POMA Proceedings of Meetings on Acoustics

Volume 45 http://acousticalsociety.org/

181st Meeting of the Acoustical Society of America

Seattle, Washington

29 November - 3 December 2021

Noise: Paper 4aNSa3

Simulation and application of Bayesian dose uncertainty modeling for low-boom community noise surveys

William Jeffrey Doebler
Structural Acoustics Branch, NASA Langley Research Center, Hampton, VA, 23517, USA; william.j.doebler@nasa.gov

Aaron B. Vaughn, Kathryn Ballard and Jonathan Rathsam
NASA Langley Research Center, Hampton, VA, USA; aaron.b.vaughn@nasa.gov, kathryn.ballard@nasa.gov, jonathan.rathsam@nasa.gov

In dose-response modeling, failing to account for dose uncertainty can cause artificial flattening of the estimated slope of the dose-response curve. Previous analyses of NASA sonic boom community noise survey data utilized a Bayesian multilevel logistic regression model, which did not account for dose uncertainty. The current work extends the model to account for either classical or Berkson dose uncertainty. The extended model is applied to two simulated dose-response datasets to illustrate conditions under which the dose uncertainty term does and does not correct for the artificial flattening introduced by dose uncertainty. Finally, the extended model is applied to two previous NASA sonic boom community noise surveys. The resulting dose-response curve slope for the average participant is 5 to 10% steeper, but the difference in the noise dose that elicits a 5% highly annoyed response is small (less than 1 dB). The difference remains insignificant when producing population summary dose-response curves. Commentary is included on applicability to future X-59 low-boom community noise survey data modeling and analysis.

Published by the Acoustical Society of America

https://doi.org/10.1121/2.0001592 Proceedings of Meetings on Acoustics, Vol. 45, 040004 (2022)

1. INTRODUCTION

NASA and its team of contractors led by Harris, Miller, Miller, and Hanson (HMMH) are currently preparing for the X-59 Community Response Tests.[1,2] In these community noise survey flight tests, the supersonic X-59 aircraft will fly over US communities, and residents will be recruited to rate their perception of the low-boom it produces. The low-boom loudness and perceptual data will be combined to produce noise-annoyance or "dose-response" curves, which will be provided to regulatory bodies for their use in reevaluating the current prohibition on overland civilian supersonic flight.

The analysis methods for modeling dose-response curves are currently being investigated. Noise-annoyance data from two previous community noise surveys, WSPR[3] and QSF18,[4,5] were modeled by Lee et al. using multiple models including Bayesian multilevel logistic regression (BMLR) with a single slope for the population and random intercept for each individual. However, all models assumed no uncertainty in the dose.[6,7] The literature is clear that when a logistic dose-response model is fit to a dataset with uncertainty in dose it often leads to flattening (i.e., attenuation) of the dose-response curve.[8-11] Therefore, the goal of this work is to assess the impact of failing to account for dose uncertainty and investigate alternate models that do account for dose uncertainty.

In this work, the Lee et al. BMLR model is extended to include a term to account for dose uncertainty, either as classical or Berkson additive uncertainty.[12,13] For classical uncertainty, the estimated dose is assumed to be the true dose plus an independent perturbation. Here, the true dose is the actual noise level that a participant experienced (which is unknowable in field tests without instrumenting the participant), and the estimated dose is computed based on a combination of noise measurements and modeled noise calculations. For sonic boom measurements, classical-type uncertainty can arise from a microphone system's self-noise interfering with a sonic boom pressure signal. For Berkson uncertainty, the true dose is assumed to be the estimated dose plus an independent perturbation. Berkson-type uncertainties were previously used to describe errors in radiation dosimetry studies where a mean radiation dose was assigned to a group of individuals whose true dose varied about that value. For sonic boom measurements, Berkson uncertainty can arise from estimating a mean dose value for a participant when the true dose varies around the estimated dose due to the impact of local atmospheric turbulence on the sonic boom level. Dose uncertainty could be a combination of both classical and Berkson uncertainty, for example, there is classical error associated with the microphone measurements of the sonic booms and Berkson error associated with assigning an estimated noise metric to a person whose true dose was likely affected by atmospheric turbulence or their individual listening environment. In this work, only one type of uncertainty is assumed in each model.

These extended models are first fit to simulated datasets whose true underlying dose-response curve is known. The known true doses are then perturbed by a draw from a normal distribution, and the models are fit again, illustrating the attenuating effect of dose uncertainty as well as the ability in some cases for the extended models to correct this flattening. Finally, the extended models are applied to the WSPR and QSF18 noise-annoyance data, resulting in slightly steeper dose-response curves than those without accounting for uncertainty. The dose uncertainty in both studies was estimated by Page et al.[3-5]

2. BAYESIAN DOSE-RESPONSE MODELS

In this Section, three Bayesian multilevel logistic regression model formulations are presented and applied to noise-annoyance data using the R computer program[14] and its rjags package, which interfaces with the Just Another Gibbs Sampler software to perform the Markov chain Monte Carlo sampling.[15,16] In the classical or frequentist paradigm, the collected data are considered to be random draws from a model with fixed parameters. Estimation of model parameters often takes the form of optimization. In the Bayesian paradigm, a full joint probability distribution for data and model parameters must be specified. Estimates of model parameters are obtained by means or quantiles of the posterior distribution. More details are available in Lee et al. (2020).[7]

A. LEE ET AL. BAYESIAN MULTILEVEL LOGISTIC REGRESSION MODEL

The Lee et al. BMLR model assumes that the probability that a participant will be highly annoyed by a noise dose can be modeled using a logistic function. The population of individual participants is assumed to be normally distributed about an "average participant." However, the rate of onset of annoyance for the population is assumed to be constant.

The Lee et. al BMLR model, which will be referred to as the "Original Model," is specified by

$$\begin{aligned}
&H_{ij} \sim Bernoulli(p_{ij}) \\
&p_{ij} = logit^{-1}\left(\beta_{0i} + \beta_1 \mathrm{Dose}_{ij}^{estimated}\right) \\
&\beta_{0i} \sim N(\beta_0, \sigma^2) \\
&\beta_0 \sim N(0{,}100) \\
&\beta_1 \sim N(0{,}100) \\
&\sigma^2 \sim InverseGamma(0.01{,}0.01).
\end{aligned} \tag{1}$$

Here, the probability p_{ij} that participant i will respond as highly annoyed ($H_{ij} = 1$) or not ($H_{ij} = 0$) to boom j, which has a noise dose $\mathrm{Dose}_{ij}^{estimated}$, is modeled as the inverse logit[a] of a linear function of variables with a random intercept β_{0i} that is unique to participant i and a slope β_1 that is assumed to be shared across all participants. The random intercepts are assumed to be normally distributed: $N(\beta_0, \sigma^2)$. Assignment of prior distributions to model parameters completes the Bayesian model specification. The fixed effects parameters β_0 and β_1 (parameters describing the average participant) are assigned noninformative normal prior distributions. The σ^2 parameter describing the variability of the random intercepts is assigned an inverse gamma noninformative prior distribution. The parameters β_0 and β_1 relate to the dose that elicits annoyance 50% of the time d_{50} for the average participant by $d_{50} = -\beta_0/\beta_1$. Thus, $d_{50,i}$ for an individual participant is $d_{50,i} = -\beta_{0i}/\beta_1$. β_1 can also be conceptualized as the rate of onset of annoyance, and it has units that are the inverse of the dose (e.g., PL dB^{-1}). The expected value of β_0, β_1, and σ^2 resulting from the posterior Markov chain Monte Carlo draws are reported throughout this work to describe the population.

B. EXTENDED MODEL WITH CLASSICAL DOSE UNCERTAINTY TERM

The BMLR model extended to include a term to account for additive normal classical dose uncertainty is given by

$$\begin{aligned}
&H_{ij} \sim Bernoulli(p_{ij}) \\
&p_{ij} = logit^{-1}\left(\beta_{0i} + \beta_1 \mathrm{Dose}_{ij}^{true}\right) \\
&\mathrm{Dose}_{ij}^{true} \sim Unif(-100{,}200) \\
&\mathrm{Dose}_{ij}^{estimated} \sim N\left(\mathrm{Dose}_{ij}^{true}, \sigma_{dose\ uncertainty}^2\right) \\
&\beta_{0i} \sim N(\beta_0, \sigma^2) \\
&\beta_0 \sim N(0{,}100) \\
&\beta_1 \sim N(0{,}100) \\
&\sigma^2 \sim InverseGamma(0.01{,}0.01) \\
&\sigma_{dose\ uncertainty} = \text{known value(s) in dB}.
\end{aligned} \tag{2}$$

The green text in Eq. (2) indicates changes made to the Original Model in Eq. (1). Here, the prior true dose distribution, $\mathrm{Dose}_{ij}^{true}$, is assumed to be uniform between -100 and 200 dB, which are arbitrary small and large limits on the dose. The estimated dose, $\mathrm{Dose}_{ij}^{estimated}$, which is data supplied to the model, is assumed to be normally distributed about the true dose $\mathrm{Dose}_{ij}^{true}$ with standard deviation $\sigma_{dose\ uncertainty}$. The $\sigma_{dose\ uncertainty}$ is a known value that is supplied to the model, which is estimated from the noise dose estimation procedure (examples of this are given in Section 4). The main difference between this Extended Model and the Original Model is that it uses the true dose in place of the estimated dose in the participant logistic curve. In this work, $\sigma_{dose\ uncertainty}$ is assigned a single value meaning the dose uncertainty is assumed to be constant for all doses. However, it is possible to allow $\sigma_{dose\ uncertainty}$ to take multiple values if the uncertainty is different for each dose, i.e., $\sigma_{dose\ uncertainty} \rightarrow \sigma_{dose\ uncertainty,ij}$.

[a] $logit^{-1}(x) = \frac{1}{1+e^{-x}}$

C. EXTENDED MODEL WITH BERKSON DOSE UNCERTAINTY TERM

The BMLR model extended to include a term to account for additive normal Berkson dose uncertainty is given by

$$\begin{aligned}
&H_{ij} \sim Bernoulli(p_{ij}) \\
&p_{ij} = logit^{-1}(\beta_{0i} + \beta_1 \text{Dose}_{ij}^{true}) \\
&\text{Dose}_{ij}^{true} \sim N\left(\text{Dose}_{ij}^{estimated}, \sigma_{dose\ uncertainty}^2\right) \\
&\beta_{0i} \sim N(\beta_0, \sigma^2) \\
&\beta_0 \sim N(0,100) \\
&\beta_1 \sim N(0,100) \\
&\sigma^2 \sim InverseGamma(0.01,0.01) \\
&\sigma_{dose\ uncertainty} = \text{known value(s) in dB}.
\end{aligned} \tag{3}$$

The green text in Eq. (3) again indicates changes made to the Original Model in Eq. (1). Here, the true dose Dose_{ij}^{true} is assumed to be normally distributed about the estimated dose $\text{Dose}_{ij}^{estimated}$, which is data supplied to the model, with standard deviation $\sigma_{dose\ uncertainty}$. Again, the main difference between this Extended Model and the Original model is that it uses the true dose in place of the estimated dose in the participant logistic curve. It is also notable that the computational time to fit this Extended Model using R with the rjags package is considerably lower than the classical uncertainty Extended Model in the previous section. This is not entirely surprising because the Berkson formulation does not have an additional call to $\text{Dose}_{ij}^{true} \sim Unif(-100,200)$ for each Gibbs sampler iterate that is present in the Classical formulation.

3. SIMULATION APPROACH

To assess the effectiveness of the extended BMLR models in correcting dose-response curve flattening due to dose uncertainty, the models are fit to simulated data. The use of simulated data allows exact knowledge of the true underlying dose-response curve, the true doses, and the exact dose uncertainty. The general simulation approach outlined here is to generate responses (highly annoyed or not) to known true doses from a known true underlying dose-response curve. Then, fit the models to this unperturbed dataset. Next, perturb the doses by a draw from a normal distribution with known standard deviation to simulate additive normal dose uncertainty. Finally, fit the models to the perturbed data, and see if the true underlying dose-response curve is recovered. The simulation dataset generation and model fitting process is illustrated with a flow diagram in Figure 1.

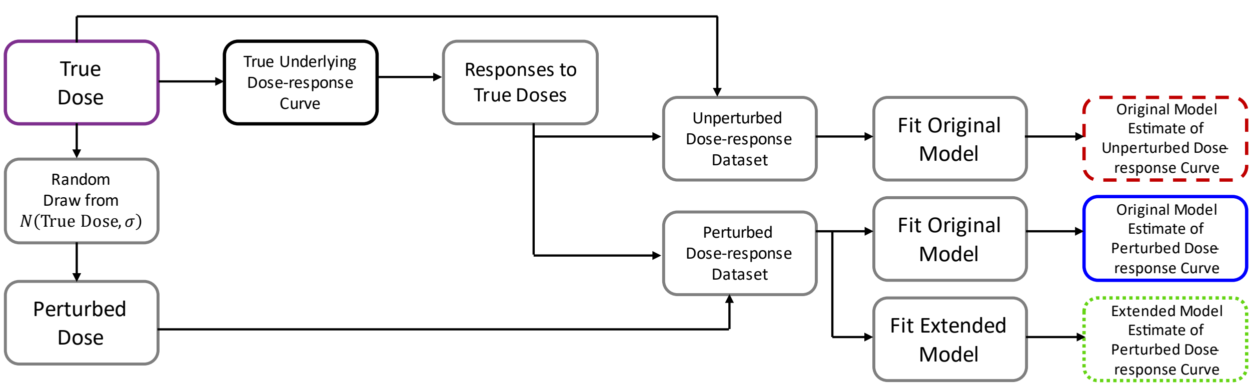

Figure 1. Simulation Approach Flow Diagram.

While the Original and Extended Models of Eq. 1-3 are written in a multilevel framework, it is not necessary to simulate a collection of many participants to illustrate the effectiveness of the extended models in correcting the flattening effect caused by dose uncertainty. For simplicity, the models can be adjusted to be used for a non-multilevel scenario, e.g., where there is only a single participant or where a collection of participants only responded once. (This is ordinary logistic regression with a term to account for additive normal dose uncertainty.)

Two datasets with differing dose ranges were simulated to illustrate the effectiveness of the extended models. They will be referred to as Simulation 1: Fully-Sampled Dose-response Curve and Simulation 2: Sparsely Sampled Dose-response Curve. Simulation 1 is used to show when the Extended Models are able to recover the true underlying dose-response curve. Simulation 2 is used to show when the Extended Models are unable to recover the true underlying dose-response curve.

The true underlying dose-response curve used to generate the responses for the simulated datasets is given by

$$\begin{aligned} &H_j \sim Bernoulli(p_j) \\ &p_j = logit^{-1}(\beta_0 + \beta_1 \text{Dose}_j^{true}) \\ &\beta_0 = -15.7 \\ &\beta_1 = 0.154 \text{ PL dB}^{-1}. \end{aligned} \tag{4}$$

Given that the dose is measured in Perceived Level (PL),[17] the PL dose that elicits high annoyance 50% of the time is $d_{50} = -\beta_0/\beta_1 \cong 102$ dB, which is near the PL of the Concorde's sonic boom.[18] The choices of β_0 and β_1 are somewhat arbitrary and were selected for the purpose of illustrating the effectiveness of the extended models in the context of realistic sonic boom levels. These values for β_0 and β_1 for the simulation dataset correspond to the β parameters from the Original Model fit to a subset of participant responses from the WSPR dataset. The subset includes only those participants who responded both highly annoyed and not annoyed at least once during the test. These β parameters were chosen because there are few highly annoyed responses when using all WSPR participants, and the effectiveness of the Extended Model improves when there are more highly annoyed responses (as shown in Section 4).

For both simulations, annoyance responses were generated from the survey participant model in Eq. 4 for each of the 201,000 true doses. In Simulation 1, the doses spanned a 200 dB range centered on d_{50} in 1 dB increments. There are 1,000 responses at each of the 201 dose values. For Simulation 2, the doses spanned a 20 dB range from 70 to 90 dB, which is similar to that which the X-59 will produce,[19] in 0.1 dB increments. Again, there are 1,000 responses at each of the dose values. The simulated responses to the true doses will be referred to as the Unperturbed Dataset for both Simulations.

Next, the doses in these datasets were perturbed to simulate dose uncertainty. The Perturbed Dataset for each simulation was generated by adding a random value from a normal distribution with a mean of 0 and standard deviation of 8 dB to the true dose. The value of 8 dB is arbitrary but was chosen to illustrate the flattening effect, which occurs despite the large number of participant responses. For context, the estimate of the dose uncertainty standard deviation in QSF18 was 4.9 dB, so the simulated uncertainty is significantly greater than what was observed in the field for the PL metric.

4. SIMULATION RESULTS

The Original and Extended Models were fit to the Unperturbed and Perturbed Datasets outlined in the previous section to illustrate the effectiveness of the models to recover the true underlying dose-response curve. In the following section, the "Extended Model" refers to either the Classical or Berkson extended model as they have identical results. Only one of these (Berkson) is shown in the Figures for visual clarity.

A. SIMULATION 1: FULLY SAMPLED DOSE-RESPONSE CURVE RESULTS

Simulation 1 consists of data from the fully sampled dose-response curve. When the Extended Model was fit to the Simulation 1 data, it recovered the true underlying dose-response curve. Figure 2 shows the results of fitting the

1. Original Model to the Unperturbed Dataset (red dashed)
2. Original Model to the Perturbed Dataset (blue)
3. Extended Model to the Perturbed Dataset (green dotted)

to the fully sampled dose-response curve. The true underlying dose-response curve is also plotted (black), but it is covered by the red dashed and green dotted curves. The colors correspond to those in the flow diagram (Figure 1). The difference between the red dashed and blue curves shows the flattening effect of dose uncertainty. While the flattening effect is small in this dataset, there is about a 4 dB underestimation in the dose that elicits high annoyance 5% of the time. This underestimation could matter in a practical sense, e.g., if a noise limit were set using the attenuated blue curve, it would place unnecessary difficulty on aircraft designers to reduce the noise. Thus, the Extended Model is preferred as it can recover the true underlying dose-response curve.

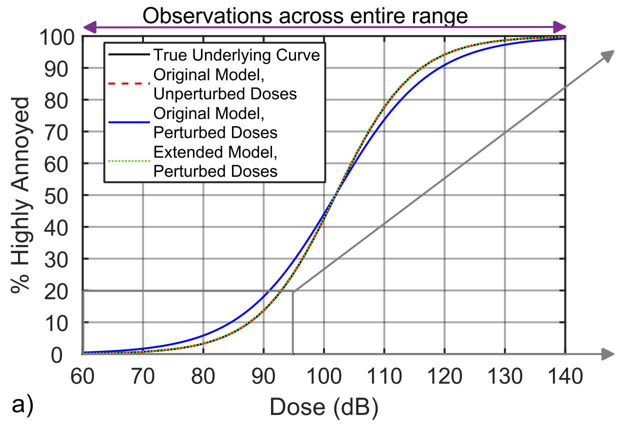

Figure 2. Fully sampled dose-response curve with varying combinations of the Original and Extended Models fit to the Unperturbed and Perturbed data. (The Black, Red Dashed, and Green Dotted lines are overlapping.)

B. SIMULATION 2: SPARSELY SAMPLED DOSE-RESPONSE CURVE RESULTS

Simulation 2 consists of data from the sparsely sampled dose-response curve, i.e., the doses range from 70 to 90 dB. When the Extended Model was fit to the Simulation 2 data, it was unable to recover the true underlying dose-response curve. Figure 3 shows the results of fitting the

1. Original Model to the Unperturbed Dataset (red dashed)
2. Original Model to the Perturbed Dataset (blue)
3. Extended Model to the Perturbed Dataset (green dotted)

to the sparsely sampled dose-response curve. The true underlying dose-response curve (black) is also plotted. The colors correspond to those in the flow diagram (Figure 1). In this case, neither model fit to the Perturbed Dataset is able to recover the true underlying dose-response curve, and there is minimal difference in the resulting erroneous fits (blue and green). Because the Original Model fit to the Unperturbed Dataset is able to recover the true underlying curve, the erroneous green and blue curves are due to the uncertainty in the dose.

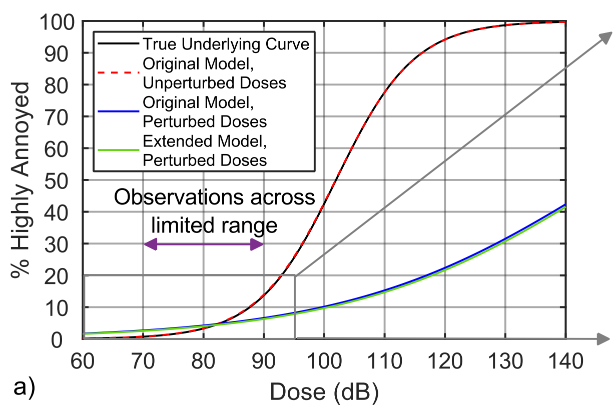

Figure 3. Sparsely sampled dose-response curve with varying combinations of the Original and Extended Models fit to the Unperturbed and Perturbed data.

C. DISCUSSION OF SIMULATION RESULTS

The two simulations showed scenarios where the Extended Model is and is not effective in recovering the true underlying dose-response curve. When a participant's dose-response curve is sufficiently sampled, their true underlying dose-response curve can be recovered despite dose uncertainty. When a participant's dose-response curve is not sufficiently sampled, their true underlying dose-response curve cannot be recovered. In community noise testing, it is possible that there will be participants from both groups, fully and sparsely sampled, as well as the spectrum in between. In a multilevel framework, the Extended Model will recover the dose-response curve of those participants in the fully sampled group and have minimal impact on the sparsely sampled group. Thus, the Extended Model is expected to perform no worse than the Original Model and may perform better the more fully sampled participants exist in the dataset.

5. APPLICATION OF BMLR MODELS TO SONIC BOOM SURVEY DATA

In this Section, the Original and Extended Models are applied to real sonic boom noise-annoyance data from two previous NASA community noise flight tests.

A. WSPR

The WSPR community noise survey flight test was conducted over Edwards Air Force Base in 2011. The ~1 square mile (2.6 square km) test area was exposed to 110 sonic boom events over the course of two weeks. The estimated PL dose ranged from 63 to 106 dB. High annoyance was reported 133 times out of 1981 (6.7%) survey responses from the 49 survey participants. The low number of highly annoyed responses indicates most participants' dose-response curves are likely in the "sparsely sampled" regime.

The test area was instrumented with 12 noise monitors, which recorded the sonic booms. The PL of the booms were computed. A distance-weighted measurement interpolation scheme was used to estimate the doses for each participant. A leave-one-out method was used to estimate the uncertainty in the doses at each noise monitor for each boom by computing the standard deviation of the difference between the estimated and measured sonic boom level at the monitor which was left out. Table 22 in the WSPR report includes the standard deviation of the difference between the estimated and measured noise level, and the relevant data are reproduced here in Table 1. The observation weighted average of the standard deviation from each of the 12 monitors is given in the last row of the Table. Thus, when applying the extended BMLR model to the WSPR dataset, $\sigma_{dose\ uncertainty} = \text{known value(s) in dB} = 3.7\text{ dB}$.

Table 1. Standard deviation of difference between estimated and measured dose in WSPR.

Noise Monitor Name	Standard Deviation of PL Difference (dB)	Number of Observations
Alpha	3.3	85
Bravo	3.7	84
Charlie	3.2	83
Delta	3.8	79
Echo	3.8	74
Foxtrot	4.3	73
Golf	3.7	79
Hotel	3.5	69
India	3.5	85
Juliet	3.8	82
Kilo	3.7	77
Lima	3.6	86
Observation Weighted Average	3.7	-

The Original and Extended Models were fit to the WSPR noise-annoyance data using 2 chains, 4,000 burn-in and 400,000 posterior draws, mirroring Lee et al. (2020). Table 2 shows the expected values and standard deviation [SD] of the model fit parameters, and Figure 4 shows the dose-response curve of the average participant for the Original (solid) and Extended (dashed) Models. Only one Extended fit is shown for visual clarity because the models extended to include Classical and Berkson terms have nearly identical results. The expected value of the slope term β_1 is roughly 5% steeper for Extended Models, indicating there is minor flattening correction. There is not a significant difference, though, as there is overlap between the Original and Extended models' β_1

posterior distributions. There is also little practical difference between the Original and Extended model fits. There is only a 0.3 dB difference in the dose that elicits annoyance 50% of the time d_{50}, and only about 1 dB difference in the dose that elicits high annoyance 5% of the time.

Table 2. Model fit parameters for the WSPR test.

Model	β_0 [SD]	β_1 [SD]	σ [SD]	$d_{50} = -\beta_0/\beta_1$ (dB)
Original Model	-19.6 [1.7]	0.156 [0.015]	3.60 [0.77]	125.6
Extended Model with Classical Dose Uncertainty Term	-20.7 [2.0]	0.164 [0.018]	3.82 [0.85]	126.2
Extended Model with Berkson Dose Uncertainty Term	-20.6 [2.0]	0.164 [0.018]	3.82 [0.85]	125.6

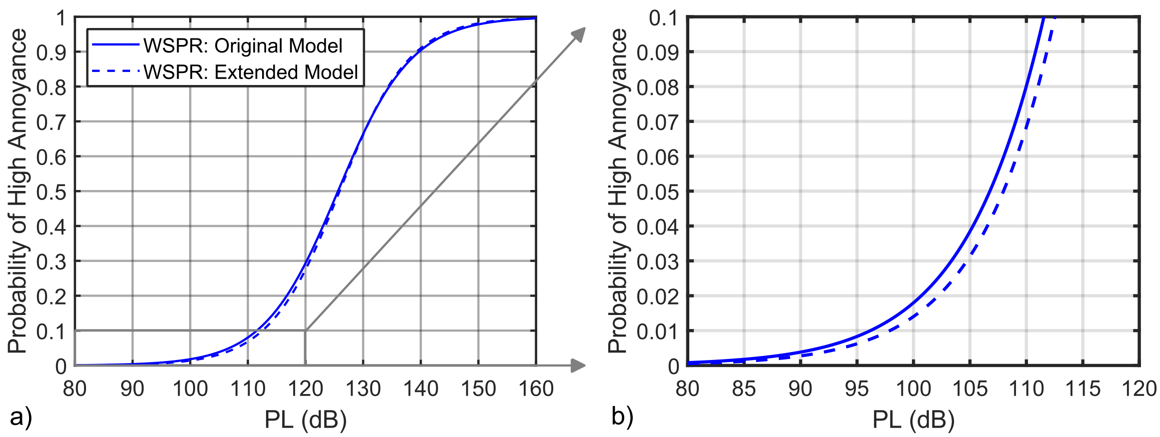

Figure 4. Dose-response curve of the average WSPR participant modeled with the Original and Extended Models.

B. QSF18

The QSF18 community noise survey flight test was conducted over Galveston, TX in 2018. The ~60 square mile (155 square km) test area was exposed to 52 sonic boom events (although 3 events had no valid survey responses) over the course of eleven days. The estimated PL dose ranged from 56 to 90 dB. High annoyance was reported 47 times out of 4998 survey responses (0.9%) from 371 participants. This indicates most participants' dose-response curves are likely in the "sparsely sampled" regime.

The test area was instrumented with 12 noise monitors, which recorded the sonic booms. A method combining measurements and PCBoom[20,21] predictions was used to estimate the doses for each participant. Details of the dose estimation method can be found in [4-5]. A leave-one-out method was used to estimate the uncertainty in the doses at each noise monitor for each boom by computing the standard deviation of the difference between the estimated and measured sonic boom level at the monitor which was left out. Table 6-10 in the QSF18 report gives the standard deviation of the difference between the estimated and measured noise level, and the relevant data are reproduced here in Table 3. The observation weighted average of the standard deviation from each of the 12 monitors is given in the last row of the Table. Thus, when applying the extended BMLR model to the QSF18 dataset, $\sigma_{dose\ uncertainty} = \text{known value(s) in dB} = 4.9\ \text{dB}$.

Table 3. Standard deviation of difference between estimated and measured dose in QSF18.

Noise Monitor Name	Standard Deviation of PL Difference (dB)	Number of Observations
Alpha	4.6	50
Bravo	5.6	48
Charlie	4.3	51
Delta	5.1	47
Echo	4.0	47
Foxtrot	4.4	48
Golf	4.6	42
Hotel	7.0	35
India	3.6	28
Juliet	6.8	30
Kilo	4.9	46
Lima	4.6	4
Observation Weighted Average	4.9	-

The Original and Extended Models were fit to the QSF18 noise-annoyance data[b] using 2 chains, 4,000 burn-in and 400,000 posterior draws, mirroring Lee et al. (2020). Table 4 shows the expected value and standard deviation [SD] of the model fit parameters, and Figure 5 shows the dose-response curve of the average participant for the Original (red, solid) and Extended Model (red, dashed) along with the WSPR results (blue) repeated from the previous Section. Only one Extended Model fit is shown for visual clarity because the models extended to include Classical and Berkson terms have nearly identical results. The expected value of the slope term β_1 is roughly 10% steeper for the Extended Models, indicating there is minor flattening correction. There is not a significant difference, though, as there is overlap between the Original and Extended models' β_1 posterior distributions. There is also little practical difference between the Original and Extended Model fits. There is only a 1.2 dB difference in the dose that elicits annoyance 50% of the time d_{50}, and only about 0.2 dB difference in the dose that elicits annoyance 5% of the time.

Table 4. Model fit parameters for the QSF18 test.

Model	$\boldsymbol{\beta_0}$ [SD]	$\boldsymbol{\beta_1}$ [SD]	$\boldsymbol{\sigma}$ [SD]	$\boldsymbol{d_{50} = -\beta_0/\beta_1}$ (dB)
Original Model	-19.0 [2.4]	0.153 [0.028]	2.61 [0.54]	124.2
Extended Model with Classical Dose Uncertainty Term	-20.9 [3.3]	0.169 [0.036]	2.80 [0.61]	123.7
Extended Model with Berkson Dose Uncertainty Term	-20.7 [3.2]	0.169 [0.036]	2.77 [0.58]	122.5

[b] The QSF18 dose-response data is available at https://doi.org/10.1121/10.0001021

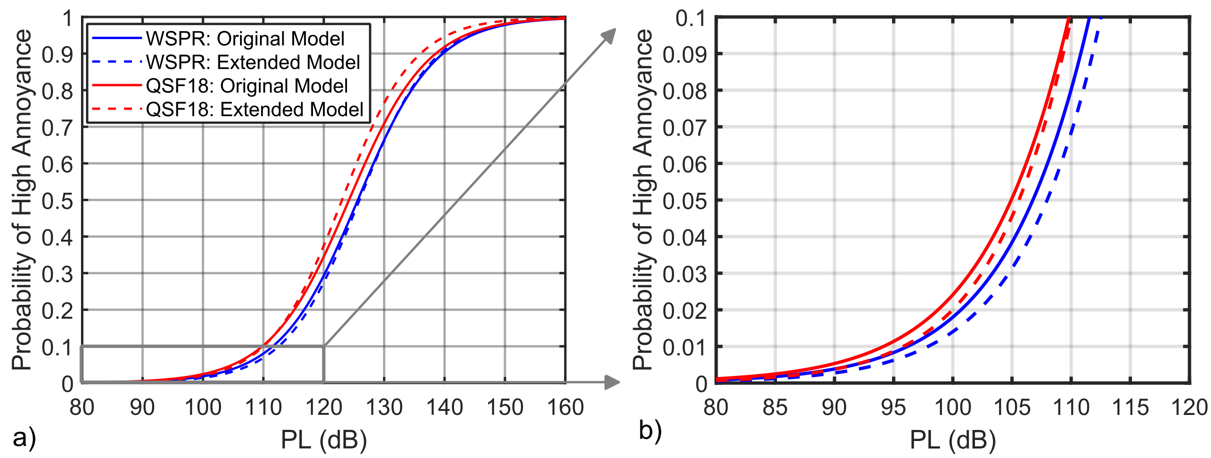

Figure 5. Dose-response curves of the average QSF18 and WSPR participant modeled with the Original and Extended Models.

C. POPULATION AVERAGED RESULTS

The results and dose-response curves in Figures 4 and 5 are for the average participant, but the population average curve is more relevant for the regulatory community. Lee et al. used a pointwise average method, which averages the individual dose-response curves from each of the posterior draws to estimate the population average curve and credible interval. Figure 6 shows the comparison of the population summary curves for the Original (black) and Extended (red) Models for WSPR and QSF18. The 95% credible intervals are shown as the dashed lines and overlap for the two models for both the WSPR and QSF18 datasets. The differences in results from the Original and Extended Models are also insignificant in a practical sense, as the difference in the dose that elicits annoyance from 5% of the population is approximately 0.5 dB for WSPR and 0.7 dB for QSF18.

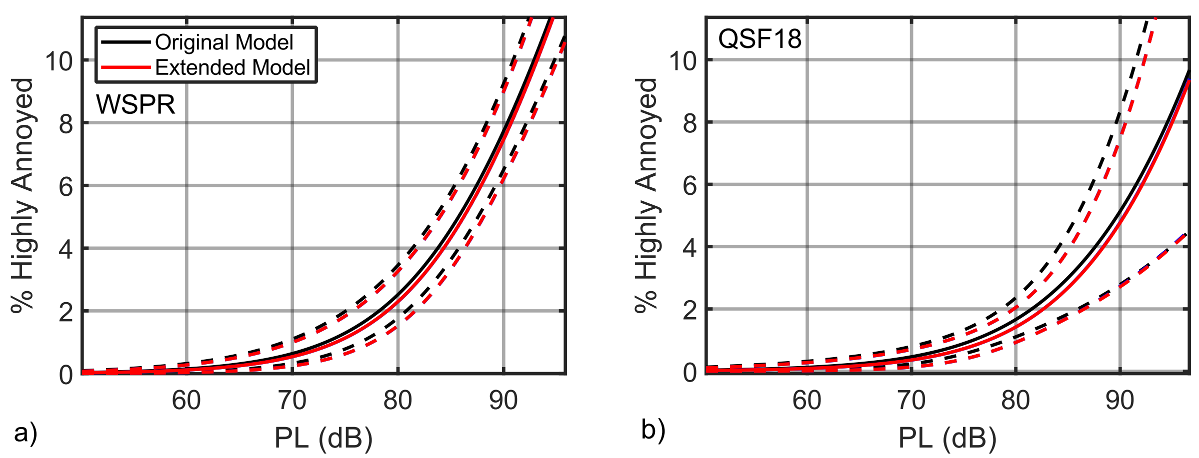

Figure 6. Population summary dose-response curves for a) WSPR and b) QSF18 with the Original and Extended Models. Dashed lines denote 95% credible intervals.

D. DISCUSSION OF SONIC BOOM SURVEY DATA RESULTS

While the expected value of the slope term β_1 for both the WSPR and QSF18 datasets is slightly steeper when using the Extended Models, there is little practical difference in the population summary dose-response curves. The majority of participants fall into the "sparsely sampled" dose-response curve category, which may be the reason that there is little difference when the Extended Model is used. Using a fixed value for the dose uncertainty standard deviation instead of estimating the uncertainty for each dose may also be a contributor. Testing these hypotheses is left for future analyses.

For future X-59 low-boom community response testing, the estimated dose range is expected to be similar to these sonic boom community noise pilot studies, so the majority of participants may again fall into the "sparsely sampled" dose-response curve category. Thus, alternative methods for accounting for dose uncertainty in dose-response modeling may be necessary. However, the X-59 tests will likely be longer than WSPR and QSF18, so there will be more opportunities to sample each participant. The X-59's low-boom noise exposure region is also predicted to be more uniform in terms of loudness level than the F-18 low-boom dives in WSPR and QSF18, so this will likely result in lower dose uncertainty. It is also likely that the noise monitors will be clustered into groups to better estimate the uncertainty in dose due to atmospheric turbulence and ambient noise.[22] The uncertainty will also likely be estimated for each dose instead of using a single value of $\sigma_{dose\ uncertainty}$, which was a simplifying assumption implemented in this work. Future work could be to apply such a technique to estimate the dose uncertainty for each dose to the WSPR and QSF18 datasets, but the goal for this work was to exercise these extended models on simulated and field data.

6. SUMMARY AND CONCLUSIONS

Previous analyses of NASA sonic boom community noise-annoyance data were modeled assuming the noise dose was known without uncertainty. Dose uncertainty is known to cause attenuation or flattening of the dose-response curve. The Lee et al. Bayesian multilevel logistic regression model was extended to include a term to account for additive normally distributed dose uncertainty with zero mean and known standard deviation in a classical or Berkson framework.

Two simulated datasets were generated to demonstrate the cases where the Extended Models can and cannot correct for this uncertainty-induced flattening of the estimated dose-response curve. When an individual participant's dose-response curve is sufficiently sampled, the Extended Models are able to correct the flattening effect. However, if the participant's dose-response curve is sparsely sampled, the Extended Model is unable to correct the flattening effect.

Finally, the model was applied to two NASA sonic boom community noise surveys, WSPR and QSF18. There are few observations of high annoyance in these datasets, so the majority of participants likely fall into the "sparsely sampled" category. However, the expected value of the slope term β_1 increased by 5 to 10% when the data are fit using the extended models compared to the original Lee et al. BMLR model that does not account for uncertainty. This did not translate to a practical difference in the population average dose-response curve with the dose eliciting high annoyance in 5% of the population shifting by less than 1 dB with overlapping credible intervals.

For X-59 community response testing, it is plausible that a similar result could arise due to the similar dose range to these pilot tests, which does not necessarily mean that dose uncertainty can be neglected. The process to estimate dose uncertainty in the X-59 community response tests will be more thorough than the pilot tests, and alternative dose-response models or approaches to accounting for uncertainty may still be needed.

ACKNOLWEDGEMENTS

The authors appreciate the support from the NASA Commercial Supersonic Technology Project and those involved in reviewing the work including Nathan Cruze, Alexandra Loubeau, Ran Cabell, and Steve Bauer.

REFERENCES

[1] X-59 Mission Webpage, *NASA,* (2022); nasa.gov/X59

[2] L.P. Ozoroski, et al., *NASA Community Test Workshop 2*, (2021); ntrs.nasa.gov/citations/20210025956

[3] J. A. Page, et al., "Waveforms and Sonic Boom Perception and Response (WSPR): Low-Boom Community Response Program Pilot Test Design, Execution, and Analysis," *NASA Contractor Report*, NASA/CR-2014-218180, (2014); ntrs.nasa.gov/citations/20140002785

[4] J. A. Page, et al., "Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas Risk Reduction for Future Community Testing with a Low-Boom Flight Demonstration Vehicle," *NASA Contractor Report*, NASA/CR-2020-220589/Volume I, (2020); ntrs.nasa.gov/citations/20200003223

[5] J. A. Page, et al., "Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas Risk Reduction for Future Community Testing with a Low-Boom Flight Demonstration Vehicle [Appendices]," *NASA Contractor Report*, NASA/CR-2020-220589/Volume II, (2020); ntrs.nasa.gov/citations/20200003224

[6] J. Lee, J. Rathsam, and A. Wilson, "Statistical Modeling of Quiet Sonic Boom Community Response Survey Data," *NASA Technical Memorandum*, NASA/TM-2019-220427, (2019); ntrs.nasa.gov/citations/20190033466

[7] J. Lee, J. Rathsam, and A. Wilson, "Bayesian statistical models for community annoyance survey data," *J. Acoust. Soc. Am.* **147**(4), 2222-2234 (2020); doi: 10.1121/10.0001021

[8] R. D. Horonjeff, "An examination of dose uncertainty and dose distribution effects on community noise attitudinal survey outcomes," *J. Acoust. Soc. Am.* **150**(3), 1691-1701 (2021); doi: 10.1121/10.0005949

[9] R. D. Horonjeff, "Mathematical characterization of dose uncertainty effects on functions summarizing findings of community noise attitudinal surveys," *J. Acoust. Soc. Am.* **151**(4), 2739-2750 (2022); doi: 10.1121/10.0010311

[10] C. Kirisits, C. Lechner, and H. Kirisits, "Impact of uncertainties related to noise indicator determination on observed exposure-effect relationship," *Noise Health* **20**(96), 212-216 (2018); doi: 10.4103/nah.NAH_57_17

[11] R. J. Carroll, et al., "Chapter 3: Linear Regression and Attenuation," *Measurement Error in Nonlinear Models 2nd Edition*, Chapman and Hall/CRC 41-64 (2006); doi: 10.1201/9781420010138

[12] J. Berkson, "Are there Two Regressions?" *J. Am. Stat. Assoc.* **45**(250), 164-180 (1950); doi: 10.1080/01621459.1950.10483349

[13] D. W. Schafer and E. S. Gilbert, "Some Statistical Implications of Dose Uncertainty in Radiation Dose-Response Analyses," *Radiation Res.* **166**(1), 303-312 (2006); doi: 10.1667/RR3358.1

[14] R Core Team, "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*; R-project.org

[15] M. Plummer, A. Stukalov, and M. Denwood, "Package 'rjags': Bayesian Graphical Models using MCMC," (2021); cran.r-project.org/web/packages/rjags/rjags.pdf

[16] M. Plummer, "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," *Proc. 3rd Intl. Workshop on Distrib. Stat. Comp.* **124**, 1-10 (2003).

[17] S. S. Stevens, "Perceived level of noise by Mark VII and decibels (E)," *J. Acoust. Soc. Am.* **51**, 575–601. doi: 10.1121/1.1912880

[18] W. J. Doebler and J. Rathsam, "How loud is X-59's shaped sonic boom?" *Proc. Mtgs. Acoust.* **36**, 040005 (2019); doi: 10.1121/2.0001265

[19] W. J. Doebler and A. Loubeau, "The noise dose range of the X-59 estimated from propagation simulations," *J. Acoust. Soc. Am.* **150**, A208 (2021); doi: 10.1121/10.0008134

[20] J. A. Page, K. J. Plotkin, and C. Wilmer, "PCBoom Version 6.6 Technical Reference and User Manual," *Wyle Report*, WR 10-10, (2010)

[21] J. A. Page, et al., "PCBoom Version 7.1 User's Guide," *NASA Technical Memorandum*, NASA/TM-20205007703, (2020); ntrs.nasa.gov/citations/20205007703

[22] J. Klos, "Recommendations for Using Noise Monitors to Estimate Noise Exposure During X-59 Community Tests," *NASA Technical Memorandum*, NASA/TM-20205007926 Corrected Copy, (2020); ntrs.nasa.gov/citations/20205007926