconv=xx", it is manually tuning convergence with a delta of xx.

PROC LOESS); in applying LOESS to the model prediction rather than to the observed dichotomous responses, we can account for the covariates and correlation structure in the data. The LOESS algorithm is a nonparametric regression technique that takes a window around every dose value and accounts for all observed responses in that window, aggregating them in a weighted fashion, where the weights are inversely related to the distance from the specific dose value. The amount of smoothing in SAS is estimated from the data with a generalized cross validation approach. A detailed discussion of smoothing is presented in Appendix S. In summary, a LOESS smoother was applied to individual model predicted data, and a second LOESS smoother was applied to aggregated data, as follows.

Specific smoothing parameters for single event and daily does analyses are provided in Table S-1 and Table S-14, respectively. While this produces a smoother predicted probability that an individual would respond to a noise dose with an (ordinal) rating that corresponds to the Highly Annoyed (binary) rating it may not in all cases induce a strictly monotone increasing shape. For the sake of visualization, yet another LOESS smoother is applied to the smoothed predicted probabilities to yield the desired effect, this time in R, where the amount of smoothing incorporates neighborhoods with 2/3 of the data by default. This consistently yields smooth, monotone increasing functions for the dose-response fitted model. This approach, which applies smoothing twice, uses the optimized smoothing parameter as determined by GCV, as opposed to a single- smoothing approach which could be considered more subjective with regard to parameter selection.

In summary, the process is to: (1) fit logistic regression to the HA/not HA data; (2) estimate the probability of high annoyance for each observation using the original design matrix; (3) smooth the predicted points using LOESS for visuals and determine the smoothing parameter based on GCV in SAS; and (4) smooth again using LOESS but using the default of 0.66 for the smoothing parameter. Existing information is insufficient to quantify the uncertainty introduced by the multi-layered approach used. Figure 6-37 provides an example of the variation of LOESS bin width as a function of level for step (3), where the smoothing parameter is based on GCV in SAS. As shown, the bin widths are smallest, typically around 1 to 3 dB, in the middle of the range of levels, where the quantity of data points is greatest. At the extremes, bin width is maximum, typically around 8 to 10 dB, due to the more sparse data points in those regions. Figure 6-38 provides an example of the variation of LOESS bin width as a function of level for step (4), where the smoothing parameter is defaulted to 0.66, a much larger neighborhood. Again, the bin widths are smallest in the middle of the range of levels, where the quantity of data points is greatest, typically around 12 to 14 dB. At the extremes, bin width is maximum, due to the more sparse data points in those regions, typically around 19 to 20 dB.

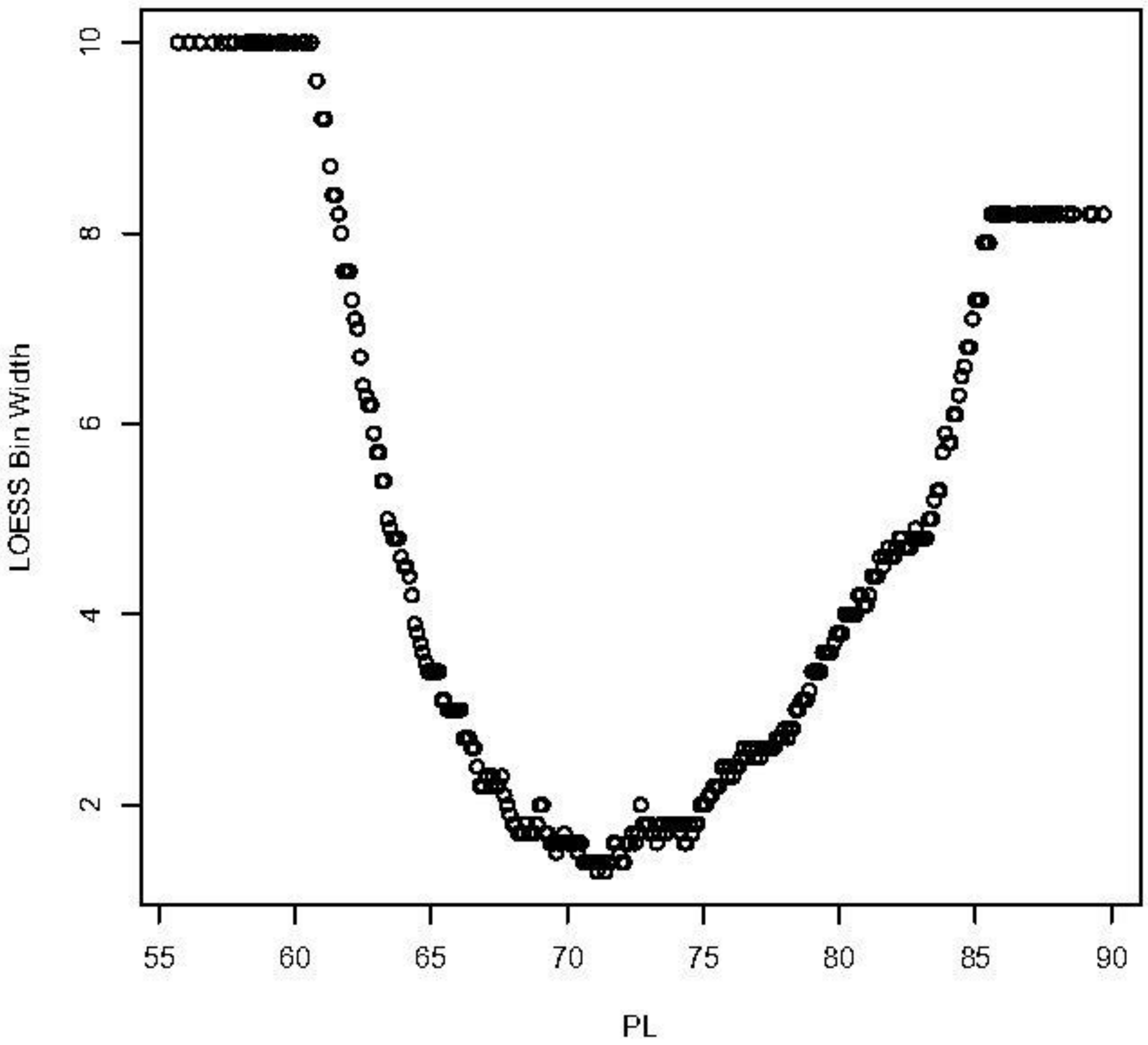


Figure 6-37 Example LOESS bin width as a function of level for first smoothing [step (3)], where the smoothing parameter is based on GCV in SAS

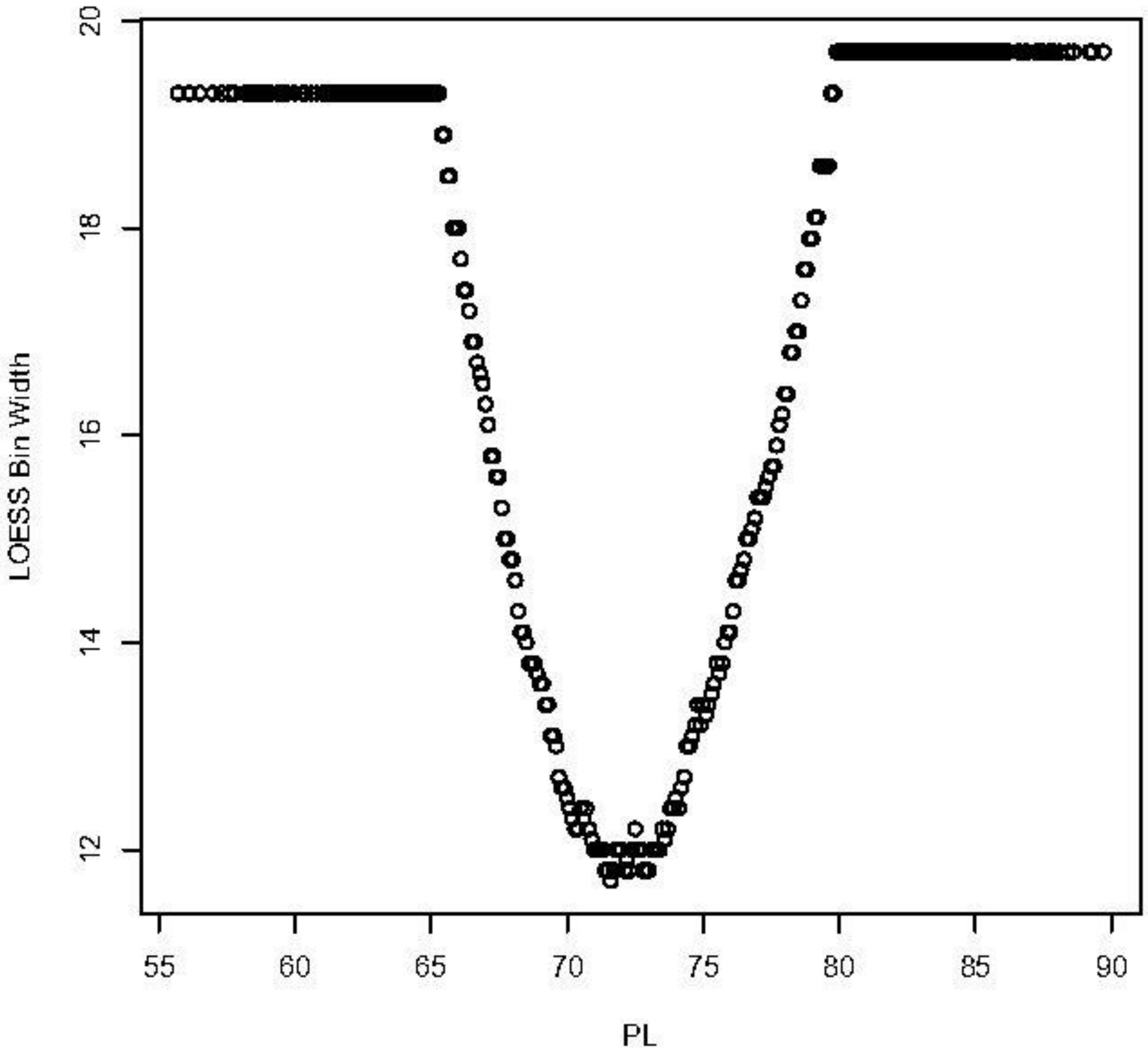


Figure 6-38 Example LOESS bin width as a function of level for second smoothing [step (4)], where the smoothing parameter is defaulted to 0.66

These models and LOESS smoothers were run on all metrics, each calculated in all six ways (each of the two windows, and for each of the three ambient thresholds) for both single event annoyance and daily summary annoyance Table 6-16 through Table 6-23 present an abbreviated version of the SAS output from a single analysis for single event annoyance, using the PL metric calculated with the 650ms window and using a 5dB ambient threshold. The analysis fits the random coefficients logistic regression model described above to the full set of 5634 SE reports that have recorded noise dose and a full set of covariates. The discussion following these tables gives a succinct description of the most salient pieces of information to take away; a full discussion on what information is contained in the tables, and details on their uses and calculations, can be found in the PROC GLIMMIX documentation (SAS Institute Inc. 2018. SAS/STAT® 15.1 User's Guide. Cary, NC: SAS Institute Inc.).

Table 6-16 Model information

Model Information	
Data Set	QSF18.NOISE13_PL_NO0
Response Variable	HA
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	PARTICIPANT_ID
Estimation Technique	Residual PL
Degrees of Freedom Method	Between-Within

Table 6-17 Sample size

Number of Observations Read	5796
Number of Observations Used	5634

Table 6-18 Response tabulation

Response Profile		
Ordered Value	**HA**	**Total Frequency**
1	HA	61
2	Not HA	5573
The GLIMMIX procedure is modeling the probability that HA='HA'.		

Table 6-19 Fit statistics

Fit Statistics	
-2 Res Log Pseudo-Likelihood	47161.91
Generalized Chi-Square	1144.02
Gener. Chi-Square / DF	0.20

Table 6-20 Global covariance parameter estimate

Covariance Parameter Estimates			
Cov Parm	**Subject**	**Estimate**	**Standard Error**
Intercept	PARTICIPANT_ID	2.2955	0.5900

Table 6-21 Parameter estimates and inference

Solutions for Fixed Effects								
Effect	**quad**	**group**	**gender**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept				-63.5613	35.5455	360	-1.79	0.0746
quad	QUADRANT A			0.2046	0.4916	360	0.42	0.6775
quad	QUADRANT B			-1.4376	0.7645	360	-1.88	0.0609
quad	QUADRANT C			-17.9527	7078.25	360	-0.00	0.9980
quad	QUADRANT D			0	.	.	.	.
group		Email - No Reminder		-0.09779	0.6939	360	-0.14	0.8880
group		Email - Reminder		-0.3990	0.5564	360	-0.72	0.4738
group		Text - No Reminder		-0.3869	0.8657	360	-0.45	0.6552
group		Text - Reminder		0	.	.	.	.
gender			Female	0.1506	0.4927	360	0.31	0.7600
gender			Male	0	.	.	.	.
birth_year				0.02624	0.01821	360	1.44	0.1504
HH				-0.06356	0.2190	360	-0.29	0.7718
long_liv				-0.00265	0.01657	360	-0.16	0.8731
hab				-0.5571	0.2903	360	-1.92	0.0558
sens				1.0735	0.3564	360	3.01	0.0028
PL_num				0.09426	0.02252	5260	4.19	<.0001

Table 6-22 Omnibus hypothesis testing

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
quad	3	360	1.57	0.1951
group	3	360	0.20	0.8935
gender	1	360	0.09	0.7600
birth_year	1	360	2.08	0.1504
HH	1	360	0.08	0.7718
long_liv	1	360	0.03	0.8731
hab	1	360	3.68	0.0558
sens	1	360	9.07	0.0028
PL_num	1	5260	17.52	<.0001

Table 6-23 Individual level covariance parameter estimates (abbreviated)

Solution for Random Effects Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	PARTICIPANT_ID 100258	-0.5464	1.2499	5620	-0.44	0.6620
Intercept	PARTICIPANT_ID 100948	-0.07940	1.4593	5620	-0.05	0.9566
Intercept	PARTICIPANT_ID 137709	-0.1687	1.4052	5620	-0.12	0.9044
Intercept	PARTICIPANT_ID 141784	-0.07433	1.4638	5620	-0.05	0.9595
Intercept	PARTICIPANT_ID 142241	2.9963	0.7299	5620	4.10	<.0001
Intercept	PARTICIPANT_ID 213700	3.0882	0.6733	5620	4.59	<.0001
Intercept	PARTICIPANT_ID 214043	-0.5159	1.2525	5620	-0.41	0.6804
Intercept	PARTICIPANT_ID 215176	-0.4292	1.2808	5620	-0.34	0.7375

Table 6-16 provides some basic summaries of the model fitting options, and Table 6-17 notes that of the 5796 reports with a noise dose, 5634 also have full sets of covariates for analysis. Table 6-18 outlines the scarcity of HA responses, and Table 6-19 provides some fit statistics; notably, the last line in Table 6-19 estimates the residual dispersion, and these fit statistics should not necessarily be used to judge the adequacy of a model or to compare models, even those that are nested. Table 6-20, showing the global covariance parameter estimate, corresponding to a leading column of all 1s in the Z matrix of the model formulation, indicates that there is some significant correlation (as is typical, one can simply compare the estimate to the standard error and compare this ratio to a standard Normal distribution) between responses without parsing out individual level differences in this correlation structure; skipping to Table 6-23 (much abbreviated) in the output (corresponding to dummy-coded 0/1 columns in the Z matrix for individual ID effects) indicates that for the majority of respondents (not all shown for sake of brevity), the individual level correlation structure does not differ from the globally estimated one, while a small handful do have a slightly different pattern (the same comparisons described for Table 6-20 can be used here). Table 6-21 and Table 6-22, in between these covariance parameter estimate tables, show the estimates, standard errors, and hypothesis testing information (t and F test statistics and p-values) for the effects of the demographic covariates and noise dose.

The main conclusion that can be drawn is that the noise dose appears to account for a lot of variability in the annoyance, and the only other covariate to reach the conventional 5% significance level is the noise sensitivity scale, while the habituation scale is close at p=.0558. Figure 6-39 shows the original, LOESS smoothed visualization of the model predicted dose-response relationship of propensity for annoyance as a function of noise (in red), along with the twice smoothed (in black) dose-response curve for single event annoyance, using the PL metric calculated with the 650ms window and 5dB ambient threshold. The caption of Figure 6-39 notes the SAS PROC LOESS smoothing value used to generate the red curve; the smoothing for all other metrics is listed in Appendix S.

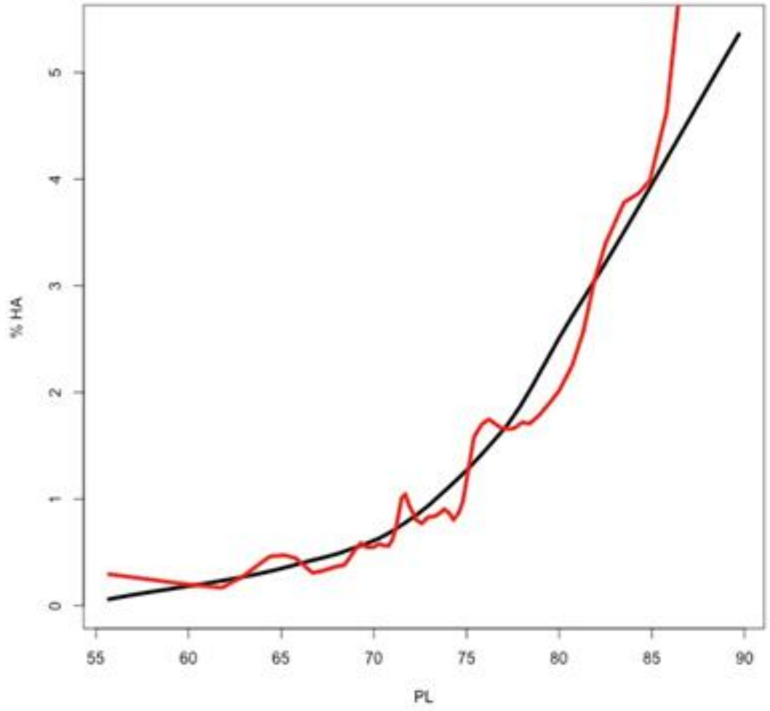


Figure 6-39 Once smoothed (red) and twice smoothed (black) dose-response curves for single event annoyance using the PL metric calculated with the 650ms window and 5dB ambient threshold. The PROC LOESS smoothing parameter is 0.1067625

It is also of interest to consider how this smoothed model fit adheres to the observed data. The percentage of respondents who were highly annoyed at a particular dose was calculated. Since each respondent gets an individualized dose calculation (depending on their locations at the event times), no two people will necessarily be exposed to exactly the same dose. Individuals with similar dose events were considered as a group. The observations were binned into neighborhoods with similar doses. Both ambient noise in the field and the distribution of the data was considered in selecting the increment in dB level used to define the bin width for similar dose. Consideration was given to utilizing a 2 dB step as a noticeable difference for the bin width, but that resulted in insufficient data within neighboring bins. Due to the limited number of HA responses to bin, and acknowledging that there was ambient noise in the field environment, the decision was made to use a 4 dB bin width, which allowed for sufficient data points within neighboring bins. The bins were defined using a fixed length and have variable numbers of events in each bin. In Figure 6-40, the same smoothed dose-response curve is shown from Figure 6-39, with both confidence bounds, which give a range of plausible values for %HA at a specific dose that one can believe with 95% confidence, and the raw annoyance data from the binned dose levels. For the nine bins, Table 6-24 itemizes the number of reports and HA responses per dose bin.

Table 6-24 Similar dose single event binning

Bin Midpoint (PLdB)	Number of Reports	Number of HA Responses
55.7	15	1
59.7	145	0
63.7	446	4
67.7	1198	3
71.7	1547	6
75.7	1165	12
79.7	684	16
83.7	478	17
87.7	92	1

The large outlying value in the first bin is seemingly a consequence of having a very small number of total dose values in this range, and one individual who rated the event as HA. Outside of this, the smoothed curve fits the data very well.

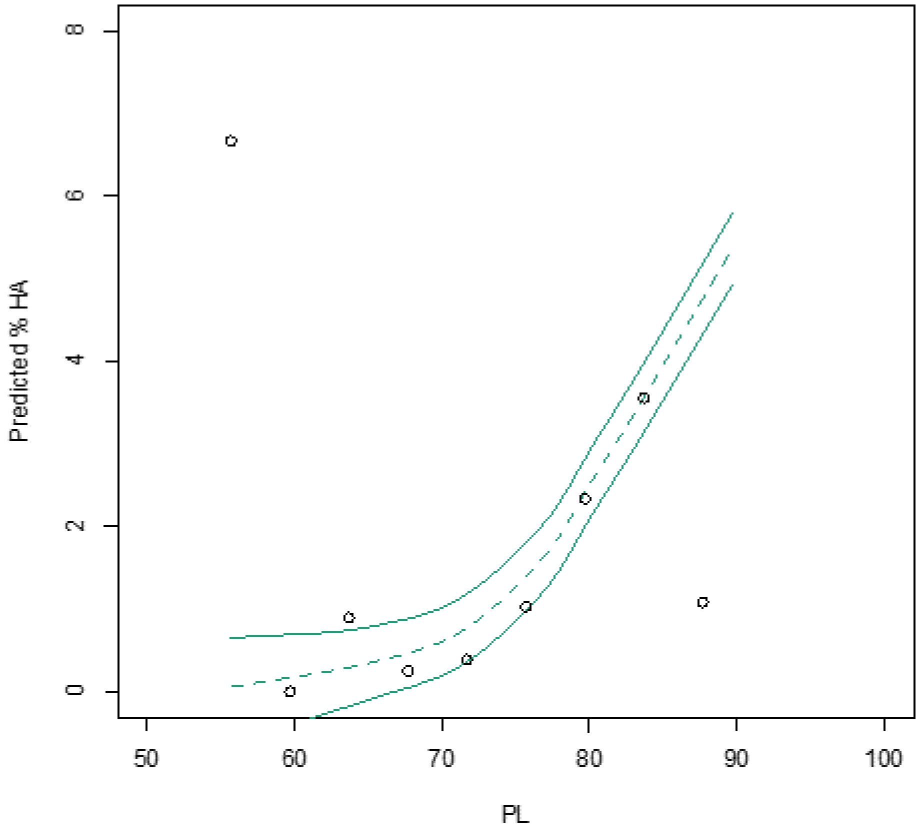


Figure 6-40 Smoothed dose-response curve for SE reports with PL (650ms, 5dB ambient), with confidence bounds on the LOESS smoother and raw annoyance data

Similar analyses were performed for all of the metrics used to calculate the noise dose; Appendix S contains 11 plots like the one above for the remaining 11 metrics (using the same window and ambient conditions for metric calculation). In addition, tables providing the percentage highly annoyed, number of highly annoyed, and number of data points are provided for each frequency bin for each of the single event metrics in tables in Appendix S. A table providing the smoothing parameters used for each single event meteric is also provided in Appendix S. In order to compare the metric relationship with its propensity to register an HA response, all of the smoothed model fits were plotted together (Figure 6-41). All of the metrics have the same general shape of relationship, suitably shifted to reflect the nature of the metric. Notably, all curves demonstrate a similar knee in the curve, indicating that there exists a certain metric level beyond which annoyance starts to increase more sharply. As an example, the PL shape seems to indicate that at levels around 75, one can expect %HA to start growing above 1%.

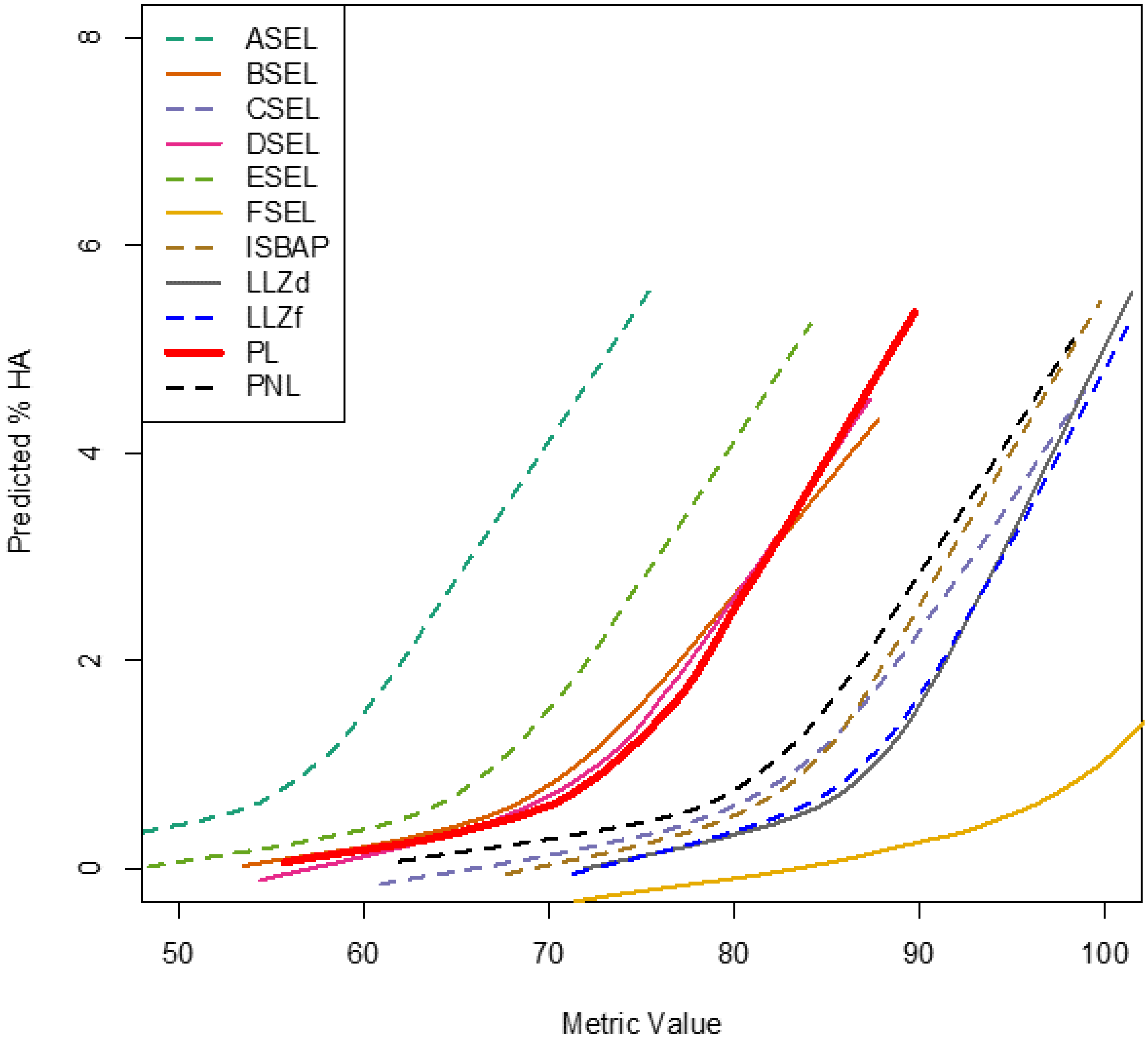


Figure 6-41 Smoothed dose-response curve for single event annoyance for all metrics calculated with the 650ms window and 5dB ambient threshold

6.3.1.2 Daily Summary Analysis

Section 6.2.1.2 notes that the initial subjective dataset included 3411 DS reports. Not all of these 3411 reports can be used in developing a dose response relationship. In order to be included, a respondent must be locatable by any of the several methods described in Section 6.1.2 and therefore have a measurable daily noise dose. For example, for the PLDN metric at the 5dB ambient threshold using the 650ms window, 2131 reports had usable noise dose data. In addition to a noise dose, the analysis also incorporated the same background survey variables as in the SE analysis. After including only those reports also with complete background information, the final analysis is conducted on 2058 reports. The model used for the DS analysis is identical to the model used for the SE analysis as described in Section 6.3.1.1, so it will not be described further here. The total usable sample sizes for the DS analyses are given in Table 6-25 below.

Table 6-25 Total usable sample size for different metrics

Metric	Sample Size
PLDN	2058
DNL	1979
BDNL	2056
CDNL	2059
DDNL	2056
EDNL	2059
FDNL	2056
DailyLLZd	2053
DailyLLZf	2053
DailyISBAP	2056
DailyPNL	2054

This section describes the results from the single analysis for DS annoyance, using the PLDN metric calculated with the 650ms window and using a 5dB ambient threshold. The analysis fit the random coefficients logistic regression model described above to the full set of 2058 DS reports that have recorded noise dose and full covariate information. An abbreviated version of the SAS output from this model run is in Table 6-26.

Table 6-26 Model information

Model Information	
Data Set	QSF18.NOISE13_PL_NO0
Response Variable	HA
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	PARTICIPANT_ID
Estimation Technique	Residual PL
Degrees of Freedom Method	Between-Within

Table 6-27 Sample size

Number of Observations Read	2131
Number of Observations Used	2058

Table 6-28 Response tabulation

Response Profile		
Ordered Value	**HA**	**Total Frequency**
1	HA	17
2	Not HA	2041
The GLIMMIX procedure is modeling the probability that HA='HA'.		

Table 6-29 Fit statistics

Fit Statistics	
-2 Res Log Pseudo-Likelihood	18187.43
Generalized Chi-Square	304.90
Gener. Chi-Square / DF	0.15

Table 6-30 Global covariance parameter estimate

Covariance Parameter Estimates			
Cov Parm	**Subject**	**Estimate**	**Standard Error**
Intercept	PARTICIPANT_ID	1.7981	0.8806

Table 6-31 Parameter estimates and inference

Solutions for Fixed Effects								
Effect	**quad**	**group**	**gender**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept				-20.6639	54.7683	374	-0.38	0.7062
quad	QUADRANT A			1.3982	0.7940	374	1.76	0.0790
quad	QUADRANT B			-1.6322	1.4734	374	-1.11	0.2687
quad	QUADRANT C			-5.3246	40.9255	374	-0.13	0.8966
quad	QUADRANT D			0	.	.	.	.
group		Email - No Reminder		-1.2251	0.8233	374	-1.49	0.1376
group		Email - Reminder		-1.3742	0.8812	374	-1.56	0.1197
group		Text - No Reminder		-1.8767	1.1955	374	-1.57	0.1173
group		Text - Reminder		0	.	.	.	.
gender			Female	0.8655	0.7653	374	1.13	0.2588
gender			Male	0	.	.	.	.
birth_year				0.007655	0.02810	374	0.27	0.7855
HH				0.04255	0.3576	374	0.12	0.9054
long_liv				0.005822	0.02245	374	0.26	0.7955
hab				-1.1036	0.4533	374	-2.43	0.0154
sens				0.6497	0.4317	374	1.50	0.1332
PL				0.07115	0.05523	1670	1.29	0.1978

Table 6-32 Ominibus hypothesis testing

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
quad	3	374	2.34	0.0731
group	3	374	1.43	0.2343
gender	1	374	1.28	0.2588
birth_year	1	374	0.07	0.7855
HH	1	374	0.01	0.9054
long_liv	1	374	0.07	0.7955
hab	1	374	5.93	0.0154
sens	1	374	2.26	0.1332
PL	1	1670	1.66	0.1978

Table 6-33 Individual level covariance parameter estimates

Solution for Random Effects						
Effect	**Subject**	**Estimate**	**Std Err Pred**	**DF**	**t Value**	**Pr > |t|**
Intercept	PARTICIPANT_ID 100258	-0.1080	1.2796	2044	-0.08	0.9327
Intercept	PARTICIPANT_ID 112870	-0.4968	1.1934	2044	-0.42	0.6772
Intercept	PARTICIPANT_ID 115081	-0.06350	1.3030	2044	-0.05	0.9611
Intercept	PARTICIPANT_ID 126677	1.3269	1.1417	2044	1.16	0.2453
Intercept	PARTICIPANT_ID 126947	-0.02199	1.3265	2044	-0.02	0.9868
Intercept	PARTICIPANT_ID 142241	2.5066	0.9901	2044	2.53	0.0114
Intercept	PARTICIPANT_ID 144270	-0.1104	1.2754	2044	-0.09	0.9310
Intercept	PARTICIPANT_ID 443892	-0.00027	1.3407	2044	-0.00	0.9998
Intercept	PARTICIPANT_ID 747832	-0.01485	1.3314	2044	-0.01	0.9911
Intercept	PARTICIPANT_ID 753190	1.8517	1.0242	2044	1.81	0.0708

Table 6-26 provides some basic summaries of the model fitting options, and Table 6-27 notes that of the 2131 reports with a noise dose, 2058 also have full sets of covariates for analysis. Table 6-28 outlines the scarcity of HA responses (only a total of 17 HA reports of the 2058 available), and Table 6-29 provides some fit statistics. Table 6-30, showing the global covariance parameter estimate, indicates that there is some correlation between DS responses without parsing out individual level differences in this correlation structure. Skipping ahead to Table 6-33 (much abbreviated) the output indicates that for the majority of respondents (not all shown for sake of brevity), the individual level correlation structure does not differ from the globally estimated one, while a small handful (e.g. ID 142241) do have a different pattern. Table 6-31 and Table 6-32, in between these covariance parameter estimate tables, show the estimates for the effects of the demographic covariates and noise dose.

The conclusion to be drawn is that the noise dose does not have a statistically significant relationship with the annoyance outcome. This is driven in part by the lack of HA reports from which to estimate such a relationship. The only covariate to reach the conventional 5% significance level is the habituation scale, while the noise sensitivity scale for this analysis is not even close to significance with a p=.1332.

Section 6.3.1.1 described a detailed process used to smooth out individual-level differences in the relationship between dose and annoyance. However, as the DS analysis yields no significant relationship between dose and annoyance for any of the metrics, there is no noticeable non-monotonicity to the fitted curves for the daily summaries. Nevertheless, for completeness one can still plot the aggregated model fit with confidence bounds to see how well the observed data adhere to the fitted curve. As such, the percentage of respondents who were highly annoyed at a particular dose was calculated. Since each respondent gets an individualized dose calculation (depending on their daily dose given their locations throughout the day), no two people will necessarily be exposed to exactly the same daily dose. Individuals with similar daily dose levels were considered as a group. The observations were binned into neighborhoods with similar doses. Due to the limited number of HA responses to bin, as with the SE analysis, the decision was made to use a 4 dB bin width, which allowed for sufficient data points within neighboring bins. The binned data for the PLDN metric is summarized in Table 6-34, and plotted in Figure 6-42 below over the smoothed dose-response (dashed line) with confidence bounds (solid lines). The raw data fits the estimated curve quite well. Note that the even more infrequent HA DS responses gives lower estimated probabilities of registering an HA report, and as a result, the LOESS curve dips below 0 for the lowest dose values.

Table 6-34 Binned PLDN data

Bin Midpoint	Number of Reports	Number of HA Responses
7.3	3	0
11.3	39	0
15.3	44	0
19.3	141	0
23.3	310	0
27.3	392	3
31.3	418	2
35.3	427	9
39.3	342	3

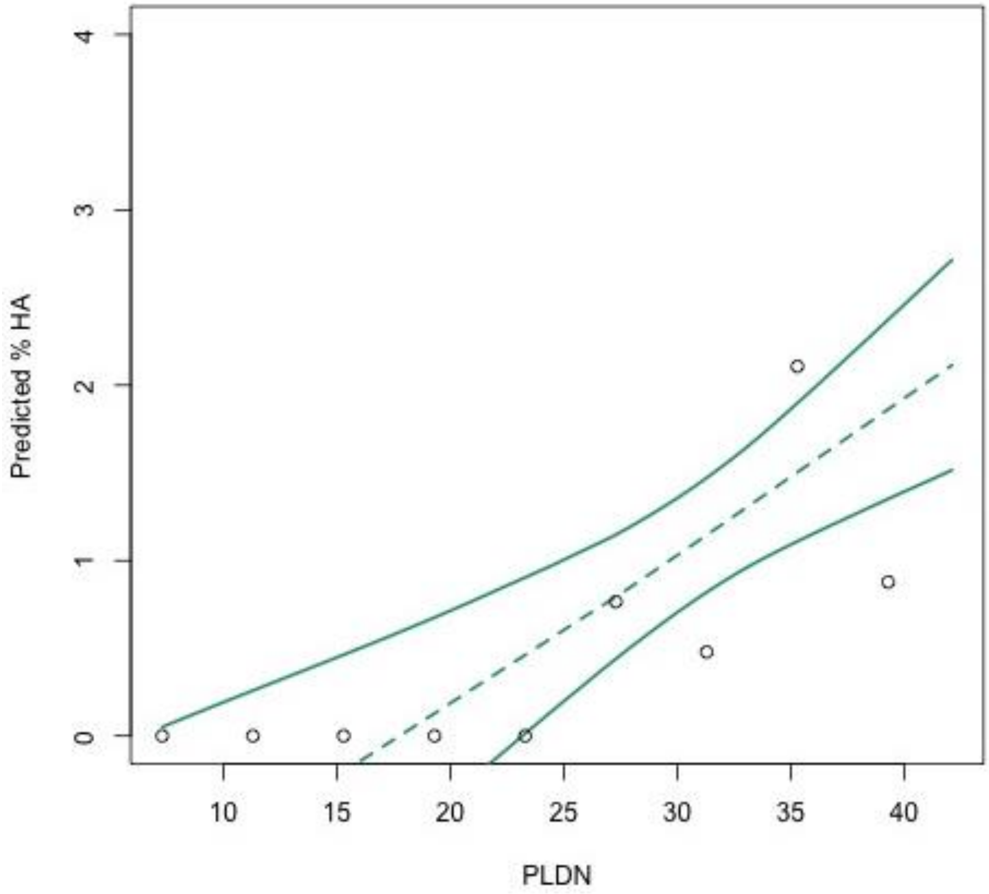


Figure 6-42 Smoothed dose-response curve for DS reports with PLDN (650ms, 5dB ambient) with confidence bounds on the LOESS aggregator and raw annoyance data

Similar analyses were performed for all of the metrics used to calculate the noise dose; Appendix S contains results for the remaining 10 metrics (using the same window and ambient conditions for metric calculation). In addition, tables providing the percentage highly annoyed, number of highly annoyed, and number of data points are provided for each cumulative dose bin for each of the daily summary metrics in tables in Appendix S. A table providing the smoothing parameters used for each daily summary metric is also provided in Appendix S. In order to compare their relationship with the propensity to register an HA response, all of their aggregated model fits were plotted together (Figure 6-43). Figure 6-43 shows that all of the metrics have the same general shape of relationship, suitably shifted to reflect the nature of the metric.

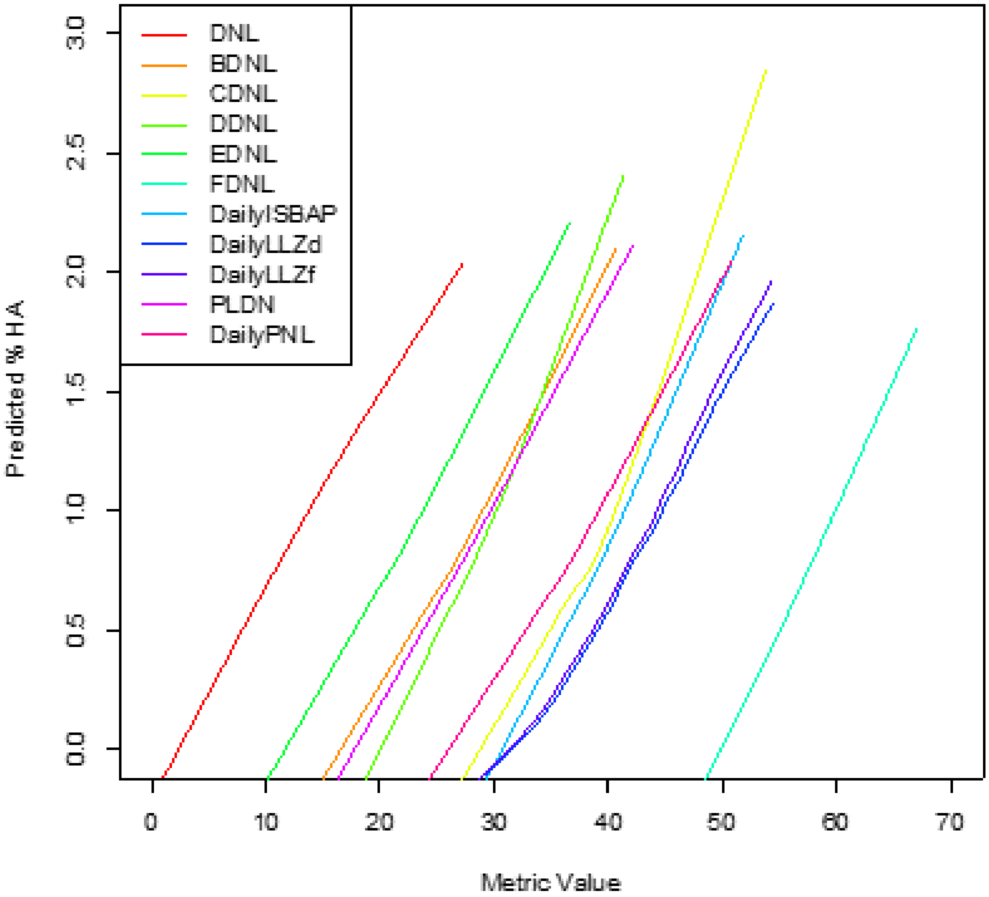


Figure 6-43 Smoothed dose-response curve for DS annoyance for all metrics calculated with the 650ms window and 5dB ambient threshold

6.3.2 Statistical Correlation and Dose Response Analysis Lessons Learned

Data collection during QSF18 was successful. Though the dose response relationship was significant, it was much less pronounced than in previous studies (see Section 6.4), ostensibly due to the very different noise source in the current study. While the DS data analysis was also successful in the sense that a meaningful model was estimated and analyzed, it did not reveal a statistically significant association between noise dose and cumulative annoyance. After also fitting models to the 5-point annoyance scale to see if any significant patterns hold with that type of response distribution, another way to think about this would be considering ways to modify future efforts to discover a significant association on the dichotomous scale. However, it is important to consider that it is possible that individuals who are exposed to realistic cumulative levels of noise due to thumps from a viable commercial aircraft might simply not notice/be annoyed by this noise source. Essentially, this is what the data suggests. While one must always be careful in interpreting null findings (e.g. a lack of statistical significance does not imply that a null hypothesis is true), even if the association existed and was masked by, say, low power of the analysis, one must wonder if it is practically meaningful. Since none of the estimated %HA for DS were above 2%, which is very much in line with the observed scatter, it is possible that any real relationship existing in this range of tested cumulative levels would not yield meaningful %HA.

6.4 Comparison with Previous Studies

The QSF18 sonic thumps are lower in level than levels used in past sonic boom research. The observed annoyance response falls significantly below the majority of annoyance data for prior field tests, as indicated below. For the QSF18 field test, the respondents went about their normal day, moving freely, and each respondent had an individualized dose calculation based on field measurements and noise modelling estimations. To analyze the QSF18 data, respondents with similar dose events were considered as one group, or bin. This was done by binning the annoyance response observations with a similar noise dose. Both ambient noise in the field and the distribution of the data was considered in selecting the increment in dB level used to define the bin width for similar noise dose. Due to the limited number of HA responses to bin, and acknowledging that there was ambient noise in the field environment, the decision was made to use a 4 dB step, which allowed for sufficient data points within neighboring response bins. Most prior field tests assumed that respondents were in one location, typically at home and the noise dose was considered to be uniform across a small geographic region.

Community annoyance to sonic boom overflights has been assessed by prior research studies over the past several decades. Two early studies were conducted at Edwards Air Force Base and in Oklahoma City. The National Academy of Science, National Research Council, Committee on Hearing Bioacoustics and Biomechanics [CHABA, 1981] released a report on community response to high-energy impulsive noise that included military impulsive noise as well as sonic booms. The CHABA findings and procedures for assessing high-energy impulsive noise were incorporated into ANSI S12.9 Part 4 [1996], with similar documentation in ISO 1996-1. Comparisons of some of these prior findings to the CHABA data [Maglieri, *et al.*,2014] are shown in Figure 6-44.

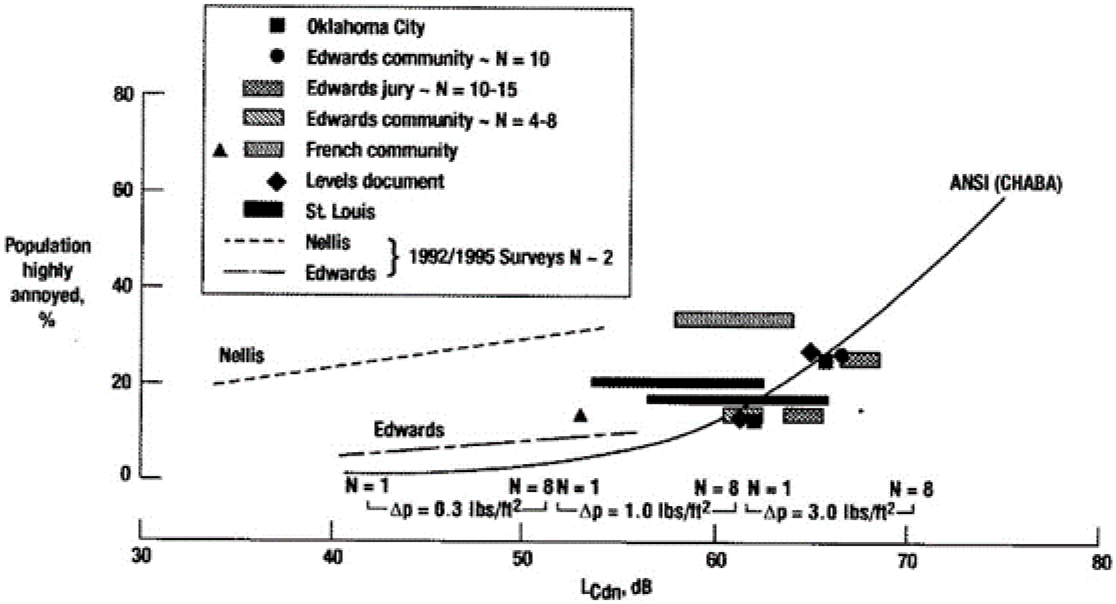


Figure 6-44 Earlier studies compared to CHABA and ANSI. Source: [Maglieri, *et al.*,2014]

The NASA sponsored Waveform and Sonicboom Perception and Response (WSPR) Program conducted a field study of subjective response to noise from multiple low-amplitude sonic booms. The test was conducted as an assessment of data acquisition and analysis methods. The WSPR experiment involved exposing subjects living in the Edwards Air Force Base (EAFB) housing area to two weeks of low-amplitude sonic booms while recording their responses via structured surveys. EAFB is an active base with frequent supersonic operations so the test area is also exposed regularly to other non-WSPR sonic boom events, and the community residents are considered to be acclimated to sonic booms. A schedule of sonic boom exposure was designed covering a CDNL range from 41 to 67 dB. The WSPR dataset spanned the exposure levels of prior research efforts. It was anticipated that future tests would include levels that were much lower in level than past efforts. This could result in response data that was skewed to the left and appropriate for analysis by non-parametric methods. As such, the statistical analysis included the use of a select set of non-parametric methods to establish their potential application for future efforts. Subjective data was collected before, during and after the test period to support the analysis and assess the methods of data collection. Survey instruments consisted of a baseline survey, a single event survey and a daily summary survey.

The data from the WSPR low boom field test provides a measure of the acceptance of low booms in an acclimated community. A comparison with the findings of previous studies, several of which are summarized [CHABA, 1996] indicate that the annoyance levels for sonic booms at CDNL levels of exposure below 60 dB are generally lower. The WSPR analysis relating percent highly annoyed (%HA) to the cumulative noise showed high correlation for the cumulative noise metrics with the %HA response. Kendall's Tau-b correlations indicated that the five modes of single event annoyance ranked interference as the strongest driver of annoyance (.76), followed by startle (.70), loudness (.55), vibration (.45) and rattle (.42) and the four modes for cumulative daily annoyance also ranked interference highest (.75) followed by loudness (.64), vibration (.49) and rattle (.47).

The range of planned cumulative metric values for QSF18 is compared with the values presented in WSPR [Page, *et al*., 2014] and with the findings of previous studies, several of which are summarized in later CHABA report [CHABA, 1996]. The 1996 CHABA report presented findings from five prior studies. The CHABA reported cited findings from two sonic boom studies, Oklahoma City [Borsky, 1965], and NASA [Fields *et al*., 1994], and three blast noise studies, from Ft. Bragg [Schomer, 1981], Ft. Lewis [Schomer, 1985], and Sweden reported by [Rylander and Lundquist, 1996]. Table 6-35 presents the range of cumulative metrics for the CHABA [1996] data compared to WSPR 2011 data [Page, *et al*., 2014] and the QSF18 CDNL data.

Table 6-35 Comparison of CDNL impact across prior field tests

Source	Test	Approx. CDNL
F-18 LBDM	Page *et al.*,2018 Galveston TX	20.7 – 48.7
F-18 LBDM and some conventional booms	Page *et al.*, 2011 EAFB	47.4 – 56.9
Sonic Boom	Borsky 1965, OK City	54 – 64
Sonic Boom	Fields *et al.*, 1994 Nellis AFB	38 -56
Artillery	Schomer 1981, Ft. Bragg	58 – 70
Gunfire	Sweden Rylander Lundquist, 1996	41 – 68
Artillery	Schomer, 1985, Fort Lewis	51 – 65

The previous studies typically recruited respondents to be in one location for the majority of the test, and it was assumed that respondents received a similar noise dose across the noise footprint in the community. The noise data from other prior research tests provided the CDNL as a function of test day, assuming similar dose across all respondents. In contrast, an individualized noise dose was determined for QSF18 respondents. Respondents with similar dose events were considered as one group, or bin.

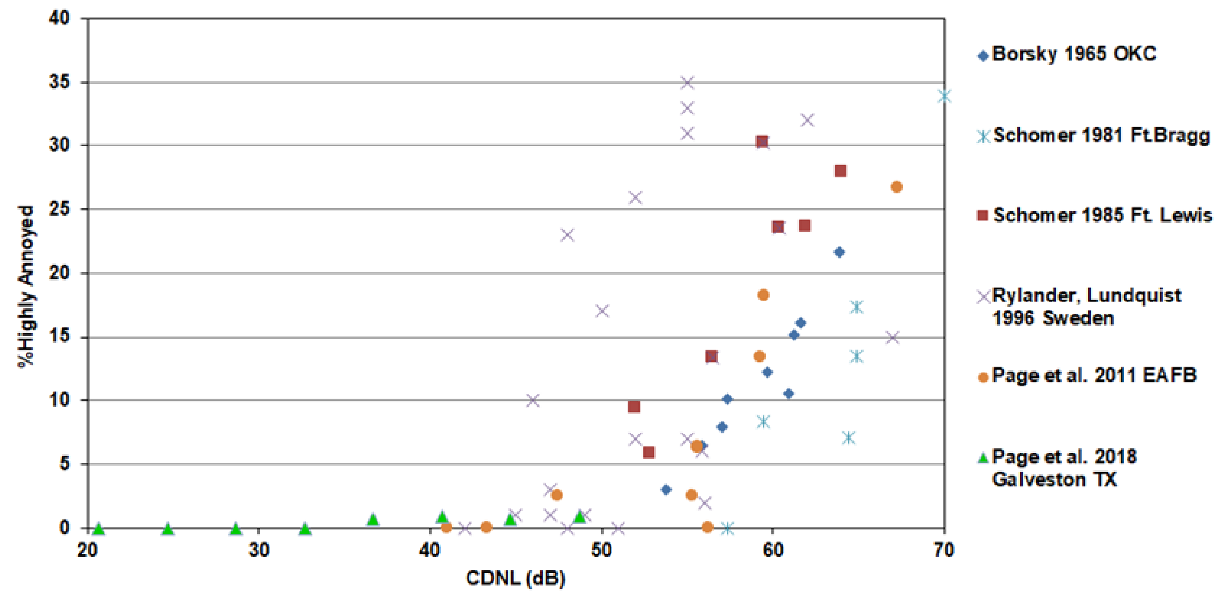


Figure 6-45 Percent highly annoyed for QSF18 (Page *et al.*) data vs. previous studies

Annoyance data was gathered for both the WSPR and CHABA prior tests, with the WSPR 2011 and the QSF18 daily levels shown in Table 6-35 expressed in terms of the percent highly annoyed as a function of the yearly averaged metric C-weighted Day-Night Level (CDNL) [Page *et al.*, 2014]. The annoyance ratings for WSPR are significantly lower than was observed in Fields [1994] or Rylander and Lundquist [1996] but are consistent with the data from the other past researchers. The QSF18 data overlapped the lower end of the WSPR data and was significantly below all of the prior data sets.

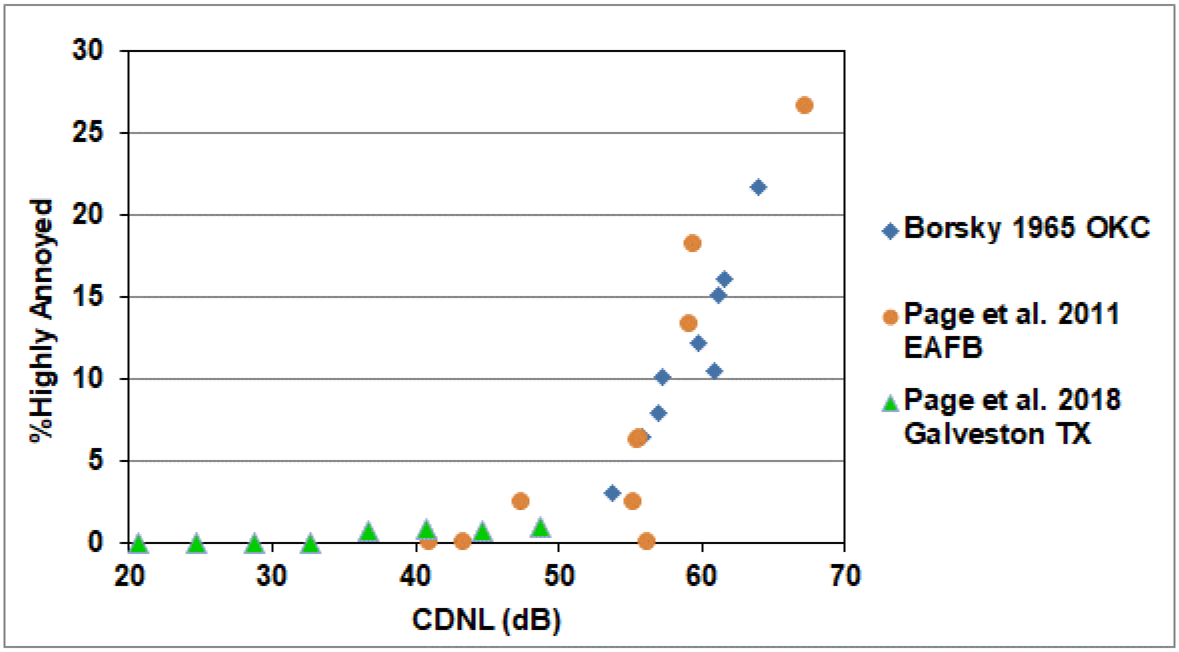


Figure 6-46 Percent highly annoyed for QSF18 data vs. Borsky (OKC) and WSPR (EAFB)

The WSPR 2011 and QSF18 teams used noise measurements obtained during the same period as the social surveys, while some of the prior studies relied on measurements from different time periods or from predicted levels. Figure 6-46 compares the percent highly annoyed for QSF18 data versus Borsky and WSPR [as cited in Page, *et al*. 2014]. The QSF18 levels fall significantly below the levels previously observed in both WSPR and prior tests, for both the level and the percent highly annoyed.

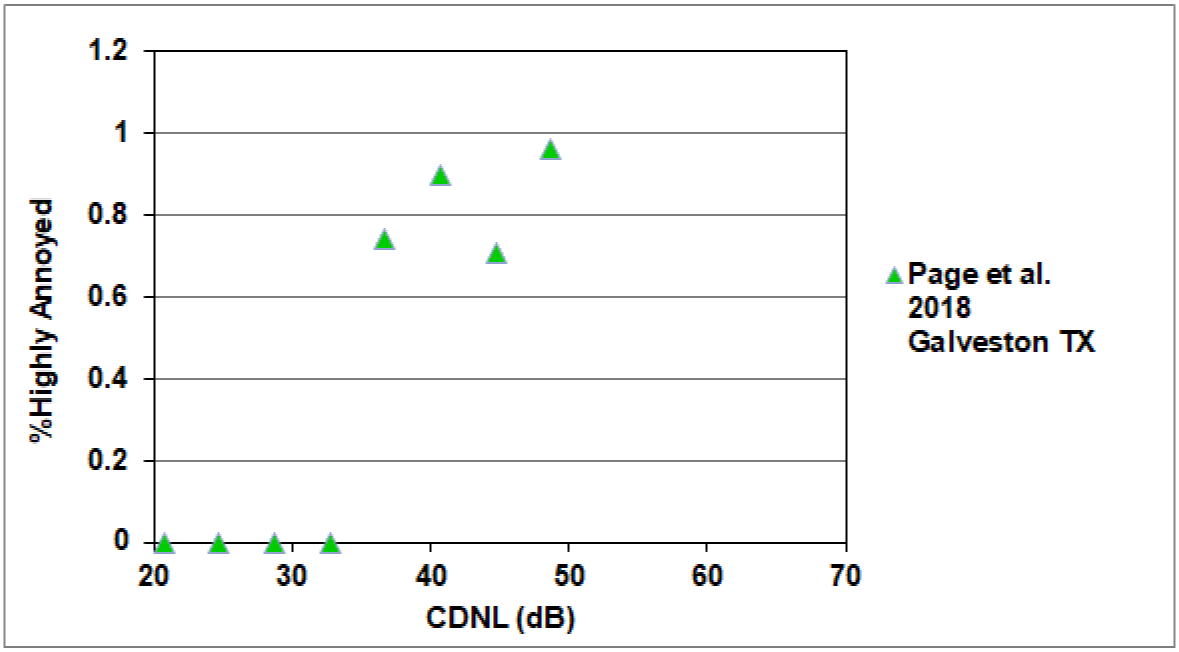


Figure 6-47 Percent highly annoyed for QSF18 data

To facilitate comparison of the QSF18 sonic thump data with prior tests, the QSF18 data was plotted against the WSPR low boom data [Page, *et al*., 2014] and the Borsky [1964] data from Oklahoma City as cited in CHABA (1996). This comparison is presented in Figure 6-46. The Borsky sonic boom and the WSPR low boom data present a similar shape to the curve. The QSF18 is so much lower in level and annoyance that it falls below the majority of both of those data sets. A close up of the QSF18 data, presented in Figure 6-47, shows that the annoyance ranged was less than 1% Highly Annoyed.

6.5 Noise Exposure and Community Response Databases

A full set of databases containing noise exposure and community response data for QSF18 has been assembled and delivered to NASA. These databases, which provide cleaned and processed data as discussed in this section, differ from the measurement data archive, which provides raw test data and acoustic metrics calculated at sensor locations, as discussed in Section 5.7. This section provides an overview description of the structure and contents of of the noise exposure and community response databases. Full details, with description of structure and contents, and file nomenclature, are provided in the Description of Data included in the database archive. Figure 6-48 and Figure 6-49 show the directory structure and file names. In cases where large numbers of files exist for flight and meteorological data, the file naming details are similar for each, and a representative file name is shown. To assist using the data, certain subdirectories contain readme files that provide specific information for interpreting the data.

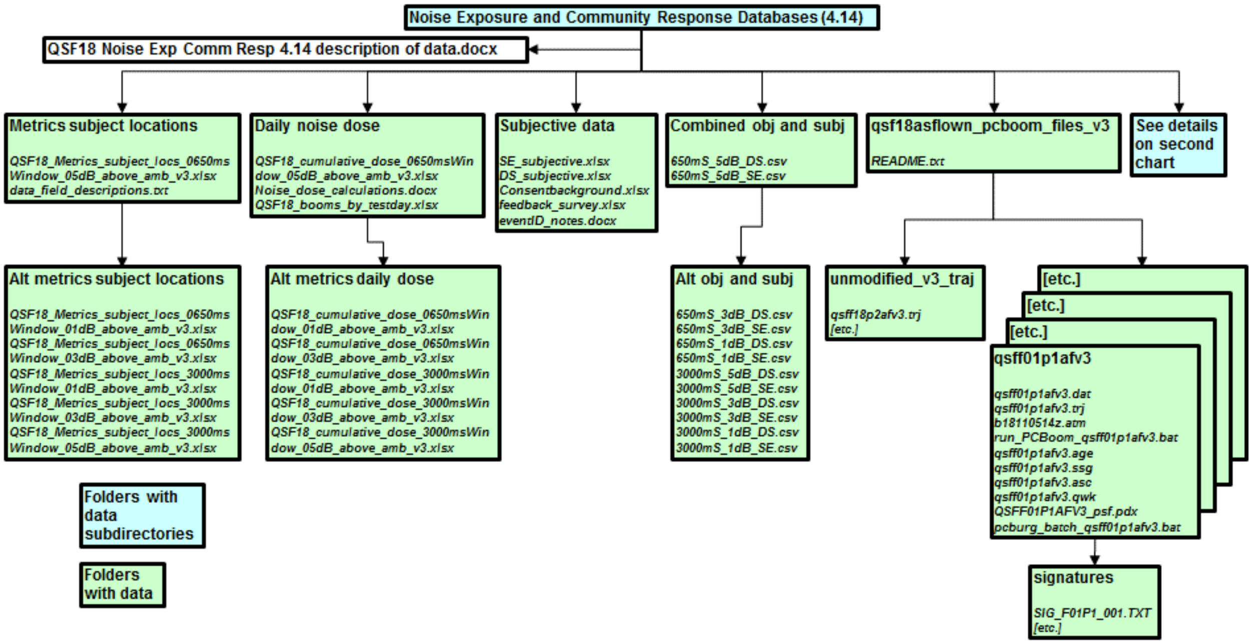


Figure 6-48 QSF18 noise exposure and community response databases structure (1 of 2)

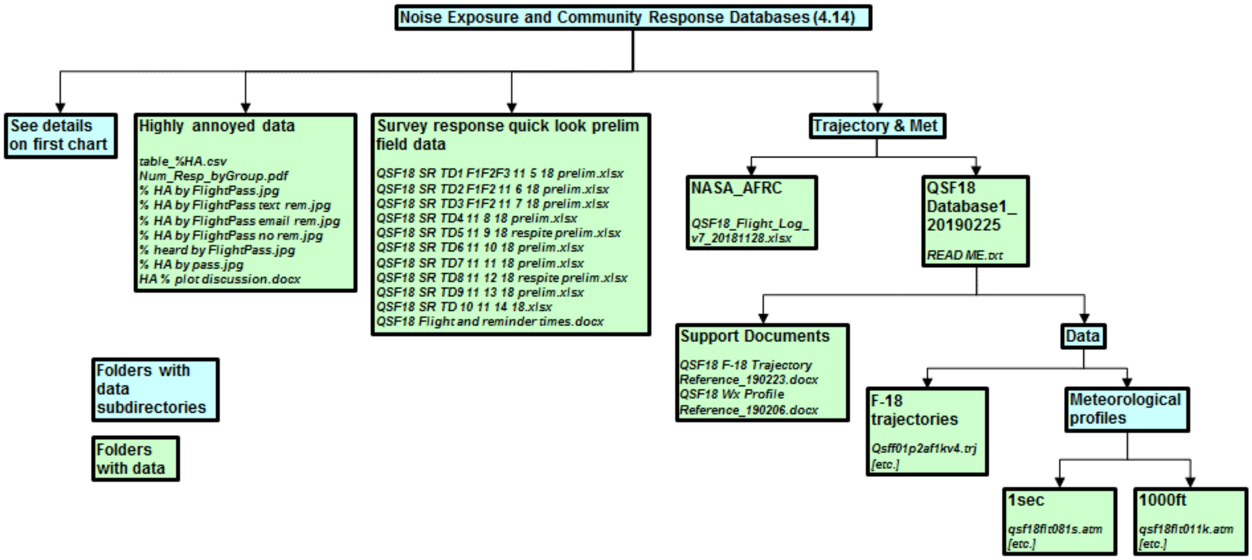


Figure 6-49 QSF18 noise exposure and community response databases structure (2 of 2)

As shown, the set of folders underlying the top level consists of:

- Acoustic metrics mapped to subject locations
- Daily noise dose
- Cleaned up subjective data resulting from post test review
- Combined objective and subjective datasets
- PCBoom input and output files for model using as-flown aircraft trajectories
- Detailed data on highly annoyed participant responses, with interpretive plots
- Survey response quick look preliminary field data – note that these are the data that were assembled on site each day while the testing was ongoing. Due to the fact that participant responses can be delayed, these quick look data are not final – the data are not necessarily complete. The quick look data are included in this database to provide analysts the opportunity to understand how the test data gathering proceeded, which offers insight into test execution.
- Aircraft flight trajectory and meteorological data

7. QSF18 Conclusions and Recommendations

7.1 Summary of Findings

As evidenced by QSF18, community outreach and recruitment requires significant NASA engagement, including efforts to proactively respond to participant questions during recruitment, throughout test execution and follow up phase.

The AFRC Pre-Test and the QSF18 test showed that instrumentation upgrades to leverage cellular connectivity proved successful and reliable. The analysis of field acoustic data (including waveforms and metrics) immediately following each event provided adequate quality checks on instrumentation. A steady stream of objective and subjective data flowing to the team on an event by event basis provided vital information for decision making regarding noise dose and operational waypoint planning.

Logistics and site preparation went well, but ambient noise in urban areas makes extracting sonic boom metrics problematic, suggesting more emphasis on advance site scouting and low noise monitor sites where possible in urban areas for future tests. Ultimately QSF18 data processing was successful by employing manual event waveform identification techniques, however this approach is untenable for wider scale testing. There are still open questions about the details of windowing and spectral subtraction of background noise that need to be addressed, especially if the metric of the ambient is greater than the recorded event. Thus an agreed upon method of addressing the ambient issue must be in place prior to X-59 overflights.

The nature of the F-18 low boom dive maneuver introduced regions beyond cutoff (down track from the dive point in addition to lateral cutoff) where noise from the F-18 was generated and responded to by participants. These sounds were of a different nature and had a longer duration, which necessitated different metric analysis procedures (depending on the participant location at the time of the event). While this was anticipated for QSF18 it is appropriate to expect such situations at lateral cutoff for future X-59 testing due to the unpredictable motion of participants during the course of their daily routine. Procedures for handling “non-primary footprint” sonic thump events, specifically metric analysis and incorporation into the dose-response analysis, should be examined in closer detail.

While the F-18 LBDM proved a useful noise source surrogate for the X-59 for QSF18 risk reduction test, it was difficult to deliver the desired PL metric values on the community. This is due in part to shortfalls in the propagation algorithms which didn’t adequately account for the effect of clouds but is also due to the complexity of the upper air meteorological profile in a humid coastal region.

The process to determine the subject noise dose for single events and cumulative daily levels used model guided interpolation of empirical metric values. The QSF18 test was an extreme test of this process given the nature of the F-18 low boom dive footprint. During future X-59 testing when steady level flight is expected, this process should be simplified in terms of longitudinal versus lateral variability in noise footprint, but will likely require supplemental techniques for incorporation of meteorological variability and turbulence-induced uncertainty into the dose-response correlations.

Survey geolocation worked reasonably well. Participant locations at the time of events were determined from latitude and longitude coordinates or from addresses in single event survey responses. When single event responses did not exist or could not be linked to an event, the location of participants was determined, where possible, from the times participants indicated they were at home or at work in their daily summary. It was learned that background survey reported locations for home and work should be checked for validity prior to accepting participants into the program and that survey fields for other single event addresses need more structure and checking when location services are not available. Also, since daily summaries serve as a critical fallback when single event responses do not exist, more effective techniques are required to ensure participant compliance with daily summary reports.

Although the software for subjective data gathering was tested several times in advance, including during the AFRC Pre-Test, several issues came to light regarding survey response submittals (or lack thereof due to open browser windows) in the analysis phase. The dynamic nature of technology (internet browser, mobile device capabilities and location services) and the evolving topic of personal privacy (and the use of geo-tracking technology) will ensure this topic remains a high risk challenge.

Post test statistical analyses of QSF18 data may be summarized as follows:

- Single event dose-response relationships were established for the metrics considered showing a positive correlation between noise level and %HA response.
- The correlation between cumulative daily dose and percent highly annoyed response was statistically insignificant for QSF18. This finding is presumably driven by the lack of HA reports from which to estimate such a relationship in addition to low noise levels of the sonic thumps.
- Reminders to participants resulted in significantly higher single event response rates among that group, however the opposite was true for daily summaries, in that response rate was higher for those who did not receive single event reminders.

Past studies have suggested the target cruise loudness of the X-59 of 75 PLdB will find community acceptance. If not, response to lower levels laterally would be available if the X-59 loudness decreases with lateral distance. If the lateral pattern is essentially constant or increases then lower PL levels would require flights at higher altitudes which could be performance limited. Thus, it is of great importance to establish the X-59 lateral patterns at cruise for each of the selected metrics to determine whether lower noise exposures are possible.

7.2 Experimental Design Lessons from QSF18 Applied to the Conceptual X-59 Test Plan

This section considers the operational aspects of the 2016 LBFD study on the selection of the six communities to be overflown with the X-59 with the intent of applying the insight, experiences, and lessons learned from three sonic boom flight experiments, the NASA LBFD Community Response Pre-Test at AFRC (Appendix G), SonicBat flight tests conducted at AFRC & KSC [Bradley *et al.* 2018], and the LBDM tests at Galveston TX [Page *et al.,* (presented herein)] that have been accomplished since the development of the X-59 Test Plan.

The 2016 Site selections involved a variety of operational concerns. It is of interest to revisit three of them, the climatology, flight planning, and the community ambient and background noise levels. These three concerns were revisited with an appreciation of the scale of three flight experiments and analyses that investigated human responses to aircraft-generated sonic booms. Figure 1-1, in the introduction to this report, provides a comparison of the scales of these experiments that include WSPR in 2011, the AFRC pre-test in 2017, and the QSF18 test in 2018, to the anticipated X-59 test in 2023 and beyond. It can be seen that the QSF18 flight experiment was large in comparison to WSPR and the AFRC pre-test but is small in comparison to the anticipated X-59 community overflights.

7.2.1 Climatology

Climate considerations revolved around two requirements, first to ensure the total participant population and geographic areas selected are representative of the entire United States, and second to expose the X-59 boom signature to a wide range of temperature, humidity, and lower level turbulent conditions than has been experienced to date by N-wave type aircraft.

The six sites were chosen from the five climate zones, as shown in the Figure 7-1, as defined by Building America [Baechler *et al.,* 2013]. Included are Cold, Marine, Hot-Humid, Mixed-Humid, and Hot-Dry. The Mixed-Dry and Very Cold climate zones were not used for site selection due to their relative small size and lack of large population areas. The final selection included two sites in the cold climate zone (Upstate NY and MI) and one each in Hot-Dry (CA), Hot-Humid (FL), Mixed-Humid (VA) and Marine (WA).

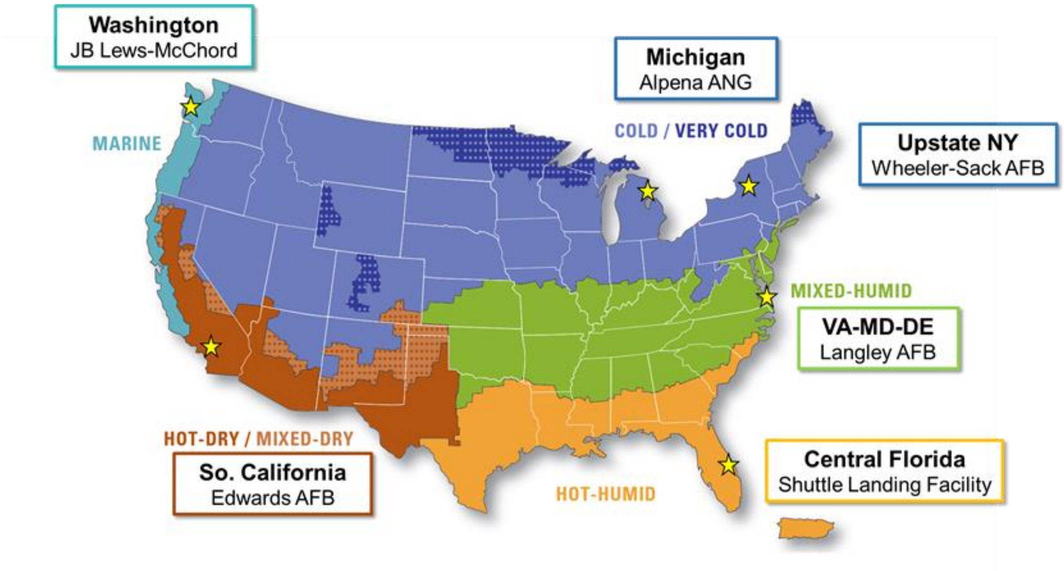


Figure 7-1 X-59 test sites

The factors that led to the selection of these six sites are described in detail in Appendix A. Note section 2.0 and in particular figures 2-1 and 2-2. In short, the lack of an unpopulated area to place the transition focus boom footprint was the primary reason communities in the United States interior were not chosen.

The selection of these five climate zones was fortunate in that it provides the opportunity to establish a much needed data base regarding the influence of the atmosphere on low boom shaped signatures. The present data base, being gathered over the past 60 years, is for N-wave type signatures. The majority of these data were obtained in the hot-dry regions of the country, and thus are not representative of many other areas of the country, let alone worldwide. This data base has shown that both the “macro” (pressure, temperature, and winds) and “micro” (atmospheric absorption e.g. humidity and the molecular relaxation of O_2 and N_2) influences of the atmosphere along with cloud cover and turbulence play a significant role in altering the boom signature. Experiments and analysis, e.g. Bradley *et al.* [2018] and Kanamori *et al.* [2017], suggest that the low boom shaped signature of the X-59 will not be as sensitive to these atmospheric influences. However, little, if any, information is available regarding the influence of atmospheres associated with cold-dry and cold-damp climates.

7.2.2 Flight Planning

The planned flight trajectories described in the 2016 study for each of the six sites have been designed to put the focus and climb region over water and only have the constant-speed cruise portion of the trajectory’s boom footprint (the carpet region) on land. The carpet region to which the test site will be exposed has two aspects: the footprint along the length of the flight path trajectory and the behavior of the footprint laterally from the flight path out to and beyond the lateral cutoff due to atmospheric refraction. Historically boom overpressures from N-wave aircraft are a maximum under the aircraft and decrease with increasing lateral distance [Maglieri *et al.,* 2014]. The LBFD is expected to display a similar pattern. However, depending upon the vehicle design, the lateral spread pattern of overpressure could be fairly uniform out to cutoff [Morgenstern *et al.,* 2012) while the Perceived Levels (PL) increase slightly, as shown in Figure 7-2.

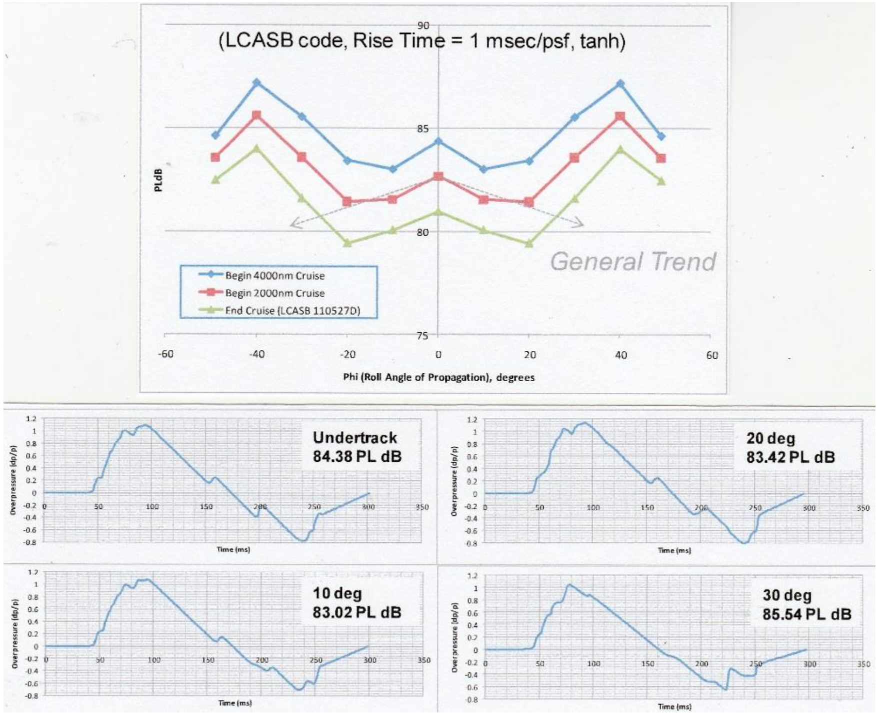


Figure 7-2 Lateral spread of boom footprint

Other studies show that the calculated PLs remain uniform across the ground footprint out to lateral cutoff resulting in the same exposure and possibly the same response among the entire community. This poses a concern regarding the lateral behavior of other metrics included in the study. Do they behave in a similar manner, or do they increase or decrease, and what role will all this play regarding subjective responses?

Past studies have suggested that the target cruise PL of the X-59 of 75 dB will find community acceptance. If not, response to lower PLs levels laterally would be available if the X-59 PL's decrease with lateral distance. If the lateral pattern is essentially constant or increases then lower PLs would require flights at higher altitudes, which would position the aircraft in the ozone concentration region. Additionally, it may be that an increase in altitude is not within the aircraft performance envelope. Thus it is of great importance to establish the X-59 lateral patterns at cruise for each of the selected metrics to determine whether lower noise exposures are possible.

7.2.3 Ambient & Background Noise

In the context of this study ambient noise is defined as the noise associated with a particular community or location surrounding a test site measuring station (rural, urban, commercial) and background noise (TV, radio, traffic noise, airplane flyover) is that noise at the respondents location that may result in the thump not being heard. Since background noise can vary based on the test subject's activities at the time of the thump, it cannot be assumed that a non-response is equivalent to a response of "not annoyed" if the thump was audible. The perception of the thump may be masked by respondent activities or other environmental noise. The present practice is: if an individual does not respond, that data point is considered "non-response". The QSF18 included testing of the push/reminder notifications to provide responses that indicated they did NOT hear the boom, thus providing legitimate data.

With regard to the ambient noise, in essentially all of the past sonic boom flight tests, including community overflights of St. Louis (1962), Oklahoma City (1964), and the National Sonic Boom program at Edwards Air Force Base (EAFB) (1966-1967), the sonic boom overpressures ranged from about 1.0 psf to 3.0 psf. At such levels the ambient noise levels at the test sites were of little or no concern in terms of influencing the measured boom signatures. At the AIAA SciTech 2018 meeting an oral presentation by NASA on the Preliminary Design Status and Low-Boom Flight Demonstration (LBFD) Project Update showed that to acquire a boom signature in the 75 PL dB range, overpressures will need to be in the 0.40 psf range. In order to produce PLs of 75 dB or lower for the Galveston tests using the F-18 LBDM, signature overpressures as low as 0.1 psf were measured. At such low boom levels the ambient noise at the test sites have a great influence on the measured signature depending upon how far its spectrum was submerged in the ambient noise spectrum.

WSPR 2011 explored the issue of how to address ambient noise when the metrics calculated for ambient noise are close to that of sonic booms. If the ambient levels were not at least 1 dB less than the boom metric, then that particular recording was considered too contaminated to use for further analysis. It was determined that if a boom is lower in amplitude and has energy comparable to the ambient, then the only way to remove the ambient energy from the metric calculations is to subtract it from the energy spectrum before calculating the metrics. For the Galveston tests, the metrics used in the analysis do not have the ambient subtracted. To determine whether an event was excessively influenced by the ambient, the metric for the event is compared to the metric for the ambient. Of the 476 recorded events, there were 387 that had a PL difference of 1dB above the ambient, 314 that were 3 dB above the ambient, 266 that were 5 dB above the ambient, and 143 that were 10 dB or more above the ambient.

Although the ambient noise level could be well above the boom signature, the boom may still be observed by the test subjects whether they are outdoor or indoor. Regarding the indoor case, it is the boom signature that excites the structure, not the ambient noise. However, for both indoor and outdoor situations the response of the test subject must be related to the boom signature and the associated metrics and not the measured event that consists of the ambient and the boom signature. Subtraction of the ambient from the measured event to obtain the boom signature may not be applicable to all cases depending on how far it is submerged in the ambient and if turbulence has altered the boom signature.

It is appropriate to present a view of the role the ambient noise in the communities to be overflown by the X-59 may play in influencing its low boom shaped signature. It is assumed that two notional LBFD's both generate ground boom signatures having PL's of 78.2 dB and 69.8 dB and overpressures of about 0.40 psf and periods of about 100 msec (Figure 7-3). Note that the signature having a PL=69.8 dB, shown on the right, has a sinusoidal shape. This shape approaches the optimum boom signature, which is a sine wave without the abrupt change in pressure from ambient on the front shock and on the return to ambient pressure at the rear shock. The noise spectra of these two notional signatures are shown in Figure 7-4 with the noise spectra of three areas having ambient noise typical of a rural setting, an urban setting [Albert & Decato 2017], and a national park environment [National Park Service 2012]. Also shown is the ambient noise spectrum during one of the supersonic passes in the Galveston tests.

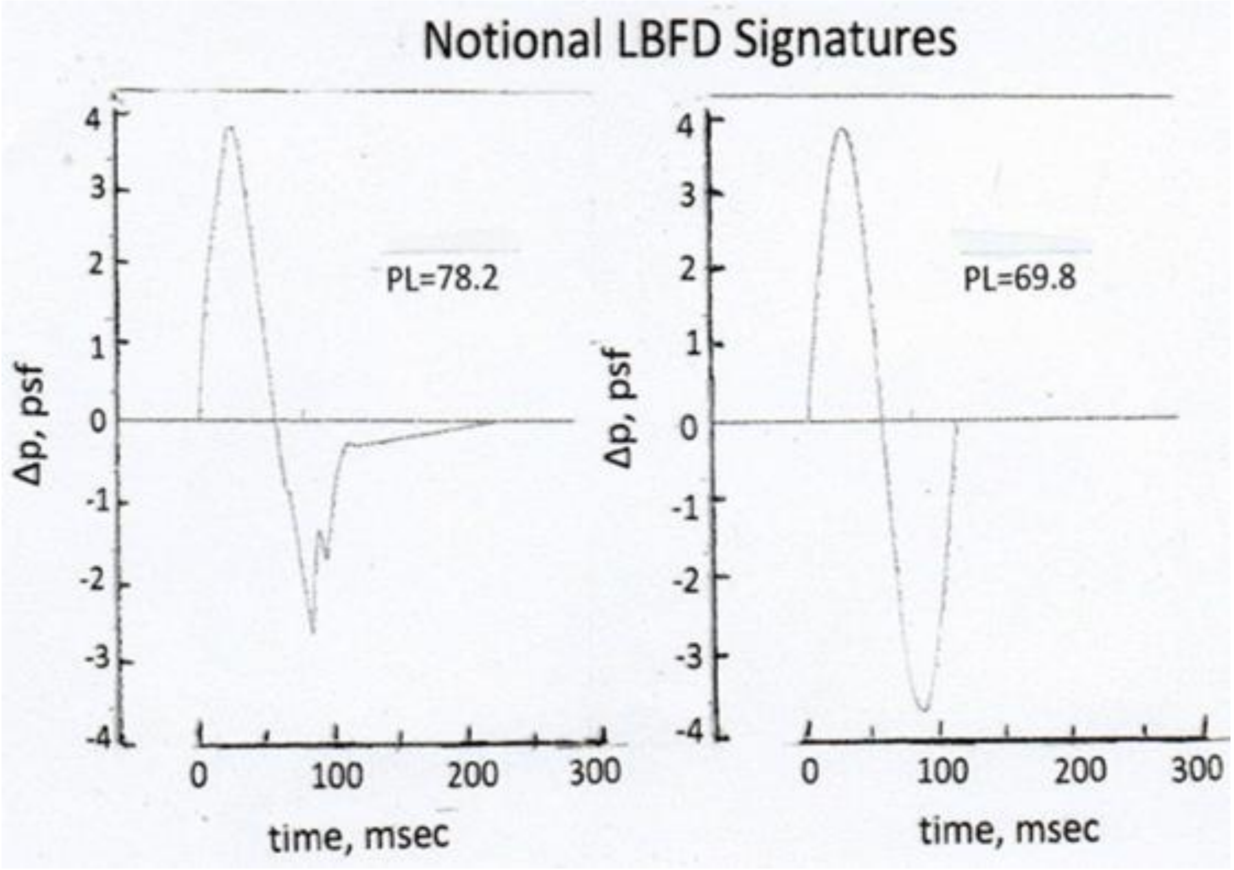


Figure 7-3 Notional LBFD signatures

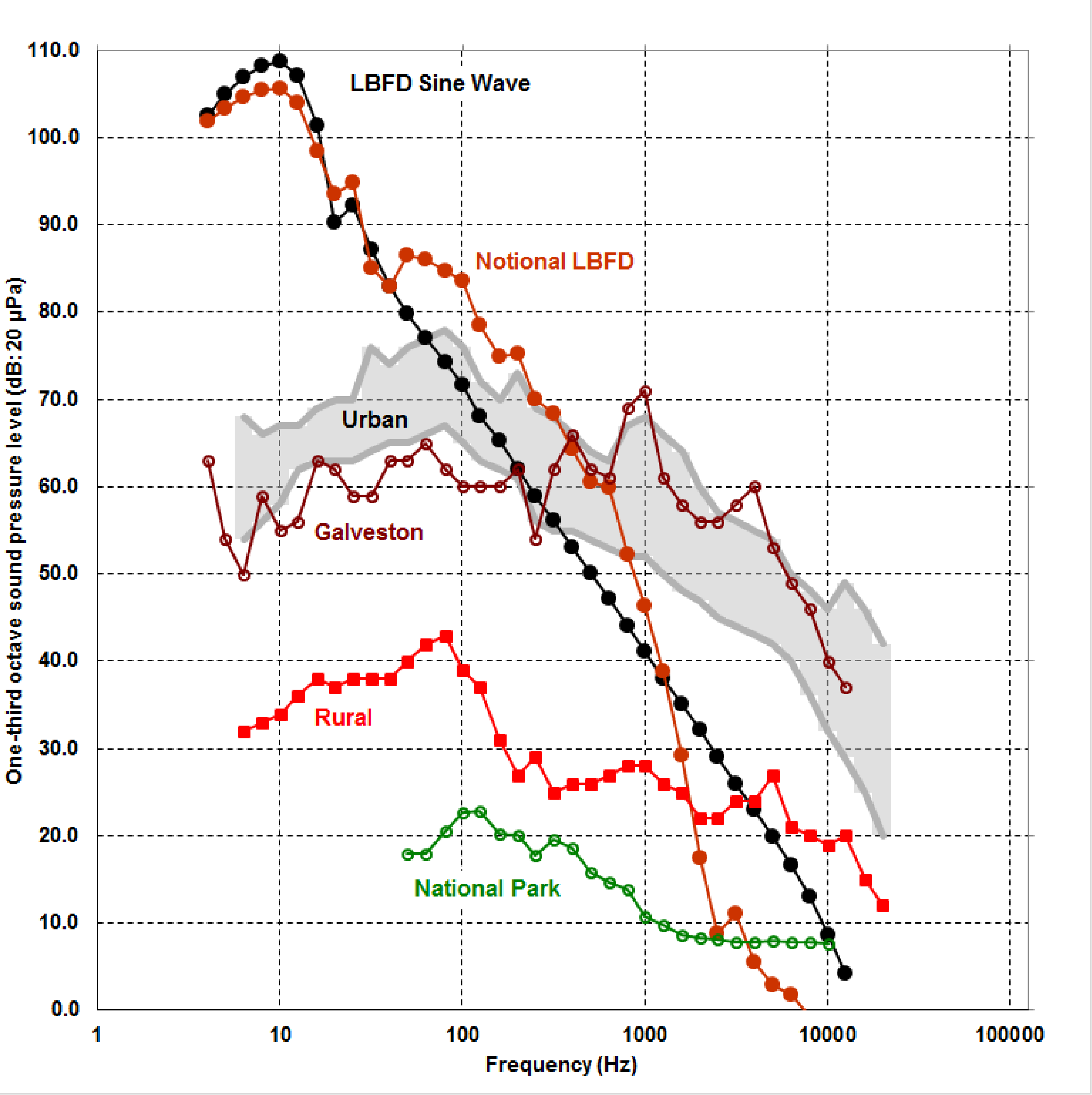


Figure 7-4 LBFD signatures and comparison to ambient noise

If one assumes the spectrum of the X-59 cruise signature closely follows the LBFD Sine Wave spectrum, at about 100Hz its noise level coincides with the urban ambient level. Above about 500Hz the ambient dominates the measured signature spectrum. Even higher ambient levels are expected in business/commercial and industrial/manufacturing areas or areas close to highly trafficked roadways. The possible dominance of the ambient noise on the measured signature spectrum could result in calculated metrics that do not represent the noise the subjects were exposed to and may even result in the loss of data.

An accepted method of addressing the ambient noise issue must be in place well before the X-59 takes to flight.

7.3 Proposed Further Risk Reduction Activities

The future LBFD test design development parameters were defined in 2015 and are itemized in Table 7-1. The X-59 vehicle design has been refined and these vehicle performance driven requirements have evolved since then. Tighter integration of the sonic boom performance of the X-59 is needed with flight test planning activities.

Table 7-1 NASA LBFD testing guidelines

Manned aircraft; public airspace
Day and night flight operations
Runway length > 9000 ft
At least 2 community exposures , 20+ minutes apart
Closest community < 125 nmi. from base of operations
Exposure (boom carpet) ~50 n.mi. long by ~35 n.mi. wide
Supersonic range up to 350 n.mi
Take-off and landing sites up to 500 n.mi apart
Under-track ~75 PLdB off-track ~70-75 PLdB; 85 PLdB possible
Cruise: level flight, Mach~ 1.6, ~50kft
Acceleration focus: Mach ~1.2, ~35kft, ~2° climb
Op Tempo: 3 flights/~9 hrs <or> 4 flights/~12 hrs
Deployments limited to < 1 month
Three deployments per year for 2 years

During the planning and preparation of the LBFD test plan during phase 1, strong interdependencies were identified between the elements identified in Figure 7-5. These factors strongly influence site selection and identification of prominent communities targeted for recruitment. The goal Is to ensure that the aggregated recruitment yields a U.S. representative distribution across such parameters as:

- Demographics
- Meteorological
- Seasonal
- Geographic, including considerations for focus placement and avoidance

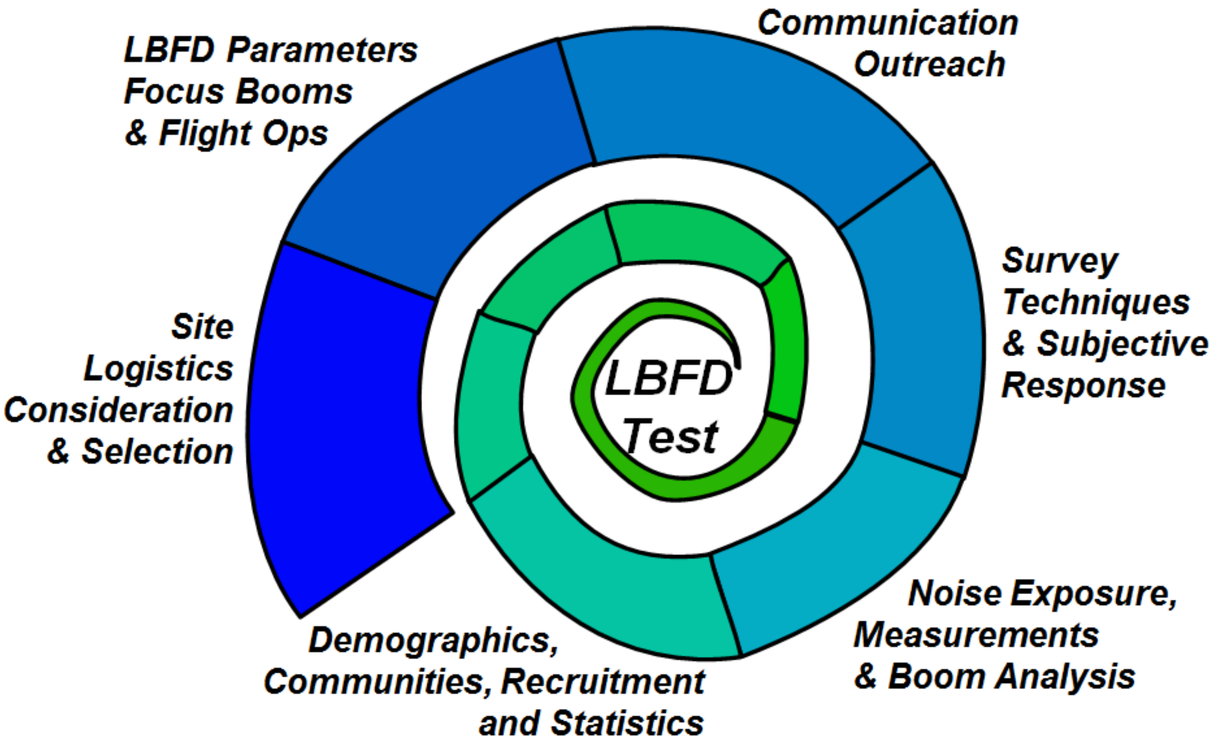


Figure 7-5 Spiral LBFD test design process

One example of the experimental design interdependencies is that the lateral boom distribution directly controls the variation in dose across the test area, hence X-59 performance capabilities (for different flight speeds or altitudes) need to be considered as part of the design process. This allows for an achievable range in single event metric levels, and (for a given operational tempo and number of flights and daily events) the range in cumulative daily dose. It was learned from the QSF18 data analysis that the range of single event levels yielded a positive dose-response relationship, while the cumulative daily dose range did not. Will this provide an adequate dose response dataset for the FAA and ICAO?

One must also consider the geographic nexus between prominent communities and the combination of the flight operation patterns and lateral boom distribution. This is in essence the recruitment stencil/strategy, for which requirements need to defined in more detail. Each of the future X-59 tests will have a different recruitment geometry, by virtue of being in different geographic areas, strongly driven by the influence of home and work locations and potential flight operational patterns. Additional investigation is needed to identify a suitable target U.S. representative distribution.

The future drivers of testing technology, procedures and preparations include greatly expanded geometric breadth and a significant ramping up of the testing operational tempo with multiple tests per year expected. This suggests the need to develop procedures to conduct Quality Assurance/Quality Control (QA/QC) of incoming data streams and develop near real-time data analysis tools.

Conducting subject testing in urban areas introduces a number of difficulties due to the higher levels of background noise. These affect not only the acoustic measurements and extraction of metrics for computation of the dose, but also affect the perception and audibility of the thump sounds. A number of activities and analyses can leverage the existing QSF18 data and help inform and prepare for X-59 community testing including:

- How to account for different levels of ambient data over a large geographic area with different community environments
- Interaction between metric calculation and ambient details and the model guided interpolation process for dose determination
- Spectral techniques for performing metric calculations
- Metric calculations (windowing, etc.)
- Propagation modeling (clouds, coastal meteorology, humidity effects)
- Dose-response modeling and analysis
- Heard / Not-Heard and nexus with ambient noise
- Audibility threshold influence on dose determination
- Ambient level criteria and guidelines for X-59 instrumentation placement
- Review feedback and comments in the database
- Examination of multiple response comments and test understanding
- Survey techniques for quantifying ambient noise influence

Survey instrumentation development needs to stay abreast of current technology. Some of the survey responses suggested a lack of participant understanding of the single event noise evaluation. Further examination of this data could provide input on how to better clarify instructions, including possible interactive (automated) online training for participants. Georeferencing of the subjective data identified a need to refine the survey instrument address gathering mode when device geolocation was ineffective.

The QSF18 analysis highlighted the need to improve survey compliance among the participants. Techniques and methods to improve survey compliance need to be developed and tested for single event and daily summary surveys.

Another item that came to light during the QSF18 test is the need to develop an effective X-59 subjective and objective linked database and establish protocols for development of an archival data set. The methods by which this data can be shared effectively among researchers and ultimately delivered to NASA, while maintaining compliance with IRB and OMB protocols, need to be investigated. Language for the use of the data (the participant data agreement) needs to be developed and the IRB informed consent language adapted to accommodate the data and the eventual archive. For example: “We may use your research information for other research studies or may share your information here or at other institutions for future research efforts without additional informed consent.” Potential dataset and archive options include the following types of datasets:

- Fully Identifiable data: all the data gathered except respondents’ identity
- Partially de-identified data: Include the lat/long location data but remove the home addresses to protect household identify (limits ability to fully use the dataset.)
- Fully de-identified data: This would include noise dose and response data, but without the location associated with the dose. Limits ability to fully use the dataset.

Development of hardware performance requirements for wide-scale acoustic testing is important. Current analysis could utilize the QSF18 and other datasets to conduct comparisons among the different systems that are currently being employed to measure sonic booms including the SBUDAS, SPIKE, Brigham Young University (BYU) sensors, and the Volpe CARS (Volpe Center Acoustic Recording System). Parameters to be explored include the following:

- Hardware configuration, microphone orientation
- Weatherproofing
- Power
- Networking / HUMS
- Automated event recording triggering
- Near-real time analysis capabilities

As has been described in this document, it is often difficult to extract a sonic boom signal from the ambient noise. Tools to automatically identify the thump events are necessary. This work could commence leveraging the QSF18 (and other) data sets.

As well as a prevalent topic for QSF18, identification of and treatment of sounds beyond cut off is important. This will likely occur for X-59 as well. It is feasible that lab testing could be employed to further develop protocols for handling such sounds. Some of the activities that should be considered include:

- Lateral cutoff sounds / metrics
- Ambient testing w/ subjects to understand SNR for shaped booms
    - Impact on site selection criteria
    - Metric analysis procedures

Presently there is some active research in this area under the FAA ASCENT program which might be leveraged. This work is categorizing and exploring metrics and subjective response to sounds from Mach cut off operations.

Future flight testing (e.g. CarpetDIEM or Acoustic Validation of X-59) should also be considered. Upcoming data from flights could be analyzed for the following purposes:

- Incorporation of stochastic turbulence modeling In the dose quantification and response analysis
- Sensor development, hardening, reliability, and network testing
- Opportunity for data input for near real-time run stream testing
- Gathering additional test data for ICAO Supersonic Task Group (SSTG), Procedures Subgroup (PrSG) including data for certification method exploration

One potential method for integrating all of these elements together, including the near real-time analysis, is development and execution of LBFD flight simulation campaigns. These would be techniques (without actual flights) to exercise the various real-time data analysis protocols, QA/QC procedures and data analysis streams. The task could be conducted initially over the course of a flight, then a day, then an entire campaign. Data could be seeded using existing QSF18 data, or with other simulated data inserted, designed to test exception handling and other considerations. The various steps for preparing an LBFD flight simulation campaign are described in Figure 7-6. This approach allows for the planning and systematic development of the tools, analysis procedures and data flows, and provides a means for incremental test and refinement, while practicing with the X-59 team members prior to the first X-59 community test. Such a simulated LBFD test can also serve as an effective tool to focus, in terms of functionality and timing, the research efforts and activities described above.

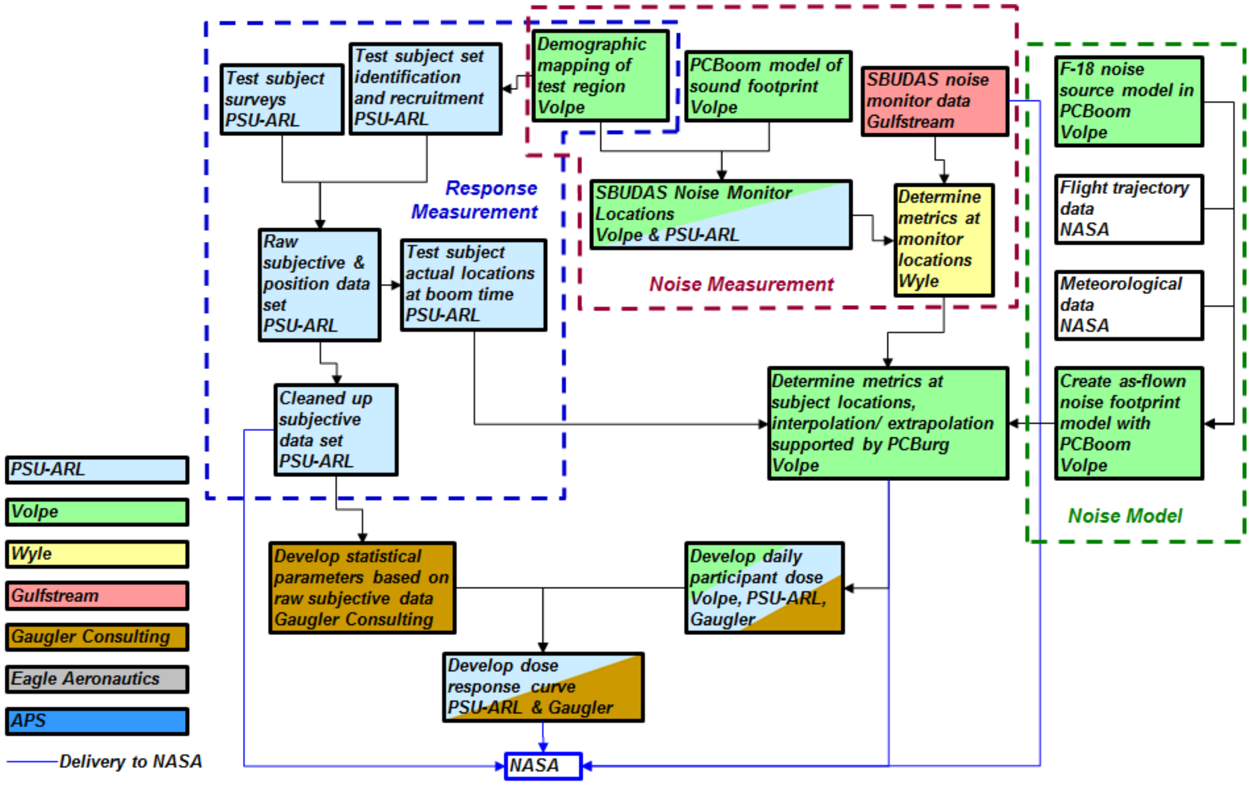


Figure 7-6 Near real-time analysis stream

References

Albert, D.G. and Decato, S.N. 2017: "Acoustic and Seismic Ambient Noise Measurements in Urban and Rural Areas". Applied Acoustics, Vol 119, pp. 135-143, April.

Baechler, M., Gilbride, T., Cole, P., Hefty, M., & Ruiz, K., 2013. "Building America Best Practices Series, Volume 7.2, High-Performance Home Technologies: Guide to Determining Climate Regions by County," U.S. Department of Energy, PNNL-17211, Rev. 2, November.

Borsky, P.N., 1965. "Community Reactions to Sonic Booms in the Oklahoma City Area," AMRL-TR-65-37, Wright-Patterson Air Force Case, Ohio, Aerospace Medical Research Laboratory.

Bradley *et al*., 2018. "Sonic Booms in Atmospheric Turbulence: The Influence of Turbulence on Shaped Sonic Booms.".

CHABA (1996). Fidell, S. (editor), "Community Response to High-Energy Impulsive Sounds: An Assessment of the Field since 1981," Committee on Hearing, Bioacoustics and Biomechanics, National Research Council, Wash. D.C., National Academy Press.

Fields, James M.Reactions of Residents to Long-Term Sonic Boom Noise Environments. NASA CR201704, June 1997.

Fields, J.M., 1994. "A Review of an Updated Synthesis of Noise/Annoyance Relationships," NASA CR-194950, National Aeronautics and Space Administration, Washington D.C.

Fields, J.M.; Ehrlich, G.E.; and Zador, P. (2000). Theory and Design Tools for Studies of Reactions to Abrupt Changes in Noise Exposure. NASA CR-2000-210280. National Aeronautics and Space Administration, Washington, D.C.

Fields, J.M., R.G. De Jong, T. Gjestland, I.H. Flindell, R.F.S. Job, S. Kurra, P. Lercher, M. Vallet, T. Yano, R. Guski, U. Felsheur-Suhr, and R. Schumer. (2001). "Standardized General Purpose Noise Reaction Questions for Community Noise Surveys: Research and a Recommendation." Journal of Sound and Vibration 242:641-679.

Haering Jr., Edward A., James W. Smolka, James E. Murray, and Kenneth J. Plotkin, 2005. "Flight Demonstration of Low Overpressure N-Wave Sonic Booms and Evanescent Waves," Proc. of 17th International Symposium on Nonlinear Acoustics; AM. Inst. Of Physics, Melville, NY, pp. 838, 647-650, July.

Hilton, D. and Newman, J., 1966. "Instrumentation Techniques for Measurement of Sonic-Boom Signatures", *J. Acoust. Soc. Am.* **(39)**, 5B.

Hobbs, C. (2012). "Auto Boom Finder Program (ABF)," Wyle Technical Note TN 12-30, Arlington, VA.

Kanamori, Masashi; Takahashi, Takashi; Naka, Yusuke; Makino, Yoshikazu; Takahahashi, Hidemi; Ishikawa, Hiroaki (2017). . "Numerical Evaluation of Effect of Atmospheric Tubulence on Sonic Boom Observed in D-SEND#2 Flight Test." AIAA 2017-0278.

Luz, George, August 2005, Noise Sensitivity Rating of Individuals, Sound and Vibration.

Maglieri, Domenic J., Percy J. Bobbitt, Kenneth J. Plotkin, Kevin P. Shepherd, Peter G. Coen, David M. Richwine, 2014. Sonic Boom, Six Decades of Research, NASA/SP-2014-622, NASA Langley Research Center, Va.

Maglieri, D.J. and Hubbard, H.H., 1959. "Some Ground Measurements of the Shock Noise from Airplanes in Steady Flight at Altitudes from 25,000 to 45,000 Feet." NASA TN D-48 September 1959.

Maglieri, D.J. and Hubbard, H.H., 1961. "Ground Measurements of the Shock-Wave Noise from Supersonic Bomber Airplanes in the Altitude Range from 30,000 to 50,000 Feet," NASA TN D-880, July 1961

Morgenstern, J. M., M. Buonanno, N. Nordstrud, 2012. "N+2 Low Boom Wind Tunnel Model Design and Validation." AIAA 2012-3217, 30th AIAA Applied Aerodynamics Conference, New Orleans, LA, 25 - 28 June.

NASA, 2015. Request for Information for Strategic and Acquisition Strategy Planning for the Commercial Supersonic Technology Project and Low Boom Flight Demonstration. Solicitation: LOWBOOM-20160602, June. https://www.fbo.gov/index?s=opportunity&mode=form&id=ba3821e08cb4f4542f7f2a6217607bff&tab=core&tabmode=list& Accessed 06 March 2016.

National Park Service 2012, Natural Sounds and Night Sties Division, August 21 ,2012 memo to Peter Taylor, Forest

Environmental coordinator, Superior National Forest-March 2011 Royal Lake measurement.

Nykaza, Edward T., Valente, Dan, MacAllister, Bruce, Hodgdon, Kathleen, Gaugler, Trent, Krecker, Peg, Luz George. August 2014. An Investigation of Community Attitudes Toward Blast Noise: Final Report, SERDP Project WP-1546.

Page, Juliet and Downs, Robert. "Sonic Boom Weather Analysis of the F-18 Low Boom Dive Maneuver", Joint Meeting of the Acoustical Society of America and the European Acoustic Association, Paper 2pNSb8, 25-29 June 2017.

Page, J., K. Hodgdon, C. Hobbs, C. Wilmer, P. Krecker, C. Koenig, T. Holmes, R. Cowart, T. Gaugler, D. Shumway, J. Rosenberger, and D. Phillips, 2014. "Waveforms and Sonicboom Perception and Response (WSPR): Low-boom Community Response Program Pilot Test Design, Execution and Analysis", NASA/CR-2014-218180, NF1676L-18285 Technical Report, March 2014.

Petrinovich, L., 1984. "A Two Factor Dual-Process Theory of Habituation and Sensitization," In H. Peeke & L. Petrinovich (Eds) Habituation, Sensitization, and Behavior, New York: Academic Press, Inc., pp 17-55.

Rachami, J., Page, J.A., Zhou, L., Kim, B., 2009. "Environmental Modeling of Advanced Vehicles in NextGen", AIAA 2009-6981.

Rylander, R., B. Lundquist, 1996. "Annoyance caused by noise from heavy weapons and shooting ranges," *J. Sound and Vibration,* **192**(1), pp. 199-206.

Salamone, Joe, 2009. "Recent Sonic Boom Propagation Studies at Gulfstream Aerospace," AIAA 2009-3388, May.

Schomer, 1981. "Community Reaction to Impulse Noise: Initial Army Survey", Construction Engineering Research Laboratory, CERL-TR-N-108, June.

Schomer, P.D., 1985. "Assessment of Community Reaction to Impulse Noise," *J. Acoust. Soc. Am.* **77**(2).

University of Wyoming, 2016. "Upper Air Sounding Data", http://weather.uwyo.edu/upperair/sounding.html Accessed March 2016.

U.S. Census Bureau, 2000. "Demographic Profiles: Census 2000."

U.S. Census Bureau, 2009. "Annual State Resident Population Estimates for 6 Race Groups (5 Race Alone Groups and One Group with Two or more Race Groups) by Age, Sex, and Hispanic Origin: April 1, 2000 to July 1, 2008", (released May 14, 2009).

U.S. Census Bureau, 2010. "Demographic Profiles: Census 2010." http://www.census.gov/2010census/data/ Accessed 07 March 2016.

Weinstein, N. D. 1978. "Individual Differences in Reactions to Noise: A Longitudinal Study in a College Dormitory". Journal of Applied Psychology, 63(4), 458-466., http://dx.doi.org/10.1037/0021-9010.63.4.458.

Appendices

A. NASA Low Boom Flight Demonstrator Conceptual Test Plan for Community Response Testing Risk Identification and Proposed Risk Mitigation Activities (Phase 1 report)

The LBFD conceptual test plan and risk reduction report is provided in the separate appendix file for this report.

B. Site Selection Grids and Community Demographics

The site selection grids and community demographics are provided in the separate appendix file for this report).

C. Sonic Boom Weather Analysis of the F-18 Low Boom Dive Maneuver

"Sonic Boom Weather Analysis of the F-18 Low Boom Dive Maneuver" [Page & Downs, 2017] provides a description of the PCBoom sonic boom propagation results and interpretive techniques for assessing potential coastal sites for conducting dose-response testing using the F-18 dive maneuver. This briefing is provided in the separate appendix file for this report.

D. QSF18 Detailed Test Plan for Community Response Testing in Galveston Texas

The QSF18 detailed test plan is provided in the separate appendix file for this report.

E. Supplemental Meteorological Analysis and Go/No-Go Criteria

The supplemental focused analyses of meteorology and go/no-go criteria are provided in the separate appendix file for this report.

F. Armstrong Flight Research Center Waveforms and Sonic boom Perception and Response Risk Reduction (WSPRRR) Test Plan

The AFRC detailed test plan is provided in the separate appendix file for this report.

G. NASA Low Boom Flight Demonstrator Community Response Pre-Test Armstrong Flight Research Center May 8-12, 2017

The results of the AFRC test are provided in the separate appendix file for this report.

H. QSF18 OMB Application

The QSF18 OMB Application is provided in the separate appendix file for this report.

I. QSF18 IRB Documentation

The QSF18 IRB Application is provided in the separate appendix file for this report.

J. QSF18 Survey Instruments Outline

The outline (text format) of the survey instruments used for QSF18 is provided in the separate appendix file for this report.

K. QSF18 Survey Instruments Screen Shots

The screen shots of the survey instruments used for QSF18 are provided in the separate appendix file for this report.

L. Background Survey Summary Details

Detailed participant demographic data obtained from surveys, plus derived noise habituation and sensitivity scales, are provided in the separate appendix file for this report.

M. PCBoom Best Practices

Lessons learned regarding best practices for using PCBoom that emerged from this effort are provided in the separate appendix file for this report.

N. Locating Participants at Time of Sonic Thumps

The methodology for locating participants at the time of sonic thumps is provided in the separate appendix file for this report.

O. Calculating Metrics at Participant Locations

The methodology for calculating metrics at participant locations is provided in the separate appendix file for this report.

P. Daily Noise Dose Calculation

The process for calculating daily dose is provided in the separate appendix file for this report.

Q. QSF18 Measured Sonic Booms Across the Area

Plots which display the calculated footprint's peak overpressure contours overlaid on the study area are provided in the separate appendix file for this report.

R. Flight and False Reminder Records

Details regarding flights and false reminders during QSF18 are provided in the separate appendix file for this report.

S. QSF18 Supplementary Statistics

Supplementary statistics for the analysis of QSF18 are provided in the separate appendix file for this report.

T. QSF18 Noise Dose Comparison

Noise dose comparisons for the QSF18 test events are provided in the separate appendix file for this report.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD-MM-YYYY)*
01-05-2020

2. REPORT TYPE
Contractor Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas
Risk Reduction for Future Community Testing with a Low-Boom
Flight Demonstration Vehicle - Volume I

5a. CONTRACT NUMBER

5b. GRANT NUMBER
NNL15AA00C

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Juliet A. Page; Kathleen K. Hodgdon; Robert P. Hunte; Dwight E. Davis; Trent A. Gaugler; Robert Downs; Robert A. Cowart; Domenic J. Maglieri; Christopher Hobbs; Gary Baker; Matthew Collmar; Kevin A. Bradley; Brian Sonak; Diana Crom; Christopher Cutler

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
110076.02.07.02.03

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, Virginia 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/CR-2020-2204589/Volume I

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-
Subject Category 71
Availability: NASA STI Program (757) 864-9658

13. SUPPLEMENTARY NOTES
Langley Technical Monitor: Jonathan Rathsam

14. ABSTRACT

The Quiet Supersonic Flights 2018 (QSF18) Program was designed to develop tools and methods for demonstration of overland supersonic flight with an acceptable sonic boom, and collect a large dataset of responses from a representative sample of the population. Phase 1 provided the basis for a low amplitude sonic boom testing in six different climate regions that will enable international regulatory agencies to draft a noise-based standard for certifying civilian supersonic overland flight. Phase 2 successfully executed a large scale test in Galveston, Texas, developed well documented data sets, calculated dose response relationships, yielded lessons, and identified future risk reduction activities.

15. SUBJECT TERMS
Galveston; community survey; dose-response relationship; low boom; sonic boom; sonic thump

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: `help@sti.nasa.gov`)
					19b. TELEPHONE NUMBER ***(Include area code)***
U	U	U	UU	190	(757) 864-9658