NASA/TM–20230008716

Multilevel Logistic Regression with Random Slope for Community Annoyance Survey Data

Aaron B. Vaughn and Nathan B. Cruze
Langley Research Center, Hampton, Virginia

July 2023

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/TM–20230008716

Multilevel Logistic Regression with Random Slope for Community Annoyance Survey Data

Aaron B. Vaughn and Nathan B. Cruze
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

July 2023

Available from:

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

This page intentionally left blank.

Table of contents page

Table of Figures

Table of Tables

Nomenclature and Acronyms

DIC	Deviance information criterion
GEE	Generalized estimating equations
HA	Highly annoyed
MLR	Multilevel logistic regression
MOR	Multilevel ordinal regression
PL	Perceived Level
QSF18	Quiet Supersonic Flights 2018
WSPR	Waveforms and Sonic Boom Perception and Response

Executive Summary

This paper documents recent dose-response modeling work comparing the results of a Bayesian multilevel logistic regression (MLR) model with a fixed slope to one with a random slope using data acquired during the Waveforms and Sonic Boom Perception and Response (WSPR) and Quiet Supersonic Flights 2018 (QSF18) tests. Previously reported dose-response modeling efforts of WSPR and QSF18 data have used a MLR model with a fixed slope term. A random slope may more accurately depict the dose-response relationship of individuals in the efforts to produce a population summary dose-response curve. Results described here for the WSPR and QSF18 data indicate minimal difference between the modeling methods. The simpler fixed slope model is preferable for these data, but these results do not preclude consideration of a random slope term in modeling efforts of future X-59 community test data.

I. INTRODUCTION AND BACKGROUND

Multilevel models are presently recommended to model dose-response data for upcoming X-59 community tests. Given the longitudinal data to be collected, accounting for within-subject correlation with a multilevel model is a natural choice as the model contains participant-specific parameters. Within the framework of a multilevel model, there are additional options to consider, such as using ordinal responses versus dichotomizing for binary responses or setting the slope parameter to be fixed versus random. A fixed slope assumes each participant has the same rate of annoyance while a random slope allows each individual's curve to reflect a unique rate of annoyance (Cruze et al., 2022; Fitzmaurice et al., 2012). While the end product in the dose-response modeling efforts is a population summary curve, a random slope may provide more realistic individual curves and aid in the development of the population summary curve.

Community response data from previous NASA field studies provide an opportunity to prepare analysis methods for future X-59 community response data. Waveforms and Sonic Boom Perception and Response (WSPR) was conducted in 2011 near Edwards Air Force Base in California with 50 recruited participants. Quiet Supersonic Flights 2018 (QSF18) was conducted in 2018 in Galveston, Texas with 500 recruited participants. Both studies consisted of an F-18 performing a low-boom dive maneuver to generate quiet supersonic noise signatures to which participants provided their annoyance responses to each flyover event.

In the present context of dose-response modeling of community response to quiet supersonic noise signatures, a Bayesian multilevel logistic regression (MLR) with a fixed slope was introduced in Lee et al. (2019) where they investigated MLR among seven candidate models using WSPR single-event dose-response data. The top two performing models in terms of posterior predictive checking and deviance information criterion (DIC) were the MLR and multilevel ordinal regression (MOR). The idea behind posterior predictive checking is to check whether data replicated using the model are similar to observed data, and DIC compares the relative fit of models with preference given to lower DIC values. The MLR and MOR models were then applied by Lee et al. (2020) to QSF18 single-event dose-response data. Lee et al. (2020) used a fixed slope model, though "[Lee et al. (2019)] considered a random intercept and random slope model for [WSPR data]. Based on the DIC (Spiegelhalter et al., 2002), [Lee et al. (2019)] selected the simpler model, and [Lee et al. (2020)] used this to inform [their] model for the QSF18 data."

Building upon this work, Vaughn et al. (2022) compared MLR with generalized estimating equations (GEE) for single-event and cumulative dose-response modeling of WSPR and QSF18 data. Random slope was considered but not included based on conclusions from the work of Lee et al. (2019): "A participant-specific slope term could be used; however, minimal differences are observed relative to the fixed slope formulation, as noted by Lee et al. (2019) [(WSPR data)]. A possible reason for this is the relatively few observations per participant relative to the total number of participants. Regardless, for both model and computational simplicity, a single slope term is estimated while the intercepts vary."

The purpose of this paper is to document recent dose-response modeling efforts that compares the results of a Bayesian MLR model with a fixed slope to one with a random slope using WSPR and QSF18 single-event dose-response data. These data serve to demonstrate methodology and develop rationale for the inclusion or rejection of a random slope term in future X-59 dose-response modeling efforts.

II. DOSE-RESPONSE DATA

The present analysis only considers single-event dose-response data from WSPR and QSF18. The WSPR data are from the Wyle dataset with 1992 dose-response pairs from 49 participants, as reported in Lee et al. (2019) and Vaughn et al. (2022). Annoyance responses for WSPR were collected using an 11-point numerical scale ranging from 0 to 10. The responses are dichotomized where responses greater than or equal to 8 are considered to be "highly annoyed" (HA) and the remainder "not highly annoyed" as prescribed by Fields et al. (2001) and ISO/TS 15666 (2021). This WSPR single-event dose-response dataset is available with the Lee et al. (2019) publication at https://ntrs.nasa.gov/citations/20190002702.

The QSF18 data contain 4998 dose-response pairs from 371 participants, as reported in Lee et al. (2020) and Vaughn et al. (2022). A 5-point verbal response scale was used where *very annoyed* and *extremely annoyed* are considered to be HA as prescribed by Fields et al. (2001) and ISO/TS 15666 (2021). These QSF18 single-event dose-response data are available with the Lee et al. (2020) publication at https://doi.org/10.1121/10.0001021.

III. METHODOLOGY

The present analysis only considers Bayesian MLR models. The first model is the random intercept and fixed slope model, referred to hereafter as the fixed slope model. This model is implemented as described in Section 2.3.2 of Lee et al. (2019), Section D.1 of Lee et al. (2020), and Section 3.A of Vaughn et al. (2022). More details regarding this model are available at the aforementioned references. The model uses the following nomenclature:

- H is the binary response
- p is the probability of high annoyance
- $i \in 1, \dots, S$ is the set of participant indices
- $j \in 1, \dots, n_i$ is the set of observation indices for participant i, where n_i indicates the toal number of responses from subject i

The fixed slope model is given in Equation 1 as follows:

$$
\begin{aligned}
&H_{ij}|p_{ij} \sim \text{Bernoulli}(p_{ij}) \\
&p_{ij}|\beta_{0i}, \beta_1 = \text{logit}^{-1}(\beta_{0i} + \beta_1 PL_{ij}) \\
&\beta_{0i}|\beta_0, \sigma_0^2 \sim N(\beta_0, \sigma_0^2) \\
&\beta_0 \sim N(0,100) \\
&\beta_1 \sim N(0,100) \\
&\sigma_0^2 \sim \text{InverseGamma}(0.01,0.01)
\end{aligned}
\quad (1)
$$

The first three lines of Equation 1 are the fixed slope MLR model, and the last three are the noninformative prior distributions for model parameters β_0, β_1, and σ_0^2.

The second model is the random intercept and random slope model, referred to hereafter as the random slope model. The model is given in Equation 2, which resembles Equation 1 with new or redefined variables noted in red as follows:

$$
\begin{aligned}
&H_{ij}|p_{ij} \sim \text{Bernoulli}(p_{ij}) \\
&p_{ij}|\beta_{0i}, \beta_{1i} = \text{logit}^{-1}(\beta_{0i} + \beta_{1i} PL_{ij}) \\
&\beta_{0i}|\beta_0, \sigma_0^2 \sim N(\beta_0, \sigma_0^2) \\
&\beta_{1i}|\beta_1, \sigma_1^2 \sim N(\beta_1, \sigma_1^2)
\end{aligned}
\quad (2)
$$

$$\beta_0 \sim N(0,100)$$
$$\beta_1 \sim N(0,100)$$
$$\sigma_0^2 \sim \text{InverseGamma}(0.01,0.01)$$
$$\sigma_1^2 \sim \text{InverseGamma}(0.01,0.01)$$

The first four lines of Equation 2 constitute the random slope MLR model, and the last four are the noninformative prior distributions for model parameters β_0, β_1, σ_0^2, and σ_1^2.

Both models are implemented in the R statistical programming language with model parameters estimated using Markov chain Monte Carlo (MCMC) sampling with the software Just Another Gibbs Sampler (JAGS) Version 4.3.0 (Plummer, 2003). Two chains are used for modeling each dataset. Following Lee et al. (2019) for the WSPR data, 80,000 posterior draws are used after deleting the first 1,000 burn-in samples. Following Lee et al. (2020) for the QSF18 data, 400,000 posterior draws are used after deleting the first 4,000 samples. No thinning is used in either case and the number of posterior draws is deemed sufficient for convergence diagnostics. Additional information regarding the computational runtime is included in Appendix A.

In order to marginalize the MLR results into a population summary curve, a pointwise averaging is performed of the posterior participant-level curves as was done in Lee et al. (2019), Lee et al. (2020), and Vaughn et al. (2022). This is not a pointwise averaging of the curve probabilities but rather an average across the distribution of posterior draws for each participant. From these distributions, 95% credible intervals are computed using the 0.025 and 0.975 quantiles. This is an in-sample approach. Implementing an out-of-sample approach of marginalizing by integrating over random effects as described by Pavlou et al. (2015) would require further considerations to account for the additional random effect, perhaps along the lines of what is proposed in Hedeker et al. (2018).

IV. RESULTS

Results of this comparison study consist of figures depicting the population summary dose-response curves and tables noting the parameter estimates and deviance information criterion (DIC) values. Population summary dose-response curves are plotted with pseudodata that depict the average annoyance across observations and are plotted in 1-dB bins.

A. WSPR

Comparisons between the fixed slope and random slope MLR models with WSPR data are shown in Figure 1. The two models produce nearly identical population summary curves at PL values greater than 80 dB. Below 80 dB, the random slope model deviates with higher %HA values. Estimated parameter values for the fixed slope model are given in Table 1 and for the random slope model in
Table 2. There are no significant differences between the estimated parameter values. The DIC values in Table 3 slightly favor the fixed slope model but are not significantly different. The σ_1^2 estimate is essentially zero, which may be evidence of attempting to fit a model that is too complex to be properly supported by the data (see related discussion in Bates et al., 2018).

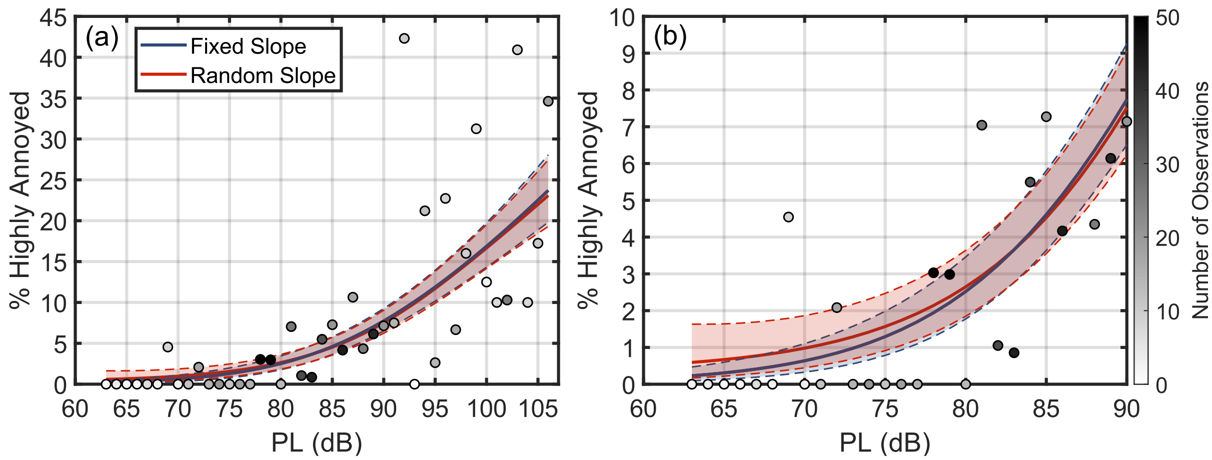

Figure 1. Dose-response population summary curves from MLR for WSPR showing the (a) full dose-response range and (b) zooming in on the lower dose-response range. Pseudodata are depicted as dots with shading noting the number of observations

Table 1. Parameter estimates for the fixed slope MLR model with single-event WSPR data. Values for β_0 and β_1 match those from Table 2.3 of Lee et al. (2019).

	Mean	SD	0.025 quant.	0.25 quant.	Median	0.75 quant.	0.975 quant.
β_0	-19.6	1.8	-23.2	-20.8	-19.6	-18.4	-16.4
β_1	0.156	0.015	0.127	0.145	0.155	0.166	0.186
σ_0^2	13.7	6.2	5.8	9.4	12.3	16.4	29.4

Table 2. Parameter estimates for the random slope MLR model with single-event WSPR data.

	Mean	SD	0.025 quant.	0.25 quant.	Median	0.75 quant.	0.975 quant.
β_0	-19.6	1.9	-23.4	-20.9	-19.6	-18.3	-15.9
β_1	0.156	0.017	0.123	0.146	0.153	0.169	0.186
σ_0^2	13.7	6.4	5.7	9.3	12.3	16.3	29.8
σ_1^2	0.0001	0.00053	5.3E-20	1.2E-12	5.8E-10	3.2E-07	0.00121

Table 3. DIC values for MLR models with single-event WSPR data.

	Fixed Slope	Random Slope
DIC	440	444
Mean Deviance	405	399
Mean PD	35	45

B. QSF18

Comparisons between the fixed slope and random slope MLR models with QSF18 data are shown in Figure 2. The two models produce essentially identical population summary curves and credible intervals. Estimated parameter values for the fixed slope model are given in Table 4 and for the random slope model in Table 5. There are no significant differences between the estimated parameter values. The σ_1^2 median estimate is again essentially zero, though the mean is larger. The

DIC values in Table 6 are minimally different, though interestingly the random slope model has a slightly lower value, though not significant enough to differentiate the two models. Again, as observed in the WSPR data, the near zero σ_1^2 estimate may suggest an overparameterization of the dose-response model.

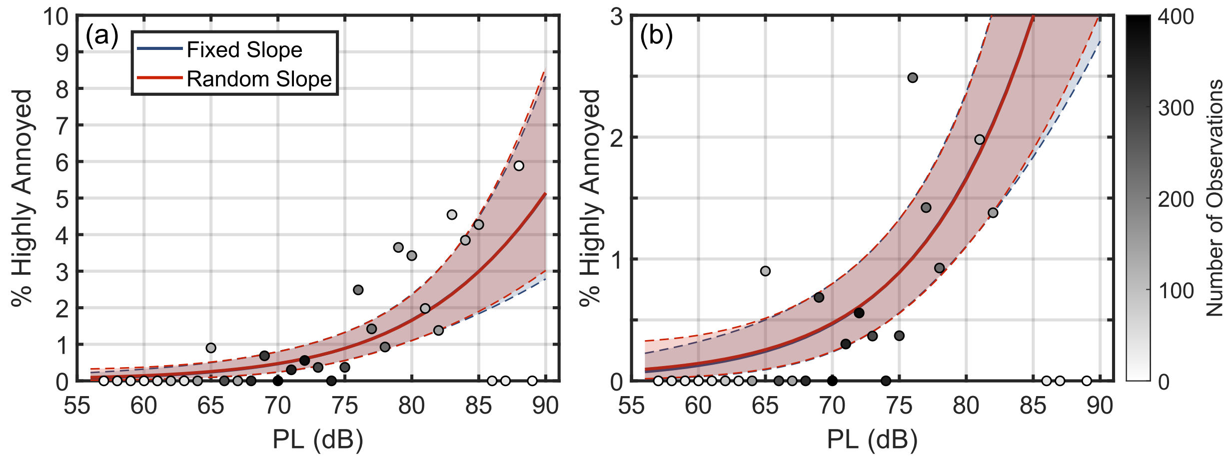

Figure 2. Dose-response population summary curves from MLR for QSF18 showing the (a) full dose-response range and (b) zooming in on the lower response range. Pseudodata are depicted as dots with shading noting the number of observations with points at (56,33.3) and (90, 16.7) not included to focus visualization on the summary curves.

Table 4. Parameter estimates for the fixed slope MLR model with single-event QSF18 data. Values for β_0 and β_1 match those from Table 1 of Lee et al. (2020).

	Mean	SD	0.025 quant.	0.25 quant.	Median	0.75 quant.	0.975 quant.
β_0	-19.0	2.4	-24.1	-20.6	-19.0	-17.3	-14.5
β_1	0.153	0.029	0.098	0.133	0.152	0.172	0.211
σ_0^2	7.1	3.0	3.0	5.0	6.6	8.6	14.6

Table 5. Parameter estimates for the random slope MLR model with single-event QSF18 data.

	Mean	SD	0.025 quant.	0.25 quant.	Median	0.75 quant.	0.975 quant.
β_0	-18.8	2.4	-23.7	-20.6	-18.4	-17.1	-15.2
β_1	0.150	0.028	0.112	0.124	0.144	0.168	0.201
σ_0^2	7.3	3.1	3.1	5.1	6.7	8.8	14.9
σ_1^2	0.0011	0.00329	2.4E-18	1.1E-12	1.2E-08	8.5E-05	0.0112

Table 6. DIC values for MLR models with single-event QSF18 data.

	Fixed Slope	Random Slope
DIC	353	351
Mean Deviance	283	270
Mean PD	70	81

V. CONCLUDING REMARKS

This paper compared results of a Bayesian MLR model with a fixed slope to one with a random slope using single-event dose-response data from WSPR and QSF18. The key takeaway is that there is minimal difference in the dose-response curves between the two models for both datasets. These results are in line with the justification for using the MLR model with a fixed slope in previous studies of WSPR and QSF18 data. Rationale supporting this choice include the DIC that generally favors the simpler fixed slope model and the near-zero estimates for the variance component associated with random slopes in both datasets. This preference for a fixed slope model with WSPR or QSF18 data does not preclude the use of a random slope in modeling efforts of future X-59 community test data. The suitability of random slope models for X-59 data is discussed in the next section.

While not pursued here, maximum likelihood methods for fitting random slope and random intercept models are available in popular statistical software packages and libraries (e.g., `glmer` in the `lme4` package (Bates et al. 2015) or `glmmPQL` in the `MASS` library (Ripley et al., 2022) in R statistical software). Cruze et al. noted a potential limitation of the `glmer` function (Cruze et al., 2022, pg. 13) for fitting models with multiple random effects, as the methods available in the function offer only a coarse approximation of the likelihood function. In short, depending on the data set in question, fitting the more complex random slope model may necessitate accepting an inappropriately coarse approximation, or else adopting other software in order to apply maximum likelihood methods for more complicated multilevel models like the random slope model.

A. Future Work

Community tests with the X-59 will consist of more supersonic flyover events than WSPR or QSF18 and include an incentive structure to retain participants, so the data may be better suited for a random slope model. Potential future work to assess the value of the random slope model or more complicated models for the X-59 community response data includes:

1) Further investigation of structures that permit nonzero estimates for variance components in models with multiple random effects. Given the type and quantities of data collected during WSPR and QSF18, the multilevel ordinal regression and the second stage of the 2-stage logistic regression outlined in the X-59 Community Response Testing Survey Analysis Plan (SM-02) may have sufficient structure to support dose-response models with multiple random effects.

2) Development of strategies to marginalize over multiple random effects and assess the sensitivities of the population average dose-response curve to estimates of multiple variance components.

These steps can help provide insights into the benefits of more complicated model structures and further develop the rationale for analysis of X-59 data once community testing is initiated.

References

Bates., D., Kliegl, R., Vasishth, S., and Baayen, R. H., (**2018**). “Parsimonious mixed models,” arXiv: 1506.04967.

Bates, D., Machler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4,” J. Stat. Softw., **67**(1), 1–48.

Cruze, N. B., Ballard, K. M., Vaughn, A. B., Doebler, W. J., Rathsam, J., and Parker, P. A., (**2022**). “Comparison of likelihood methods for generalized linear mixed models with application to Quiet Supersonic Flights 2018 data,” Technical Report No. NASA/TM-20220014998 (Last viewed 06/02/2023)

Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., Lercher, P., Vallet, M., Yano, T., Guski, R., Felscher-Suhr, U., and Schumer, R., (**2001**). “Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation,” J. Sound Vib. **242**(4), 641–679.

Fitzmaurice, G., Laird, N., and Ware, J. (**2012**). Wiley Series in Probability and Statistics, Applied Longitudinal Analysis, 2nd ed. (Wiley, Hoboken, NJ), pp. 1–752.

Hedeker, D., de Toit, S. H. C., Demirtas, H., and Gibbons, R. D., (**2018**). “A note on marginalization of regression parameters from mixed models of binary outcomes,” Biom. J. **74**(1), 354–361.

ISO (**2021**). ISO/TS15666:2021, *Acoustics—Assessment of Noise Annoyance by Means of Social and Socio-Acoustic Surveys* (International Organization for Standardization, Geneva; Switzerland).

Lee, J., Rathsam, J., and Wilson, A. (**2019**). “Statistical modeling of quiet sonic boom community response survey data,” Technical Report No. NASA/TM-2019-220427 (Last viewed 06/02/2023)

Lee, J., Rathsam, J., and Wilson, A. (**2020**). “Bayesian statistical models for community annoyance survey data,” J. Acoust. Soc. Am. **147**(4), 2222–2234.

Opsomer, J., Wivagg, J., Jodts, E., Ferg, R., Erciulescu, A., Stollery, P., Lympany, S., and Page, J. (**2022**) DRD SM-02 X-59 Community Response Testing Survey Analysis Plan [Updated Final]. HMMH Report No. 312430.001.004

Pavlou, M., Ambler, G., Seaman, S., and Omar, R. Z. (**2015**). “A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes,” BMC Med. Res. Methodol. **15**(1), 1–6.

Plummer, M. (**2003**). “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling,” in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria, **2**(1), 1–10.

Ripley, B., Bates, D. M., Hornik, K., Gebhardt, A., and Firth, D. (**2022**). “Package ‘MASS’: support Functions and Datasets for Venables and Ripley’s MASS”. https://cran.r-project.org/web/packages/lme4/lme4.pdf (Last viewed 06/07/2023)

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. **(2002)**. “Bayesian measures of model complexity and fit,” J. R. Stat. Soc.: Ser. B **64**(4), 583–639.

Vaughn, A. B., Rathsam, J., Doebler, W. J., and Ballard, K. M. (**2022**). "Comparison of two statistical models for low boom dose-response relationships with correlated responses," Proc. Mtgs. Acoust. **45**(1), 040001.

Acknowledgements

Work conducted in support of the NASA Commercial Supersonic Technology Project.

Appendix A: Computational Timing

Bayesian methods can be computationally intensive, conceivably leading to long runtimes. Appendix A of Vaughn et al. (2022) describes computational efficiencies that can be gained by running R code using *runjags* with parallel chains and utilizing binomial distributions. This builds upon the work of Lee et al. (2020) in which the R code ran using *rjags* with successive chains and Bernoulli distributions. The binomial distribution only gains efficiency if participants provide multiple responses at the same dose bin. For single events, if there are on the order of 17 1-dB dose bins and 86 single events, there may be modest computational speed to be achieved. On the other hand, parallel chains are generally a straightforward implementation to gain faster runtimes, regardless of dose binning.

Relative runtimes are given in Table 7 for the fixed and random slope models to demonstrate the potential additional runtime that a random slope model can incur. These tests were run in the following computing environment: Microsoft Windows 10 operating system with an Intel® CoreTM i9-9880H CPU @ 2.30GHz, 2304 Mhz, 8 Cores, 16 Logical Processors, and 32 GB of RAM. In Table 7, the previous fixed slope runtime refers to the runtime of the code used in Lee et al. (2019) or Lee et al. (2020), and the current fixed and random slope runtimes are the code used in the present analysis. The current fixed slope timing is about 2 to 3 times faster than the previous, primarily due to the parallel running of chains. When a random slope term is introduced, the runtime nearly reverts back to that of the previous fixed slope model. For future X-59 community test data, code efficiencies should be considered to ensure relatively quick turnaround for data processing when running a model with a random slope term or potentially any additional covariates.

Table 7. Relative runtimes for the previous and current MLR models for single-event dose-response data with a fixed and a random slope term.

Dataset	Previous Fixed Slope Runtime	Current Fixed Slope Runtime	Current Random Slope Runtime
WSPR	7.6 min	2.6 min	7.2 min
QSF18	1.6 hr	0.75 hr	1.2 hr