The relationship between aircraft noise exposure and day-use visitor survey responses in backcountry areas of national parks

Amanda Rapoza,[a)] Erika Sudderth, and Kristin Lewis
U.S. Department of Transportation, Volpe National Transportation Systems Center, 55 Broadway, Cambridge, Massachusetts 02142, USA

(Received 4 May 2014; revised 13 August 2015; accepted 19 August 2015; published online 14 October 2015)

To evaluate the relationship between aircraft noise exposure and the quality of national park visitor experience, more than 4600 visitor surveys were collected at seven backcountry sites in four U.S. national parks simultaneously with calibrated sound level measurements. Multilevel logistic regression was used to estimate parameters describing the relationship among visitor responses, aircraft noise dose metrics, and mediator variables. For the regression models, survey responses were converted to three dichotomous variables, representing visitors who did or did not experience slightly or more, moderately or more, or very or more annoyance or interference with natural quiet from aircraft noise. Models with the most predictive power included noise dose metrics of sound exposure level, percent time aircraft were audible, and percentage energy due to helicopters and fixed-wing propeller aircraft. These models also included mediator variables: visitor ratings of the "importance of calmness, peace and tranquility," visitor group composition (adults or both adults and children), first visit to the site, previously taken an air tour, and participation in bird-watching or interpretive talks. The results complement and extend previous research conducted in frontcountry areas and will inform evaluations of air tour noise effects on visitors to national parks and remote wilderness sites. [http://dx.doi.org/10.1121/1.4929934]

[SF] Pages: 2090–2105

I. INTRODUCTION

Decades of research have elucidated the effects of aircraft noise on residential populations near airports.[1–4] However, it has long been recognized that these effects and the corresponding residential dose-response relationships are not applicable to visitors to national parks and other natural areas as the ambient environments, aircraft overflight patterns, and population expectations in these settings are different than in residential areas surrounding airports. Spurred by the National Parks Overflights Act of 1987,[5] a number of federal agencies including the National Park Service (NPS), Federal Aviation Administration (FAA), United States Air Force (USAF), and United States Forest Service, initiated studies to collect data and examine the relationship between aircraft overflight noise and park visitor response.[6–9] The NPS and FAA studies, in particular, were structured to be directly comparable and resulted in the accumulation of a large database of frontcountry (short hike and overlook) noise exposure dose and related visitor response measurements.

Under a joint research program sponsored by FAA and NPS, researchers in the areas of acoustics, recreation management, psychology, and social science provided input into a detailed analysis of these frontcountry data. The analysis and resulting dose-response relationships showed differences between overlook and short-hike visitor evaluations of overflight noise and that evaluations are further mediated by the aircraft source type(s) and visitor ratings of the importance of natural quiet, whether they have visited the site before, and whether they visited with children.[10]

Additionally, the researchers recommended the study of backcountry sites with day- and overnight use as the highest priority for data needs.[11] Remote areas may have lower background sound levels that elevate the salience of aircraft noise, and visitors engaging in longer trips in remote areas may have higher expectations for environmental quality. Visitor surveys and associated noise exposure measurements were collected at backcountry sites with both day- and overnight use. This paper summarizes the field study and details the dose-response relationships developed for the majority subset of these data: the backcountry day-use (day-hike) visitor. Further information on the study methods, data collected and additional research topics explored can be found in the technical reports that document this effort.[12,13]

II. METHODS

A. Data collection

Data were collected at seven backcountry sites in four national parks (NP): Bryce Canyon, Glacier, Grand Canyon, and Zion. These sites (Table I) met the requirements set for the study, including average visitation length greater than 1 hr, hiking distance greater than 1 mile, audible aircraft overflights (both tour aircraft and high altitude jets), visitation density sufficient to provide the required number of surveys within the data collection period, and site topography appropriate for placement of acoustic monitors. Sites were further selected to cover a range of conditions within each of these criteria: low to high aircraft audibility, low to

[a)]Electronic mail: Amanda.Rapoza@dot.gov

TABLE I. Backcountry data collection locations.

Park	Site	Site type	Trail length (mile)	Description
Grand Canyon (GRCA)	Hermit Trail	Day and overnight hikes	9.2 one way	Strenuous trail descends 4340 ft from rim to Colorado River where campsites are available
	Grandview Trail	Day and overnight hikes	3.5 one way	Strenuous trail descends 2500 ft to Horseshoe Mesa, where campsites are available
Bryce Canyon (BRCA)	Fairyland Trail	Day hike	5.5 (loop)	Strenuous trail from Fairyland Point to Sunset Point (or reverse) descends 900 ft into Bryce Canyon
Zion (ZION)	West Rim Trail	Day and overnight hikes	14.4 one way	Moderately strenuous trail. Several campsites are available along the length of trail
	Taylor Creek Trail	Day hike	2.5 one way	Moderate, level trail
Glacier (GLAC)	Sperry Trail	Day and overnight hikes	6.2 one way	Strenuous trail ascends 3300 ft to Sperry Chalet where cabin accommodations and campsites are available. Branch trails provide day- and overnight-hike options to Snyder Lake Campground and Fish Lake
	Hidden Lake Trail	Short and day hikes	3.0 one way	Moderate trail ascends 500 ft to overlook, then descends 780 ft to Hidden Lake

high visitation density, and relatively short (2–4 hr) to long (multi-day) visit durations. For instance, Hermit Trail at Grand Canyon NP is characterized by high noise exposure, low visitation, and long visit durations, while Taylor Creek Trail at Zion NP is characterized by low noise exposure, high visitation, and short visit durations.

The data collected were (1) visitor surveys administered on-site at the conclusion of each visit, (2) visitor location data collected by GPS-based tracking devices, and (3) acoustical data collected by stationary monitors placed throughout the study area. The 15–20 day data collection efforts at each location were led by members of the research team from the Volpe Center, Resource Systems Group, Inc., and Harris Miller Miller and Hanson, Inc.

1. Visitor survey instruments

Three visitor survey instruments were utilized for this research. The surveys are consistent with previously used FAA and NPS dose-response and soundscape study instruments.[14] The survey instruments each offer alternative approaches to eliciting visitor responses. Administering them in parallel at a series of sites enables comparison and cross-calibration of results. The surveys contain identical introductory and demographic information sections, differing in the core section containing questions on aircraft and non-aircraft sounds and their effect on visitor experiences. Section II A 1 contains an overview of each survey instrument, while Sec. II A 2 contains additional detail on the survey questions.

The "human response to aviation noise survey version 1" (HR1) is an adaptation of the NPS/FAA/USAF "aircraft overflight studies visitor survey" used in the prior national parks frontcountry dose-response research. It is designed to understand the effects of aircraft noise on the visitor experience and uses direct queries regarding aircraft sounds.

The "human response to aviation noise survey version 2" (HR2) is an adaptation of the NPS "understanding and managing soundscapes in national parks: visitor use survey." It is designed to understand the audibility and acceptability of a variety of sounds within the soundscape.[15] Respondents are asked to identify and rate audible sounds during their visit from a list of ten, including both anthropogenic (aircraft, vehicles, voices, etc.) and natural (insects, birds, wind-in-trees, etc.) sources.

The "audio clip survey" (AC) is an adaptation of surveys designed to support the formulation of indicators and standards of quality for human-caused sounds in park management frameworks. Respondents are asked to evaluate aircraft sounds presented in a series of five audio clips, selected from a pool of 49 available clips.[16] Each clip contains a binaural recording of the loudest portion of a single aircraft overflight (helicopter, propeller aircraft or high altitude jet) within the context of natural sounds recorded in a national park setting. Aircraft sounds are overlaid on a single, low-level [approximately 20 dB(A)] recording of natural sounds, providing a consistent basis of limited, identifiable sounds (such as bird calls) to give respondents a context within which to rate each overflight. Each clip is 36 s in length, containing 26 s of overflight (with a fade-in/fade-out applied for a realistic-sounding beginning and end) and 5 s of natural sounds before and after the overflight. Clips are presented to respondents via circumaural headphones, embedded within the survey tool and automatically queued.

Audio clip research methods have been used in several laboratory and field-based soundscape assessment studies.[17–19] They provide a cost-effective means of collecting data, allowing for study at locations where aircraft overflights do not occur as well as precise control of noise exposure. However, audio clip studies may yield different results from surveys based on real exposures. Ratings of single overflights (as represented in a short audio clip) may not be similar to ratings of multiple, intermittent overflights, as experienced in a park setting with actual overflights. In addition, visitors are exclusively focused on listening and visual stimuli associated with actual overflights are not replicated in an audio clip.

Upon conclusion of the audio clip listening exercise, respondents are asked to evaluate *in situ* aircraft noise heard during the visit. This question was included to test the ability of this survey to provide data consistent with the HR1 survey instrument. Comparisons between these responses (HR1 and AC) were performed to identify any bias or sensitization to aircraft noise introduced during the listening exercise. It was placed after the listening exercise as the audio clip ratings were the main focus of this survey and given highest priority. The results presented in this paper include responses from the *in situ* portion of the audio clip survey but not detail on responses to the audio clips themselves.

2. Core survey questions

Each survey contains a core section of questions pertaining to aircraft and non-aircraft sounds and their effect on visitor experiences. The questions utilize evaluative dimensions of annoyance, interference with particular visit aspects, and/or acceptability. The "annoyance" and "interference with natural quiet and the sounds of nature" dimensions, detailed in the following text, are evaluated within the dose-response framework.

Annoyance (HR1 and AC). During your time at [site], how much did noise from airplanes, jets, helicopters, or other aircraft bother, disturb, or annoy you? (not at all, slightly, moderately, very, extremely).

Annoyance (HR2). How much did the sound of airplanes, jets helicopters, or other aircraft please or annoy you during your time at [site]? (extremely please, very please, moderately please, slightly please, neutral, slightly annoy, moderately annoy, very annoy, extremely annoy).

The HR1/AC annoyance question is formulated from recommendations of the International Commission on the Biological Effects of Noise (ICBEN) and informed by many socio-acoustic community surveys on aircraft noise.[20,21] The direct, closed-ended question is recommended as the primary measure of reactions to noise in residential areas. It uses a five-point unipolar scale based on the understanding that reactions to transportation noise are overwhelmingly either negative or neutral. In contrast, the HR2 survey uses a nine-point, bipolar scale, which has the advantage of including a natural positive/negative breakpoint and flexibility for use with both negatively and positively perceived sounds. Intensity points are identically labeled (slightly, moderately, very, and extremely) to retain consistency with HR1.

Interference with natural quiet—HR1 and HR2. How much did the sounds from aircraft interfere with each of the following aspects of your visit at [site]?… Appreciation of the natural quiet and sounds of nature at the site (not at all, slightly, moderately, very, extremely).

This question is one of a series, measuring visitors' subjective impression of the extent to which aircraft sounds affected their experiences. The questions use a five-point, unipolar scale, consistent with the understanding that aircraft sounds may interfere with these experiences. This question was not included in the audio clip survey due to survey length concerns.

3. Survey administration

Survey administration consisted of pre- and post-visit interceptions, structured to collect both survey and visitor location time-history data. The pre-visit interception allowed researchers to accurately note visitor entry times and to distribute uniquely numbered tickets (one per visitor) and GPS-based tracking devices (one per visitor group). The GPS tracking devices (Qstarz Q1000XT) recorded latitude, longitude, and altitude at 1-s intervals for the duration of the visit, enabling correlation of a visitor's location with the locations of the acoustic monitors. In the event that group members separated, the numbered tickets provided an extra measure of entry and exit time-tracking. Not all visitors were pre-intercepted and provided with these tracking mechanisms; approximately 30% of respondents began their visit outside of hours during which the entry point was staffed.

Surveys were administered on-site at the conclusion of each visit to all English-speaking visitors at least 18 yr of age. To maximize sample size, no further visitor sampling strategies (e.g., interviewing every third visitor) were employed. The surveys were administered on tablet PCs in a web-based interface. Personnel cycled through the three surveys continuously as they were distributed, so each was administered to an approximately equal number of respondents. Visitors within the same group generally received different surveys.

4. Acoustical monitoring

Stationary monitors were deployed throughout the study area to measure acoustical and meteorological data. Monitor locations were chosen based on site topography and nominal tour aircraft overflight altitudes and routes (where available), such that the estimated difference in sound exposure level between adjacent monitors for individual aircraft overflights would be less than 6 dB. Monitors were deployed at a height of 5 ft above the local ground surface in locations free of localized noise sources or reflective surfaces and representative of the environment experienced by hikers (but out of direct view). As a general rule-of-thumb, monitors were spaced at intervals of no more than 1 mile lateral and 1000 ft vertical. The monitors continuously collected: (1) 1-s, A-weighted sound levels and associated one-third octave-band unweighted spectra (20–20 000 Hz), (2) digital audio recordings, and (3) 1-s meteorological data (barometric pressure, humidity, precipitation, temperature, wind speed, and direction). Periodic system checks and calibrations were performed every 2–5 days. The main components of each monitor are: (1) pre-polarized, [1/2]-in. electret-condenser microphone (G.R.AS. Model No. 40AQ), (2) sound level meter/real-time analyzer, preamplifier, and 4-in. diameter windscreen/environmental shroud[22] [Larson Davis™ (LD) model 831, LD Model PRM831, and LD Model EPS2108, respectively], (3) audio recorder (Roland R-05), and (4) weather station (Vaisala Model WXT520), all powered by a 100 amp-hour $LiFePO_4$ battery.

In conjunction with sound level monitoring, teams of trained field observers documented the source(s) of audible sounds in the vicinity of selected monitors during daytime hours, creating a temporal record of audible sounds corresponding to the measured sound level data. Field observer logging takes full advantage of human binaural hearing capabilities, allows identification of simultaneous sound sources (e.g., two aircraft), and closely matches the experience of park visitors. Sounds were categorized in three primary acoustic states: aircraft, human (non-aircraft), and natural. Aircraft were subcategorized as helicopters, propeller-aircraft, and high-altitude jets. Where possible, aircraft operations were also identified as air tour, general aviation, commercial aviation, or military. Human sound subcategories included road vehicles, construction, voices, or domestic animal noises. Natural sound sub-categories included wind in trees, water, insects, and birds.

B. Data reduction and processing

Acoustic data were processed to identify and remove incomplete and low quality data, i.e., data collected during: (1) times where wind caused unacceptable levels of flow noise around the microphone shroud (greater than 11.2 miles per hour),[23] (2) system malfunction or input overload, and (3) contamination by field personnel activities. Adjustments applied to the data include: (1) sound level calibration, (2) microphone frequency response, (3) windscreen frequency response, and (4) instrumentation noise floor. Data remaining after these processing steps were used for computation of the respondent noise exposure dose information. Respondents for whom more than 50% of the available data were removed due to any combination of acoustical or visitor location issues (unknown location or outside of study area) are not included in the final dose-response pool (20% of total respondents). The majority of data removals were due to high-wind conditions.

For retained respondent data, removal of up to 50% of the noise exposure information has the potential to bias cumulative metrics such that visitor exposure is underestimated, resulting in an overestimate of the visitor response at a given sound level. Conversely, incorporating the high-wind data has the potential to bias the measurement such that the exposure is overestimated, resulting in an underestimate of the visitor response. As it is desirable to use as much of the respondent data as possible, an overestimate of visitor response is seen as the more desirable outcome.

A contiguous set of sound-source logs were created to correlate the observed acoustic state with the sound level data from each monitor. As field observer logs covered only a portion of the measurement locations and time periods, sound source logs were prepared post-measurement to fill-in missing time periods. These logs were constructed through either post-measurement "office" listening or visual review of spectrograms. Post-measurement listening is the preferred method as it has been found to produce results similar to field logging.[24,25] For this method, trained personnel listen to the audio recordings through noise-cancelling headphones in a controlled indoor environment. As this process is quite time consuming, visual review of spectrograms is used to expedite the process during periods when aircraft overflight sound levels are generally 20 dB greater than ambient sound levels. Potential events identified as aircraft during visual review were confirmed through audio review to ensure the accuracy of this method. The sound source logs were then converted to a hierarchical structure for comparability with previous studies. In the hierarchy, helicopter overflights are given the highest priority, followed by propeller-driven aircraft overflights, commercial jet overflights, non-aircraft human sounds (i.e., autos or voices), and last natural sounds.[7]

A complete noise exposure time history for each respondent was constructed using the 1-s sound-level records and corresponding sound-source designations from the nearest monitor throughout the duration of the visit. Monitor proximity was determined from the GPS-based respondent position information.[26] Summary aircraft noise exposure metrics were then computed for each respondent based on the entire duration and timeframe of each individual visit.[27] These include (1) sound exposure level (L_{AE}) from all aircraft sounds during the visit, (2) equivalent sound level due to all aircraft sounds during the visit, normalized to the duration aircraft were audible (L_{AeqTac}), (3) equivalent sound level due to all aircraft sounds during the visit, normalized to the respondent's visit duration (L_{AeqTresp}), (4) maximum A-weighted sound level due to aircraft during the visit (L_{ASmx}), and (5) time-based percentage of the respondent's visit during which aircraft were audible (%TAud). Aircraft audibility was based on the complete set of sound source logs from field observation, office listening, and visual review as discussed in the preceding text. Metrics were computed in aggregate for all aircraft and separately for each aircraft-type (helicopters, propeller-aircraft, and jet aircraft). Non-aircraft human and natural sound metrics computed for each respondent include: (1) median or 50th percentile of human source sound levels ($L_{50\mathrm{Hum}}$) and (2) median or 50th percentile of natural source sound levels ($L_{50\mathrm{Nat}}$).

In addition, aircraft detectability level metrics were calculated from the 1-s one-third octave-band sound level records, when aircraft were audible, as noted in the sound source observer logs. Detectability level [10log(d′) or D′L] is computed from the root-mean-square sum of the signal-to-noise ratios across one-third-octave bands, adjusted for bandwidth and frequency-specific human hearing characteristics.[28,29] From the 1-s, aircraft-source, D′L values, summary detectability metrics analogous to the A-weighted metrics were computed for each respondent. These include (1) detectability exposure level ($\mathrm{D'L_E}$, analogous to L_{AE}), (2) equivalent detectability level ($\mathrm{D'}L_{\mathrm{eq}}$, analogous to L_{AeqTresp}), (3) maximum one-second detectability level ($\mathrm{D'}L_{\max}$), and (4) percent time "noticeable" or the percentage of time during the visit where $\mathrm{D'}L \geq 17$.[30]

C. Data analysis procedure

The primary analysis goal was to develop quantitative dose-response relationships. Consistent with the frontcountry analysis, the "annoyance" (*Annoy*[31]) and "interference with natural quiet" (*Interfere*) response variables (as discussed in Sec. II A 2) were selected for the backcountry analysis. Each set of response data (*Annoy*/*Interfere*) were dichotomized three ways for analysis, resulting in six separate dose-response relations. The three dichotomizations represent visitors who did (1) or did not (0) experience "slightly or more" (SorMore), "moderately or more" (MorMore), or "very or more" (VorMore) annoyance or interference with natural quiet from aircraft noise during their visit.[32] These dichotomies may each provide useful information for evaluating impacts. For example, the VorMore dichotomy is consistent with federal criteria for determining significant aircraft noise impacts near civil airports[4] while the SorMore and MorMore dichotomies may be preferable for assessing impacts in protected natural areas.

Multilevel logistic regression analysis was used to estimate parameters describing the functional form of the relationship among visitor responses, aircraft dose metrics, and additional mediator variables.[33] Logistic regression confines the resulting curves (and their 95% confidence regions) to lie between zero and unity (100%). The multilevel aspect can properly separate out the site differences that are inherent in the data sampling. It accounts for both variability among visitors (individual level variation) and among sites (group level variation). Analyzing site as the "random" component of the multilevel regression avoids underestimates of prediction uncertainty, which can happen when site-to-site variabilities are not included in the regression equation.[10] Each model of the dichotomous response variables includes "site" as a random component, plus additional dose and mediator variables for each visitor

$$\Pr(y_i = 1) = \text{logit}^{-1}\left(\alpha_{j[i]} + \beta_0 + \sum_{d=1}^{M} \beta_d x_{id}\right), \quad \text{for } i = 1, ..., n, \tag{1}$$

where $j[i]$ indexes the site (from 1 to 7) corresponding to visitor i, and x_{id} is the value of predictor d for visitor i in a model with M predictor variables. Site is analyzed as the "random" component of the multilevel regression

$$\alpha_j^{site} \sim N(0, \sigma_{site}^2), \quad \text{for } j = 1, ..., 7. \tag{2}$$

Potential models for the backcountry day-hike dataset were evaluated by first selecting the combination of dose variables, then the combination of mediator variables, that result in models with the lowest Akaike Information Criteria (AIC) values. AIC is a commonly utilized means for model selection. AIC assesses the model "goodness of fit" using the likelihood function, while applying a penalty that increases as the number of estimated parameters increases, to discourage model overfitting.[34] For model selection, models with the lowest AIC values are preferred, but models with similar AIC values may not be significantly different. Thus in some cases several "best" models were identified using the model selection process.

The models of *Annoy* and *Interfere* responses were evaluated separately as *Annoy* is included in all three surveys (HR1, HR2, and AC) while *Interfere* is included in only two surveys (HR1 and HR2). Models for the three different dichotomies of the visitor responses (SorMore, MorMore, and VorMore) were fit for each response. In the first step, models including all single dose variables and combinations of dose variables, along with *Survey Type* (AC, HR1, and HR2)[35] and visitor ratings of the *Importance of Natural Quiet* or *Importance of calm/peace* were evaluated.[36] AIC values were used to select the model that minimized information loss (the model with the lowest AIC value). To identify the best combinations of dose variables for all three dichotomizations of the *Annoy* and *Interfere* responses, the relative probabilities of all models for a given response were calculated[37] and compared with the model with the lowest AIC value. Models with a relative probability of greater than 0.05 compared to the model with the lowest AIC value were retained as candidate models for the mediator evaluation step (step 2).

From the group of candidate dose models identified, a single combination of dose variables which resulted in the best overall fit (where the relative probability was greater than 0.05 for all three model dichotomizations and both response variables) was utilized for the mediator evaluation step for each response variable (*Annoy* and *Interfere*). Table II lists the mediator variables and associated survey questionnaire item or visit-based factor (such as visit duration). Mediator variables were added individually and in combination to each candidate *Annoy* and *Interfere* response models. Mediators were retained if they resulted in models with average AIC values (across all three dichotomizations) lower than the AIC value of the model from step one. To identify the most important mediators, in instances where average AIC values decreased only marginally (less than 1–2 units), mediators were not retained if the p values of regression coefficients were not significant for at least one dichotomization. This procedure[38] identified the combinations of dose and mediator variables that best predict visitor responses across all three dichotomizations of the *Annoy* and *Interfere* responses simultaneously, with the objective of ensured consistency in the format of the final models to simplify implementation.

III. RESULTS

The collected data consist of approximately 3200 completed day-hike visitor experience surveys and associated noise-exposure dose measurements from seven sites at four national parks. Day-hike visits ranged from 1 to 5 hr, averaging 3–4 hr. Of these surveys, 2320 *Annoy* responses and 1580 *Interfere* responses were matched with acoustic dose and mediator data and used to develop the reported dose-response models.

The aircraft noise exposure dose information (computed from all aircraft over the duration of each individual visit) for respondents at each study site is depicted in box-whisker plots of sound exposure level (Fig. 1) and percent time audible (Fig. 2). These plots depict the median exposure (center line), interquartile range (25–75 percentile, box), 1.5 times the interquartile range (whiskers), points outside this range and number of data points (n). Notable is the variation between sites: relatively high exposure and percent time audible at Hermit Trail, in contrast to high exposure but low percent time audible at Hidden Lake and both low exposure and percent time audible at Taylor Creek.

As concluded from prior research, the relative contributions of different aircraft-types plays an important role in

TABLE II. Mediator variables.

Variable short name	Definition/survey basis
Early start	Visitors who started hike/visit before 9 a.m
Duration visit	Duration of visit, described as continuous variable in minutes and log10(min)
Visited site before	"Is this your first visit to _____ trail?" (yes, visited site before or no, first visit)
Importance of…	"How important was it that your time on the _____ trail provide you with the opportunity to…." (see items a–e below) Choices: Not at all, slightly, moderately, very or extremely Response dichotomy: Very or extremely responses coded "yes"; not at all, slightly, moderately responses coded "no"
Importance of view scenery	a. View the natural scenery?
Importance of natural quiet	b. Enjoy the natural quiet and sounds of nature
Importance of history	c. Appreciate the history and cultural significance of the site
Importance of calm/peace	d. Experience a feeling of calmness, peace, or tranquility
Importance of adventure/challenge	e. Experience a sense of adventure or challenge
Activity	Which of the following activities did you take part in during your time <"on the" (day/multi-day hike trail)/"at" (Overlook/Cultural Resource Study Site)> <site>? (items a–f below)
View scenery	a. Viewing the scenery
Picnic/meal	b. Picnicking or having a meal
Watch birds	c. Watching birds
View wildlife	d. Viewing wildlife (other than birds)
View a sunrise/sunset	e. Viewing a sunrise or sunset
Talk	f. Attending a ranger-led talk, walk, or campfire program OR Attending some other demonstration, talk, or organized activity or performance
Adults only	Indicates the presence of children under the age of 16 in the visit group, based on interviewer observations. (yes, only adults, or no, group includes children)
Never air tour	Have you ever taken a scenic air tour over <park> or any other park? (YES = NEVER TAKEN AIR TOUR, OR No = taken air tour)
Residence	Where do you live? a. United States b. Another country
Group tour	Were you or your personal group part of some larger commercial, educational, or other organized group of visitors? (yes /no)

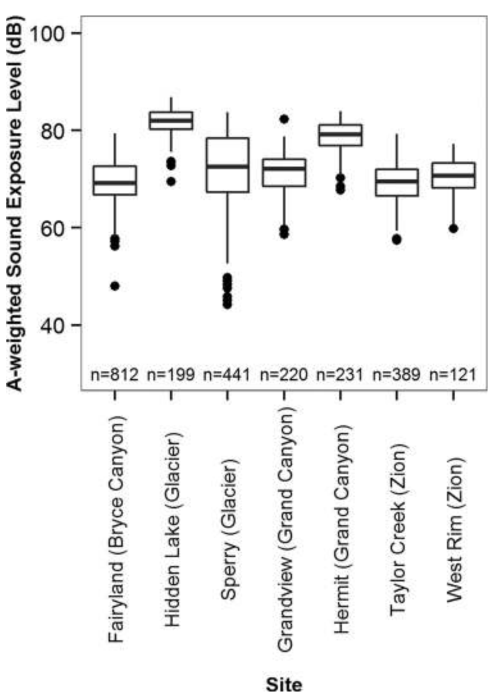

FIG. 1. Box-whisker summary of respondent aircraft A-weighted sound exposure level (L_{AE}) dose (all aircraft, summed over the entire visit duration) by measurement site. Depicted are the median exposures (center line), interquartile ranges (25–75 percentile, box), 1.5 times the interquartile range (whiskers), points outside this range and number of data points (n).

understanding variation in visitors' responses. Noise exposures at Grand Canyon's Hermit and Grandview trails and Glacier's Sperry and Hidden Lake trails included contributions from helicopters, while noise exposures at Bryce Canyon's Fairyland and Zion's Taylor Creek and West Rim Trails did not. Helicopters were audible for less than 50% of the visit duration at Sperry and Hidden Lake Trails but greater than 50% at Hermit Trail. Visitors at sites dominated by helicopters also experienced higher equivalent sound levels compared to visitors at other sites. Contributions from propeller aircraft did not vary greatly from site to site. Contributions from jet aircraft were low at Sperry and Hidden Lake Trails. The measured natural ambient sound levels at these sites ranged from approximately 20 dB(A) at barren locations to 50 dB(A) at locations where wind-foliage interactions, streams, and waterfalls were predominant natural noise sources.

A. Between-survey response comparison

Prior to formal analysis and development of dose-response relationships, responses from the similar questions on aircraft noise (reports of aircraft heard, annoyance, and interference with natural quiet) in separate survey instruments were compared to determine the appropriateness of merging the information collected within these surveys for analysis.

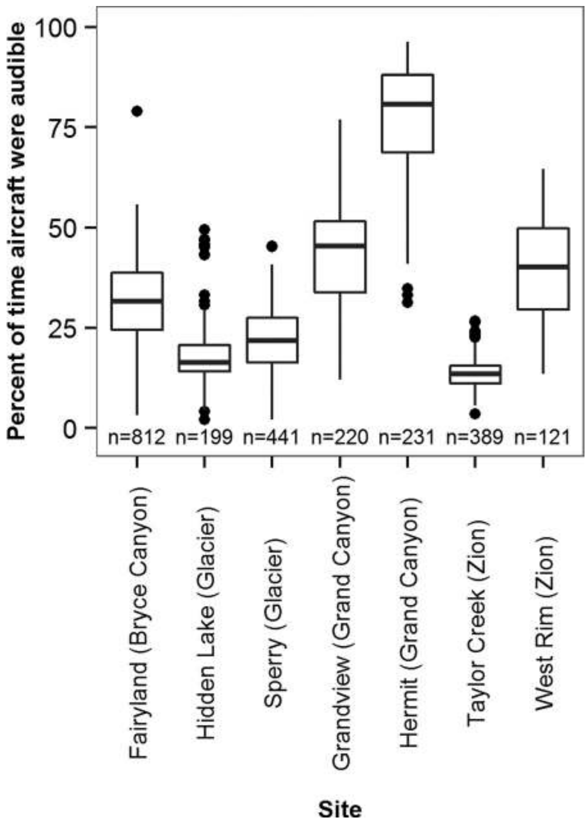

FIG. 2. Box-whisker summary of respondent aircraft percent time audible (%TAud) dose by measurement site. Depicted are the median exposures (center line), interquartile ranges (25–75 percentile, box), 1.5 times the interquartile range (whiskers), points outside this range and number of data points (*n*).

The initial screening question in the sounds section of each survey asks if aircraft were heard during the visit. HR1 and audio clip respondents are queried directly, while HR2 respondents are queried indirectly by including aircraft within a list of ten sound sources, both natural and anthropogenic. Note that in the AC survey, this question is placed *after* respondents have completed the audio clip rating exercise. On average, 5% fewer visitors reported hearing aircraft using the indirect format than the direct format (statistically significant with a p-value of 0.02). Differences were also examined as a function of sound exposure level (due to all aircraft during the visit) in three bins [<65, 65–75, and >75 dB(A)]. In all cases, fewer respondents reported hearing aircraft using the indirect format than the direct format; 10% fewer at sound exposure levels less than 65 dB(A), and 5% fewer at sound exposure levels between 65 and 75 dB(A) and above 75 dB(A). Due to the partitioning of the data, these binned differences are not statistically significant (p-values of 0.08, 0.12, and 0.08, respectively) but show that as expected respondents of indirect questions about hearing aircraft are less likely to report hearing aircraft at lower sound exposure levels.

Questions regarding (1) annoyance due to aircraft noise (HR1 and audio clip)/sounds (HR2) and (2) interference with natural sounds and natural quiet due to aircraft noise (HR1 and HR2) are subsequently presented for those reporting they heard aircraft. The annoyance question is identical in the HR1 survey and AC and allows for negative or "not at all" evaluations of aircraft noise on a five-point scale, while the HR2 survey includes allowances for positive, neutral, and negative ratings of each noise source on a nine-point scale. Table III shows the proportion of visitor ratings in each category for each survey. Reported percentages are based on all respondents, including those who did not hear aircraft. As such, totals do not equal 100%. Differences in response distributions are observed between the nine-point bipolar survey and the five-point unipolar surveys in the moderately, slightly, and not at all/neutral *Annoy* response categories. The most common response category on the five-point scale is "slightly" (21%–23% of respondents), while "neutral" is the most common response on the nine-point scale (20% of respondents). Notable also in Table III is the similar trend in response distributions between the AC and HR1 surveys.

The differences observed between surveys may be in part due to response bias where a person responds to questionnaire items on some basis other than what the items were specifically designed to measure. Extreme and non-extreme response styles are commonly discussed response biases and are among the most prevalent and problematic in attitudinal studies, occurring in a subset of nearly every respondent population. It is common with all Likert-types scales, although there is some research that suggests that the bipolar scale may evoke more extreme-response style behavior than the unipolar.[39] Those with a non-extreme response style tend to favor the middle categories and avoid the extremes of a rating scale.[40] For these, the mid-scale "neutral" point may be favored on the nine-point scale while the extreme-end "not at all" may be less favorable on the five-point scale. In contrast, the very and extremely categories, at the extreme of both response scales, are viewed similarly and contain similar percentages of respondents (8%–9%) across all three surveys.

Another common phenomenon is the influence of the "no-opinion" option. The neutral choice on the nine-point scale may present a no-opinion option to respondents, while the five-point scale lacks a no-opinion option and can be viewed as a forced-choice scale. Some suggest presenting a no-opinion option is undesirable as it allows respondents to skip the cognitive work necessary to form an opinion. Others

TABLE III. Percentage of *Annoy* responses by rating category and survey (5- and 9-point scales). Reported percentages are based on the total number of responses for each survey.

Rating	HR1 (%)	HR2 (%)	Audio clip (%)
Extremely pleased	NA	1	NA
Very pleased	NA	1	NA
Moderately pleased	NA	3	NA
Slightly pleased	NA	3	NA
Not at all (HR1/audio clip) or Neutral (HR2)	18	20	15
Slightly annoyed	23	13	21
Moderately annoyed	12	8	17
Very annoyed	5	4	6
Extremely annoyed	4	4	3

TABLE IV. Percentage of *Interfere* responses by rating category and survey (5-point scale). Reported percentages are based on the total number of responses for each survey.

Rating	HR1 (%)	HR2 (%)
Not at all	19	8
Slightly	14	15
Moderately	10	11
Very	6	6
Extremely	5	5

state that without a no-opinion option, those truly without an opinion will default to a rating from the middle of the scale, making results less accurate.[41] Both of these biases would affect the neutral response category on the nine-point scale and the slightly and moderately response categories on the five-point scale.

The question concerning interference with natural sounds and natural quiet due to aircraft noise is included only in the HR1 and HR2 surveys and is identical in both. Ratings are collected on a five-point unipolar scale. Table IV shows the proportion of visitor ratings in each category for each survey. The distribution of responses was nearly identical for any degree of interference (slightly to extremely). Responses differ in the percentage of "not at all" ratings, where there are fewer reports of aircraft heard and non-responses to this item in the HR2 survey.

Although there are notable differences between surveys, all responses to the questionnaire items within the three surveys regarding *in situ* noise exposure are utilized in the development of dose-response relationships. The benefits of utilizing all of the data collected were seen to outweigh the limitations as much of the variation between surveys can be accounted for by introducing a survey-type variable in the regression models.

B. Dose variable model fits

Step 1 of the model-fitting process identified a group of models having comparatively low AIC values for the *Annoy* and *Interfere* responses. For this analysis, "comparatively low" is defined as having a probability of greater than 0.05 relative to the model with the lowest AIC value for at least one dichotomization. Tables V and VI summarize these models (not all models are shown, only those with comparatively low AIC values), showing the included dose variables included mediator (either *importance of calm/peace* or *importance of natural quiet*), AIC values, and relative probability. From this group of models, a single combination of dose variables had the best overall fit where the relative probability was greater than 0.05 for *all three* model dichotomizations and both response variables (model 1). This model includes metrics of A-weighted sound exposure level (L_{AE}), percent time audible (%TAud), and percent aircraft-type energy contributions ($P_{EnHelos}$, $P_{EnProps}$). This single model was used for the mediator variables analysis step to limit the potentially large number of models with different dose and mediator combinations that could be considered and ultimately to simplify practical implementation.

Tables V and VI show that there are a number of dose metric combinations that result in models with similarly low AIC values. The majority of these combinations include one measure of cumulative sound exposure and one measure of either percent time audible or natural ambient. For the *Annoy* SorMore response, models 5 and 6 have the lowest AIC values and include $D'L_E$ rather than L_{AE}. Models 3 and 4 also have comparatively low AIC values (relative to those not tabulated) and use a base-10 logarithmic transformation of %TAud (rather than %TAud). For the *Interfere* SorMore response, model 2, which includes the *Importance of natural quiet* mediator rather than the *Importance of calm/peace* mediator, has the lowest AIC value. Other models with low AIC values include models 4, 8, and 10, which include $L_{AeqTresp}$ (rather than L_{AE}) and $L_{50,Nat}$ or log10(%TAud) rather than %TAud. For the *Interfere* MorMore response, model 13, which includes L_{AeqTac}, has the lowest AIC value. Other models with low AIC values include models 7, 9, 11, and 12, which include L_{ASmx}, $L_{AeqTresp}$ (in place of L_{AE}), and L_{50Nat} (in place of %TAud). For the *Interfere* VorMore response, models 2, 3, 7, 9, and 11–13 result in low AIC values. These models include L_{ASmx}, $L_{AeqTresp}$, or L_{AeqTac} in place of L_{AE}, and L_{50Nat} or log10(%TAud) rather than %TAud.

C. Mediator variables

The addition of a number of mediator variables to the dose models identified in step 1 resulted in models with lower AIC values. The mediator variables that reduced the average AIC value and had significant coefficient estimates for the *Annoy* response (see Table VII) are *Survey type*, *Importance of calm/peace*, *Adults only*, *Never air tour*, and

TABLE V. Results of dose variable model fitting, showing those models having comparatively low AIC values for the *Annoy* response. The relative probability (Rel Prob) represents the relative likelihood of the model compared to the model with the lowest AIC value. A relative probability value of 1.00 indicates the model with the lowest AIC value.

			Slightly or more		Moderately or more		Very or more	
Model No.	Dose variables	Mediator	AIC	Rel Prob	AIC	Rel Prob	AIC	Rel Prob
1	L_{AE}, %Taud, $P_{EnHelos}$ $P_{EnProps}$	Importance of calm/peace	2480.3	0.33	1646.8	1.00	894.1	1.00
2	L_{AE}, %Taud, $P_{EnHelos}$ $P_{EnProps}$	Importance of natural quiet	2480.2	0.35	1651.6	0.09	899.9	0.06
3	L_{AE}, log10(%TAud), $P_{EnHelos}$ $P_{enProps}$	Importance of calm/peace	2481.4	0.19	1653.9	0.03	900.0	0.05
4	L_{AE}, log10(%TAud), $P_{EnHelos}$ $P_{enProps}$	Importance of natural quiet	2481.3	0.20	1658.5	0.00	905.7	0.00
5	$D'L_E$, $P_{EnHelos}$, $P_{enProps}$	Importance of calm/peace	2478.9	0.67	1675.3	0.00	915.3	0.00
6	$D'L_E$, $P_{EnHelos}$, $P_{enProps}$	Importance of natural quiet	2478.1	1.00	1679.4	0.00	920.1	0.00

TABLE VI. Results of dose variable model fitting, showing those models having comparatively low AIC values for the *Interfere* response. The relative probability (Rel Prob) represents the relative likelihood of the model compared to the model with the lowest AIC value. A relative probability value of 1.00 indicates the model with the lowest AIC value.

			Slightly or more		Moderately or more		Very or More	
Model #	Dose variables	Mediator	AIC	Rel Prob	AIC	Rel Prob	AIC	Rel Prob
1	L_{AE}, %Taud, $P_{EnHelos}$ $P_{enProps}$	Importance of calm/peace	1757.4	0.05	1435.7	0.12	947.7	1.00
7	$L_{Aeq,Tresp}$, %TAud, $P_{EnHelos}$ $P_{enProps}$	Importance of calm/peace	1762.1	0.01	1435.7	0.12	949.1	0.50
9	$L_{Aeq,Tresp}$, $L_{50,Nat}$, $P_{EnHelos}$ $\mathrm{P}_{enProps}$	Importance of calm/peace	1760.4	0.01	1435.1	0.17	951.7	0.14
3	L_{AE}, log10(%TAud), $P_{EnHelos}$ $P_{enProps}$	Importance of calm/peace	1760.7	0.01	1441.2	0.01	951.8	0.13
11	L_{Asmx}	Importance of calm/peace	1782.4	0.00	1432.6	0.58	951.9	0.12
12	L_{ASmx}, $L_{50,Nat}$	Importance of calm/peace	1779.9	0.00	1433.4	0.39	952.2	0.11
2	L_{AE}, %Taud, $P_{EnHelos}$ $P_{enProps}$	Importance of natural quiet	1751.4	1.00	1440.9	0.01	952.4	0.10
13	$L_{Aeq,Tac}$, $P_{EnHelos}$ $P_{enProps}$	Importance of calm/peace	1816.6	0.00	1431.5	1.00	953.3	0.06
8	$L_{Aeq,Tresp}$, %Taud, $P_{EnHelos}$ $P_{enProps}$	Importance of natural quiet	1754.0	0.27	1441	0.01	954	0.04
10	$L_{Aeq,Tresp}$, $L_{50,Nat}$, $P_{EnHelos}$ $P_{enProps}$	Importance of natural quiet	1753.8	0.30	1439.4	0.02	955.4	0.02
4	L_{AE}, log10(%TAud), $P_{EnHelos}$ $P_{enProps}$	Importance of natural quiet	1754.8	0.18	1446.4	0.00	956.6	0.01

TABLE VII. Coefficient estimates for each predictor, their standard error and significance (*p*-value) for the *Annoy* response model.

		Annoyance		
Response dichotomization	Predictor	Coefficient estimate	Standard error	*p*-value
Slightly or more	Intercept	−5.618	0.814	0.000
	L_{AE}	0.043	0.012	0.000
	%TAud	0.013	0.004	0.000
	$P_{EnHelos}$	0.018	0.003	0.000
	$P_{EnProps}$	0.006	0.002	0.009
	Survey HR1	−0.067	0.121	0.579
	Survey HR2	−0.851	0.129	0.000
	Importance of calm/peace	0.311	0.129	0.016
	Visited site before	0.485	0.144	0.001
	Adults only	0.435	0.138	0.002
	Never air tour	−0.070	0.163	0.667
	Watch birds	0.280	0.106	0.008
Moderately or more	Intercept	−9.175	1.186	0.000
	L_{AE}	0.077	0.017	0.000
	%TAud	0.008	0.005	0.081
	$P_{EnHelos}$	0.019	0.004	0.000
	$P_{EnProps}$	0.013	0.004	0.000
	Survey HR1	−0.230	0.153	0.133
	Survey HR2	−0.713	0.163	0.000
	Importance of calm/peace	0.502	0.177	0.004
	Visited site before	0.450	0.173	0.009
	Adults only	0.117	0.177	0.510
	Never air tour	−0.820	0.258	0.002
	Watch birds	0.213	0.134	0.113
Very or more	Intercept	−11.474	1.701	0.000
	L_{AE}	0.076	0.023	0.001
	%TAud	0.022	0.003	0.000
	$P_{EnHelos}$	0.021	0.004	0.000
	$P_{EnProps}$	0.017	0.006	0.003
	Survey HR1	−0.123	0.222	0.580
	Survey HR2	−0.309	0.230	0.180
	Imp of calm/peace	0.688	0.284	0.015
	Visited site before	0.572	0.237	0.016
	Adults only	0.153	0.275	0.577
	Never air tour	−0.920	0.412	0.026
	Watch birds	0.279	0.191	0.143

Watch birds. Note that *Survey type* is a categorical variable with three values (AC, HR1, or HR2). In this analysis, the AC survey is the reference value and separate mediator coefficients are estimated for the HR1 and HR2 surveys. The remainder of the mediators are binary (yes/no) variables, where a no response is the reference value, and the coefficient can be used to estimate the difference in annoyance or interference for visitors with a yes response.

Using Table VII, an assessment of the relative magnitude and p values of the mediator coefficients in each dichotomization was conducted. The *Survey type* HR2 mediator is significant (p-value < 0.05) in the SorMore and MorMore dichotomizations (reflecting the response biases discussed earlier). The *Adults only* and *Watch birds* mediators are not significant in the MorMore and VorMore dichotomizations, while the *Never air tour* mediator is relatively small and not significant in the SorMore dichotomization. Reflecting the similarities between responses to the HR1 survey and AC, the *Survey type* HR1 mediator does not show significance but is included as it is a category of the survey type factor.

The mediator variables that reduced the average AIC value and had significant coefficient estimates for *Interfere* response (see Table VIII) are *Importance of calm/peace*, *Adults only*, *Never air tour*, and *Talk*. Using Table VIII, an assessment of the relative magnitudes and p-values of the mediator coefficients in each dichotomization shows that the *Adults only* mediator is not significant for the MorMore and VorMore dichotomizations, while *Never air tour* and *Talk* are not significant for the SorMore and MorMore dichotomizations.

D. Final dose-response relationships

Tables VII and VIII summarize the values of the regression coefficients, their standard error and p-values for the final model identified for each dichotomization of *Annoy* and *Interfere* responses. Figures 3 and 4 show the associated dose-response relationships between increases in noise exposure and visitor response depicted using the sound exposure level dose variable (L_{AE}). The three individual curves in the plots (solid lines) represent the three dichotomies of visitor response (SorMore, MorMore, and VorMore); dashed lines represent the 95% confidence intervals. As only the primary dose variable (L_{AE}) is explicitly depicted, the %TAud, $P_{EnHelos}$, and $P_{EnProps}$ doses are represented using a function[42] relating each to L_{AE}. These functions are substituted into the dose-response equation, thereby permitting the visualization of their effect through the primary dose variable. For these figures, values of the survey-based mediator variables were held constant at the average values for the 2011 days-hike survey data as presented in Table IX for the overall dataset and by site. The grey dots in Figs. 3 and 4

TABLE VIII. Coefficient estimates for each predictor, their standard error and significance (p-value) for the *Interfere* response model.

		Interference with Natural Quiet		
Response dichotomization	Predictor	Coefficient estimate	Standard error	p-value
Slightly or more	Intercept	−7.282	0.952	0.000
	L_{AE}	0.070	0.014	0.000
	%TAud	0.015	0.004	0.001
	$P_{EnHelos}$	0.017	0.003	0.000
	$P_{EnProps}$	0.003	0.003	0.184
	Importance of calm/peace	0.227	0.150	0.131
	Adults only	0.358	0.157	0.023
	Never air tour	−0.384	0.200	0.055
	Talk	0.666	0.343	0.052
Moderately or more	Intercept	−7.070	1.147	0.000
	L_{AE}	0.057	0.016	0.000
	%TAud	0.008	0.006	0.176
	$P_{EnHelos}$	0.014	0.004	0.001
	$P_{EnProps}$	0.002	0.003	0.549
	Importance of calm/peace	0.608	0.186	0.001
	Adults only	0.201	0.184	0.273
	Never air tour	−0.422	0.240	0.078
	Talk	0.588	0.355	0.098
Very or more	Intercept	−8.618	1.546	0.000
	L_{AE}	0.060	0.022	0.006
	%TAud	0.019	0.005	0.001
	$P_{EnHelos}$	0.016	0.004	0.000
	$P_{EnProps}$	0.009	0.005	0.059
	Importance of calm/peace	0.759	0.257	0.003
	Adults only	−0.187	0.230	0.417
	Never air tour	−0.755	0.353	0.032
	Talk	1.091	0.396	0.006

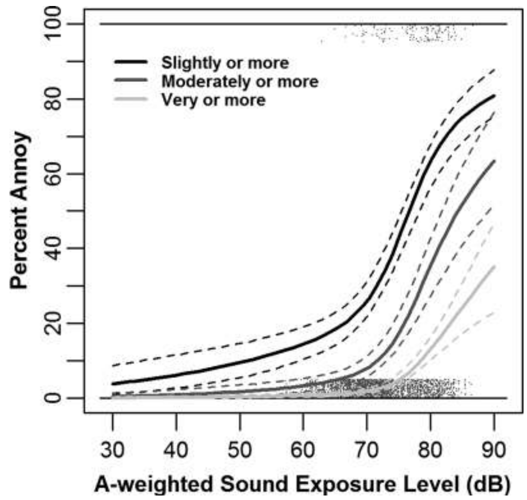

FIG. 3. A-weighted sound exposure level (L_{AE}) dose-response relationships (solid lines) and 95% confidence intervals (dotted lines) for the *Annoy* response for dichotomies of slightly or more, moderately or more, and very or more. Grey dots represent individual yes and no data points within the slightly or more dichotomy.

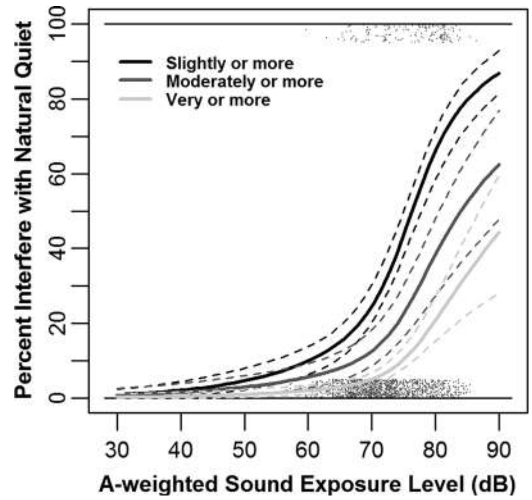

FIG. 4. A-weighted sound exposure level (L_{AE}) dose-response relationships (solid lines) and 95% confidence intervals (dotted lines) for the *Interfere* response for dichotomies of slightly or more, moderately or more, and very or more. Grey dots represent individual yes and no data points within the slightly or more dichotomy.

represent the individual dose-response data points for the SorMore dichotomy. Yes responses (slightly, moderately, very, or extremely) are shown near the top of the graphic, and No responses (not at all) are shown near the bottom. These points are jittered to aid in visualization.

E. Effects of changes in values of dose and mediator variable values

Figures 3 and 4 depict how changes in the primary dose variable influence predicted visitor response; less obvious is how changes in the values of the mediators influence predicted visitor response. To visualize these prediction sensitivities, the value of a single mediator is varied while holding others constant. For example, influences due to the $P_{EnHelos}$ value can be quantified by varying this input value and holding the values of the remaining variables constant. This process can be repeated for any number of variations of input values. Following are two examples. In each example, variables are held constant at their median values within the dataset.

Figures 5 and 6 depict how a change in %TAud from 25% (dotted line) to 75% (solid line) can shift the dose-response relationships and influence predicted visitor response. These figures depict the dose-response relationships between increases in noise exposure and visitor response using the sound exposure level dose variable (L_{AE}) at two specific values of %TAud: 25% (dotted line) and 75% (solid line). For these figures, values of $P_{EnHelos}$, $P_{EnProps}$, and the survey-based mediator variables were held constant at the average values for the 2011 days-hike survey data, as presented in Table IX. Figure 5 depicts the change for the *Annoy* response at the MorMore level, showing that an increase from 25%TAud to 75%TAud increases the predicted proportion of visitors reporting annoyance by 2% (i.e., the vertical offset) at an L_{AE} of 71 dB(A) (the median for the day-hike dataset). This change can alternately be expressed in terms of a horizontal (decibel value) offset of −5 dB(A), or the decrease in L_{AE}, which will result in equal proportions of visitors reporting annoyance at the MorMore level. Figure 6 depicts the change for the *Interfere* response at the MorMore level, showing that an increase from 25%TAud to 75%TAud increases the predicted proportion of visitors experiencing interference with natural quiet by approximately 5%. Correspondingly, a decrease of 7 dB(A) in L_{AE} would result in equal proportions of visitors

TABLE IX. Average values of mediator variables for respondent population by site and overall.

	Fairyland (BRCA)	Taylor Creek (ZION)	West Rim (ZION)	Grand-view (GRCA)	Hermit (GRCA)	Sperry (GLAC)	Hidden Lake (GLAC)	Overall
Average time audible (%)	32	14	35	42	77	23	19	31
Average helicopter energy (%)	0	0	0	6	80	86	96	31
Average prop energy (%)	36	55	30	39	12	11	4	30
Adults only (%)	81	71	87	79	89	83	79	81
Importance of calm/peace (%)	86	84	89	86	87	83	84	79
Visited site before (%)	9	11	28	18	13	16	22	13
Never air tour (%)	88	86	87	90	89	90	91	89
Watch birds (%)	44	26	30	38	41	27	28	36
Talk/presentation (%)	3	1	2	5	6	6	6	3

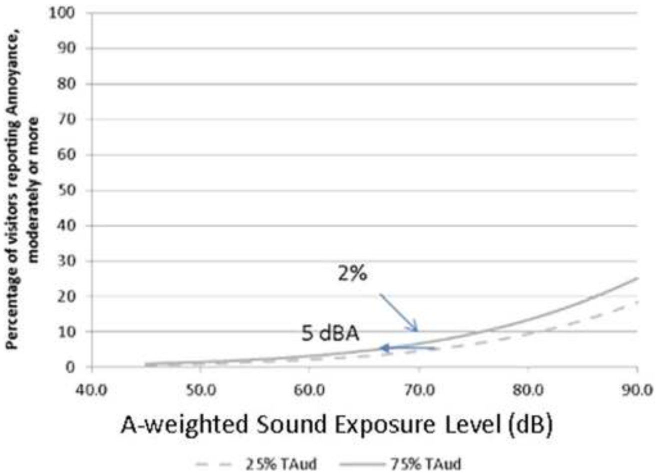

FIG. 5. (Color online) Variation in A-weighted sound exposure level (L_{AE}) dose-response relationships caused by change in aircraft percent time audible (%TAud) from 25% (dotted line) to 75% (solid line). Depicted for *Annoy* response at the moderately or more level.

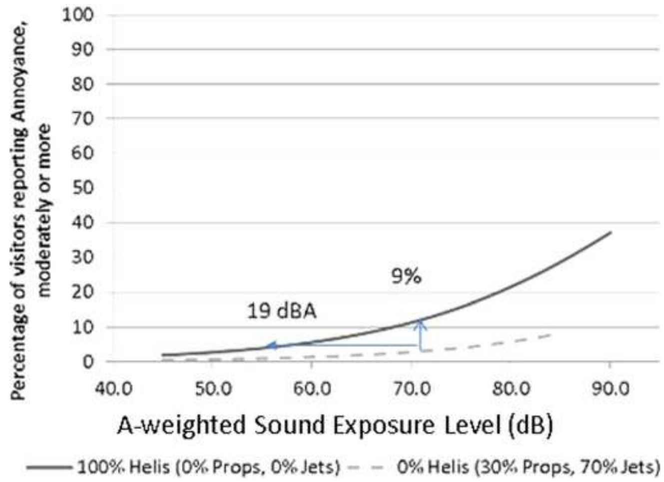

FIG. 7. (Color online) Variation in A-weighted sound exposure level (L_{AE}) dose-response relationship caused by change in percent helicopter energy ($P_{EnHelos}$) from 0% (dotted line) to 100% (solid line). Depicted for *Annoy* responses at the moderately or more level.

experiencing interference with natural quiet at the median values of the dataset.

In a similar fashion, Figs. 7 and 8 depict how a change in the percentage of sound energy from helicopters ($P_{EnHelos}$) from 0% (dotted line) to 100% (solid line) can shift the dose-response relationships and influence predicted visitor response. Note that for the 0% helicopter case, the noise exposure dose was assumed to consist of 30% propeller aircraft and 70% jet aircraft energy (reflective of the collected dataset). In an example using the *Annoy* model at the MorMore level (Fig. 7), an increase from 0% $P_{EnHelos}$ to 100% $P_{EnHelos}$ increases the predicted proportion of visitors experiencing annoyance by 9% at an L_{AE} of 71 dB(A) (the median for the day-hike dataset). The horizontal offset for this increase in $P_{EnHelos}$ represents a 19 dB(A) decrease in L_{AE}. In the example using the *Interfere* model at the MorMore level (Fig. 8), an increase from 0% $P_{EnHelos}$ to 100% $P_{EnHelos}$ increases the predicted proportion of visitors experiencing interference with natural quiet by 18% at an L_{AE} of 71 dB(A). The horizontal offset for this increase in $P_{EnHelos}$ represents a 24 dB(A) decrease in L_{AE}.

Tables X and XI summarize the sensitivities of all the included mediator variables for each of the three dichotomizations using the methods depicted in Figs. 5–8. These tables show both the vertical offset (change in the predicted proportion of visitors reporting *Annoy* or *Interfere*) and the horizontal offset (decrease in L_{AE}, which will result in equal proportions of visitors reporting *Annoy* or *Interfere*) at an L_{AE} of 71 dB(A) (the median for the day-hike dataset). Together with Table IX, these statistics inform the selection of variables most important for predicting visitor response to aircraft noise. For example, knowledge of the proportion of noise exposure due to helicopters is important as this variable can vary between sites and changes in this variable can greatly alter the shape of the dose-response curve.

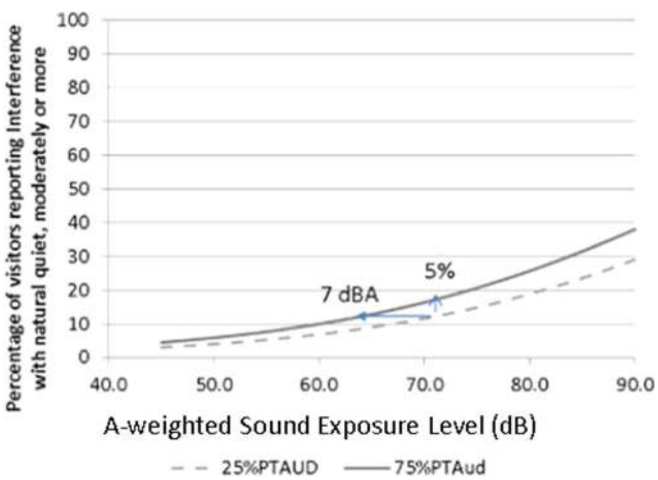

FIG. 6. (Color online) Variation in A-weighted sound exposure level (L_{AE}) dose-response relationships caused by change in aircraft percent time audible (%TAud) from 25% (dotted line) to 75% (solid line). Depicted for *Interfere* response at the moderately or more level.

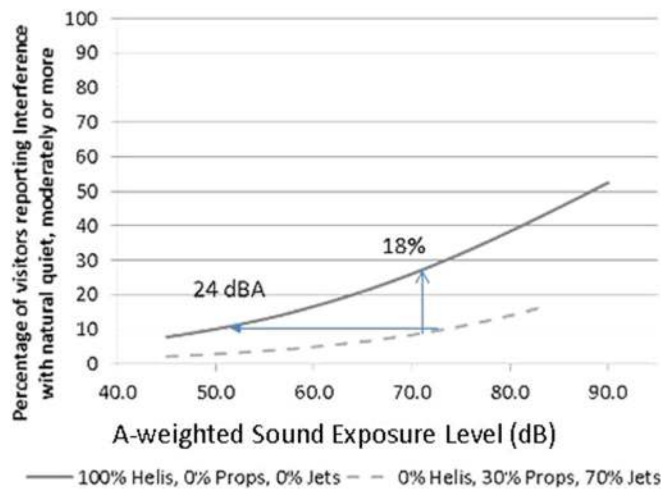

FIG. 8. (Color online) Variation in A-weighted sound exposure level (L_{AE}) dose-response relationship caused by change in percent helicopter energy ($P_{EnHelos}$) from 0% (dotted line) to 100% (solid line). Depicted for *Interfere* responses at the moderately or more level.

TABLE X. *Annoy* model sensitivity in terms of vertical (%) and horizontal (dB) offset due to changes in mediator values.

	Slightly or more		Moderately or more		Very or more	
Predictor: Change	(%)	[dB(A)]	(%)	(dB)	(%)	(dB)
%TAud: 25% to 75%	15	−15	2	−5	2	−14
$P_{EnHelos}$: 0% to 100%	36	−41	9	−25	2	−28
$P_{EnProps}$: 0% to 100%	10	−14	5	−17	2	−22
Importance of calm/peace: no to yes	5	−7	2	−7	1	−9
Visited site before: no to yes	11	−11	2	−6	1	−8
Adults only: no to yes	8	−10	1	−2	0	−2
Never air tour: no to yes	−2	2	−6	11	−2	12
Watch birds: no to yes	7	−7	1	−3	0	−4

IV. DISCUSSION

The goal for this dose-response data collection and analysis was to gain further understanding of backcountry visitor response to aircraft noise. This research resulted in the accumulation of a large dose-response dataset for backcountry day-hike visitors. Dose-response models were identified for the annoyance (*Annoy*) and interference with natural quiet (*Interfere*) responses for three visitor response dichotomizations (SorMore, MorMore, and VorMore). A key goal of this type of study is to identify dose metric combination(s) that provide the most predictive power. The models shown to minimize information loss for this dataset include dose metrics of A-weighted sound exposure level (L_{AE}), percent time audible (%TAud), and percent helicopter and propeller aircraft-type energy contributions ($P_{EnHelos}$, $P_{EnProps}$).

Detectability exposure level ($D'L_E$) was investigated in this study as, unlike L_{AE}, it is contingent on background sound level conditions. Although overall it was not as effective in predicting responses, it was the strongest predictor for the *Annoy* response at the slightly or more level. This is consistent with earlier evidence that there is a strong relationship between annoyance and the detectability of low level aircraft sounds.[28] Additionally, the appearance of the natural ambient sound level (L_{50Nat}) as a factor for interfere ratings suggests that interference with natural quiet and the sounds of nature is as expected, related to the level of natural sounds.

Mediator variables included in the final models include *Importance of calm/peace*, *Adults only*, and *Never air tour*. For the *Annoy* response, the model also includes *Survey type*, *Site visit before* and *Watch birds* mediators. For the *Interfere* response, the model also includes a mediator variable for participation in an interpretive *Talk*. Many of these additional mediator variables (*Importance of calm/peace*, *Never air tour*, *Talk*, and *Watch birds*) are not available within the front country dataset as they were not included in the 1990s survey questionnaires. In particular, the significance of the *Importance of calm/peace* variable corroborates earlier research suggesting that this is an important value in national park settings.[43] The significance of the *Talk* and *Watch birds* "activity" variables suggests participation in specific activities can increase visitor sensitivity to aircraft noise. In this analysis, mediator variables were evaluated with the goal of identifying a single model across all three dichotomizations of the *Annoy* or *Interfere* response. Alternative approaches, including model averaging, could be utilized to select the best models for each dichotomization independently, resulting in unique models for each response.

A key aspect in the design of the backcountry study is the use of multiple survey instruments. Of particular interest is the HR2 survey, designed to avoid bias resulting from direct questions related to aircraft noise by instead querying on a variety of sounds. HR2 survey respondents were significantly less likely to report hearing aircraft. No difference in response distributions between surveys was found for the *Interfere* question (identically scaled in HR1 and HR2). However, visitor response does vary significantly between the surveys for the *Annoy* response as the response distributions (Table III) and magnitude of the *Survey type* coefficient estimates indicate (Table VII). HR2 survey respondents reported significantly less annoyance at the slightly and moderately levels at a given noise dose when compared to HR1 or AC survey respondents. The difference may be due in part to response bias introduced by the different ratings

TABLE XI. *Interfere* model sensitivity in terms of vertical (%) and horizontal (dB) offset due to changes in mediator values.

	Slightly or more		Moderately or more		Very or more	
Predictor: change	(%)	(dB)	(%)	(dB)	(%)	(dB)
%TAud: 25% to 75%	15	−11	5	−7	6	−16
$P_{EnHelos}$: 0% to 100%	37	−24	23	−23	11	−22
$P_{EnProps}$: 0% to 100%	6	−5	2	4	5	−15
Importance of calm/peace: no to yes	7	−7	5	−10	2	−13
Adults only: no to yes	5	−5	2	−4	1	−3
Never air tour: no to yes	−6	5	−4	7	−2	12
Talk: no to yes	7	−10	8	−10	8	−18

scales (five-point unipolar in audio clip and HR1, nine-point, bipolar in HR2). The slightly or more and moderately or more relationships are most affected by the response biases discussed, while the very or more relationship is unaffected. Although unaffected, this relationship may have limited utility assessments as there are few reports of very or extreme annoyance at low noise exposures. The selection of one or the other survey type and dichotomization may ultimately be based on management goals and requirements.

In using the dose-response relationships developed in this study, an analyst must be cognizant of the sensitivity of this model to the variety of potential model inputs that act as mediators on the underlying dose-response relationship. For example, knowledge of the percent time audible of aircraft and proportion of noise exposure due to helicopters and fixed-wing aircraft is important as these variables can vary widely between sites, and changes in these variables can significantly alter the shape of the dose-response curve.

The backcountry data and dose-response models are intended to complement the frontcountry data and models.[10] However, there are a number of key aspects that differ between the studies that must be noted. First, a number of data collection protocols in the current study were implemented to address the added complexity of backcountry research and improve upon prior research. These include the pre-visit survey intercept, use of GPS-based tracking devices, and multiple tablet-PC-based survey instruments. Second, the noise exposure composition differed between studies and between sites in the current study. In the frontcountry study, the majority of the visitors' noise exposure was attributable to helicopter air tour overflights. In the current backcountry study, 50% of respondents had noise exposures mostly attributable to helicopter air tour overflights and the remaining 50% had noise exposure attributable to general aviation and high altitude commercial overflights. Although the models are structured to account for differences due to aircraft-type, predictions from these models may be influenced by these underlying data differences.

The model identified for the backcountry data differs from that identified for the frontcountry data, most notably in the noise dose metrics.[10] The frontcountry model includes dose metrics of $L_{AeqTresp}$, $P_{EnHelos}$, and $P_{EnProps}$ and the interaction term $P_{EnHelos} * P_{EnProps}$. In contrast, the backcountry model identified herein includes dose metrics of L_{AE}, %TAud, $P_{EnHelos}$, and $P_{EnProps.}$ There is some similarity in the dose metrics found in these models as the $L_{AeqTresp}$ dose (frontcountry model) is derived from components of L_{AE} and visit duration; absent is the duration of aircraft sounds included within the %TAud dose in the backcountry model. This indicates that the total noise exposure is important in both situations, while the relative duration of the aircraft exposure exhibits more significance in the backcountry model where visits are longer and visitors are more immersed in the natural setting.

For comparative purposes only, a regression was fit to the frontcountry data using the dose combination of L_{AE}, %TAud, $P_{EnHelos}$, and $P_{EnProps}$ as identified for the backcountry data. Figures 9 and 10 depict the resulting backcountry and frontcountry dose-response relationships (solid lines) and 95% confidence intervals (dotted lines) for the A-weighted sound exposure level dose and *Annoy* response, using the moderately or more dichotomization and HR1 survey-type (as this was the survey used within the frontcountry research). The backcountry relationship has been extrapolated for this comparison; areas where no data are available are greyed-out. Intuitively, one might expect that a greater percentage of backcountry respondents would report annoyance due to lower ambient sound, levels, longer duration of exposure, and further immersion in the natural soundscape. Note, however, that these differences have, to some extent, been incorporated in the dose-response relationships through the importance of natural quiet mediator and percent time audible dose (exposure duration and ambient sound levels are both factors in this metric).

At sound exposures above approximately 75 dB(A), a greater percentage of day-hike respondents are predicted to

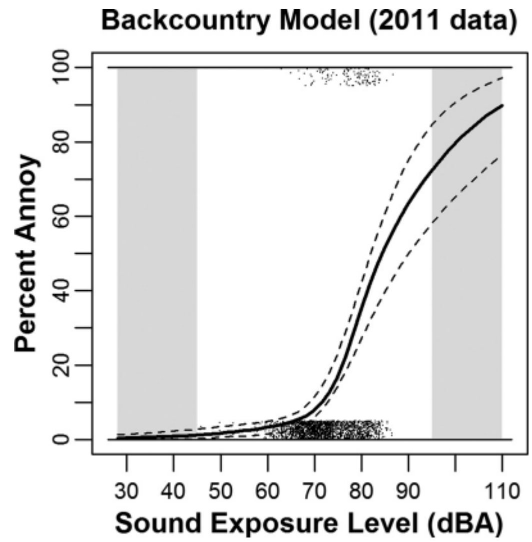

FIG. 9. Backcountry (day-hike) dose-response relationship depicted for the A-weighted sound exposure level, moderately or more dichotomization, and HR1 survey-type. Shaded areas represent extrapolated portions of the curve. Grey dots represent individual yes and no data points.

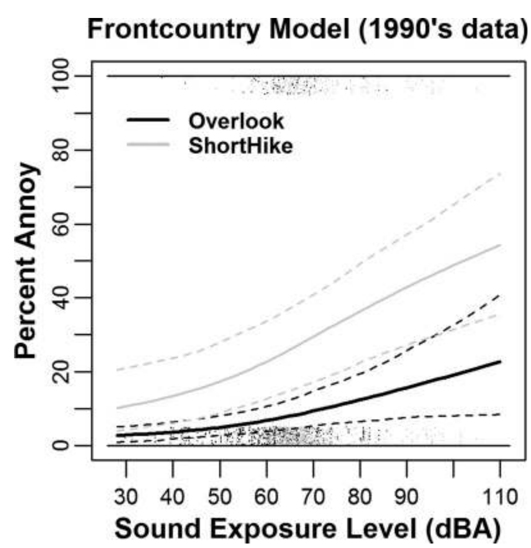

FIG. 10. Frontcountry (short hike and overlook) dose-response relationships depicted for the A-weighted sound exposure level and moderately or more dichotomization. Grey dots represent individual yes and no data points.

report annoyance. The models yield similar predictions in the region between 70 and 80 dB(A), where the majority of day-hike data are centered and predictions most accurate. Predictions are less intuitive at sound exposures below 70 dB(A), where a greater percentage of frontcountry short-hike respondents are predicted to report annoyance. However, the confidence intervals surrounding the day- and short-hike relationships overlap through the majority of the data range, and neither relationship is based on significant amounts of data at sound exposures below 50 and above 85 dB(A). Therefore it is likely that these behaviors are an artifact of data limitations and predictions outside the available data range should be used with caution. Analysis of a combined front- and backcountry dataset could produce a single model and yield further insights but would be limited by the smaller number of dose and mediator variables measured in the frontcountry surveys.

V. CONCLUSIONS

The data and analysis presented herein represent a significant step forward in understanding the effects of air tour and aircraft noise on park visitors. The combinations of dose and mediator variables that best predict visitor response to aircraft noise for backcountry day-use visitors were identified from a wide range of potential explanatory factors. Particular visitor valuations and activities while at national parks were found to be important predictors of response to aircraft noise. The dose-response relationships presented herein complement and extend previous research conducted in frontcountry areas. They can be used to predict visitor response to specific air tour operational scenarios and can be used as a tool to assist the evaluation of potential impacts of air tour noise on visitors to national parks and other protected areas. Additional analysis of a combined front- and backcountry dataset could further elucidate differences between the site-types and derive a single model (combination of dose metrics and mediator variables) that could be applied across all site types.

ACKNOWLEDGMENTS

The authors of this report wish to express their sincere gratitude to all who helped made this a successful study. Coordination and support were provided by the FAA, Western-Pacific Regional Office, and Office of Environment and Energy as well as the NPS, Natural Sounds and Night Skies Division. FAA representatives include Barry Brayer, Pete Ciesla, and Keith Lusk; Western-Pacific Regional Office, Rebecca Cointin, Raquel Girvin, Ph.D., Bill He, Ph.D., and Jake Plante, Ph.D.; Office of Environment and Energy, Headquarters. NPS representatives include Karen Trevino, Kurt Fristrup, Ph.D., Frank Turina, Ph.D., Shan Burson, and Rick Ernenwein. The following panel of researchers was invaluable to this effort, providing expertise, advice and leadership during research development, survey design, and data collection efforts: Grant Anderson, Bill Borrie, Ph.D., Gregg Fleming, Joshua Hassol, Ph.D., Aaron Hastings, Ph.D., Richard Horonjeff, Steve Lawson, Ph.D., Cynthia Lee, Peter Newman, Ph.D., Britton Mace, Ph.D., Robert Manning, Ph.D., Nicholas Miller, Christopher Roof, and Erica Ryherd, Ph.D.

[1]T. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**(6), 377–405 (1978).
[2]S. Fidell, D. S. Barber, and T. J. Schultz, "Updating dosage-effect relationship for the prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233 (1991).
[3]H. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[4]Federal Interagency Committee on Noise (FICON), *Federal Agency Review of Selected Airport Noise Analysis Issues* (U.S. Department of Defense, Washington, DC, 1992), pp. 2-1–2-7.
[5]National Parks Overflights Act: Public Law 100-91 (1987).
[6]G. S. Anderson, R. D. Horonjeff, C. W. Menge, N. P. Miller, W. E. Robert, D. Rossano, G. Sanchez, R. M. Baumgartner, and C. McDonald, "Dose-response relationships derived from data collected at Grand Canyon, Haleakala and Hawaii Volcanoes National Parks," NPOA Report No. 93-6, Harris Miller Miller and Hanson, Lexington, MA (1993).
[7]G. G. Fleming, C. J. Roof, A. S. Rapoza, D. R. Read, J. C. Webster, and P. C. Liebman, "Development of noise dose/visitor response relationships for the national parks' overflight rule: Bryce Canyon National Park Study," Federal Aviation Administration. Report No. FAA-AEE-98-01, U.S. Department of Transportation, Washington, DC (1998).
[8]N. P. Miller, G. S. Anderson, R. D. Horonjeff, R. H. Thompson, R. M. Baumgartner, and P. Rathbun, "Mitigating the effects of military aircraft overflights on recreational users of parks: Final report," Harris Miller Miller and Hanson, Burlington, MA (1999).
[9]S. Fidell, L. Silvati, R. Howe, and K. S. Pearsons, "Effects of aircraft overflights on wilderness recreationists," J. Acoust. Soc. Am. **100**(5), 2909–2918 (1996).
[10]G. S. Anderson, A. S. Rapoza, G. G. Fleming, and N. P. Miller, "Aircraft noise dose-response relationships for national parks," Noise Control Eng. J. **59**(5), 519–540 (2011).
[11]A. S. Rapoza, C. S. Y. Lee, and J. Hassol, "Proceedings of the second workshop on human response to aviation noise in protected natural areas," U.S. Department of Transportation, Volpe National Transportation Systems Center, Cambridge, MA (2009). Available online at http://ntlsearch.bts.gov/tris/record/ntl/35897.html (Last viewed 2/26/2015).
[12]A. S. Rapoza, K. C. Lewis, E. A. Sudderth, C. S. Y. Lee, and J. Hassol, *Human Response to Aviation Noise Development of Dose-Response Relationships for Backcountry Visitors. Study Methods* (U.S. Department of Transportation, Volpe National Transportation Systems Center, Cambridge, MA, 2014), Vol. I.
[13]A. S. Rapoza, K. C. Lewis, E. A. Sudderth, C. S. Y. Lee, and J. Hassol, *Human Response to Aviation Noise Development of Dose-Response Relationships for Backcountry Visitors. Results and Analyses* (U.S. Department of Transportation, Volpe National Transportation Systems Center, Cambridge, MA, 2014), Vol. II.
[14]See supplementary material at http://dx.doi.org/10.1121/1.4929934 for a summary document containing a representation of the visitor survey instruments.
[15]S. Lawson, B. Kiser, K. Hockett, and A. Ingram, *Research to Support Backcountry Visitor Use Management and Resource Protection in Halekala National Park: Study Completion Report* (Virginia Polytechnic Institute and State University, College of Natural Resources, Blacksburg, VA, 2008).
[16]Clips are chosen by a computer sampling algorithm with a partially randomized design to minimize bias due to aircraft, aircraft type, selection of clips, or order in which the clips are played.
[17]E. Pilcher, P. Newman, and R. Manning, "Understanding and managing experiential aspects of soundscapes at Muir Woods National Monument," Environ. Manage. **43**, 425–435 (2009).
[18]B. L. Mace, P. A. Bell, R. J. Loomis, and G. E. Haas, "Source attribution of helicopter noise in pristine national park landscapes," J. Park Recreat. Admin. **21**(3), 97–119 (2003).
[19]B. L. Mace, G. C. Corser, L. Zitting, and J. Denison, "Effects of overflights on the national park experience," J. Environ. Psychol. **35**, 30–39 (2013).
[20]J. M. Fields, R. G. DeJong, T. Gjestland, I. H. Flindell, R. F. S. Job, S. Kurra, P. Lercher, M. Vallet, and T. Yano, "Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation," J. Sound Vib. **242**, 641–679 (2001).

[21]ISO/TS 15666:2003, "Acoustics—Assessment of Noise Annoyance by means of social and socio-acoustic surveys" (International Standards Organization, Geneva, Switzerland, 2003).
[22]Previous measurements of this type utilized ultra-sensitive short-term monitoring systems with a two-stage windscreen (a 20-in. diameter fabric-covered outer stage with a 3.5-in. diameter foam inner stage) and a 1-in. microphone. It is expected that the two-stage system and the system described perform similarly at wind speeds less than 5 m/s. Data at wind speeds greater than 5 m/s were eliminated in both studies (see Ref. 23).
[23]A. S. Rapoza, J. M. MacDonald, A. Hastings, C. Scarpone, C. S. Y. Lee, and G. G. Fleming, "Development of improved ambient computation methods in support of the National Parks Air Tour Management Act," U.S. Department of Transportation, Volpe National Transportation Systems Center, Cambridge, MA, 2008), pp. 79–81. Available online at http://ntlsearch.bts.gov/tris/record/ntl/37847.html (Last viewed 2/26/2015).
[24]S. Ambrose and C. Florian, *Acoustic Measurements in Arches National Park, Canyonlands National Park, Hovenweep National Monument, and Natural Bridges National Monument, 2000–2007* (Sandhill, Castle Valley, UT, 2009), p. 62.
[25]S. Ambrose, *Sound Levels in the Primary Vegetation Types in Grand Canyon National Park, July 2005* (Sandhill, Castle Valley, UT, 2006), p. 62.
[26]Position information for respondents without a tracking device (approximately 30% of total) was estimated from average hike times of similar visits and noted destination and start time.
[27]ANSI S1.1-1994, *Acoustical Terminology* (Acoustical Society of America, New York, 1994).
[28]S. Fidell, S. Teffeteller, R. Horonjeff, and D. M. Green, "Predicting annoyance from detectability of low-level sounds," J. Acoust. Soc. Am. **66**(5), 1427–1434 (1979).
[29]N. P. Miller, G. S. Anderson, R. D. Horonjeff, C. W. Menge, J. C. Ross, and M. Newmark, "Aircraft noise model validation study," HMMH Report No. 295860.29, Harris Miller Miller and Hanson, Burlington MA (2003), pp. 167–179.
[30]N. H. Reddingius, "User's Manual for the National Park Service Overflight Decision Support System," BBN Report No. 7984, U.S. Department of the Interior, National Park Service, Denver CO (1994), pp. 95–97.
[31]References to shortened response and mediator variable names are noted in italics throughout this paper.
[32]Those respondents who received an aircraft noise exposure dose but indicated they did not hear aircraft were coded as "not at all" annoyed or experiencing interference. In addition, HR2 survey responses on the positive or pleasing aspect of the nine-point response scale were coded as "not at all" annoyed.
[33]A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, New York, 2007), pp. 301–324.
[34]H. Akaike, "A new look at the statistical model identification," IEEE Trans. Automat. Control **19**, 716–723 (1974).
[35]Note that survey type is a categorical variable with three values (AC, HR1, or HR2). During regression analysis, the AC survey is the reference and coefficients are estimated for the HR1 and HR2 survey types.
[36]These mediators are included as they were previously found to strongly influence visitor response to a given dose.
[37]Relative probability = exp[(AICmin-AICi)/2].
[38]Alternate model selection procedures and criteria may be used and may lead to different results.
[39]G. Moors, N. Kieruj, and J. Vermunt, "The effect of labeling and numbering of response scales on the likelihood of response bias," Sociol. Method. **44**(1), 369–399 (2014).
[40]E. Wetzel, C. H. Carstensen, and J. R. Bohnke, "Consistency of extreme response style and non-extreme response style across traits," J. Res. Personal. **47**(2), 178–189 (2013).
[41]D. S. Tull and D. Hawkins I, *Marketing Research: Measurement and Method* (Macmillan, New York, 1993).
[42]To most accurately represent the data collected, a linear relationship was used between %TAud and LAE (slope1, intercept = −40), while logistic relationship was utilized for PEn's to constrain the percent energy between 0% and 100% (PEnHelos slope at midpoint = 0.27, midpoint = −20, PEnProps slope at midpoint = −0.06, midpoint = 3.7).
[43]M. A. Tarrant, G. E. Haas, and M. J. Manfredo, "Factors affecting visitor evaluations of aircraft overflights of wilderness areas," Soc. Nat. Resources **8**(4), 351–360 (1995).