ICA 2019 AACHEN

PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Uncertainty Analysis of Environmental Sound: Analysis of a Series of Experiments

John FENLON[1]; William WHITFIELD[2]

[1] Dept. of Statistics, University of Warwick, Coventry, CV4 7AL, U.K.

[2] Noise.co.uk Ltd., Bretford, Warwickshire, CV23 0JU, U.K.

ABSTRACT

A series of experiments used attended and unattended instruments at two primary sites: (i) adjacent to the A428 (a second-tier road), and (ii) adjacent to the West Coast mainline railway. These are both rural sites, so that the traffic / rail noise will tend to dominate even at some distance from the noise source. Various configurations of monitors at different distances from the noise source, on different days and for different times were tested, and sound measurements were integrated over 5-minute (road) or 1-minute (rail) periods. At each site a series of measurements were taken during July and August in two separate years, which included several 4-day sets, using two different types of noise meter.

The underlying aim of the project was to develop a simple 'uncertainty' budget for the measurement of environmental noise. The primary interest was in determining day-to-day variation, differences between instruments (at the same position), and differences between measurements at different distances from the source. One set of experiments also tested the difference between measurement positions chosen by different engineers.

Keywords: Environmental sound, Uncertainty

1. INTRODUCTION

All measurement is subject to uncertainty. The British standard requires us to "understand the uncertainty", in terms of understanding the qualitative and quantitative factors affecting the outputs. ISO 17025 (1) requires an uncertainty evaluation for all measurements, and tasks laboratories to identify all components of variation and make a reasonable estimate of their uncertainty to ensure that results can be assessed fairly. A general discussion of these ideas is given in an earlier paper (2). Two important aspects of uncertainty are repeatability and reproducibility, good definitions of which can be found in (3).

At the outset it is important to make two important points: one is that the uncertainty is a statement of the variability associated with measurement and is not intended as a buffer to increase or decrease the measured value. Although uncertainty is measured using the standard deviation (s.d.) of quantitative observations, there is no intention of creating confidence intervals. The second point relates to the approach taken in regular environmental site surveys, which are usually conducted in benign weather conditions, and at positions close to the noise source, so that the more deterministic corrections made for meteorological and topographical factors are not particularly relevant.

2. BACKGROUND

In preparing an uncertainty budget, one should consider all the potential sources of variation. The following factors can be considered stochastic (of a random nature): day-to-day variation, variation between operators, variation between noise meters (whether of the same or different types), and instrument measurement error. It is a truism that sound measurement taken on any two days will differ, so whilst a simple

[1] J.fenlon@warwick.ac.uk
[2] bill@noise.co.uk

random sample may be representative, assuming it is taken on a normal (or representative) day, we do not know by how much it may vary from day to day. For road samples, the day-to-day variation represents not only the difference in traffic between days, but also differences in climatic conditions, temperature, humidity, etc. (4) explores this statistically by looking at historic data sets of road noise measurements taken over several days. The other factors reflect types of measurement error. Some of them are often not considered – there is a view that once a monitor has been calibrated then it should record with a very limited measurement error; that all instruments (when calibrated) should effectively record the same signal; and that operators should not make much difference! Empirical statistical studies suggest that this is far from the truth, and that these factors can be very important. There are further factors, often regarded as deterministic, though there will surely be stochastic components associated with them: meteorological conditions, choice of monitor position on site, distance from noise source, terrain (between monitor position and noise source), seasons or periods (e.g. school term or vacation), and while there are methods for dealing with these, their impact under 'normal' conditions and at sites close to the noise source are likely to be swamped by the stochastic effects.

Statistically, a random sample of days for a particular site would normally cover most uncertainties, e.g. different instruments, operators, weather conditions and the day-today variation itself. This must be regarded as the 'gold standard', though, it is extremely unlikely that randomized, replicated data can be practically (or economically) observed.

To assess the uncertainty associated with sound measurement, some estimate of the standard deviation of a typical sample is necessary. This can be effected in two major ways: one is to generate a data set comprising a sample of days, the other is to set up some form of paired measurement experiment (see, e.g. 5). Generating 24-hour measurements on a random set of days is unlikely to ever be practical and continuous sampling of contiguous days is suggested as the best alternative (e.g. by leaving a monitor *in situ* for several days e.g. a good starting point for road traffic noise would be to take a continuous 5-day sample Monday to Friday). An example of this use of 'historical' data is given in (4). The use of multiple samples (i.e. several 24-hour samples) allows an unbiased estimate of the standard deviation of $L_{AEq,16h}$ or $L_{AEq,24h}$, giving a measure of uncertainty that incorporates primarily the day-to-day variation, though it may well include some assessment of meteorological variability as well as general environmental variability. Throughout, we concentrate on L_{AEq} as the main focus of these investigations has been on façade design.

A particular problem in environmental monitoring is that the formal measure of 'day-time' or 24-hour noise using $L_{Aeq,16h}$ or $L_{Aeq,24h}$ respectively produces a single statistic which has no corresponding standard deviation. This raises the question of subsampling. For rail noise, on a busy track with a regular timetable, we will show that this is a reasonable proposition; for road noise, however, subsampling will lead to bias (and therefore extra uncertainty). In the latter case, the only way of assessing 'true' day-to-day variation is to take several 24-hour samples, so that a standard deviation (s.d.) of the multiple daily means can be determined. Ellison *et al.* (3) refers to standard uncertainty as the s.d. for quantitative measurements; see also Chapter 4 of (6). To assess uncertainty, some replication of the mean is necessary.

Historically, measurement uncertainty has tended to focus on instrument accuracy and precision, whilst ignoring stochastic variation. Suppose, for example, that we were to set up a fixed monitor at a specific location close to some environmental noise source (e.g. a road) and measure the noise levels every day. We would anticipate that the profiles of any two days would not be the same; we would expect the compound means ($L_{Aeq,24h}$) to vary from day to day, with potentially different patterns on different days, in different seasons, and in different weather conditions. In fact, day-to-day variation is compounded of several of these various factors which are not easily separable. If we were to do the same exercise, but in a different way: sending a different engineer out each day to a specified location to make the same measurements, then we would introduce a whole new tranche of variation including positional variation, operator variation and instrument variation. Clearly, there are limitations to what can be achieved at a single observational site. For example, in the project reported here the rail and road sites were rural, so that sound was not blocked or diverted by the built environment as it would be in an urban setting. Nevertheless, the extended samples on the A428 have provided valuable estimates of day-to-day variation (perhaps the largest component of uncertainty), as has the analysis of historic data from noise.co.uk's database (4).

This paper describes an effort to assess that uncertainty in a series of experiments conducted in a rural location in central Warwickshire (UK). Over two successive years, two students supported by the Institute of Physics (IoP) were employed by noise.co.uk to measure traffic and rail noise at locations close to the company's headquarters. In what follows we describe two approaches that noise.co.uk has taken in assessing uncertainty. One involved repeated 24-hour monitoring of a site on several days over a two-month period, the second involved a series of paired (and extended) experiments conducted during the summers of 2012 and 2013.

3. EXPERIMENTS

3.1 Experimental set-up

A series of experiments used attended and unattended instruments at two primary sites:

- adjacent to the A428 (a second-tier road)
- adjacent to the West Coast mainline railway

These are both rural sites, so that the traffic / rail noise will tend to be dominant even at some distance from the noise source. The approach to the road is flat, but the train line runs on an elevated section. Various configurations of monitors at different distances from the noise source and for different times were tested, and sound measurements were integrated over 5- minute (road) or 1-minute (rail) periods. At each site a series of measurements were taken during July and August in both 2012 and 2013. Two basic methods of sampling were employed: extended sampling, when monitors were left for several days adjacent to the road or railway track, and daytime sampling (between 0900 and 1800) when monitors were usually paired. Generally, groups of monitors were set up on a transect perpendicular to the road or track, at measured distances (usually 10, 20, 40 and 80m) from a control monitor positioned as close as possible to the sound source. At other times different configurations of monitors were used, e.g. paired monitors (both similar and different meters) at the same and different positions from the noise source. In one series of experiments, pairs of monitors were set up by different operators, choosing their own positions for a given location.

It was realised quite early on that not all experiments were going to work (see next section) and there was an element of *ad hoc* sequencing of individual experiments rather than a formally structured series. One of the main aims of the programme was to explore the various components of uncertainty as defined above, but also to explore the nature of the data to see if some more general aspects of its structure could be determined.

3.2 Data screening

Not all experiments worked, in the sense that the sound profile did not look consistent (e.g. excessive spiking), or there appeared to be some drift in the data; alternatively, with two or more monitors the data sometimes appeared incompatible. The most important thing with any set of data is to visualise it as a first check of its validity. If downloading to Excel, this is done simply by capturing the time and LA_{Eq} and producing a simple time-plot or a pair of time plots. A 'spiky' response suggests sudden loud noise, which may be real, but can often be an artefact, such as a bird singing close to or perching on top of a monitor – monitors are often sited close to hedges! With two or more monitors, the monitors are synchronised at set-up, but there are periods when the data were clearly not consistent, e.g. a correlation (x-y) plot shows non-collinear data. It is usually advisable to 'top-and-tail the data, i.e. remove the initial and final few observations, as they often pick up set-up noise. A particular problem is that outliers seriously affect the L_{AEq} mean because of its reliance on exponentiation. Several of these ideas will be illustrated in the visual presentation.

3.3 Statistical methods

Most of the statistical methods used are fairly elementary: e.g. summary statistics, regression and analysis of variance, with particular analyses driven by visual interpretation of the data sets (see, e.g. 3 or 7).There are two primary types of data: independent sets that are to be summarised, such as day-to-day records for which some measure of variability is required, or series of data (usually paired, though sometimes parallel sets with more than two series) that can be subjected to analysis of variance methods; the second set comprises correlated data, i.e. data which are commensurate, such as observations taken on more than one monitor at the same time, monitors being co-positioned or displaced.

In an earlier presentation (2) some simple linear models were presented in which the difference between pairs was examined at each time point, and those differences aggregated. An alternative approach, adopted here, simply determines the L_{AEq} for the two separate series and compares them. Whilst these estimates will not comply with standard measures, e.g. 16 or 24h, their difference can be a reasonable measure of uncertainty. For the rail data one can focus on events rather than an L_{Aeq} estimate. Note that, from the point of view of obtaining an unbiased estimate of both mean and variation (uncertainty) the data should comprise a random sample of days. This is unlikely ever to be practicable, and the use of sequences of days is probably the best substitute.

4. RESULTS

4.1 Day-to-day variation

Some of the results for the road data have been described in some detail in (8). A set of nineteen 24-hour 5-minute L_{AEq} samples of the A428 taken on different days during the summers of 2018 and 2019. Two samples showed 'atypical' characteristics, and were omitted. Of the remainder, the mean $L_{AEq,16h}$ was 59.2 dB(A) with a standard deviation (s.d.) of 2.44 dB(A). The ensemble of series is shown in Figure 1 with a heavy dark line marking the average for every 5-minute interval.

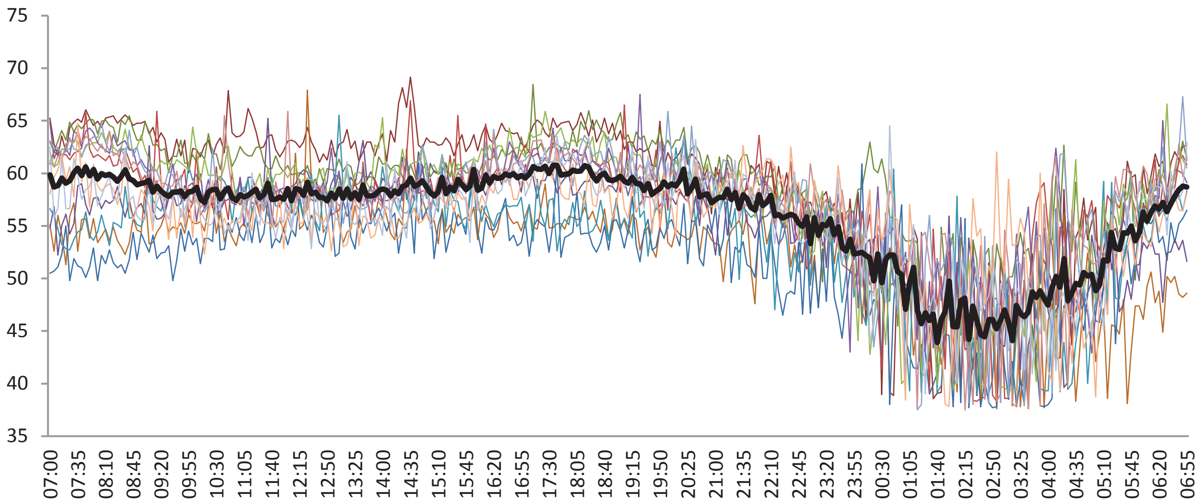

Figure 1: Combined sound pressure level data from 17 x 24 hour samples on the A428 during July and August of 2012 and 2013. The variability of each sample around the overall mean (heavy line) is shown.

Figure 2 shows a typical plot of integrated 1-minute sound pressure train noise (LA_{Eq}) over a 24h period for the section of rail line sampled. There is little traffic between midnight and 06.30, and the minimum LA_{Eq} noise level is slightly lower during the night, c. 35dB(A) relative to c. 40 dB(A) during the day. There are clearly troughs in the afternoon where the sound level is lower. Looking at the peaks there appear to be different levels. Although there are different train types their sound profiles are not hugely different. It was initially thought that the differences in sound levels were most likely to be due to (a) two trains crossing, and (b) a train passage intersecting two measurement periods. However, a subtler difference has also emerged, which appears to be due to a difference in noise levels from the 'up' and 'down' tracks. The site is elevated, so that such differences might be more pronounced.

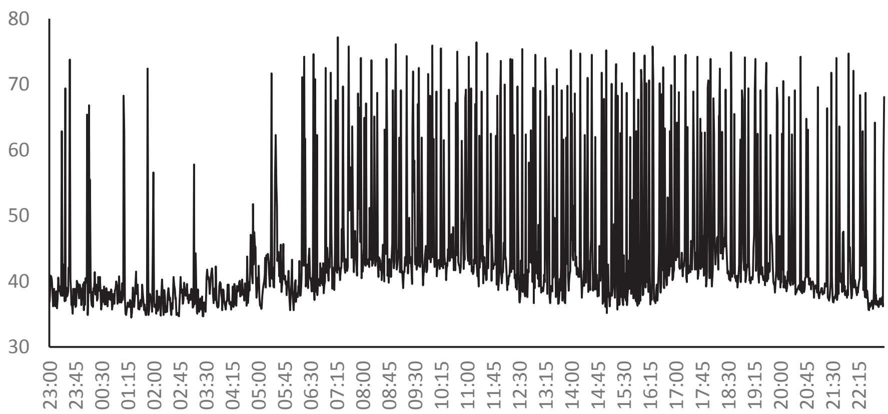

Figure 2: $LA_{Eq,1min}$ observations on the West-coast main line between Coventry and Rugby (Warwickshire) on Sunday 15 July 2012.

Trains pass along this section of track at speeds up to 200 kph. At such speeds, passage times are relatively short with a triangular profile (9). Using 1- minute time intervals individual trains can effectively be identified. This gives the possibility of considering train noise as a simple bi-modal response, i.e. a train event or not, rather than averaging over the complete set of observations. Figure 3 shows a histogram of the 1-minute measurements for 24 hours (14 July 2012). The plot is clearly bi-modal – one might claim multi-modal – although the twin-peak model will suffice.

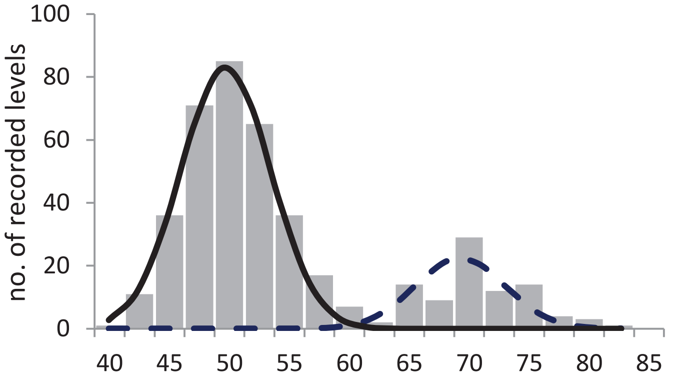

Figure 3: Histogram of 1-minute data for 15-July 2012 @ 10m

The distribution centred around 50 dB represents the background noise and comprises about 80% of the measurements, while the shallower peak (close to 70dB) with a much wider distribution represents train noise. Two approximate normal distributions are superimposed on the Figure to demonstrate the distinct nature of the two noise components, which separate almost exactly. Figures 2 and 3 are typical of the measurements taken over the dozen or so full 24 hour measurements taken over the two sample years. This 'neat' separation might not be expected in suburban or urban locations where other noise interference might be anticipated.

There is clearly some periodicity in the data in Figure 1, and analysis of several consecutive 1-hr or 2-hr samples taken on different days suggests that the difference between days is relatively small. For sound measurement close to busy rail lines where a regular, repeating timetable is in place (as here), sub-sampling may be appropriate as a proxy for full 16-hour sampling. Figure 4 illustrates four consecutive 2-hour profiles from Figure 1, clearly demonstrating the similarities between the different samples, but also showing the quite distinct differentiation between the noise of passing trains and the background ambient noise.

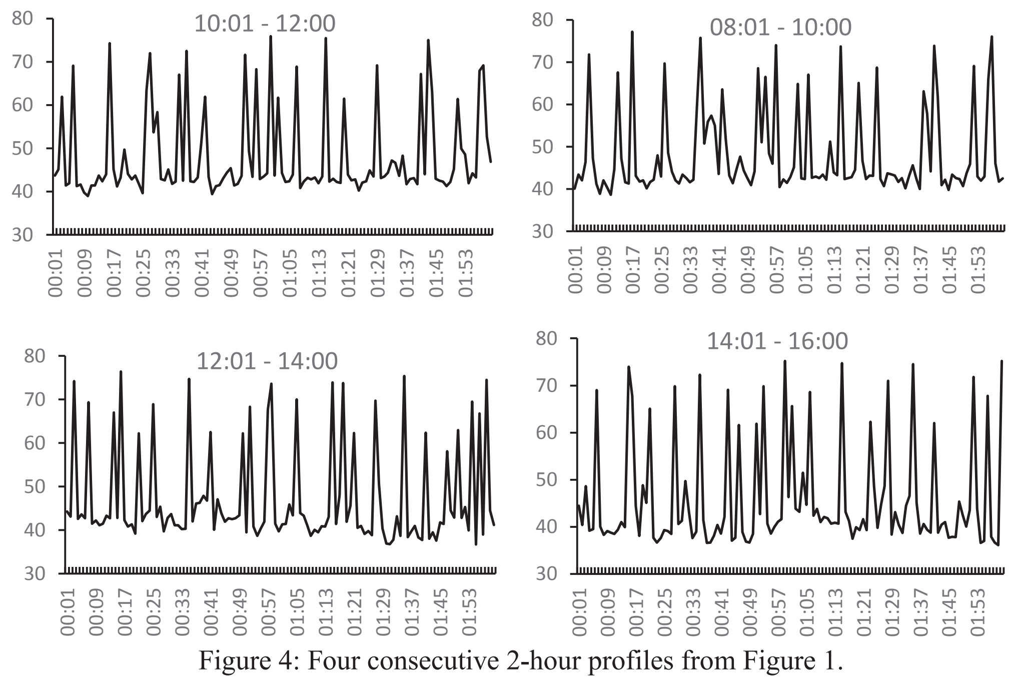

Figure 4: Four consecutive 2-hour profiles from Figure 1.

Extracting the values above 60 db(A) in Figure 4 (essentially the peaks) and integrating them (i.e. forming the log-exponential mean) for the four 2-hour sequences gives L_{AEq} values of 80.1, 81.2, 81.1 and 81.2 (s.d.=0.535), demonstrating the potential effectiveness of subsampling. The corresponding $L_{AEq,2hr}$ values were 72.8, 77.1, 77.1, 77.0 (s.d.=3.257). In a set of data run through a week in July 2013, a set of 25 consecutive 2-hour samples had a s.d. of 0.905. A more extended set of 29 2-hour samples values in 2012 had an s.d. of 2.674.

4.2 Observations at different distances

Various experiments were set up in 2012 and 2013 with instruments simultaneously at distances of 10, 20, 40 and 80m from the target road or rail track. Figure 5 shows a plot of simultaneous measurements at 10, 20 and 80 m from the railway track. The three plots are superimposed, showing the decline in sound level moving away from the track. Note that this decline only holds for train noise data; the background or baseline data (in the absence of train movement) is much the same for all three distances, as might be expected. The $L_{AEq,1000\text{-}2000}$ values are 62.3 dB(A) at 10m, 60.3 at 20m and 53.0 at 80m. Given the differences between the distances we would (theoretically) anticipate a difference of 3 dB(A) between profiles at double distance, so 3 dB(A) from 10 to 20m, and 6 dB(A) between 20 and 80m. Superimposed time-series plots, and simple scatter-plots are very helpful in validating (checking) the data. An example using three monitors at 20, 40 and 80m from the A428 road is shown in Figure 6. Here, the $L_{AEq,16h}$ values are, respectively, 53.0, 51.1 and 49.8 dB(A), the differences between them being somewhat less than the theoretical differences. Considerable variation was noted on different days, and there was a tendency for samples at 80m or more from the target to be more diffuse, an observation noted in correlation plots.

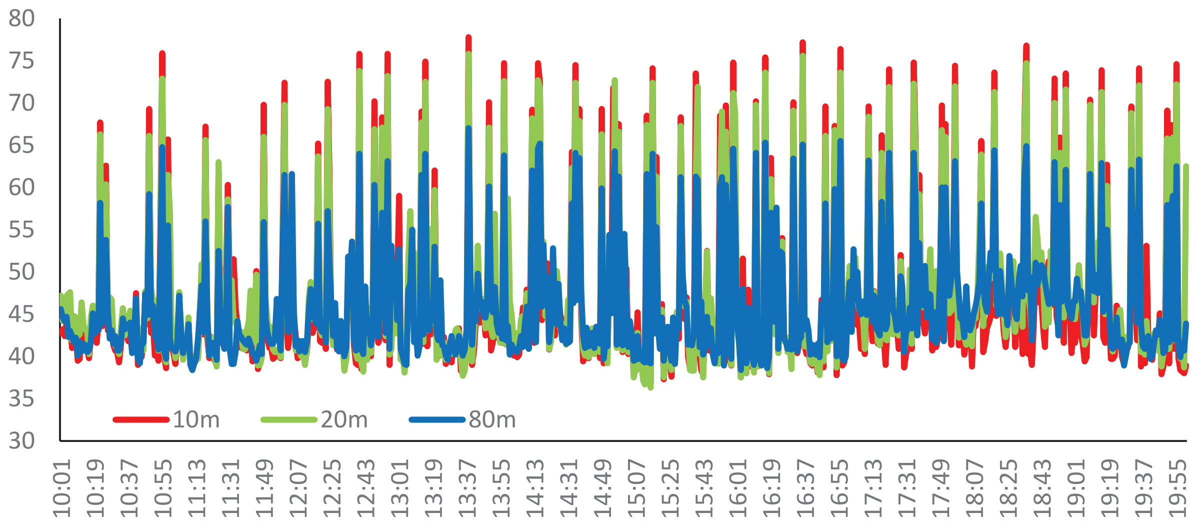

Figure 5: Simultaneous measures at 10, 20 and 80m from West Coast railway on 15 July 2012

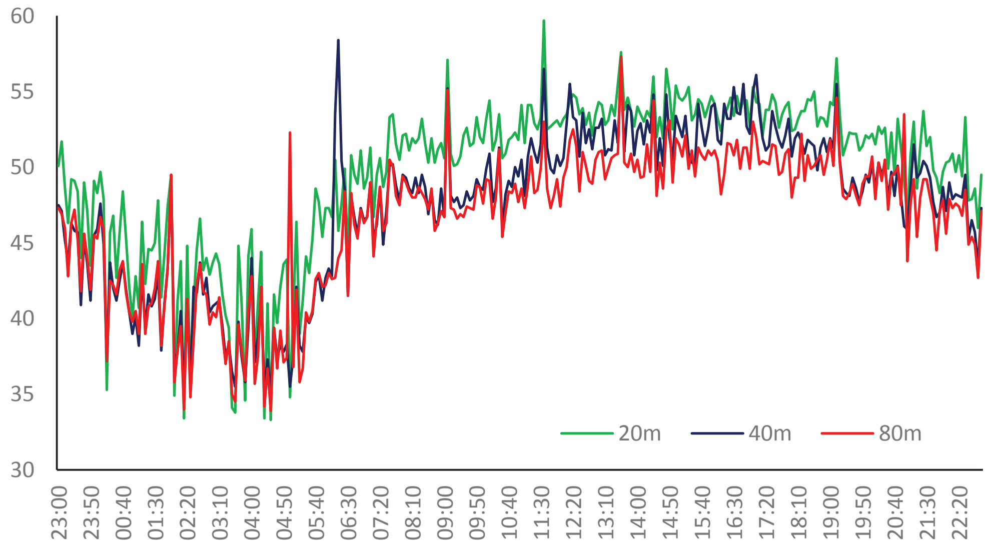

Figure 6: Superimposed plots of 24-hour 5-minute LAEq data for three monitors at 20, 40 and 80m from the A428 (Saturday 11 August 2012)

4.3 Between instrument comparisons

In the 2013 series of experiments, identical instruments (a pair of 360s or a pair of 140s) were set up alongside each other at distances of 10m, 20m, 40m and 80m from the Control point. The comparison of any pair was done on the same day, generally between 09:30 and 17:00, though distance comparisons were staged through the programme. On occasions the two sets of data do not match, in which case the data are simply discarded. Various tests were also undertaken to compare different instruments. The method of analysis for these data conforms to Model (1) of (3); the 'average' estimate of instrument error according to the model was close to 0.5 dB, and there was not much evidence of significant differences between instruments, so the value above represents a generic instrument error, a measure of repeatability as suggested above. Model (1) compares almost exactly with the difference between estimates of $L_{AEq,t}$, where t represents the period of measurement: 58.8 and 59.3 respectively for Instruments 1 and 2.

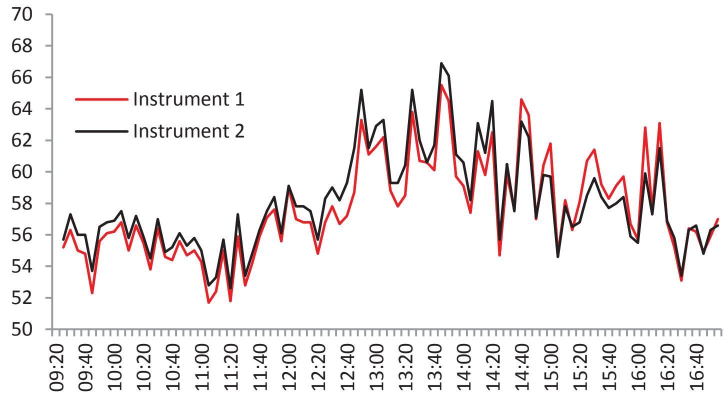

Figure 7: Profiles of two identical meters placed together at 10m from the A428 (25 July 2013)

4.4 Operator variation

In another suite of experiments there were 11 pairs of data associated with the six engineers' instructions at the two sites. As for the instrument comparisons, the data here consist of sets of paired observations from the two instruments, but we may think of the errors in Model (1) relating to a compound uncertainty comprising both instrument and operator error. If we assume the instrument errors to be 0.5 dB (as above) then the operator error is of the order of 1.0 dB(A).

5. DISCUSSION

Suppose that we were to set up a fixed monitor at a specific location close to some environmental noise source (e.g. a road) and measure the noise levels every day. We would anticipate that no two days' profiles would be the same, and we would expect the compound means ($L_{Aeq,24h}$) to vary from day to day, with potentially different patterns on different days, in different seasons, and in different weather conditions. In fact, day-to-day variation is compounded of several of these various factors which are not easily separable. If we were to do the same exercise, but in a different way: sending a different engineer out each day to a specified location to make the same measurements, then we would introduce a whole new tranche of variation including positional variation, operator variation and instrument variation. So that, day-to-day variation is compounded of several factors that are not easily separable. The suite of experiments described here allow us to consider different 'components' of the variation. Some of these 'components of variation' can be distilled from multiple sets of observations with a particular type of linear model, but there are limitations to what can be achieved at a single observational site. For example, in the project reported here the rail and road sites were rural, so that sound was not blocked or diverted by the built environment as it would be in an urban setting. Nevertheless, the extended samples on the A428 have provided valuable estimates of day-to-day variation (perhaps the largest component of uncertainty), as has the analysis of historic data from noise.co.uk's database (4).

6. CONCLUSIONS

In conclusion, the estimated overall uncertainty associated with an $L_{AEq,16h}$ road noise is around 2.5 dB, and similarly that for the uncertainty associated with train event measurement is of the same order. Measures of instrument to instrument variation are generally lower and not necessarily dissimilar for different instrument types. However, there do appear to be occasional marked, unexplainable differences. In the tests reported here, operator effects were generally quite limited although it is often difficult to separate operator and instrument effects. Positional effects for road noise were not too dissimilar from the deterministic adjustments frequently made, i.e. 3 dB for distance doubling, though this did not appear to be the case for train events at the site reported on here. This could well be due to the fact that the track is at an elevated position relative to the measuring instruments. Finally, it should be remembered that components of variation are additive in quadrature, so that an uncertainty of 2 contributes four times the variation of an uncertainty of 1. It also means that an overall large uncertainty is not much affected by smaller components of variance.

REFERENCES

1. ISO 17025: 2005 General requirements for the competence of testing and calibration laboratories, ISO, Geneva, Switzerland.
2. Fenlon, J.S. & Whitfield, W.A. (2017). Measurement Uncertainty in Environmental Noise Surveys: Comparison of Field Test Data, ICSV24 (London), 1480-7.
3. Ellison S.L.R, Barwick, V.J. and Duguid Farrant T.J. (2009). Practical Statistics for the Analytical Scientist, RSC Publishing, Cambridge, UK.
4. Fenlon, J.S. & Whitfield, W.A. (2019a). Uncertainty analysis of environmental sound: a statistical analysis of 100 sites with measurement over several days. ICSV26 (Montreal) (in press).
5. Clarke, G.M. and Kempson, R.E. (1997). Introduction to the Design and Analysis of Experiments, Arnold, London, UK.
6. Kirkup L and Frenkel R.B. (2006). An Introduction to Uncertainty in Measurement, Cambridge University Press, Cambridge, UK.
7. Snedecor & Cochran (1980). Statistical Methods, 7th ed., University of Iowa Press, Ames, Iowa, USA.
8. Fenlon & Whitfield (2019b). Design of experiments for determining uncertainty in Environmental Measurement. Inter.noise 2019 (Madrid) (in press).
9. Smith, Fenlon & Whitfield (2015). IOA Proceedings, Harrogate.