

Alternative analysis of sleep-awakening data

Grant S. Anderson^{a)} and Nicholas P. Miller^{b)}

(Received 2006 May 18; revised 2006 November 03; accepted 2007 November 02)

Most research on sleep awakenings caused by aircraft noise determines the percent of awakenings from an *average person's* exposure to *single aircraft* events. Such research estimates the percent of people who would likely be awakened by a single aircraft of a given sound level, if everyone had average sensitivity to sleep awakening. However, such results do not show what percent of a composite population (of *all sensitivities*) would likely be awakened by a *full night* of aircraft events, of varying sound levels. It is this latter type of disturbance that provides a useful answer to the common question: How will altered nighttime aircraft operations affect awakenings in the communities surrounding an airport? This paper describes an alternative method for analyzing available sleep-awakening data to provide an answer to this question. The method is determined with one set of sleep-disturbance data, to yield dose-response relationships that may be applied to actual airport operations. The relationships are then validated by applying them to independent sleep-awakening data, to predict the occurrence of sleep awakening in that data set. Predicted sleep awakening (and its predicted uncertainty) matched the independent data well. © 2007 Institute of Noise Control Engineering.

Primary subject classification: 63.4; Secondary subject classification: 52.2

1 INTRODUCTION

Aircraft noise can disturb sleep. The circumstances that produce disturbance have long been studied and continue to be investigated.¹⁻⁹ In many such studies, results are expressed as the percent of awakenings likely to occur as a function of a single-event noise metric (for example Fig. 1 from Ref. 2).

Though such results provide insight into the average sensitivity to awakening from single noise events, they have two shortcomings. First, they provide no direct way to estimate percent of the population likely to be awakened from a *full night* of aircraft noise events. Second, they represent average sensitivity to awakening, losing information about the *range of sensitivities* that are likely to exist in any exposed population, and raising the possibility that estimates of total awakenings will be incorrect—since applying average sensitivity to all people can incorrectly compute total awakenings (see Sec. 6.3 below).

One method to overcome these shortcomings was developed for possible regulatory use by the European

Union, in response to the European Noise Directive.¹⁰ This directive specifies the cumulative metric L_{night} (the annual-average equivalent sound level between 11 p.m. and 7 a.m.) to assess sleep disturbance. For sleep disturbance produced by aircraft noise, this regulatory guidance addresses (1) the expected maximum number of noise-induced motilities (body movements while asleep) and (2) the increase in mean motility—both as a function of L_{night} .¹¹

This use of L_{night} , however useful for regulatory purposes, is likely to have some of the same difficulties encountered with other cumulative metrics such as the Day-Night Average Sound Level (DNL). Primarily, it is difficult (if not impossible) to relate specific aircraft-noise effects to such cumulative metrics (e.g., what percent of the population will be awakened?). So too, it is difficult to communicate to decision-makers and to the public the meaning of changes in such cumulative metrics—especially in easily understood terms such as “the expected change in percent of the population awakened.”

What the authors wish to achieve, with the alternative analysis described here, is an advance on the dose-response relationship previously recommended by FICAN (solid curve in Fig. 1) and an improvement in the utility of the results of past sleep-disturbance studies. This alternative analysis estimates the percent-

^{a)} Harris Miller Miller & Hanson Inc., 77 South Bedford Street, Burlington MA 01803; email: ganderson@hmmh.com

^{b)} Harris Miller Miller & Hanson Inc., 77 South Bedford Street, Burlington MA 01803; email: nmiller@hmmh.com

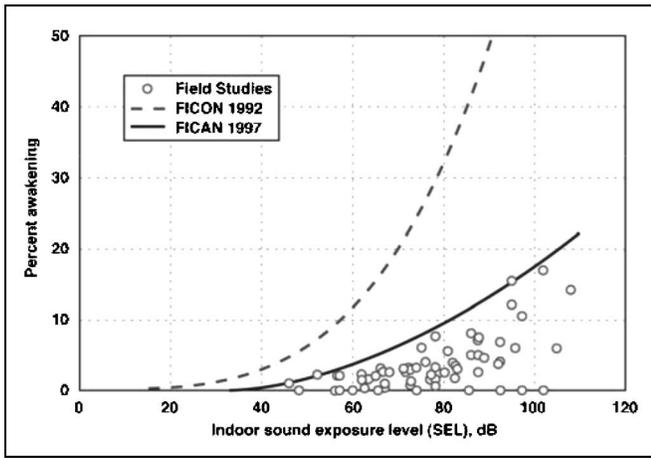


Fig. 1—Dose-response relationship for awakening from aircraft noise events: Recommended by FICAN in 1997.

age of the population awakened (once or more) by a full night of aircraft operations, taking into account the large person-to-person variability in sensitivity to noise-induced sleep awakening. For this estimate, the analysis uses Sound Exposure Levels (SELs) and their timings through the night, as computed by the Federal Aviation Administration’s Integrated Noise Model (INM) at grid points around an airport.

Section 2 of this paper summarizes the data for this alternative analysis. The U.S. Air Force provided these data, which were previously obtained by Dr. Sanford Fidell and his co-workers under contract to the U.S. Air Force and were previously reported in Refs. 3, 4, 6, and 7. The data are divided into two sets: (1) the “regression” set, used to develop the analysis method and

results, and (2) the “validation” set, used to test the results.

Section 3 describes the alternative analysis, along with its resulting dose-response relationship for single events and the generalization of that relationship to a full night of aircraft noise events. The multiple-event analysis accounts for each person’s specific exposure to aircraft SELs and their timings during the night. Moreover, the analysis allows determination of sleep awakening for population sub-groups, each with a different awakening sensitivity to aircraft noise, then the summation of total sleep awakening experienced by the entire population.

Section 4 describes the validation of these results using the validation data (same sleep studies, independent set of subjects), for which specific subject sensitivities are treated as unknowns—as they would be in application to an independent airport population. The validation tests address (1) potential bias with independent data, and (2) validity of computed uncertainties using the validation data.

Section 5 then describes how the results can be applied to a specific airport—to determine at specific INM grid points the percent of people awakened during an average night. In conclusion, Section 6 summarizes and discusses the main points of this paper.

2 DATA

2.1 Data Came from Prior Studies

Data for this paper were obtained in people’s homes by Dr. Sanford Fidell and his co-workers and have been previously reported in the acoustical literature.^{3,4,6,7,12}

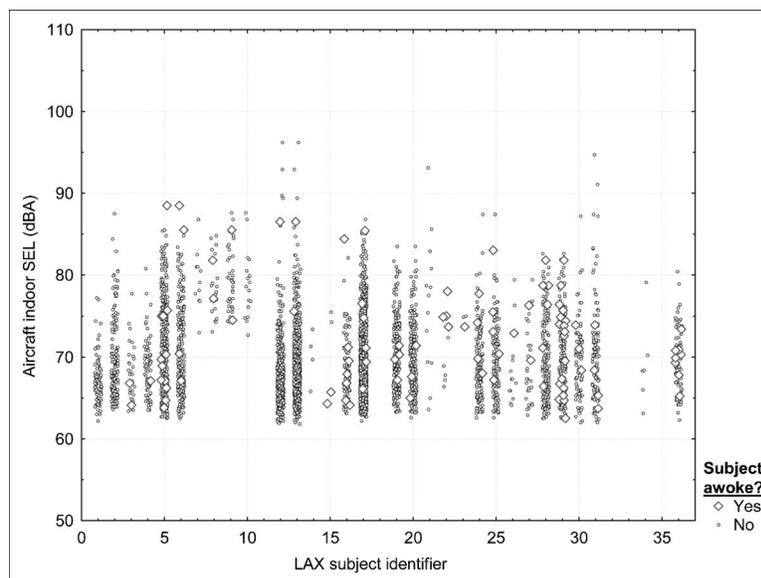


Fig. 2—Aircraft noise events experienced by subjects: LAX.

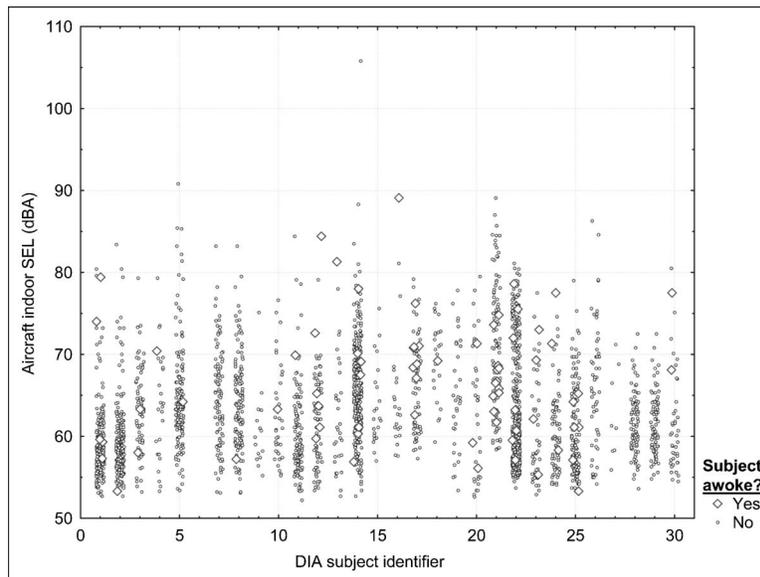


Fig. 3—Aircraft noise events experienced by subjects: DIA.

These data were collected during sleep-disturbance studies near three air facilities: Los Angeles International Airport (LAX), Denver International Airport (DIA), and Castle Air Force Base (Castle). At these facilities, aircraft noise was measured in 84 test-subjects’ bedrooms, while their “behavioral awakening” was simultaneously monitored (they pressed a button whenever they awoke).

2.2 Subject Awakening Sensitivity was Essential to this Alternative Analysis

Figures 2–4 plot each measured aircraft indoor SEL, separately for all 84 test subjects in the original studies. The SELs with large symbols awoke the subject; the small-symbol SELs did not.

As is obvious from these figures, test subjects varied significantly in their sensitivity to aircraft noise. For example, LAX Subject 29 shows high noise sensitivity, while LAX Subject 12 does not. This large subject-to-subject variability suggests “subject awakening sensitivity” as an important controlling variable in the analysis—a variable not included in prior analyses.

Use of this “sensitivity” variable, however, restricts this paper’s analysis to those subjects awakened at least once by aircraft noise. Only such subjects have sensitivities derivable from the data. Therefore, this analysis excluded all test subjects who never awoke due to aircraft noise during their entire noise study—17 out of 84, leaving 67 suitable test subjects. The excluded subjects were of three types: (1) apparently imperturb-

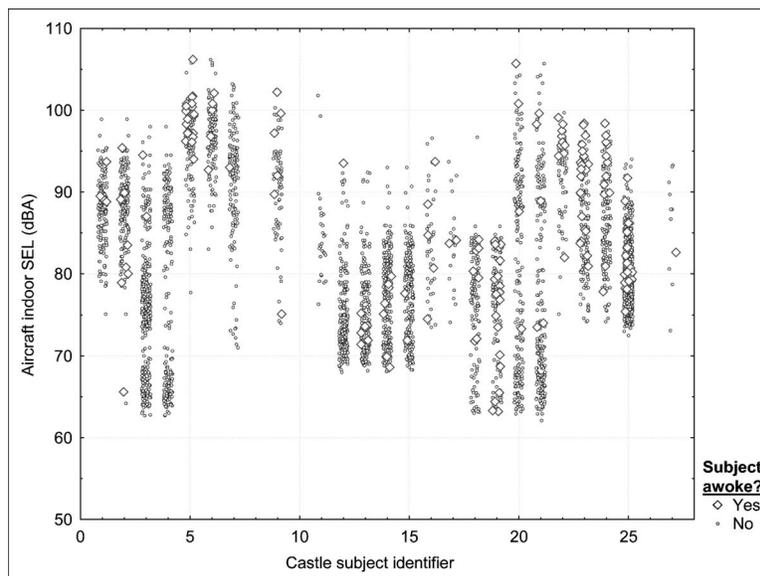


Fig. 4—Aircraft noise events experienced by subjects: Castle Air Force Base.

Table 1—Further data summary: Original studies compared to this re-analysis.

Air facility	Original study		Suitable for this paper's re-analysis (subject awoke at least once due to aircraft noise)			
	Number of subjects	Number of subjects	Number of subject-nights	Number of aircraft events	Number of aircraft events per subject-night	Aircraft indoor SEL (dB)
LAX (Los Angeles International Airport)	31	24	290	4,743	Maximum=279 Average=16 Minimum=1	Maximum=96 Average=70 Minimum=62
DIA (Denver International Airport—round 3)	29	21	359	2,432	Maximum=77 Average=7 Minimum=1	Maximum=106 Average=63 Minimum=52
Castle (Castle Air Force Base)	24	22	216	3,909	Maximum=435 Average=18 Minimum=1	Maximum=106 Average=81 Minimum=62
TOTAL	84	67	865	11,084	Maximum=435 Average=13 Minimum=1	Maximum=106 Average=72 Minimum=52

able subjects (Castle Subject 4, for example), (2) subjects not tested long enough to be awakened (LAX Subject 10, for example), and (3) subjects intermediate between these two extremes (DIA Subject 26, for example).

In total, the remaining 67 suitable subjects were tested over 865 subject-nights, while exposed to 11,084 aircraft events (averaging 13 per subject-night, 72 dB aircraft indoor SEL). Analysis data also include the timing of each aircraft event and whether or not that aircraft awoke the subject—plus additional explanatory variables such as the subject's duration of residence, age, and tiredness before retiring. Table 1 further summarizes these data.

2.3 Data were Split into Regression and Validation Subsets

Prior to analysis, test subjects were randomly split into two subsets: a “regression” subset (33 subjects, with 7,139 aircraft-noise events) and a “validation” subset (34 test subjects, with 3,945 aircraft-noise events).¹³ Table 2 summarizes these two data subsets.

Data for the regression group of test subjects determined this paper's dose-response relationships. Data for the validation group were used to test the analysis results. The regression and validation subsets did not share any test subjects.

3 ANALYSIS

3.1 Logistic Regression Determined the Single-Aircraft Dose-Response Curve

The sleep-awakening relation for single aircraft was determined from the regression subset of test subjects, using logistic regression with this functional form:^{14,15}

$$P_{awake, single} = \frac{1}{1 + e^{-Z}}$$

$$Z = \beta_0 + \beta_L L_E + \sum_{i=1}^9 \beta_i v_i + \sum_{s=1}^{32} \beta_s d_s = \text{“the logit.”}$$
(1)

From this dose-response relation, the probability of being awakened by a single aircraft is a function of the following:

- (1) L_E , the aircraft indoor SEL (dB)
- (2) A summation of terms for potentially nine additional explanatory variables (see Table 3)
- (3) A summation of the “sensitivity” terms, one for each of the 33 test subjects in the regression data—but minus one “reference” test subject, whose sensitivity is subsumed in the constant β_0 .

In this equation, each explanatory variable is multiplied by its corresponding regression coefficient, β . For the subject-sensitivity terms, each β_s is one test subject's sleep-disturbance sensitivity to aircraft noise—multiplied by a dummy variable, d_s , that is unity for that test subject and zero for all others.

Table 2—Further data summary: Regression data compared to validation data.

Type of variable	Variable	Details	Regression data			Validation data		
			Min	Avg	Max	Min	Avg	Max
Demographic	Number of aircraft events per subject	—	10	216	1158	3	116	366
Response	Behavioral awakening	0: not awakened by that aircraft noise event 1: awakened by that aircraft noise event	—	—	—	—	—	—
Explanatory	Subject number	One dummy variable and one regression coefficient needed for each subject, minus one “reference” subject.	—	—	—	—	—	—
	Indoor SEL (dB)	—	52	71	106	53	75	106
	Time since retiring (min)	—	0	168	626	0	179	602
	Subject duration of residence (months)	—	5	113	504	16	193	486
	Subject age (years)	—	25	48	79	18	51	72
	Subject tiredness: Four dummy variables needed for five tiredness categories.	Logged just before retiring on a 5-point scale: Not at all, slightly, moderately, very, extremely.	1	2.4	5	1	2.4	5
	Indoor ambient sound level (dB)	For LAX and Castle: L_{eq} of the quietest 5-minute epoch during the night. For DIA: 2-second L_{eq} just prior to each individual aircraft event.	19	40	81	20	34	61
	Number of spontaneous button pushes during the night	These are the awakenings not associated with an aircraft noise event.	0	1.5	16	0	1.4	9
			LAX	DIA	Castle	LAX	DIA	Castle
Demographic	Number of male subjects	—	8	7	2	2	2	7
	Number of female subjects	—	6	5	5	8	7	8
	Total number of subjects	—	14	12	7	10	9	15

Table 3 contains the best-fit logistic regression results—that is, the values of β in Eqn. (1) that best match the regression set’s subject awakenings during the sleep studies at these three air facilities. The first column of the table lists each explanatory variable in the regression (including the intercept, β_0). Subsequent columns of the table pertain to separate logistic regressions. Regression 1 retains only aircraft SEL and the intercept. Regression 2 retains all explanatory variables except subject awakening sensitivity. Regression 3

retains variables selected on the basis of p-value and pragmatic application considerations (readily available during typical airport applications)—and includes subject awakening sensitivity.

For each regression, the various table cells contain each variable’s regression coefficient and standard uncertainty (often called standard error) in parentheses, followed by its p-value. Plus/minus twice the standard uncertainty is the approximate 95% confidence range of the regression coefficient. For well-determined

Table 3—Logistic regression results.

Explanatory variable	Regression coefficient, then its (standard uncertainty), then its p-value					
	Regression 1: Only SEL and intercept		Regression 2: All explanatory variables except subject awakening sensitivity		Regression 3: Selected variables, including subject awakening sensitivity	
Intercept, β_0	-6.619 (0.699)	p < 0.001	-7.538 (0.972)	p < 0.01	-10.723 (1.036)	p < 0.01
Aircraft indoor SEL, L_E	+0.03888 (0.00950)	p < 0.001	+0.04444 (0.01037)	p < 0.01	+0.08617 (0.01395)	p < 0.01
Time since retiring, T_{retire} , (min)	—		+0.00336(0.00044)	p < 0.01	+0.00402 (0.00059)	p < 0.01
Subject duration of residence (months)	—		+0.00266 (0.00101)	p < 0.01	—	
Subject age (years)	—		-0.01301 (0.00944)	p=0.17	—	
Subject tiredness: Slightly	—		-1.155 (0.249)	p < 0.01	—	
Subject tiredness: Moderately	—		-0.041 (0.178)	p=0.82	—	
Subject tiredness: Very	—		+0.295 (0.165)	p=0.07	—	
Subject tiredness: Extremely	—		-0.278 (0.266)	p=0.29	—	
Indoor ambient sound level	—		-0.000024 (0.000099)	p=0.81	—	
Number of spontaneous button pushes during the night	—		+0.09415 (0.04696)	p=0.04	—	
Subject awakening sensitivity	—		—		Most sensitive: +2.646 90 th percentile: +1.688 75 th percentile: +0.871 50 th percentile: -0.038 25 th percentile: -0.947 10 th percentile: -1.764 Least sensitive: -2.928	

Note: The regression coefficients for subject awakening sensitivity are not included in the table. Instead, summary values of their distributions are provided. Fig. 5 is a histogram of all 33 coefficients derived from the regression data set.⁴²

coefficients, this range will not include zero—and correspondingly, the p-value will be less than 0.05.

3.2 Regressions 1 and 2 Provided a Baseline

Regression 1 is the sparsest regression of interest. Its SEL regression coefficient is well determined: p < 0.001, corresponding to a standard uncertainty very small compared to the coefficient, itself (0.00950 compared to 0.03888). That regression coefficient provides a baseline for comparison with Regressions 2 and 3.

Regression 2 is the most complex regression that excludes subject awakening sensitivity. Its SEL coefficient is well determined and is somewhat larger than for Regression 1. Note that several coefficients in this regression are *not* well determined: subject age, all but one of the tiredness coefficients, indoor ambient sound

level, and spontaneous button pushes (marginally well determined). Either these variables are not truly important to sleep awakening or their association is too weak to determine with this quantity of data. These variables were dropped from further consideration.¹⁶

3.3 Regression 3 Carried Forward in the Analysis

Regression 3 is carried forward in the analysis. From the regression coefficients in Table 3:

$$P_{awake,single} = \frac{1}{1 + e^{-Z}}$$

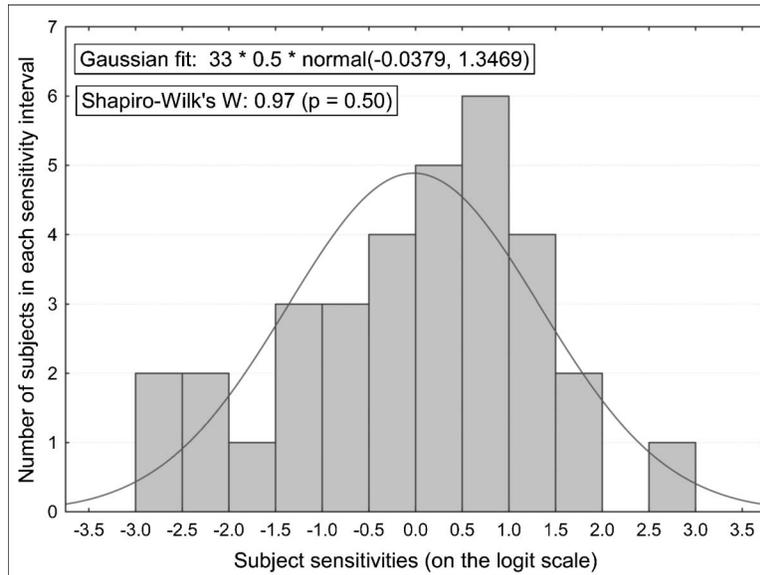


Fig. 5—Histogram of subject-awakening-sensitivity coefficients.

$$Z = -10.723 + 0.08617L_E + 0.00402T_{retire} + \sum_{s=1}^{32} \beta_s d_s, \quad (2)$$

where, for compactness, the sensitivity coefficients of Fig. 5 are not explicitly included.^{17–19} The time since retiring, T_{retire} , is retained. In addition to variables previously dropped, this regression also drops duration of residence.²⁰

As an example of this equation’s use, the probability of awakening a 50-percentile person ($\beta_{50\%} = -0.038$ from Table 3, $d_{50\%} = 1$, all other $d_s = 0$) computes to 0.05—for an aircraft indoor SEL of 80 dB that occurs four hours (240 minutes) after the person retires. The corresponding probability of “sleeping through” that aircraft is 0.95.

Figure 6 graphs the dose-response curve for a single-aircraft event, as computed with Eqn. (2). Each frame of the figure has a heavy line that is the percentage chance of awakening (100 times the probability) as a function of the aircraft indoor SEL. This line is flanked vertically by its 95% confidence bounds. For any value of SEL on the horizontal axis, we are 95% certain that these confidence bounds contain the true response between them. Mathematically, these confidence bounds depend upon the standard uncertainties of all the regression coefficients, as well as the correlations among these uncertainties (see Sec. 7 for a derivation/discussion of uncertainty, and thereby for a further explanation of this statement about uncertainty).

In addition to aircraft indoor SEL, the response to a single-aircraft event is also a function of that noise event’s “time since retiring” and of the subject’s sensi-

tivity. To show those two dependencies, the figure plots the dose-response curve for these nine combinations: (1) three different times since retiring, across the figure, and (2) three different subject sensitivities, down the figure (see the top and left labels). Finally, the histograms at the bottom of the figure show the distribution of aircraft indoor SELs in the regression data set for each of the three time periods.

Figure 6 illustrates Eqn. (2) and shows that increased awakening is associated with (1) increased SEL (as expected), (2) increased time since retiring (as found in previous analyses of these data^{3,4} and in a more recent study²¹), and (3) increased subject awakening sensitivity (no surprise, but newly quantified in this analysis). Uncertainties are very small at low SELs, due mainly to the zero lower-bound constraint on the chance of awakening. At intermediate SELs, uncertainties remain small—due to large numbers of intermediate-SEL data (see the histograms in Fig. 6). At high SELs, uncertainties are significantly larger, due to very few high-SEL data. Although harder to see in the figure, uncertainties also depend upon time since retiring and subject awakening sensitivity, though in a complex way that depends upon SEL, as well.

3.4 Regression 3 Generalized to Multiple Aircraft each Night

Application of Eqn. (2) to an entire night of aircraft requires generalization from a single aircraft to multiple aircraft. The approach used here is to determine first the probability of sleeping through *all* aircraft during the night and subtracting that probability from unity, yielding the probability of awakening *at least once* during the night.

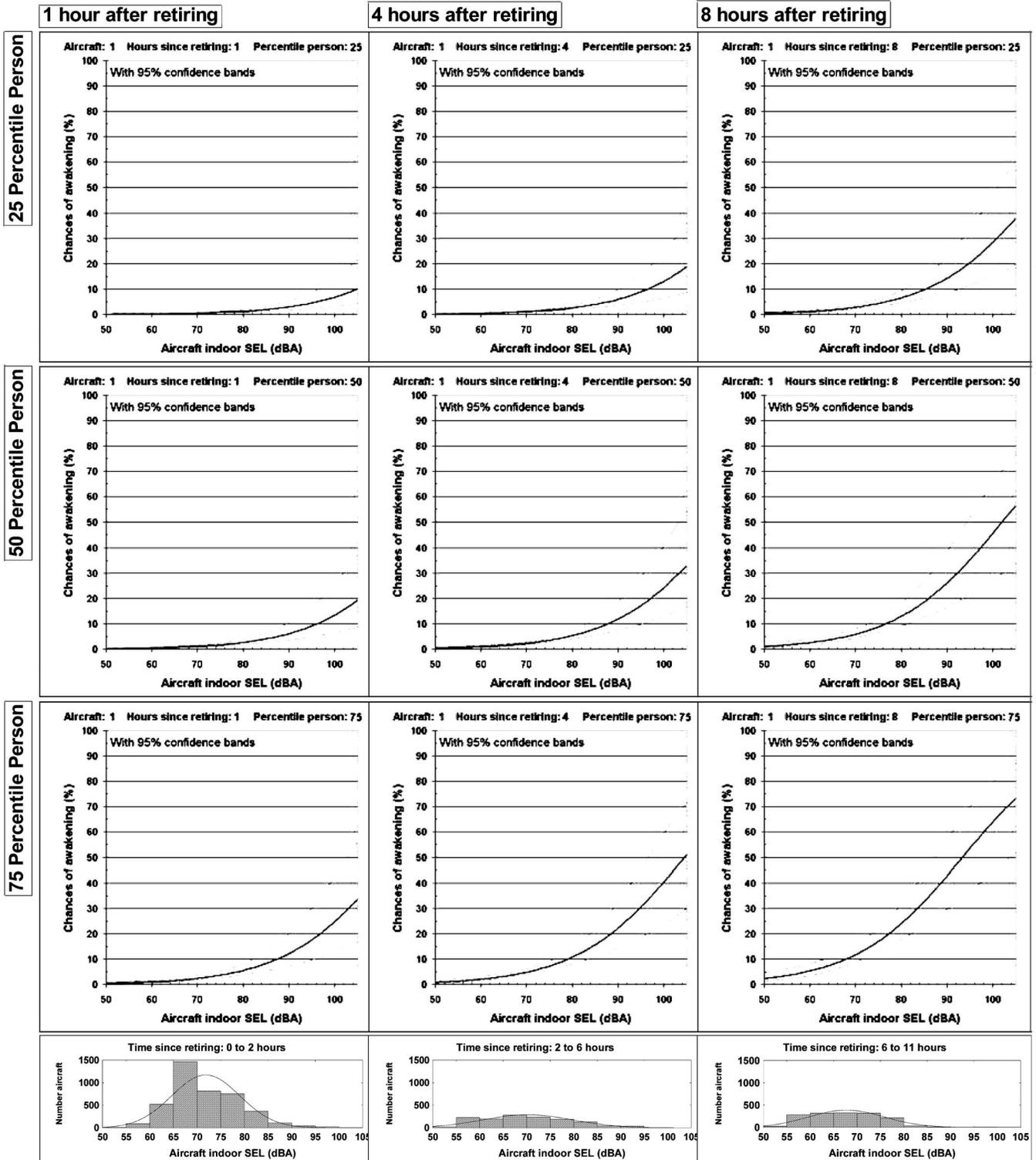


Fig. 6—Dose-response curves and their uncertainties: Single aircraft.

The probability of sleeping through all aircraft equals the joint probability of sleeping through the first one, *and* sleeping through the second one, *and* sleeping through the third one, and so on. For example, with multiple SELs of 80 dB, four hours after retiring, the probability of a 50-percentile person sleeping through the first aircraft is 0.95. Therefore, the probability of

sleeping through both the first and the second is 0.95 multiplied by 0.95 (95% of 95%), which equals 0.90. Another multiplication is required for the third aircraft, multiplying again by 0.95, to get 0.86. The more aircraft that fly over, the lower is the probability of sleeping through them all, as expected. In general:

$$P_{sleep,multiple} = \prod_{a=1}^N (p_{sleep,single})_a, \quad (3)$$

where $a=1, \dots, N$ is an index over all aircraft during the night.²² In turn, the probability of *awakening* at least once during the night follows from the reciprocal nature of “sleeping through” versus awakening:

$$\begin{aligned} P_{awake,multiple} &= 1 - P_{sleep,multiple} \\ &= 1 - \prod_{a=1}^N (p_{sleep,single})_a \\ &= 1 - \prod_{a=1}^N (1 - p_{awake,single})_a, \end{aligned} \quad (4)$$

where $p_{awake,single}$ is determined by Eqn. (2). As indicated in these two equations, p_{sleep} and p_{awake} sum to unity, for both single and multiple aircraft.

Figure 7 graphs the resulting dose-response curves for multiple aircraft. As with the previous figure, each dose-response curve is flanked vertically by its 95% confidence bounds. To show dependence upon the number of aircraft during the night (top labels), the curves in this figure are based on a time after retiring of four hours. Governed by Eqn. (4), the figure shows increased awakening for larger numbers of aircraft during the night: from one, to five, to ten aircraft across the figure. Additional awakening would occur for even larger numbers of aircraft during the night.

The curves of this figure were computed assuming the most likely value of sensitivity (the 50-percentile person)—in contrast to Fig. 6, in which the effect of sensitivity was considered known and shown explicitly. Hence, the dose-response curve in the upper left of Fig. 7 (1 person, 1 aircraft at 4 hours after retiring) is the same as in the center frame of Fig. 6. What are different between the two are the confidence limits. They are much wider in Fig. 7 because the true sensitivity of the person is unknown—as it would be in an actual application around an airport—whereas that frame in Fig. 6 is for the 50-percentile person.

What is most important to observe is that in Fig. 7, increasing the number of persons averaged and/or increasing the number of aircraft events significantly reduces the uncertainty of the relationship. As the number of aircraft increases, that uncertainty becomes less because the person’s individual response is averaged over more and more aircraft (which affect the person *independently*, no matter what the person’s sensitivity).

Lower rows in Fig. 7 show the effect of increasing the number of persons averaged. The uncertainties diminish dramatically as this number increases. Note

that the diminishing uncertainty becomes stable after approximately 100 people are averaged—that is, the bottom row of graphs (1000 persons) shows essentially the same uncertainty as the row above it (100 persons). This stability occurs when the sensitivity contribution to total uncertainty becomes less than the contribution due to uncertain regression coefficients. To reduce uncertainty further would require more sleep-study data, to obtain better-known regression coefficients. In the bottom-right panel of the figure, the combined effect of averaging people and averaging aircraft is dramatic—an uncertainty of only plus/minus several percentage points.

4 VALIDATION

The regression coefficients in Eqn. (2), as well as the confidence bounds in Figs. 6 and 7, were all developed by analysis of the “regression” subset of data. How valid are these coefficients and uncertainty computations for independent data? This section tests their validity using the “validation” subset of data, which did not contribute in any way to the regression derived above.

4.1 Bias Appears Small using Validation Data

The bias test is most straightforward. It is applied to two situations: (1) predicting awakening from single aircraft events, (2) predicting awakening from full nights of aircraft events.

4.1.1 Single aircraft events

Regression bias for single aircraft events was computed as follows. For each aircraft noise event across all “validation” subjects, the probability of awakening was computed with Eqn. (2). These computations used each aircraft’s indoor SEL and its time after the subject retired (both measured), but did not use each subject’s sensitivity to aircraft sound because that sensitivity was not known for “validation” subjects. Rather, the average (expected) sensitivity, derived for the “regression” subjects, was assumed for each validation subject.

Then all these probabilities (one for each subject-aircraft pair) were averaged and multiplied by 100, to obtain 4.7 percent. This computed “percentage chance of awakening for single aircraft events” was then compared to the percentage of events that *actually* awoke the “validation” subjects: 5.3 percent. The first row of Table 4 contains that computed/actual comparison. As shown, the computations under-estimate the actual single-event awakenings by only 0.6 percentage points.

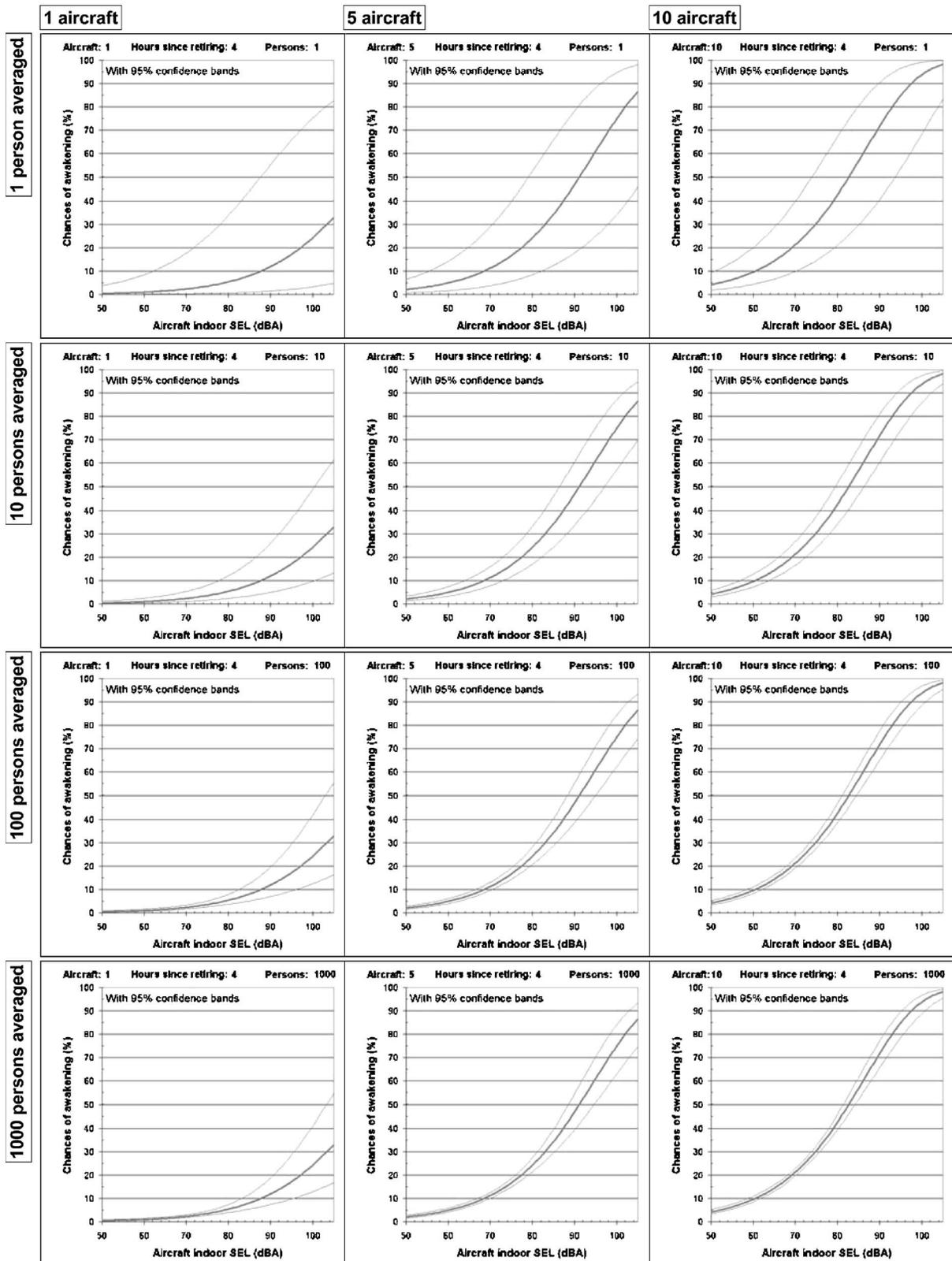


Fig. 7—Dose-response curves and their uncertainties: Multiple aircraft and multiple persons averaged (4 hours after retiring, most-likely sensitivity).

4.1.2 Full nights of aircraft events

Regression bias for full nights of aircraft events was computed as follows. Equations (2) and (4) were first

used to compute the probability that each “validation” subject would awaken at least once during each night—again assuming average sensitivity for each subject.

Table 4—Bias in chance awakened.

Situation	Chance awakened		
	Computed (%)	Actual (%)	Bias (percentage points)
Single aircraft	4.7	5.3	-0.6 percentage points
Multiple aircraft	26.3	23.7	+2.6 percentage points

Then all these probabilities (one for each subject-night) were averaged and multiplied by 100, to obtain 26.3 percent. This computed “percentage chance of awakening at least once during the night” was then compared to the percentage of subject-nights for which the subject actually awoke: 23.7 percent. The second row of Table 4 contains that computed/actual comparison. As the table shows, the regression’s multiple-aircraft bias equals only +2.6 percentage points.

4.2 Computed Uncertainties Appear Valid for Validation Data

This second validation test is more complex. It tests whether the confidence bounds in Figs. 6 and 7 are reasonable to apply to independent data (the validation data set). Of most importance is the validity of confidence bounds for multi-person averages and for multiple aircraft during the night—because when this sleep-awakening method is applied to an actual airport, it will be applied to a group of people (a census block, for example) for a full night of aircraft operations (Sec. 5, below). Testing these data-hungry averages required this paper’s validation data set to be a larger-than-normal fraction of the total data set.

For this test of uncertainty, all validation data were first grouped by subject-nights. Then these were put in random order and further grouped into sequential blocks of subject-nights. Each subject-night was kept intact so that awakening of a single subject from a full night of aircraft events could be assessed. The uncer-

tainty computations were made for 50, 90 and 95% confidence intervals, with block sizes of 3, 6, 12, 25, 50, 100 and 200 subject-nights.

Figure 8 shows the result of this uncertainty test for the 95% confidence interval, averaged 12 subject-nights at a time. The figure plots the chance of awakening at least once during the night, separately for each of the resulting 34 blocks of data. For each block of subject-nights, the solid symbol shows the computed chance of awakening, based upon the dose-response equation. The vertical bar shows the 95% confidence limits of that computation, based upon the actual SELs and timings of all aircraft over all 12 nights in that block. This confidence interval is relatively small for blocks of data with many aircraft events or with lower-SEL events, for which the awakening potential is known with more certainty.

For comparison with these computed results, the figure’s empty symbols show each block’s actual percentage of subject-nights in which the subject awoke. Of interest is how completely the computed confidence bounds capture the actual percentage awakened. In other words, what fraction of the empty symbols actually lies within their block’s computed 95% confidence interval? In the figure, the target capture rate was 95%, while the actual capture rate was less than that—85%.

For this situation of blocks of 12 subject-nights, the computed confidence limits are slightly too small. If the confidence limits were to enlarge just a little, they would capture 95% of the empty symbols—more than the 85% currently captured. This result is expected. Independent data (in this case the validation data) always never fit better than do the data that generate the regression curve and its uncertainty. In the figure, actual awakenings for *five blocks* are outside of the computed 95% confidence intervals for those blocks (5 out of 34 equals 15 percent). Close examination of the figure, however, shows that the computed confidence limits do not miss those empty symbols by very much.

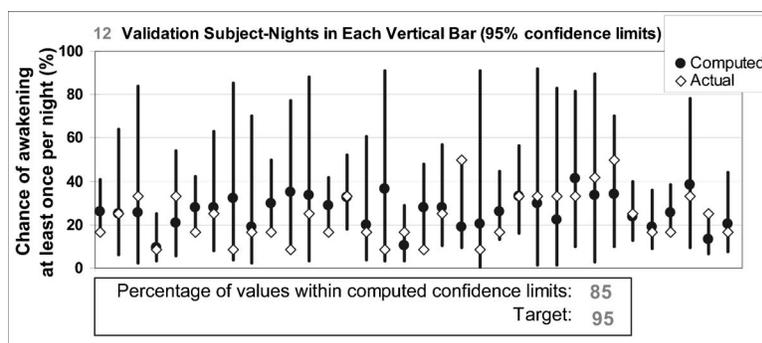


Fig. 8—Sample result of uncertainty test.

Table 5—Under-capture of uncertainty.

Confidence interval (%)	Target capture rate	Validation capture rate
95	95	93.6
90	90	86.7
50	50	46.9

In fact, none misses enough to be discerned in the figure.

Table 5 summarizes the full set of validation results for this test—all three confidence intervals, totaled over all seven block sizes. As the table shows, the under-capture of uncertainty is relatively minor, on the average. For the validation data, the uncertainty computations under-estimate the target confidence bounds by only several percent. This is strong validation that the uncertainty computations transfer closely to independent data. In other words, the sleep-awakening analysis method, with its coefficients and uncertainty bounds derived from the regression data, can be used confidently for independent populations.

5 APPLICATION TO AN AIRPORT SCENARIO

The results of the analysis in this paper lead to an algorithm that, for a given set of nighttime aircraft operations, estimates what portion of the population will be awakened at least once during the night.²³ A computer program, such as the INM, is run to determine the set of nighttime SEL values that occur at a series of grid points, each associated with a population. The succeeding calculations are iterative and are conducted at four nested levels:

- For each grid point, associated with a population
- At each grid point, by population group based on sensitivity to awakening
- For each time period of the night
- For each aircraft SEL, adjusted for outdoor-to-indoor sound level reduction.

This sequence of calculations yields both the percent of people and the number of people likely to be awakened at least once during the night at each grid point.²⁴ For airport application, the following equation combines the results of Eqns. (2) and (4), above, with the additional nesting by population group and time period of the night:

$$P_{awake,GridPoint} = \sum_{g=1}^{33} (F_g) \left\{ 1 - \prod_{i=1}^I \prod_{a_i=1}^{A_i} (1 - P_{awake,single})_{a_i,i,g} \right\}$$

$$(P_{awake,single})_{a_i,i,g} = \frac{1}{1 + e^{-Z_{a_i,i,g}}}$$

$$Z_{a_i,i,g} = -10.723 + 0.08617(L_E)_{a_i} + 0.00402(T_{retire})_i + \beta_g. \quad (5)$$

In this equation:

- (1) $P_{awake,GridPoint}$ is the probability of awakening at least once per average night, for all persons living at an INM grid point.
- (2) The summation is over the 33 population groupings, g , in Table 6 (each 0.25 logit interval from -4.00 to $+4.00$).²⁵
- (3) F_g is the fraction of the population in the g^{th} population group (see Table 6).
- (4) β_g is the sensitivity of the g^{th} population group (see Table 6).
- (5) a_i is an index for the a^{th} aircraft in the i^{th} time interval, where i ranges from one to I (see Table 7).
- (6) L_E is the aircraft indoor SEL.
- (7) $(T_{retire})_i$ is the mid-point time of the i^{th} time interval since retiring, in minutes.

Note that for N identical aircraft:

$$\prod_{a_i=1}^N (1 - P_{awake,single})_{a_i,i,g} = \left((1 - P_{awake,single})_{a_i,i,g} \right)^N \quad (6)$$

By combining identical terms in this manner, Eqn. (5) also works properly for *fractions* of an aircraft. For example, with one-third of an aircraft, N equals 1/3 in Eqn. (6)—yielding a *smaller* percent chance of awakening, compared to a full aircraft. Three such fractional aircraft contribute three such terms, each raised to the one-third power. The product of these three terms equals a single term with unity exponent, as expected.

Although fractional aircraft have no physical reality, annual-average aircraft operations at airports always include such fractions (yearly operations for a specific aircraft/track combination, divided by 365 days). The INM routinely handles such fractional aircraft when it computes yearly average DNL. For sleep awakening, terms like Eqn. (6), imbedded in Eqn. (5) where needed, permit fractional aircraft, as well.

Numbers of people can be totaled over all grid points, as a single-number comparison among alternative scenarios. Contours of percent awakened can graphically illustrate the geographic distribution of sleep awakening—with a metric that is generally understood by airport neighbors and decision-makers. The following subsections provide a more detailed prescription for the associated calculations.

Table 6—Distribution of grid-point population by sensitivity to awakening.

Population group number (g in Eqn. (5))	Sensitivity on the logit scale (β_g in Eqn. (5))	Relative sensitivity to awakening (dB)	Fraction of grid-point population (F_g in Eqn. (5))
1	-4.00	-46.4	0.065
2	-3.75	-43.5	0.110
3	-3.50	-40.6	0.180
4	-3.25	-37.7	0.284
5	-3.00	-34.8	0.435
6	-2.75	-31.9	0.644
7	-2.50	-29.0	0.919
8	-2.25	-26.1	1.269
9	-2.00	-23.2	1.691
10	-1.75	-20.3	2.179
11	-1.50	-17.4	2.711
12	-1.25	-14.5	3.260
13	-1.00	-11.6	3.787
14	-0.75	-8.7	4.250
15	-0.50	-5.8	4.608
16	-0.25	-2.9	4.827
17	+0.00	+0.0	4.885
18	+0.25	+2.9	4.777
19	+0.50	+5.8	4.513
20	+0.75	+8.7	4.119
21	+1.00	+11.6	3.632
22	+1.25	+14.5	3.094
23	+1.50	+17.4	2.547
24	+1.75	+20.3	2.025
25	+2.00	+23.2	1.556
26	+2.25	+26.1	1.155
27	+2.50	+29.0	0.828
28	+2.75	+31.9	0.574
29	+3.00	+34.8	0.384
30	+3.25	+37.7	0.248
31	+3.50	+40.6	0.155
32	+3.75	+43.5	0.094
33	+4.00	+46.4	0.055

Sensitivities on the logit scale are equally spaced. They are converted to equivalent sensitivities, in dB, by division with the L_E regression coefficient (0.08617).

5.1 INM Grid Points are the Primary Division of Exposed Population

Division of the total population around an airport into INM grid points requires that each grid-point's population receive nearly equal exposure to aircraft sound levels. Hence, grid points need to be sufficiently dense, especially at locations close to the airport—so that aircraft sound levels are nearly identical across the area within which that grid-point's population resides. Such a grid-point approach to sleep awakening is consistent with standard output available from the

INM. For each grid point, the INM provides the SEL and the associated number of occurrences (often fractions) for each modeled aircraft.

5.2 Differences in Sensitivity to Noise-Induced Awakening are Preserved

At each grid point, the population is divided into 33 sensitivity ranges. The number of people assigned to each level of sensitivity is determined by the Gaussian sensitivity distribution derived in the analysis. Table 6 gives that population distribution by relative sensitivity to noise-induced awakening—a large sensitivity range,

Table 7—Periods of night and corresponding potential to awaken.

Time-interval index (<i>i</i> in Eqn. (5))	Time interval	Increasing potential to awaken during the night (dB)
1	22:00–23:00	+1.4
2	23:00–24:00	+4.2
3	00:00–01:00	+7.0
4	01:00–02:00	+9.8
5	02:00–03:00	+12.6
6	03:00–04:00	+15.4
7	04:00–05:00	+18.2
8	05:00–06:00	+21.0
9	06:00–07:00	+23.8

Increasing potential to awaken is computed as $(0.00402/0.08617)T_{retire}$, where T_{retire} is in minutes to the mid-point of the time interval.

indeed. The number or percent likely to be awakened is computed separately for each population group.

5.3 Awakening Depends upon Time of Night

The analysis showed that awakening is more likely as time passes since retiring. Hence, during application the night is divided into nine hours, for which Table 7 gives increasing potentials to awaken. Division of the night into time intervals means that INM output needs to be computed for the same intervals. Such division of output is most easily managed by making separate runs of the INM. This division also means that an estimate of actual departure/arrival times is required as INM input.

Per Table 7, for modeling purposes we assume that everyone retires at 22:00 (10 p. m.)—for lack of actual retiring times and to be consistent with the FAA definition of nighttime (22:00 to 07:00). Unfortunately, the actual retiring times for sleep-study subjects are not derivable from the sleep-study databases.

5.4 INM and these Equations Compute Awakenings

The INM, when run for the operations that occur during the desired time interval, will produce detailed-grid output that contains the SEL values and corresponding numbers of events needed for each calculation. In the very limited example of Table 8, there are

three 747400 departures from Runway 7L, each producing a slightly different SEL (METRIC_ONE) at this particular grid point.

Use of INM’s detailed-grid output requires some care, however. Depending on the number of flight tracks and operations modeled, the detailed-grid output may contain hundreds of thousands of records. For this reason, the INM run may need to be divided into several parts.

Additionally, an overall simplification is possible by deleting from the grid output any SELs below a pre-determined threshold. Based on thresholds below which virtually no behavioral awakening occurs, it seems reasonable to limit sleep-awakening calculations to outdoor SELs of 65 dB or higher. This limit is reasonable from several perspectives. Assuming an outdoor-to-indoor noise level reduction of at least 15 dB, interior levels would be thereby limited to SELs of 50 dB or greater. This matches the data sets used for the present analysis, in which all measured interior levels exceeded 50 dB SEL. If behavioral awakenings occur below this indoor SEL value, that fact is not derivable from these data.

For comparison, Ollerhead et al. found that sleep “arousal” rates did not exceed non-aircraft noise-event arousal rates until outdoor SELs exceeded about 80 dB.²⁶ Passchier-Vermeer et al. found no increased “motility” above that of non-noise epochs at levels below 40 dB SEL indoors.²⁷ Basner et al. found no increase in motility, or changes to sleep-stage 1, beyond that occurring during the non-noise epochs, until aircraft noise maximums exceeded 35 dBA indoors (approximately an SEL of 43 dB to 45 dB).²⁸ It should be noted that increase motility and changes in sleep stage would be expected to occur at levels below those that produce behavioral awakenings, the subject of the method derived here.

5.5 Awakening Results Permit Comparison of Alternatives

Neither past research nor this current study has definitively identified adverse health effects that may be caused by noise-induced awakenings. Nevertheless, it is possible to use the awakening results produced by this method to communicate with the public and decision-makers and to compare alternative airport

Table 8—Example of INM detailed-grid output.

METRIC-ID	ACFT-ID	OP-TYPE	RWY-ID	OPS-EQUIV	METRIC-ONE
SEL	747400	D	07L	1.0000	59.1
SEL	747400	D	07L	1.0000	59.0
SEL	747400	D	07L	1.0000	58.8

Table 9—Approximate equivalencies among the explanatory variables in Regression 3.

Explanatory variable	Value of explanatory variable	Approximate aircraft-SEL equivalency
Time since retiring	Baseline:	Baseline SEL
	One hour after retiring	2.8 dB louder than baseline
	Four hours after retiring	11.2 dB louder than baseline
	Eight hours after retiring	22.4 dB louder than baseline
Subject awakening sensitivity	90 th percentile person	20 dB louder than baseline
	75 th percentile person	10 dB louder than baseline
	Baseline:	Baseline SEL
	50 th percentile person	
	25 th percentile person	10 dB quieter than baseline
	10 th percentile person	20 dB quieter than baseline

actions. Any changes in nighttime operations may be compared and rank-ordered by the results of this method. Additionally, use of contours can identify which communities around the airports are likely to experience the greatest changes in “people awakened.” Such information provides opportunities to assess the value of additional modifications to nighttime operations—to limit increases in number of people awakened, or even to decrease those numbers.

6 SUMMARY AND DISCUSSION OF RESULTS

The alternative analysis of this paper provides a revised dose-response relation for single-aircraft awakenings, along with its extension to multiple aircraft during the night. This relation predicts awakenings as a function of aircraft indoor SELs, of the timings of aircraft events during the night, and of individual awakening sensitivity. For airport application, SELs and timings are available from INM input and output, by grid point, while individual awakening sensitivity is included through intermediate computations for 33 separate population subgroups.

6.1 Inclusion of Aircraft Timings Allows Computation of Timing-Related Noise Abatement

By including aircraft timings in the dose-response relation, noise abatement can be computed when aircraft are flown *earlier* in the nighttime. For example, using the T_{retire} and the L_E regression coefficients of Eqn. (5), each hour that a flight is moved earlier during the night is equivalent to a noise reduction of 2.8 dB (see Table 9 for the effects of other times, plus the effects of subject awakening sensitivity). Hence, significant reduction of awakenings may be possible by

re-scheduling aircraft to fly earlier in the night. Note that “slow-wave sleep” occurs more in the first part of the night, thereby weakening the chance to awaken early in the night, compared to later in the night.²⁹

6.2 Ignoring Influential Mediating Variables Biases the SEL Dependence Downward

Regression 3 (in Table 3) included noise-event timing and subject awakening sensitivity, while Regression 1 ignored those two mediating variables. Their omission thereby biased the L_E coefficient downward by a factor of 2.2 (0.08617 divided by 0.03888).

Figure 9 illustrates that bias, as follows. The figure contains five solid curves from Regression 3, for various values of noise-event timing and subject awakening sensitivity. As shown, both the timing and the sensitivity coefficients in the regression displace the logistic curve sideways (as in all logistic regres-

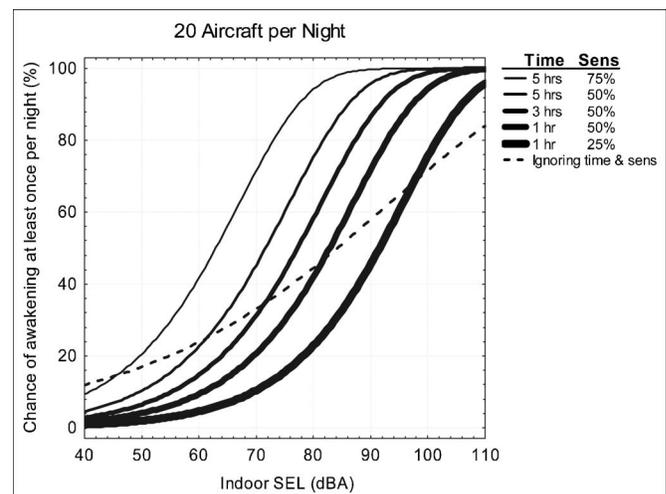


Fig. 9—Effect of ignoring influential mediating variables: time and sensitivity.

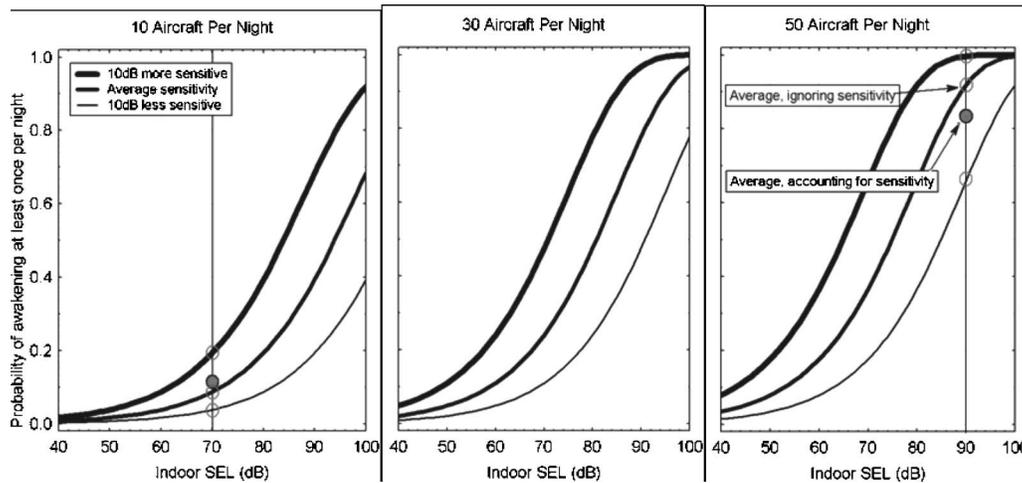


Fig. 10—Illustration of “effective” subject awakening sensitivity.

sions), thereby changing the chance of awakening upwards or downwards at any given indoor SEL. The figure also contains the single dashed line from Regression 1, independent of event timing and sensitivity. The lower slope of Regression 1 is needed for it to “cover” the full data field, horizontally, which Regression 3 covers with multiple curves.^{30,31}

Of most importance, Regression 3’s higher slope is more “relevant” than the lower slope of Regression 1. Everyone who lives near an airport has just one specific sensitivity, and each aircraft during that person’s night has just one specific timing. So each person-aircraft combination is represented by *just one* of these higher-sloped dose-response lines. Any noise reduction for that aircraft will benefit that person *according to the slope of that specific line*. For the central Regression-3 line in the figure, an SEL reduction from 82 to 72 dB would decrease the chance of awakening by 29 percentage points, from 65 to 36 percent. In contrast, Regression 1 indicates a reduction of only 12 percentage points, from 47 to 35 percent. For that person and that aircraft, Regression 1 “washes out” more than half of the benefit of noise reduction.³²

In brief, when influential mediating variables are omitted, the resulting lower slope for aircraft SEL does not mean that aircraft noise is only a “weak determinant” of sleep awakening. Instead, it means that the full SEL effect still lies hidden within the data scatter—because SEL is being asked to substitute for the ignored variables in explaining the data scatter.

6.3 Sensitivity Subgroups Eliminate Prediction Bias

Intermediate computation of awakenings, by sensitivity subgroup, predicts a more accurate (unbiased) total of awakenings than does only one prediction with “average sensitivity.” Figure 10 illustrates this fact,

with a simplified example that compares a computation *with* sensitivity distribution to a computation that *ignores* that distribution.

Three dose-response curves appear in each frame of the figure: (1) a central curve for a person with average sensitivity, (2) a curve 10 decibels to its left, for a person 10-decibels *more* sensitive, and (3) a comparable curve to its right, for a person 10 decibels *less* sensitive. Note that this horizontal displacement is symmetrical about the center curve—conforming to the sensitivity symmetry of Fig. 5 and Table 6, above.

Due to their horizontal displacement relative to one another for any particular SEL, the three curves predict different awakenings, vertically. But the vertical differences are *not* symmetrical about the center curve, because the curves are not straight lines. This non-linear vertical behavior is crucial to their awakening implications. The upper curve represents (in a one-curve, very simplified manner) all the people who are more sensitive to noise awakening than average. The bottom curve similarly represents the less-sensitive people. As shown by the intersecting vertical line in the right frame (50 aircraft per night), when these two curves compute awakening for a 90-decibel aircraft, they compute 100% of the more-sensitive people and 66% percent of the less sensitive people—which averages to 83% of everyone (the shaded circle in the figure). This value of 83%, which accounts for sensitivity (in this very simplified manner), is significantly less than 92% from the central curve, which ignores sensitivity. At least in this SEL region, for 50 aircraft per night, “average” sensitivity significantly over-estimates awakening.³³

The opposite is true in the left frame of the figure (10 aircraft per night), where the average of high and low-sensitivity curves, at SEL equals 70 dB, yields 12% awakenings (accounting for sensitivity), whereas

the average curve (ignoring sensitivity) yields a *lower* value—that is, 9%.

In short, ignoring sensitivity produces biased results. As this figure illustrates, when the computation is done separately by sensitivity subgroups (see Eqn. (5), above), the results differ from a computation that ignores sensitivity distribution (an average-person computation). The average-person bias results from an improper use of average input (average person) in a computation that is not linear. Grid-point computations around an urban airport have shown average-person biases that are larger than 10 decibels.

6.4 Uncertainty Analysis and Validation Reduce Doubt about Analysis Results

Figures 6 and 7 include 95-percent confidence bounds. Those bounds show clearly that uncertainty is very small with many aircraft per night, averaged over many people (many sleep sensitivities)—even for airport applications where SEL input and timings are not precisely known (see Sec. 7.3).

Validation with the independent “validation” data showed a relatively small potential bias in the regression results. It also showed the validity of the uncertainty analysis for independent test subjects.³⁴

Further, the results of this paper might well apply to other airports—based upon the strong consistency of noise-induced awakenings among various studies in the past, especially among those studies that used similar methods and metrics of awakening.^{3,4,6,7,11,26–28} For awakening, consistency may result from physiological consistency among humans. In contrast for annoyance, poor consistency across air facilities and communities may result from highly variable attitudes of airport neighbors towards specific airports, as well as a host of other non-physiological factors.

To determine the believability of computed results, both the Environmental Protection Agency (EPA) and the Congressional Office of Management and Budget (OMB) strongly recommend uncertainty analysis as an integral part of any environmental computation.^{35,36}

6.5 Assumption of Aircraft Independence Appears Sufficient

Equation (3) assumes that successive chances of “sleeping through” aircraft during the night are mutually independent—that is, whether a person sleeps through one aircraft is completely unrelated to whether that person slept through the previous aircraft. It is possible that this noise-event independence is not true. Depending on the timing distribution of events within the night—and particularly the time lags between them—reactions to events might be quite dependent on

each other. This can only be tested by future studies that model the impact of each noise event on subject reaction to events that follow it.

For example, people are in different “sleep stages” during different portions of the night. Awakening may be more likely in some of these sleep stages than in others, which will weaken the independence of successive chances of awakening. Moreover, sleep stages may be affected by noise-induced awakening or by spontaneous awakenings. As a result, awakening may be more likely shortly after a prior noise event—in particular, one that did not succeed in awakening but did succeed in bringing the sleeper to a higher sleep stage—especially when noise events follow one another closely in time.

The mathematics in this paper assumes that any such departures from noise-event independence will average out over the night—some sleep stages increasing, and some decreasing the chance of sleeping through each aircraft event. This paper’s validation with the independent validation data set, including validation of computed uncertainties, tends to confirm that the assumption of noise-event independence is sufficient when the timing distribution of events is the same as with the regression data. Whether this is true for other timing sequences calls for further studies and sophisticated methods. In all, this paper’s assumption of independence is not fully persuasive. Two recent papers discuss this independence issue.^{37,38}

6.6 Analysis Method Likely Overestimates Awakenings

When applied to an airport scenario, Eqn. (5) likely overestimates awakenings, because the data that underlie that equation excluded the 20 percent of sleep subjects who were never awakened by aircraft events.²³ The amount of overestimate is not possible to determine from the data, however. At one extreme, if that entire 20 percent consisted of truly “imperturbable” people, then only the remaining 80 percent are susceptible to awakening by aircraft (at least for the sound levels encountered here). For this reason, the computed percentage awakenings from Eqn. (5) should only be applied to that 80 percent of the population, thereby reducing awakenings to 80 percent of the value from Eqn. (5).

However, as discussed in Sec. 2.2 above, some of that 20 percent were simply not tested long enough to be awakened. For example, LAX subject 34 was tested by only six aircraft, with a maximum indoor SEL of only 79 dB. From such a sparse noise exposure, we certainly cannot conclude that this subject was “imperturbable.” So at the other extreme, if that entire 20 percent consisted of subjects *not tested enough to tell*,

then Eqn. (5) would produce no over-estimate at all. Those subjects would be legitimate to exclude, for lack of adequate testing.

The truth is somewhere in between these two extremes—though it can only be speculated from the exposure data in Figs. 2–4, above. Rather than speculate, we accept some intermediate amount of over-prediction—thereby over-estimating noise impact during application.

7 APPENDIX A: DERIVATION OF DOSE-RESPONSE UNCERTAINTIES

In the main body of this paper, Eqn. (2) contains the dose-response relation for single aircraft, while Table 3 contains that relation's regression coefficients and their standard uncertainties. These standard uncertainties show whether each individual regression coefficient is well determined. More important than individual-coefficient uncertainty, however, is the total uncertainty for the full dose-response curve (see the 95% confidence bounds in Fig. 6, for example). This appendix derives total uncertainties for all dose-response relations above—single aircraft and multiple aircraft, for known subject sensitivities (the regression data) as well as uncertain subject sensitivities (validation data and later application at airports).

The derivations of this appendix are based upon methods in the ISO *Guide to the Expression of Uncertainty in Measurement*.³⁹ The ISO guide specifies how to “propagate” uncertainty through any algebraic function that computes an output y from a series of statistically uncertain input, x_i . The equations for p_{awake} in this paper are such algebraic functions.

Propagation of statistical uncertainty through any algebraic function, from input to output, requires use of the following:

$$\begin{aligned} \text{When } y &= f(x_i) \quad \text{for } x_i = 1, \dots, N \\ \text{then } (u_y)^2 &= \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 (u_{x_i})^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial x_i} \right) \\ &\quad \times \left(\frac{\partial f}{\partial x_j} \right) r_{x_i x_j} u_{x_i} u_{x_j}. \end{aligned} \quad (7)$$

In this ISO equation, all partial derivatives are determined analytically from $y=f(x_i)$, while logistic regression determines u_{x_i} , the standard uncertainty of each input, and $r_{x_i x_j}$, the correlation between each pair of these uncertainties. On the left side of the equation, u_y is the resulting standard uncertainty of y .

Whenever the underlying distribution of y is Gaussian (normal), then u_y from Eqn. (7) is multiplied by a “coverage factor”—commonly 1.96—to convert it to a 95% coverage interval. In this study, however, the

underlying distribution of y is not Gaussian. Instead, y is a probability and therefore has a binomial distribution, confined between zero and unity. To similarly confine y 's 95% coverage interval requires a more-complicated conversion than this simple multiplication by 1.96. The required conversion uses the following set of equations:^{40,41}

$$\begin{aligned} n &\equiv \frac{y(1-y)}{u_y^2} \\ a &\equiv n + (1.96)^2 \\ b &\equiv -(2yn + (1.96)^2) \\ c &\equiv ny^2 \\ y_{upper} &= \frac{-b + \sqrt{b^2 - 4ac}}{2a} \\ y_{lower} &= \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \end{aligned} \quad (8)$$

In the sections that follow, Eqns. (7) and (8) derive uncertainties for the various dose-response relations in the main text of this paper.

7.1 Regression Data: Uncertainty Only in the Regression Coefficients

When the dose-response relations of this paper are used with the regression data, input uncertainty resides only in the regression coefficients—uncertain because they derive from sampled data. Of interest is propagation of uncertainty through Eqns. (2) and (4)—resulting in the 95% confidence bounds of Fig. 6 for single aircraft and Fig. 7 for multiple aircraft. Table 10 collects all the necessary input uncertainties from the logistic regression that produced Eqn. (2). The diagonal entries in this variance-covariance matrix are values of $(u_{x_i})^2$ —that is, the squares of each coefficient's standard uncertainty. The off-diagonal entries are values of $r_{x_i x_j} u_{x_i} u_{x_j}$.

7.1.1 Single aircraft

For single aircraft, Eqn. (2) is repeated here:

$$\begin{aligned} P_{awake, single} &= \frac{1}{1 + e^{-Z}} \\ Z &= \beta_0 + \beta_L L_E + \beta_T T_{retire} + \sum_{s=1}^{32} \beta_s d_s \end{aligned}$$

Table 10—Variance-covariance matrix for the logistic regression of Eqn. (2).

		Intercept	Aircraft indoor SEL	Time since retiring	Subject sensitivities (notes 1 through 4)		
		β_0	β_L	β_T	β_{s1}	...	β_{s32}
Intercept	β_0	+1.07405831	-0.01416296	-0.00011744	0	0	0
Aircraft indoor SEL	β_L	-0.01416296	+0.00019469	+0.00000022	0	0	0
Time since retiring	β_T	-0.00011744	+0.00000022	+0.00000035	0	0	0
Subject sensitivities	β_{s1}	0	0	0	Note 5	0	0
	...	0	0	0	0	Note 5	0
	β_{s32}	0	0	0	0	0	Note 5

Note 1. The 32 correlations between β_s and β_0 range between -0.04 and +0.18—adequately close to zero to ignore.

Note 2. The 32 correlations between β_s and β_L range between -0.002 and +0.0005—adequately close to zero to ignore.

Note 3. The 32 correlations between β_s and β_T range between -0.00002 and +0.000005—adequately close to zero to ignore.

Note 4. The 528 correlations among all the 32 β_s range between -0.09 and +0.03—adequately close to zero to ignore.

Note 5. The square of the standard uncertainties of all 32 β_s which closely fit the following expression: $(0.6013 - 0.0554\beta_s + 0.0202\beta_s^2)^2$.

$$Z = -10.723 + 0.08617 L_E + 0.00402 T_{retire} + \sum_{s=1}^{32} \beta_s d_s, \quad (9)$$

where β_s is each subject's individual sensitivity to noise. From Eqn. (7), the standard uncertainty of $p_{awake, single}$ then computes as follows:

$$\begin{aligned} (u_{p_{awake, single}})^2 = & \left\{ \frac{e^Z}{(1 + e^Z)^2} \right\} \left\{ (u_{\beta_0})^2 + L_E (u_{\beta_L})^2 \right. \\ & + T_{retire} (u_{\beta_T})^2 + (u_{\beta_s})^2 \\ & + 2L_E (r_{\beta_0 \beta_L} u_{\beta_0} u_{\beta_L}) \\ & + 2T_{retire} (r_{\beta_0 \beta_T} u_{\beta_0} u_{\beta_T}) \\ & \left. + 2L_E T_{retire} (r_{\beta_L \beta_T} u_{\beta_L} u_{\beta_T}) \right\}, \quad (10) \end{aligned}$$

where the various values of u and r appear in Table 10, above. Equation (8) must then be used to expand $u_{p_{awake, single}}$ to the upper and lower 95% confidence bounds in Fig. 6, above. Note that $u_{p_{awake, single}}$ and its expansion depend upon the parameters L_E , T_{retire} and subject awakening sensitivity β_s , as governed by their presence in Eqn. (10). Also note that each graph in the figure assumes one specific subject awakening sensitivity, β_s , per the labels to the left.

7.1.2 Multiple aircraft

For multiple aircraft during the night, Eqn. (2) is supplemented with Eqn. (4), repeated here:

$$p_{awake, multiple} = 1 - \prod_{a=1}^N (1 - p_{awake, single})_a. \quad (11)$$

The uncertainty in $p_{awake, single}$ propagates through this equation, to determine the uncertainty in $p_{awake, multiple}$ —assuming awakenings from all aircraft are independent—as follows:

$$\begin{aligned} (u_{p_{awake, multiple}})^2 = & \sum_{a=1}^N \left(\frac{\partial p_{awake, multiple}}{\partial (p_{awake, single})_a} \right)^2 (u_{p_{awake, single}})_a^2 \\ & + 0 \\ = & \sum_{a=1}^N \left(\frac{1 - p_{awake, multiple}}{(1 - p_{awake, single})_a} \right)^2 (u_{p_{awake, single}})_a^2 \\ = & (1 - p_{awake, multiple}) \\ & \times \sum_{a=1}^N \left(\frac{(u_{p_{awake, single}})_a}{(1 - p_{awake, single})_a} \right)^2. \quad (12) \end{aligned}$$

As a special case, when all N aircraft have same probability and the same uncertainty:

$$\begin{aligned} (u_{p_{awake, multiple}})^2 = & N(1 - p_{awake, multiple}) \\ & \times \left(\frac{(u_{p_{awake, single}})_a}{(1 - p_{awake, single})_a} \right)^2. \quad (13) \end{aligned}$$

Substitution of Eqn. (10) into either Eqn. (12) or (13) then determines the standard uncertainty the multiple-aircraft $p_{awake, multiple}$. Further substitution into Eqn. (8) determines the 95% percent uncertainty bounds for multiple aircraft for the regression set of data (not graphed above).

7.2 Validation Data: Uncertainty also in Subject Sensitivities

For the regression subset of data, all explanatory variables were measured directly and the data were analyzed to determine individual subject sensitivities. In contrast, for the validation subset of data, individual subject sensitivities are not known, because these data are never regressed to determine them. Hence, when the regression equation is applied to the validation subset of data, the chance of awakening must be computed with the average subject awakening sensitivity (-0.0379).

As is apparent, this most-likely sensitivity value will not be particularly good for many of the individual validation test subjects. In other words, if the regression equation is used to compute *individual* chances of awakening, person by person, it will be highly uncertain for the validation subjects. Mathematically, that is because the most-likely sensitivity has a large standard uncertainty—numerically equal to the standard deviation of the histogram of sensitivities from Fig. 5 (1.3469). This large standard uncertainty results in a much poorer estimate of individual validation subjects than for the regression subjects, who have known sensitivities.

When the chance of awakening is averaged over many validation test subjects, however, uncertainty is reduced due to that averaging process. Mathematically:

$$(u_{p_{\text{awake,multiple}}})_{G \text{ subjects}} = \frac{(u_{p_{\text{awake,multiple}}})_{1 \text{ subject}}}{\sqrt{G}} \quad (14)$$

With this additional factor, \sqrt{G} , supplementing Eqn. (13), the composite set of equations then computes the 95% confidence bounds in Fig. 7, above—as well as the coverage bounds in Sec. 4.2. The good agreement in that section confirms the correctness and usefulness of these uncertainty computations for independent data.

7.3 Airport Application: Input Uncertainty, Also

When Eqn. (5) is applied to an airport scenario, uncertainty also exists in the equation's input: aircraft indoor SELs and their timings during the night. This paper does not attempt to incorporate these uncertainties into the formal mathematics. Instead, we estimate the general magnitude of these uncertainties and examine graphically their influence on the chances of awakening. In these matters, we rely upon judgment and experience.

7.3.1 Uncertain aircraft indoor SEL

Indoor-SEL uncertainty depends first upon the inherent uncertainties of INM. In our experience, INM predicts DNL usually within ± 4 dB—corresponding to a standard uncertainty of 2 dB. This uncertainty range most closely relates to the *energy-average* SEL at any given location, because the energy-average SEL is what determines DNL. In contrast, sleep awakening is not a function of energy-average SEL, but depends upon SEL through the more-complicated Eqn. (2), above. Nevertheless, we take 2 dB to be the standard uncertainty of INM outdoor SELs, as well.

In addition, indoor-SEL uncertainty depends upon the variability from house-to-house of the outdoor-to-indoor sound insulation. First, it is necessary to measure the average value of this around the airport of interest, to account for gross regional differences from airport to airport. Once this average value is known, however, that average value is used to convert *all* outdoor SELs to indoor values. House by house, this conversion is uncertain, of course—perhaps by another ± 4 dB. However, desired at any grid point is the average percentage of people awakened, so this uncertain noise insulation is averaged in the calculations over all the homes at that grid point. This averaging process reduces the 2-dB standard uncertainty to a trivial amount—say, $2 \text{ dB} / \sqrt{1000} = 0.06 \text{ dB}$ or so for 1000 homes at the grid point.

7.3.2 Uncertain aircraft timing and its SEL equivalent

In addition, uncertain aircraft timing also contributes to uncertainty in the chances of awakening. Timing uncertainty depends inherently upon the quality of INM input—the Official Airline Guide for scheduled aircraft, or projections of general aviation or military activity at the air facility. In addition, timing uncertainty derives from the 1-hour time windows used to compute awakening, so that each timing is good within ± 30 minutes, at best. A total range of ± 30 minutes corresponds to a standard uncertainty of $\sqrt{(60 \text{ min})^2 / 12} \cong 17$ minutes.

Per the regression coefficients in Eqn. (2), this 17 minutes converts to an equivalent SEL uncertainty of $(17)(0.00402/0.08617) \cong 0.8 \text{ dB}$. However, the cumulative effect at any grid point is the average effect of all aircraft during the night. Some of these will pass over earlier, and some will pass over later, than assumed in the INM input. To a first approximation, the combined standard uncertainty therefore corresponds to the uncertainty of the average passby time—say, $0.8 \text{ dB} / \sqrt{10} = 0.3 \text{ dB}$ for ten aircraft per average night.

7.3.3 Total effect

These standard input uncertainties total to $\sqrt{2^2+0.06^2+0.3^2}=2.0$ dB, for a 95-percent confidence interval of ± 4 dB. Figure 7 can be used to determine how important this amount of input uncertainty is to the computation of awakening. The most relevant graph in that figure is in the right column, third from the top (10 aircraft per night and 100 people averaged at the grid point). In that graph, ± 4 dB converts to the following uncertainties in awakening: ± 2 percentage points at an indoor SEL of 55 dB, and ± 8 percentage points at an indoor SEL of 75 dB, and ± 6 percentage points at an indoor SEL of 95 dB.

At 55 and 95 dB, this input uncertainty is about the same at the regression-coefficient uncertainty shown in the graph, while at 75 dB it is two-to-three times as large. At 75 dB, the probability of awakening equals a baseline of 0.30, which might increase to 0.38 due to input uncertainty. In terms of “odds” rather than probabilities, the baseline of 30/70 would thereby increase to 38/62—an odds multiplier of $(38/62) \div (30/70)$, which equals only 1.4.

In all, once averaged over this many aircraft per night and this many people, input uncertainties do not significantly compromise the ability of the dose-response to usefully compute the percentage of people likely to be awakened at each grid point.

8 REFERENCES

1. K. S. Pearsons et al., “Analyses of the predictability of noise-induced sleep disturbance,” Report HSD-TR-89-029, Noise and Sonic Boom Impact Technology, Human Systems Division, Air Force Systems Command, Brooks Air Force Base, Texas (1989).
2. Federal Interagency Committee on Noise (FICON), “Federal agency review of selected airport noise analysis issues” (1992).
3. S. Fidell et al., “Noise-induced sleep disturbance in residential settings,” Report AL/OE-TR-1994-0131, Occupational & Environmental Health Division, Armstrong Laboratory, Wright Patterson Air Force Base, Ohio (1994).
4. S. Fidell et al., “Field study of noise-induced sleep disturbance,” *J. Acoust. Soc. Am.* 98(2) (1995).
5. Federal Interagency Committee on Aviation Noise (FICAN), “Effects of aviation noise on awakenings from sleep,” www.fican.org/pages/sleepdst.html (1997).
6. S. Fidell et al., “Effects on sleep disturbance of changes in aircraft noise near three airports,” *J. Acoust. Soc. Am.* 107(5) (2000).
7. S. Fidell et al., “Noise-induced sleep disturbance in residences near two civil airports,” NASA Contractor Report 198252, Contract NAS1-200101 (1995).
8. American National Standards Institute (ANSI), “Quantities and procedures for description and measurement of environmental sound - Part 6: Methods for estimation of awakenings associated with aircraft noise events heard in homes,” ANSI S12.9-2000/Part 6 (2000).
9. Gezondheidsraad, Health Council of the Netherlands, “The influence of night-time noise on sleep and health,” Report U 1007/WP/718-K, The Hague (2004).
10. European Parliament and Council Directive 2002/49/EC, Annex I, OJ: 189, 18.7.2002, p. 12 (2002).
11. H. M. E. Miedema, W. Passchier-Vermeer and H. Vos, “Elements for a position paper on night-time transportation noise and sleep disturbance,” TNO Inro report 2002-59 (2003).
12. Data from the original studies were provided to the current authors by Larry Finegold and Robert Lee—from records of the U.S. Air Force, sponsors of the original studies.
13. The expectation was for approximately the same number of aircraft-noise events in each of the two data sets, because the number of subjects was nearly the same. However, the random split resulted in one set having nearly double the number of aircraft-noise events as the other. That larger set was chosen for “regression,” with the smaller set retained for “validation.”
14. The GLZ module within Statistica 7.1 (StatSoft, Inc., Tulsa, Oklahoma, www.statsoft.com) was used for all logistic regressions. Interaction terms were not investigated.
15. Other functional forms from the literature are simpler than this logistic form—for example, a linear regression. However, because the underlying data distribution is binomial, rather than Gaussian, least-squares regression is not proper for these data—certainly not to determine uncertainty bounds. The binomial requirement therefore undercuts the otherwise simple analysis for linear regression. In addition, it is highly desirable to have a functional form that is asymptotic to zero and unity, because it makes more realistic the multiple-aircraft joint probabilities, which can easily approach unity. For all these reasons, a functional form that transitions smoothly between zero and unity is highly desired. The logistic function is the simplest and most widely used of such transitional functions.
16. Indoor ambient sound level was dropped for two additional reasons: (1) this parameter was differently defined/measured at the three air facilities, and (2) values for it were missing from the database for many aircraft noise events.
17. Although this regression equation contains 32 sensitivity regression coefficients, these all collapse into just two parameters—their mean and standard deviation—when regression results are used. For that reason, the regression coefficients are not susceptible to complaints of “over-fitting.” In essence, they use up just two degrees of freedom from a data set with more than 7,000 data points.
18. The histogram and its Gaussian fit also contain the “reference” subject’s sensitivity, equal to +0.00.
19. For this observed distribution of sensitivities, the Shapiro-Wilk’s W statistic equals 0.97, with a p-value of 0.50. That p-value shows no cause to reject that the observed distribution is Gaussian, and so it was accepted as Gaussian.
20. Note that duration of residence would be difficult to determine during application of these results at an airport under study. In contrast, this variable might truly contribute to each subject’s aircraft-noise sensitivity. By dropping it, residence duration is thereby subsumed into subject awakening sensitivity, which can be taken into account (through the histogram of Fig. 5) during airport application of these results.
21. M. Brink et al., “Effects of early morning aircraft overflights on sleep and implications for policy making,” EuroNoise 2006, Tampere, Finland (2006).
22. This equation is strictly true when the probabilities of awakening are independent, from one aircraft to the next. See discussion in Sec. 6.
23. In the same studies whose findings are re-analyzed here, aircraft-induced awakenings occurred at approximately half the rate of spontaneous (non-aircraft) awakenings. Although not founded in research, the authors find it reasonable that these more-numerous spontaneous awakenings are fundamentally less disruptive to the sleeper, since spontaneous awakenings are “volunteered,” rather than “forced.”
24. The authors do not contend that “awakening at least once per night” is an all-encompassing sleep-disturbance metric—only that it is one possible metric that is easily understood by airport neighbors.
25. It is pure happenstance that the number of population groups

- (33) equals the number of sleep subjects in the “regression” data.
26. J. B. Ollerhead et al., “Report of a field study of aircraft noise and sleep disturbance,” Department of Transport, Civil Aviation Authority, U.K. (1992). Arousal was determined as “any EEG-determined arousal to wakefulness” of 15 seconds or more, or movement time lasting 10 seconds or more.
 27. W. Passchier-Vermeer et al., “Sleep disturbance and aircraft noise exposure: Exposure-effect relationships,” Report TNO 2002.027, Division Public Health, The Netherlands (2002).
 28. M. Basner et al., “Effects of nocturnal aircraft noise - Volume 1: Executive summary,” Report 2004-07/E, German Aerospace Center (DLR), Institute of Aerospace Medicine, Flightphysiology Department, Cologne, Germany (2004).
 29. In addition, re-scheduling aircraft to fly earlier in the night is likely beneficial to the “recovery function” of sleep. In particular, impeded slow-wave sleep in the first part of the night can more easily be made up in the second part, rather than the other way round. Moreover, when awakened in the early morning, people will have more trouble falling asleep again, so will more likely remember their awakening and potentially convert that memory to “annoyance.”
 30. Regression 1 numerically equals Regression 3 for the average indoor SEL (72 dB), the average event timing (slightly under three hours), and the average sensitivity (50-percentile person).
 31. Each person’s dose-response curve has a particular left-right shift, depending upon that person’s sleep stage at the moment, emotional state, sub-conscious interpretation of the potential meaning of the noise, total time in bed, fatigue level, use of sleeping medication, and a host of other factors that pertain to that person, at that moment. All these matters affect the horizontal location of that person’s dose-response curve at the moment—that is, that person’s sensitivity to noise-induced awakening.
 32. Note that the sleep-study data were obtained in neighborhoods with “stable” aircraft exposure. For this reason, if aircraft noise is suddenly decreased by some action of the airport, the short-term reduction in awakening may not mirror the slope of this dose-response curve. Transitional instabilities may intervene. We expect, however, that after the neighborhood returns to aircraft-noise stability (perhaps several years?), that the SEL slope will approximate the accrued long-term reduction in aircraft awakenings. In addition, remember that the dose-response curve does not predict individual reactions, person by person, but only the distribution of reactions over large totals of people.
 33. Note that the resulting 83% is not what Eqn. (5) would predict. Instead, it is the simple average of the values from the top and bottom curves, which represent a very simplified awakening-sensitivity distribution in this explanatory figure.
 34. One reviewer stated that the sampling of the original sleep studies was not random, but was “purposive,” instead. In response, purposive samples are often used to represent larger populations. Because the “validation” data so clearly validated results from the “regression data,” we contend that the purposive nature of subject selection did not introduce significant systematic subject bias. To the extent that it did, then this analysis—as well as the analysis of the original sleep studies—may not be representative. In addition, this study used what data were available to the authors. Until the methods of this paper can be applied to additional data, the authors suggest that there is considerable value in not only the method but in the findings.
 35. A. Krupnik et al., “Not a sure thing: Making regulatory choices under uncertainty,” Resources for the Future, Inc., www.rff.org (under Contract 68-W-02-046 to the U.S. Environmental Protection Agency), Washington DC (2006).
 36. Office of Management and Budget (OMB), “Regulatory analysis as required under Section 6(a)(3)(c) of Executive Order 12866, ‘Regulatory Planning and Review,’ the Regulatory Right-to-Know Act, and a variety of related authorities,” Circular A-4, Washington DC (2003).
 37. M. Basner, “Markov state transition models for the prediction of changes in sleep structure induced by aircraft noise,” Research report 2006-07, German center for air and space travel (Deutsches Zentrum für Luft- und Raumfahrt, DLR), Institute of Aerospace Medicine, Cologne. Available at www.eid.dlr.de/me/Institut/Abteilungen/Flugphysiologie/Fluglaerm/Lit_Laermw_Reports/FB_2006-07_markov.pdf.
 38. M. Brink et al., “Bestimmung lärminduzierter Aufwachwahrscheinlichkeiten in der Nachtlärmwirkungsforschung und Anwendung entsprechender Wirkungsmodelle für Prognosezwecke,” Report to the Workshop ‘Aufwachwahrscheinlichkeit,’ Swiss Federal Institute of Technology Zurich, 21–22 April 2006 (English translation in preparation). Available at http://e-collection.ethbib.ethz.ch/ecol-pool/bericht/bericht_485.pdf.
 39. International Standards Organization (ISO), “Guide to the expression of uncertainty in measurement,” Section 5.2 for correlated input (corrected and reprinted 1995).
 40. I. S. Guttman, W. Wilks and J. S. Hunter, *Introductory engineering statistics*: Third edition, Section 9.8, John Wiley & Sons (1982).
 41. In this set of equations, n is not the total number of samples underlying the regression. Instead, it is an “effective” number of samples that underlies the regression in each explanatory-variable region. In turn, that effective number depends upon u_y in that region, as shown. This definition of n allows it to compensate for binary over-dispersion—a common occurrence.
 42. Computed R^2 values are not relevant single-number assessments of logistic regressions. Relevant single-number assessments—such as the regression’s loglikelihood and Akaike Information Criterion—are not reported here. Instead, the more-informative uncertainty bounds of Figs. 7 and 7 serve to judge the quality of the regression, as a function of specific input values of the independent parameters. As those figures show, regression uncertainty varies greatly over different ranges of those parameters.