FINDING A GOOD AIRCRAFT NOISE ANNOYANCE CURVE

Peter Brooker

Cranfield University

1. Introduction

The aim is to find a good aircraft noise annoyance Dose-Response curve, ie to measure how people's annoyance varies with their exposure to aircraft noise. The focus is on practical and robust techniques with the minimum of modelling assumptions. Useful background sources are Brooker (2004) and Fidell & Silvati (2004).

The starting point here is what an 'ideal' Dose-Response relationship might be like. An examination of some real UK data then puts this into perspective. The kinds of practical problems indicated by the real data show the potential complexity of the Dose-Response functional dependence. Moving average techniques could be useful because they provide large 'synthetic' samples – some explorations of a large dataset confirm that this would be a worthwhile approach.

2. The Ideal

Two key concepts are Response and Dose:

> Dose: physically measurable combination of noise parameters, combination of objective physical, measurable parameters about aircraft noise at a particular location

> Response: expression of annoyance, ie measure – or combination of measures – of people's feelings about aircraft noise at that location.

Figure 1 shows the ideal: there is a relationship between Dose and the Response, something like an S-shaped curve. The graph describes an average person and the dose is also some kind of average – both issues are explored further here. People are not annoyed when there is no noise, possibly 100% of people are annoyed at extremely high noise, and the two extremes have to be joined smoothly. However, it is difficult to rule out a quite complicated monotonic curve.

Policy makers want to know if there is some Dose value below which people 'are not annoyed', but the evidence is that this is a very low Dose value. There might be some value of the Dose marking the 'Onset' of rapidly increasing Response. Where one would place such an Onset is open to debate. More usually, policy makers decide upon one or more Response/Dose Standard values. Airport noise exposure contours generally use particular Standard values.
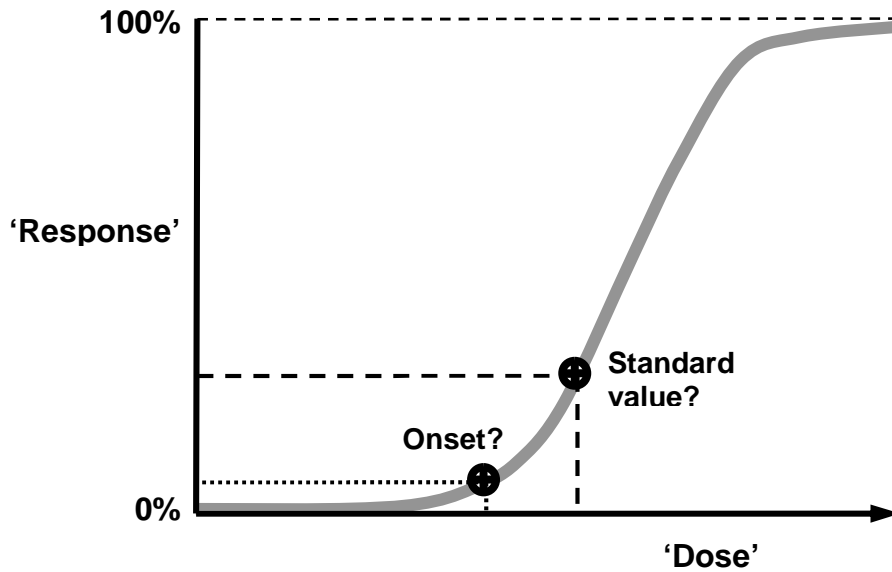
Figure 1. Ideal Dose-Response relationship

3.   Reality in the UK – and elsewhere

In reality, Dose-Response relationships are more complicated. Figure 2 is taken from the 'ANIS' data in Brooker et al (1985). This is rather middle-aged data, but serves to illustrate some issues well because the data on noise and socio-economic variables are extensive. Note that this is a dataset from a single exercise, ie the context/questionnaire is the same for all respondents.
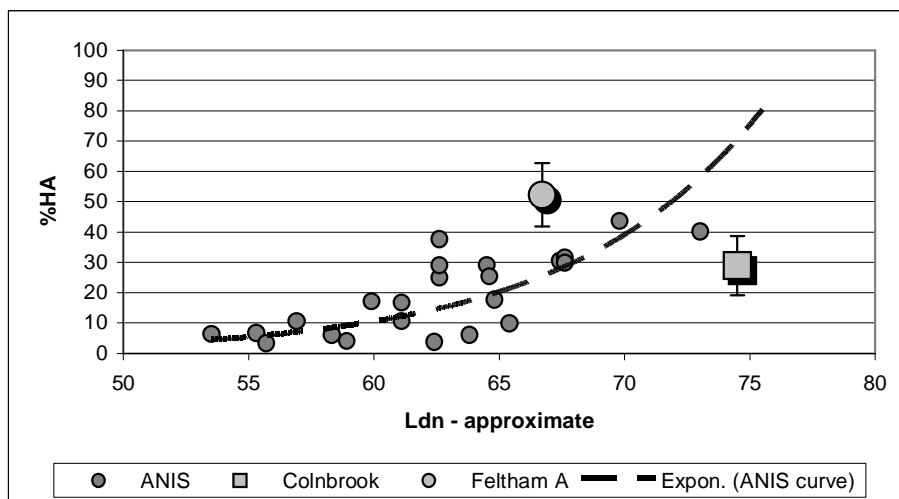

Figure 2. ANIS data on %HA and Ldn
[The large Feltham A symbol conceals a 'Feltham B' point.]

The Response in Figure 2 is taken to be the percentage of people at a particular location who say that they are 'Highly Annoyed' – %HA (Fidell & Silvati, 2004). The Dose is the noise measure Ldn. The values shown are approximate – conversions from the usual UK Leq metric. Ldn is widely used for international comparisons.

The scatter plot data in Figure 2 are for a variety of social surveyed sites near the major UK airports. The dotted line is an exponential fit to the dataset. Is there a better fit? And why are some points well above or below the fitted curve?

First, consider the Feltham A data point, a large round symbol above the fitted curve. It has also been marked with a vertical confidence band. The %HA data points in the Figures are actually approximately binomial samples. Suppose p is the proportion of HA in a population for a specific Ldn value. The survey data is a random sample of size n; if X is the count of HA in the sample, and, if n is small relative to the population size, X is an approximately binomial random variable with mean and standard deviation:

$$\mu_X = np \qquad \text{and} \qquad \sigma_X = \sqrt{np(1-p)}$$

For Feltham A, $n = 88$ and $p = 0.52$, so the standard deviation is 4.7. This produces an approximate statistical 95% confidence band of 10.4% either side of the plotted percentage point. But the typical observed standard deviation for a particular Ldn value is much larger. This cannot be a simple sampling fluctuation. The rest of the points in the Figure have similar-sized sampling confidence band sizes to the Feltham A point.

First, Feltham A, which is to the south east of Heathrow airport and mainly affected by easterly takeoffs. A possible explanation for the high point is that people are rating their disturbance when affected by the *recent* operations of the airport rather than averaging out the Leq energies (?) for the different airport runway operation modes – different 'noise climates' – over a long period. Brooker et al (1985) shows that people's surveyed reactions were more highly correlated with their exposure during the previous week than with longer periods or at the worst times. For this dataset, Heathrow operated easterly for 27.5% of the time during the summer period for the Ldn estimate in the Figure. But during the week before the social survey at Feltham A, the easterly percentage during the daytime was 67.5% – a huge difference. This meant that the Ldn value for that period was nearly four decibels *higher* than the summer average. A four decibel shift to the right on the Figure would put the Feltham A point just about on the fitted curve.

If people's annoyance attitudes vary markedly with their most recent experience, then the question is how to estimate a sensible 'average' value for their annoyance? The estimated average annoyance over a summer period would be a sum of an appropriate weighted average of the Ldn values (assuming that was the best physical measure correlating with annoyance).

For estimates of future aircraft noise contours around an airport, it would be necessary to decide upon some kind of standard mix and duration of noise climates. Note that simply using 'Worst Mode' contours would not work: it would be inequitable to (say) treat two places with the same worst mode value as suffering equivalent disturbance if one place gets it 75% of the time and the other 25% of the time.

The Colnbrook example has even more possible contributing factors. Colnbrook is to the north west of Heathrow airport. Its main noise exposure is from westerly airport operations. Similarly to Feltham A, the ANIS survey at Colnbrook was during a period with a high proportion of easterly operations. In the week before the survey,

the Ldn was slightly more than four decibels *lower* than the summer average. A four-decibel shift to the left would bring the Colnbrook point only slightly below the fitted curve.

Even the average Ldn estimates used in Figure 2 – and any of the studies discussed later – are subject to some inaccuracy. This arises from the facts that the noise estimates quoted for one location probably cover a widespread community (Brooker et al, 1985 – Appendix E); and that noise estimates involve some extrapolation from a sample data collection.

Socio-economic factors also could affect the Colnbrook responses. About 18% of the respondents in that survey had work connections with the airport. ANIS showed that this connection with the airport was the major confounding factor, to high levels of statistical significance. High noise exposure had been a feature of living in Colnbrook for some years prior to the study – particularly since the operations of types such the Trident, B747-100 and Concorde. The likelihood is that there had been some 'sensitivity sorting' in the Colnbrook housing market, so that people living there during the ANIS data collection might not be typical of the general population.

Thus, there are at least four potential factors to explain why individual data points would not be on the underlying Dose-Response curve:

> Sampling variations, dependent on the survey sample size
>
> Variations due to airport operation mode if there are annoyance 'recency' effects
>
> Employment connection with the airport tending to imply reduced annoyance
>
> Sensitivity differences, with population sorting at the highest noise exposure locations.

Housing mobility does vary considerably between countries, depending on the balance between supply and demand, and the total costs of moving home. In the USA/Canada, the legal costs of moving home are generally much less than in Europe, although the UK has low moving costs compared with continental countries.

Thus observed social survey data points have a potentially very complex functional dependence on noise, socio-economic and airport operation variables. Previous work enables identification of *prima facie* explanatory factors, but large-scale social surveys and noise estimation programmes would be necessary to measure their effects with statistical confidence, ie to 'see through' the inherent sampling variations.

4.   A 'Best' Fit: Some Attempted Solutions

Most of the past effort on developing a best fit to aircraft noise annoyance data has concentrated on combining data from different studies and then finding a curve fit that best matches the data as a whole. The usual technique is non-linear regression analysis. One of the most quoted early papers on this approach was by Schultz (1978); recent reviews are Miedema & Vos (1998) and Fidell & Silvati (2004).

Some of the curve fitting studies used datasets with air, road and rail transport noise doses. However, factors such as the airport mode of operation and employment

connection with the airport are much less important for these other kinds of transport. Road and rail operations are generally similar in nature from day to day – and not much affected by the wind direction! Major airports employ large numbers of people and have a substantial infrastructure employing many more, and most of these people tend to live within commuting distance of the airport, ie in comparatively high noise exposure locations. The following analysis uses air transport data.

The curve fitting approaches start with a scatter plot and then fit the data to the 'best' functional form. Examples of this method are Fidell & Silvati (2004) – the dataset in Figure 3: it is vital to scrutinise raw data. The average sample size in the Fidell & Silvati dataset is about 160 – but with wide variations (hence the data points shown are not of equal statistical value), corresponding to a typical 95% sampling error bar of ±7%.

The major problem is determining an appropriate fitting function f(x, a, b). If the fit between the curve and the data is not very good, then the residuals, ie the differences between the data points and the fitted curve values, will show strong patterns when plotted against their x_i values. There is not agreement about what is the best f(x, a, b) for aircraft noise annoyance as function of Ldn. Fidell & Silvati quote seven different functions.
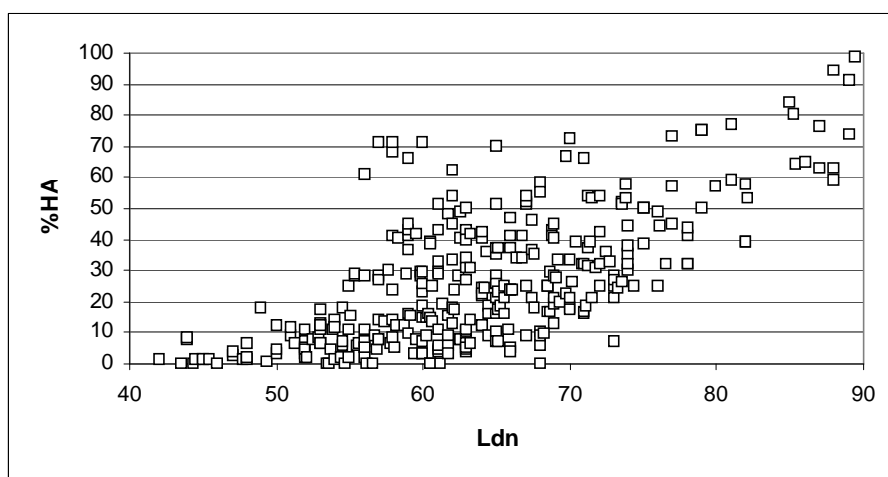

Figure 3. Fidell & Silvati Dataset.

5.   Pragmatic ways around the Inherent Problems?

If the aim is to get the best fit using the large amount of extant data, then some problems are inherent. The variability of data points around the 'true' curve arises for a complicated set of reasons. Sampling variation is obviously there: the only cure for that is much larger samples. Similarly, the fitted estimates assume some kind of average questionnaire/survey context, again based on the variations in the past set of studies. The amount of variation should be less for more recent survey work if researchers are tending to move towards some kind of best practice, but that does not solve issues arising from past work, so a current best fit is a kind of average across past studies.

Population sorting because of people's sensitivities and the effects of employment connected with the airport are intrinsic. The actual effects depend on the relevant

socio-economic factors – nature of industrial patterns near the airport, the efficiency of housing markets, etc. This limits the applicability of average results, even from large data sets, to countries with similar average housing markets (note population sorting in Amsterdam – van Praag and Baarsma, 2005). Predictions of future disturbance *assume* the average effects in the combined data set. Excluding such respondents from analyses would presumably not be ethical, because these are real people living around the airport, whose views should surely be included.

Finally, there is the problem of modal effects. These represent a defect in the existing Dose-Response model, in that the 'true' Dose is not the estimate used in the dataset. This corresponds to an error in the presumed independent variable, and so complicates the usual framework for regression analysis.

Most of these modifying factors represent variations in the surveyed airports' environs and their modes of operations. The observed statistical variations arising from these factors are properties of their past distributional characteristics, and hence would – at best – be repeated approximately in future research studies.

Is there, perhaps as pre-model processing before curve fitting is attempted, a smoothing process that properly averages the large amounts of data? Can this be done in a way that reduces sampling variations, averages out variations in study context/questionnaire, and makes an effort to 'balance out' modal effects? An approach is to apply a moving average to the data. This combines (purportedly) consistent sample data to generate much larger 'synthetic samples'. Moving averages' main use is in the analysis of time series data, but they are equally suitable as a generic smoothing operation where data has a natural sequence. The use of a moving average does *not* add modelling assumptions about the underlying curve shape.

The smoothing form adopted here is a centred moving average – a 7-point average. This effectively multiplies the data point's sample size by 7, and hence reduces the sampling standard deviation typically by about a factor of 2.6, producing a typical 95% sampling error bar of ±2.5%. An n-point average is appropriate in cases when the number of data points is much larger than n.

The first step is to rank the raw data is in ascending Ldn order. Then, let x_i be the Ldn value for the ith data point, W_i be the number of people responding in the ith survey, and y_i be the percentage of people saying they are highly annoyed in the ith survey. The 7-point centred moving average value $\hat{Y}_i$ at x_i is found from:

$$\hat{S}_i = y_{i-3}.W_{i-3} + y_{i-2}.W_{i-2} + y_{i-1}.W_{i-1} + y_i.W_i + y_{i+1}.W_{i+1} + y_{i+2}.W_{i+2} + y_{i+3}.W_{i+3}$$

$$\hat{W}_i = W_{i-3} + W_{i-2} + W_{i-1} + W_i + W_{i+1} + W_{i+2} + W_{i+3}$$

$$\hat{Y}_i = \hat{S}_i / \hat{W}_i$$

As the points in the Fidell & Silvati dataset – 326 of them – are dense in the usual Ldn range, the average spacing between the successive (ranked) x-values x_{i-3} to x_{i+3} is on average very small (probably markedly smaller than the precision of the Ldn estimation). This is a good reason for asserting that $\hat{Y}_i$ is a good estimate of the %HA value at x_i, with a sample size of $\hat{W}_i$. Moving averages should generally match

the underlying shape of Dose-Response curves, ie with smooth monotonic variations and without strong points of inflection.

6.   Moving Average Explorations

What does the Fidell & Silvati dataset generate with a 7-point centred moving average? Figure 4 shows the results, which also includes a simple cubic fit to the moving average points.
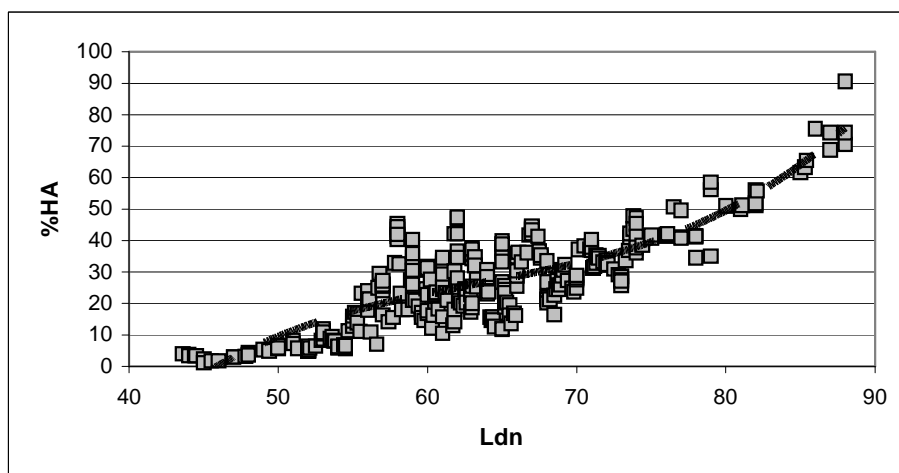
Figure 4. Fidell & Silvati 7-Point Moving Average Data, with cubic fit

The %HA values are squeezed together vertically in Figure 4, but the scatter plot is still messy. Many of the data points are well away from the fitted cubic trend line, even when allowing for the data points' vertical sampling confidence bands. There is still a great deal of statistical structure/variability in the data.

One obvious possibility is that the true Dose-Response curve is somehow changing over time (compare with Brooker, 2008). The data is therefore divided into three time-periods: Pre-1980, 1980s, and 1990s/2000s. This results in the three diagrams in Figure 5. The first and third now look very encouraging, ie much more like a smooth curve – the data is combined but otherwise unprocessed. The fact that long segments of these datasets naturally fall onto a smooth curve is strong evidence that these are near to best-fit relationships.

However, part (b) of Figure 5 is much less successful. From an examination of the studies involved, there seem to be two potential reasons why this might occur: some of the studies are of military operations and others are for 'New Effects' airports. The reason for filtering military operations data is simply that people's reactions to military aircraft might be different to that from 'equivalent' amounts of noise from civil flights (eg Waitz, 2005). A New Effects airport has no immediate definition, but it would usually be one at which large numbers of extra flights took place over affected populations in the recent past, eg because of the opening of new runways. This would have two effects on people's annoyance: first, the affected people would be responding not just the actual Ldn but also to marked increases in its recent values; second, people with high sensitivity would be getting much higher noise exposure, but would not yet able to move to a lower noise exposure location. There is ample evidence from the earliest literature – eg Fidell et al (1985) – that New Effects traffic

changes shift the Dose-Response curve markedly upwards: how large this effect is and how long it persists are complex questions to answer.

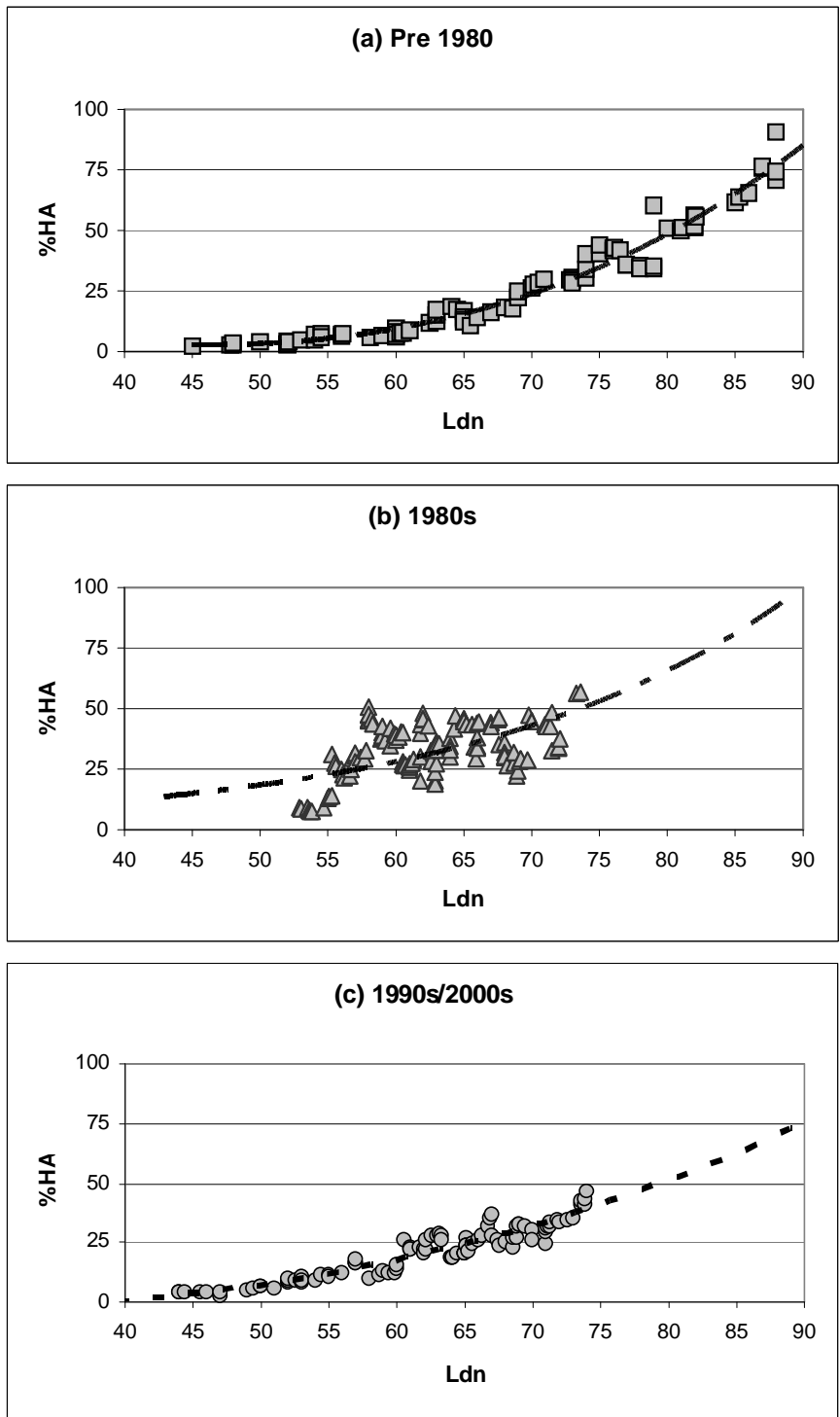**(a) Pre 1980**

(b) 1980s

(c) 1990s/2000s

Figure 5. Fidell & Silvati %HA data, time groups, 7-point moving average, fitted lines (a) and (c) cubic, (b) exponential

The obvious New Effects airports and military operations studies in the Fidell & Silvati dataset are: Vancouver Round 2, Burbank Aircraft, Orange County Aircraft, Bodø Lufthavn, U.S. Airbase, and Long Beach. Most of these studies are in the 1980s, so excluding them markedly reduces the number of data points in that period. Arbitrarily, the remaining data is divided into Pre-1990 (157 data points) and

1990s/2000s (85) data points. Figures 6 (a) and (b) show the results for this filtered data, including approximate 95% sampling bands for the moving average points.

These appear to be good fits. The structure in the processed data beyond a smooth curve variation probably arises arise from the inclusion of some datasets at New Effects airports, very strong socio-economic/industrial and airport modal effects, and survey mode differences. Note also that the quality of some of the very early data is likely to be much poorer than that of later studies (Brooker, 2008).
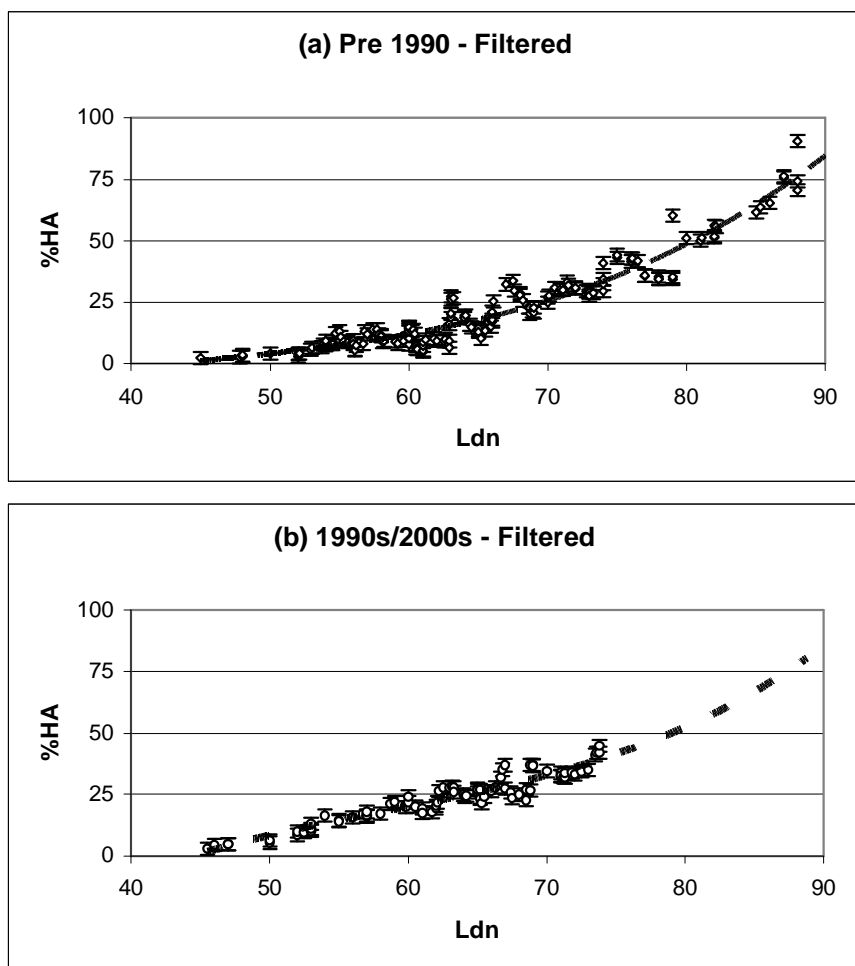

Figure 6. Fidell & Silvati %HA data in time groups, filtered (see text), approximate 95% sampling bands and cubic fitted lines

7. Conclusions

The aim has been to find a good aircraft noise annoyance Dose-Response curve, using practical and robust techniques with the minimum of modelling assumptions.

Several socio-economic/industrial and airport operation factors affected Dose-Response data. This includes 'population sorting' at higher noise exposure locations and employment connections, which are likely to reduce annoyance reactions at higher Ldn values; and airport modal effects on people's recent noise exposure experience, which will produce a defective Dose-Response relationship.

Simple moving average smoothing of the data is a useful procedure. This enables the construction of synthetic large samples – without curve modelling assumptions. It makes apparent the Dose-Response data's *underlying structure*. It is straightforward to fit simple curves to this processed data, and to indicate statistical confidence. Note that the large amount of Dose-Response data available is not sampled from a single curve, but rather from a variety of such curves. The assumption is that there is the same underlying 'mix' of characteristics in the future.

The analysis has to exclude data from new runways, etc airports. The affected people would be responding to marked Ldn *increase*s over a comparatively short time, not just the actual Ldn at the time of survey. The degree of population sorting is an issue, ie people with high sensitivity moving to a lower noise exposure location.

References

Brooker, P. (2004). The UK Aircraft Noise Index Study [ANIS]: 20 Years On. Acoustics Bulletin. May/June, 10-16. https://dspace.lib.cranfield.ac.uk/handle/1826/1004

Brooker, P. (2008). Is there good evidence that aircraft noise annoyance is changing over time? To appear in Applied Acoustics.

Brooker, P., Critchley, J. B., Monkman, D. J. & Richmond, C. (1985). United Kingdom Aircraft Noise Index Study (ANIS): Main Report DR Report 8402, for CAA on behalf of the Department of Transport, CAA, London.

Fidell, S. (2003). The Schultz curve 25 years later: A research perspective. J. Acoustical Society of America 114(6), 3007-3015.

Fidell, S. & Silvati, L. (2004). Parsimonious alternative to regression analysis for characterizing prevalence rates of aircraft noise annoyance. Noise Control Engineering Journal, 5(2), March/April, 56-68.

Miedema, H. M. E., Vos, H. (1998) Exposure response functions for transportation noise. Journal Acoustical Society of America 104, 3432-3445.

Schultz, T. J. (1978). Synthesis of social surveys on noise annoyance. Journal Acoustical Society of America 64, 377-405.

van Praag, B. M. S. & Baarsma, B. E. (2005). Using Happiness Surveys to Value Intangibles: The Case of Airport Noise, Economic Journal. 500, 224-246.

Waitz, I. A. Lukachko, S. P. & Lee J. J. (2005). Military Aviation and the Environment: Historical Trends and Comparison to Civil Aviation. AIAA Journal of Aircraft. 42(2), 329-339.