Proceedings of The Institute of Acoustics

THE UNITED KINGDOM AIRCRAFT NOISE INDEX STUDY
PART II - STATISTICAL ANALYSIS OF DISTURBANCE AND NOISE EXPOSURE

Peter Brooker and Catriona Richmond

Chief Scientist's Division
Civil Aviation Authority, London, UK

INTRODUCTION

The companion paper [1] discusses the need for a UK Aircraft Noise Index, and the planning, execution and main results of the present study. This paper examines the statistical modelling involved.

To determine an aircraft noise index it is first necessary to model the stimulus-response relationship: the 'stimulus' of exposure to aircraft noise, and the 'response' in terms of annoyance, disturbance, bother, etc expressed by the individuals affected. In general it is to be expected that annoyance etc will increase with exposure, but response is complicated - or 'confounded' - by many other factors, eg levels of exposure to other noise sources and the degree to which individuals are involved economically with the airport.

The stimulus-response model must include confounding factors for the effect of the aircraft noise stimulus to be properly understood. An index based on such a statistical relationship will not include any confounding factors. This is in part in the interests of equity, in part because the inclusion of such factors would cause difficulty in application. The confounding factors could require detailed socio-economic data or forecasts. For example, a factor relating to the economic dependence of a near-airport community on that airport would require detailed forecasts in the case of a 'green-field' airport site. In practice most confounding factors can be put in such a form as to depress adverse response; ie the index will tend to over-estimate the response.

STIMULUS AND RESPONSE

An aircraft noise stimulus can be measured by a variety of commonly accepted noise metrics, which measure noise above different 'cut-off' levels, at different times of day, averaged over different periods of time. 'Noise' here may be the peak level (measured in PNdB), numbers of noise events (or the logarithm of the number), or equivalent aircraft sound energy.

Response is measured using scales derived from the replies to the questions in the social survey [1]. A disturbance reponse on such a scale does not, intrinsically, possess 'well-behaved' qualities. Thus, there is no a priori reason to suppose that someone scoring 4 on a six-point scale is, in some sense, twice as annoyed as someone scoring 2, or that the step between scores of 2 and 3 represents the same degree of difference in response as that between scores of 3 and 4.

There appears to be no wholly objective means of establishing the validity of a given scale, either in terms of its success in measuring an attitude or in the nature of the 'arithmetic' properties given as examples above. However, several scales designed to measure the same attitude, albeit through different questions, should show a high degree of commonality. In particular there should be a strong similarity in the way they rank individuals by the degree of annoyance. (This has been tested using the non-parametric Spearman Rank Correlation coefficient [2]).

Scales of disturbance from aircraft noise may be overt or covert. Overt measures arise from questions relating explicitly to aircraft noise (mostly obviously from 'How annoyed are you by aircraft noise?'), whereas covert measures arise from questions which invite a range of responses including aircraft noise (eg 'What sorts of noises do you hear around here?'). A scale may be constructed directly in the question (eg 'On a scale of one to ten ...'), or by associating a score which matches to responses (eg very much annoyed = 3, moderately annoyed = 2, and so on), or by combining in some way the responses to several questions. A well-known example of this is the Guttman Annoyance Scale (GAS) for aircraft noise disturbance, in which the reaction to interference with five specific activities is combined with a general disturbance measure. Table 1 lists the major scales examined in this study (the complex GAS scoring procedure is given in [3]).

Individual responses tend to vary strongly. To some extent this can be attributed to confounding factors, but the 'unexplainable' natural variations between individuals are a major element. As discussed in [1], the present study uses common noise areas to assess the responses within (quite small) communities. By defining the 'average response' to a question, or by using the percentage of the community responding to a certain degree, a representative estimate of disturbance can be made. The average gives a 'typical' response: a percentile measure may be said to weight the strength of a response. Given the multiple regression techniques used here, percentile measures have the advantage that their sampling properties are better defined than those of averages of scale results.

THE UNITED KINGDOM AIRCRAFT NOISE INDEX STUDY - PART II

MODEL FORMULATION

As noted in [1], the UK has used the Noise and Number Index (NNI) to assess aircraft noise disturbance since the early 1960s. Any new index needs to be an improvement on NNI in its prediction of response. NNI is defined by:

$$NNI = L + 15\log_{10}N - 80$$

where L is the 'log' (strictly energy) average of daytime peak noise levels above 80 PNdB and N is the corresponding number of noise events. A better index would require a change in one or more of the elements in this expression - the trade-off factor of 15, the use of an 80 PNdB cut-off, a log N or N dependence - or the substitution of some L_{eq}-type expression.

The emphasis in [1] is on the relative weighting of L and log N. The study sites were chosen for their statistical efficiency in determining this trade-off, via the use of areas of high noise level/low number and low noise level/high number in preference to areas where L and log N were strongly correlated. The wide range of noise metrics examined as possible variants to the L and N of the NNI is given in [3]. The general form of the 'exposure' combination is (initially) a linear set of noise metrics with arbitrary weightings, ie a form appropriate for multiple regression analysis. These noise metrics included lower cut-offs, L_{eq} measured over different times of the 24 hours, different averaging periods and worst mode [3] measurements.

STATISTICAL MODELLING

The NNI was formulated from a social survey response measure based on GAS - the average GAS score, denoted as AVOGAS here. In the present study a wider range of disturbance measures was analysed. From the full data base comprising the completed questionnaires a compact data base for the 26 areas was constructed. 13 key disturbance questions were used to form 21 aggregated response measures for each of the areas. Each area thus had a set of disturbance measures and associated noise metrics. The response measures were then examined with the aid of the non-parametric correlation coefficient (Spearman), with the intention of reducing the 21 down to a restricted, but representative, set. A set of 4 were finally chosen and are listed in Table 1. AVOGAS is required for reasons of historical continuity; VMANN is analogous to the USA 'highly annoyed' (which equates to 'very much annoyed' - there is a <u>sub</u>-category 'extremely annoyed'); ARCNA has obvious relevance for decision-makers; ARCBOTH is a representative covert disturbance measure. The measures not chosen are eliminated by virtue of similarity to one or more of these four.

In addition to the disturbance measure and noise metrics, six confounding factors were evaluated for each community and included in the compact data base. These were: % of female respondents, % of shift workers, % of respondents with family member involved economically with airport, % with double-glazing, % of non-manual workers, and average length of residence in the area.

No specific variable was included to represent the possibility of a dependence of disturbance on the particular airport. (Thus, for example, it is possible that because of different development histories residents at Aberdeen might, at comparable noise exposure, be less disturbed than Heathrow residents.) Such a dependence was tested by an examination of multiple regression residuals.

The multiple regression analysis package used was the forward inclusion option of the routine REGRESSION from SPSS [2]. The package uses F-tests and tolerance factors to assess variable entry. Full details of the values used in the regression investigation are given in the main study report [3]. In summary, the F-test examines the 'explanatory contribution' of the possible variables to an improved regression fit; the tolerance factor tests variables for correlation with those already in the regression set. The scheme of multiple regression analysis was to start from the variables used in the NNI and then expand systematically the range of noise metrics and confounding variables available for selection. At each stage the degree of fit and structure of residuals was examined. The results for all four disturbance measures were then examined for any consistent pattern, leading to a new model for the next stage.

The following text sketches the procedure detailed in [3]. To begin with the variables used in the NNI were tested against AVOGAS. The trade-off determined between L and log N was 6.4, significantly different at the 5% level from the 15 of the NNI. Next the confounding factors were made available for selection: choosing the best gave a trade-off factor of 6.8, again significantly different from 15; the multiple regression coefficient increasing from 0.82 to 0.87.

Next, the range of possible noise variables was expanded: first to include a representative L_{eq} variable (covering the 24 hours). All variables still referred to the 3 month NNI period. At this stage an examination of residuals indicated some non-linearity against L_{eq}. Dummy 'STEP' variables, zero below given L_{eq} values and unity above, were introduced, to enable a fit to be made to the apparent jump in response appearing in the middle of the exposure range. The STEP variables permit an approximation to the ogive form common in stimulus-response studies.

The set of independent variables was then expanded to include the full set of noise metrics of [3]. All four disturbance measures were now included as dependent variables and the results examined for any consistent patterns. With such a large number of independent variables (64) there is a great danger of overfitting, ie using more variables than are necessary to give a satisfactory fit: this was apparent in one of the regressions (for VMANN) where 6 variables contribute substantially to the final regression, as compared with two or three for the other disturbance measures. A consistent pattern does appear, relying on three variables: one week 24 hour L_{eq} (W1LQ24), the percentage economically involved with the airport, and a STEP variable. The F-test level for the inclusion of other variables is now raised to the ~1% level from the ~10% used in the previous search. No further variable was required: a typical fit is illustrated in Figure 1 which shows AVOGAS, 'corrected' for the airport economic dependence, versus W1LQ24, the step in the L_{eq} variable being at about 57 dBA.

Since L_{eq} is not markedly dissimilar in form to NNI - the former corresponding approximately to an L-N trade-off of 10 - it is of interest to determine how well an L-N combination compares with an L_{eq} fit. It turns out that one week 70 PNdB cut-off variables with a trade-off of 10 (ARCNA) and 8.5 (the other measures) performs about as well as the L_{eq} variable above. The multiple regression coefficients for both sets are around 0.94. (Note that L,N refer to daytime here).

DIURNAL VARIATIONS IN RESPONSE

Although the noise components of the compact data base included L_{eq} variables for different times of the 24 hours (as well, of course, as the whole time period), none of the regressions discussed above were found to include weighted contributions from different times. This is an important issue and, given that statistical testing methods do not generally identify weak effects, further examination was required.

In the social survey the question, 'How annoyed are you by aircraft noise?', was asked with specific reference to three time periods - day (0700 to 1900), evening (1900 to 2300) and the night (2300 to 0700)- all BST. Responses can be divided according to whether the respondent was generally at home during the day or not - Table 2. Several points may be made: (a) It appears to make little difference to the response whether or not the individual was usually exposed to the daytime noise 'climate'; (b) Daytime and evening responses are very similar; night-time disturbance is much lower than either; (c) The response to the same question referring to general experience shows greater annoyance than that for any single time period, indicating perhaps that this general response is cumulative, rather than averaged or focussing on the worst period.

These findings are reinforced by restricting consideration to those areas with small differences between daytime and evening, and evening and night-time exposure (taken here as one week L_{eq}). The pattern remains the same: daytime and evening annoyance are very similar; night-time annoyance much lower. A small evening weighting of 1 or 2 dB might be justified - but a zero weighting is quite consistent given statistical fluctuations. There is no evidence in this study to suggest that any night-time weighting is required. This is in contrast to the commonly-used USA index L_{dn} in which a night weighting of 10 dB is employed. Other negative findings in the study were the lack of need for worst mode effects [3] and an absence of any marked airport dependence.

FURTHER WORK

The analysis described here covers the main features of the results. Further examination of the statistical data published in [3] is certainly possible: a better non-linear fit to the response data is possible; factor analysis of the response measures could produce interesting results.

REFERENCES

[1] P. Brooker and C. Richmond, 'The United Kingdom Aircraft Noise Index Study, Part I - Main Results'. This Journal, preceeding paper.

[2] N.H. Nie et al, 'Statistical Package for the Social Sciences (SPSS)' Second Edition, McGraw-Hill Book Company, (1975).

[3] P. Brooker, J.B. Critchley, D.J. Monkman and C. Richmond, 'United Kingdom Aircraft Noise Index Study: main report' DR Report 8402, Civil Aviation Authority (1985).

Table 1. Community Disturbance Scales

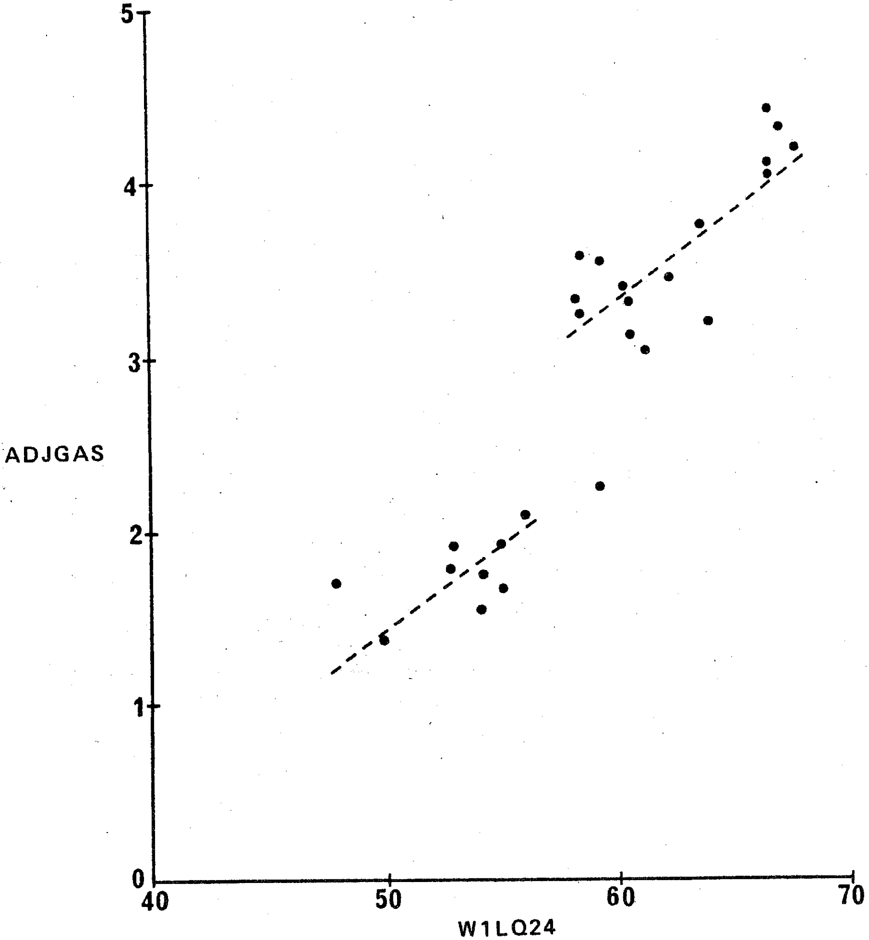
VMANN:	'Please look at this scale and tell me how much the noise of aircraft here bothers or annoys you?' VERY MUCH → VMANN = % VERY MUCH ANNOYED MODERATELY A LITTLE NOT AT ALL
AVOGAS:	The average community score on the GAS scale which is a composite of: 'How much does the noise of aircraft here bother or annoy you?' (see above). 'Do the aircraft ever: wake you up? interfere with listening to radio, TV, or Hi Fi? make the house vibrate or shake? interfere with conversation? startle you; interfere with or disturb any other activity; annoy or disturb you in any other way?' (For each) If YES: 'How much does this annoy you?' See [3] for scoring method.
ARCNA:	'All things considered, do you think that the amount of aircraft noise here is 'acceptable or unacceptable?' → ARCNA = 100 - % acceptable
ARCBOTH:	'What is the most bothersome noise you hear round here?' → ARCBOTH = % mentioning aircraft

Table 2. Disturbance at different times

Score	Day	Evening		Night		ANAS
		In	Out	In	Out	
3	18.6%	19.7%	17.7%	7.3%	6.0%	23.4%
2	23.7%	21.6%	27.7%	8.7%	7.3%	28.8%
1	26.5%	26.4%	26.9%	17.7%	18.7%	25.8%
0	31.2%	32.3%	27.8%	66.3%	67.9%	21.9%
Sample	993	963	810	985	893	2052

Key: 3 = very much annoyed, 2 = moderately annoyed,
1 = a little annoyed, 0 = not at all annoyed/don't know.
In/Out = Response from someone in/out during the day
ANAS = Question corresponding to VMANN above

Figure 1 Adjusted* values for average GAS scores, plotted against 1 week 24 hour Leq, (ADJGAS vs W1LQ24)



*The average GAS Score (AVOGAS) is 'corrected' for the economic dependence on the airport (via the multiple regression fit) to give ADJGAS [3]