MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Design of Experiments for determining uncertainty in Environmental Measurement

Fenlon, John S.[1]
Dept. of Statistics, University of Warwick
Mathematical Sciences Building, University of Warwick, Coventry, CV4 7AL, UK

Whitfield, William A.[2]
Noise.co.uk
Bretford, Warwickshire, CV23 0JU, UK

ABSTRACT

The British Standard requires us to "understand the uncertainty", i.e. to have both a qualitative and a quantitative understanding of the factors affecting the outputs, and to provide suitable estimates for them. ISO 17025 requires an uncertainty evaluation for all measurements, and states that laboratories must attempt to identify all components of variation and make a reasonable estimate of their uncertainty to ensure that results are fair. Factors such as day-to-day variation, variation between operators, between instruments and the effect of position or distance from a façade all contribute to the variability in measurement. The first of these takes account of the fact that sound measurement taken on any two days will differ. Such day-to-day variation represents not only the difference in traffic between days, but also differences in climatic conditions, temperature, humidity, etc. The other factors reflect types of measurement error. In this presentation we shall look at the statistical discipline of Design of Experiments (DoE) to see how an uncertainty budget can be developed that isolates the individual sources or components of error.

Keywords: environmental noise measurement, uncertainty, road and rail
I-INCE Classification of Subject Number: 77

1. INTRODUCTION

Environmental noise is a major factor affecting the quality of life. The British Standard requires us to "understand the uncertainty", i.e. to have both a qualitative and a quantitative understanding of the factors affecting the outputs, and to provide suitable estimates for them. ISO 17025 [1] requires an uncertainty evaluation for all measurements, and states that laboratories must attempt to identify all components of variation and make a reasonable estimate of their uncertainty to ensure that results are fair.

---

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

Uncertainty in environmental noise measurement is a combination of uncertainties that may not always be easily disentangled. Until recently, measurement uncertainty has tended to focus on instrument accuracy and precision, and has not really considered the stochastic variation inherent in the process. 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 no two days' profiles would be the same, and we would expect the means ($L_{Aeq,T}$) 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, changing some of the other conditions, e.g. 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, operator and instrument variation.

"Statistics is usually defined as the science of collecting, analysing and drawing conclusions from data. Data is usually collected through sampling surveys, observational studies, or experiments." [2]. In most disciplines, a characteristic of collected data is their uncertainty. For environmental sound measurements taken at the same site on different days (but at the same time and of the same duration) we have already noted that the two values (samples) will almost certainly be different. Part of that variation will be due to measurement: instruments, operators, specific siting of equipment, while another part will be due to the environment: including different traffic flow, weather conditions, and other factors which will be detailed more fully below. The fact is, that no single measurement can simply characterize the sound environment without some specification of the uncertainty associated with it.

2. BACKGROUND

2.1. Uncertainty

The following can be considered as stochastic uncertainties, i.e. of a fairly random nature: day-to-day variation, variation between operators, variation between noise meters (whether of the same or different types), and instrument measurement error. The first of these takes account of the fact that sound measurement taken on any two days will differ, so whilst a simple random sample may be representative, assuming it is taken on a normal 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. 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 which are sometimes 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).

Empirical adjustments are usually made for distance, e.g. if the monitoring site is not at the specified position. Such adjustments assume exact measurements and no interactions with other factors such as meteorological conditions and ground cover. Clearly, wind direction is an important factor, and the predominant prevailing wind direction will affect the sound measurement at a site. Although there are guidelines on positioning of monitors, etc., access to sites is not always straightforward, and exact positioning can be difficult. Not all of these characteristics are easily measured, and for some of them (e.g. distance, meteorological conditions, terrain) deterministic adjustments can be made using formulae taken from academic studies.

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 replicated data can be practically (or economically) observed. 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 2012 and 2013.

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. 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 [3]. The use of multiple samples (i.e. several 24-hour samples) allows an unbiased estimate of the standard deviation of $LA_{Eq,16h}$ or $LA_{Eq,24h}$, 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. In one experiment in a semi-rural location, for example, extraneous noise caused by birds and other random wildlife, farm animals and occasional mechanical machinery occurs. Measurement on contiguous days will not usually contain any measure of positional, operator or instrumental variation, as it would only involve the setting of one instrument in one position by a single operator.

In acoustics, we often think of uncertainty in terms of repeatability and reproducibility. Ellison *et al.* [4] suggest that repeatability gives an indication of short-term variation in measurement results, and is typically used to estimate the likely difference between replicate measurements. Reproducibility is a more general concept describing any conditions of measurement other than repeatability, so could describe differences due to different instruments, operators, distances and times, or a combination thereof. In environmental sound measurement terms the concept of repeatability is quite tricky to 'pin down' as parallel measurements involve instrument error, while displaced measurements on the same instrument may well involve different ambient sounds (e.g. traffic noise). It is useful to think of repeatability as being the variation (uncertainty) between two identical instruments measuring the same sound source. Reproducibility represents all other sources of uncertainty, and is addressed in more detail below.

2.2. Designed Experiments

The Design of Experiments is based on simple principles that allow inferences to be made about the interventions or treatments that are introduced in an experiment. The primary principles relate to what might be termed 'the three R's': randomisation, replication and representativeness. Randomisation and replication are important in ensuring that measurements are independent of each other, whether of the same or different interventions, and are strongly linked to the ideas of repeatability and reproducibility, common in measurement science. Representativeness need not concern us in the sense that what is being measured in environmental monitoring is exactly the effect we are measuring. Note that in the case of more formal experiments, such as 'round-robin' experiments for comparing laboratories [5], the test environment should be representative in the sense of being reproducible.

Design of Experiments, in many contexts, relates to comparisons of different interventions, but another purpose of experimentation is to study sources of variability in the response [2], i.e. differentiating and measuring components of variation, essentially the different uncertainties identified above, and this is the focus of the work described here.

The simplest type of experiment is essentially a repeatability or uniformity study in which replicates of experimental units (independent measures of a treatment / intervention) are made. In the case of $LA_{Eq,T}$, where T = 16h or 24h, this means independent samples of a site measurement. Although this can be considered as repeated sampling, it is more of an experimental study in the sense that the measurement comprises a summary of the full 16 or 24h sample. Another simple type of experiment involves two instruments set up side by side (i.e. at the same position) by the same operator on the same day. The most important aspect of such experiments is the synchronicity of the measurements: it is possible to compare different sources of variation / uncertainty, some simple (e.g. two instruments of the same type at the same position positioned by a single operator), others more complex, i.e. incorporating more than one definable source of uncertainty. Thus, for example, if a single operator sets up two instruments at distances d_1 and d_2 from a noise source (where $d_1 \neq d_2$) then the difference between the integrated measurements will contain a component of variation due to the relative distance of the two instruments from the sound source, but it will also contain a measure of the variability of the instruments. Other sources of uncertainty such as operators can be built into the experimental schedule.

Clearly, observations at the same distance should give (almost) identical observations, assuming proper instrument calibration. So, we can consider this type of experiment as giving an estimate of repeatability, the repeatability uncertainty being given by the square root of half the sample variance of the paired differences between the samples. The uncertainty for quantitative measurements is given by the standard deviation (the square root of the variance) to ensure that the measure and its uncertainty have the same dimension.

There are two types of primary 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. More statistical details of these models are given in [3]. Where independent daily sets of data are available, the principal summary statistic for 16-hour or 24-hour data is a single measure $LA_{Eq,16h}$ or $LA_{Eq,24h}$; uncertainty has to be estimated from the sample standard deviation of these separate daily values. If several sets of data are available for shorter periods (e.g. the working day), then a reasonable estimate of uncertainty should be determined by calculating values of LA_{Eq} for each day over the common temporal range, and determining their standard deviation. For rail noise, there may be a case for considering estimates based on shorter periods, when a repeating train passage timetable is available.

2.1. Measurement

For the most part we shall consider $L_{Aeq,T}$, which is appropriate for assessing environmental noise. Conventionally daytime is defined as the 16-hour period between 0700-2300 and night time is the remaining 8-hour period between 2300-0700, with $L_{Aeq,24h}$ being a conflation of the two. Modern instruments ‘average’ the sound level over a designated period, usually 5 minutes for road traffic sources (which are usually continuous but variable), and 1 minute for rail sources. In the latter case, the 1-minute interval further allows the number of rail events to be counted with reasonable accuracy during a given period. As daytime data offers the higher sound pressure levels above the prevailing background noise, the period results are log averaged to give a daytime mean value using the following:

$$y_d = 10.\log_{10}\left[\frac{1}{16}\sum_{i=8}^{23} 10^{\frac{x_i}{10}}\right]$$

where x_i corresponds to the hourly *LAeq*’s (hour ending) so y_d is the log-transformed exponential average of the sixteen hourly noise levels from 0700-2300hrs (the summation from *i* = 7 to 22 implies the inclusion of both hours, ending at 2300) . The hourly noise levels are themselves determined using an analogous formula to integrate the 5- or 1-minute integrated measures for a location over a 1-hour period. Note that the logarithmic average y_d is always greater than the arithmetic mean $\bar{x}_d \left(= \frac{1}{16}\sum_{i=8}^{23} x_i \right)$ unless all the x_i s are equal, and that the value y_d is dominated by high sound pressure levels which makes the statistic prone to influence by spurious unidentified sources with high sound pressure levels – much more so than if simple arithmetic averaging were used.

3. METHODS

3.1. Experimental

A series of experiments were conducted using attended and unattended instruments at two primary sites: adjacent to the A428 (a second-tier road), and adjacent to the West Coast mainline railway. At each site a series of measurements were taken during the months of July and August in 2012 and 2013, which included several 4-day sets, mostly using the same instruments or instrument types. Measurements were averaged over 5-minute periods for the road data, 1-minute for rail ($LA_{Eq,5}$ or $LA_{Eq,1}$). Both sites (road and rail) are quite close to the **noise.co.uk** premises. In both cases experiments involved monitors at varying distances from the environmental source (10, 20, 40, 80m), paired monitors of similar or different types, measurements on opposite sides of the road. Pairing of monitors can be considered as di-located (at different distances), or co-located. Sometimes these experiments were set up by the same operator, and sometimes by different operators. In another set of experiments, monitors were set up according to the instructions of two experienced engineers at a specific location, based on a topographical assessment of the site to be monitored.

3.2. Data validation

Two factors make analysis of the type of data collected here difficult: one is the simple fact that an outlying high value can markedly affect the L_{AEq} measure. As an example, suppose we have 50 measurements of 40 dB(A), then the log-exponential average is simply 40 dB(A). Now let us suppose that there are 50 ‘random’ measurements whose mean is 40 dB(A), but whose standard deviation is 2.5 dB(A) (the values will cover, approximately, the range 35 to 45 dB(A)) – these have an ‘average’ of just over 41 dB(A) – not hugely different from the constant measurement. However, if we add just one value of 60 dB(A) to 49 values of 40 dB(A), the ‘average’ jumps to nearly 45 dB(A), a dramatic change compared to the simple arithmetic mean of 40.4 dB(A). The second problem is the disparity between pairs of monitors set up to record the same environmental phenomenon, which is often caused by some interfering effect such as a gust of wind, a raindrop, or even an insect or bird adjacent to or actually on the microphone. It is also the case that, despite careful configuration, calibration, etc., the two records may just diverge! Visualisation or some simple statistical diagnostics allows some filtering of the data in the sense of (a) editing out clear start and end effects, and flagging inconsistencies, e.g. anomalous intermediate value, or sudden drops in level, and (b) noting incompatibilities of paired and multiple data sets – where simultaneous distance sampling has been undertaken, the profiles should be very highly correlated.

4.3. Statistical methods

Most of the statistical methods used in the analyses presented here are fairly elementary: e.g. summary statistics, such as the mean and standard deviation, etc., regression and analysis of variance, with particular analyses driven by visual interpretation of the data sets. Basic statistical text books such as [6], or more specialised books on measurement and analysis, e.g. [5], show typical examples.

In the case of paired experiments (i.e. instruments at the same position, or displaced instruments) we can produce estimates of differences between synchronous samples, or estimates of means at distances from the noise source. An important summarising method is analysis of variance (anova), see, for example [7], where differences between sets of estimates can be combined to produce overall means corresponding to different sets, and a measure of the variability of those means which is essentially our measure of uncertainty, reflecting the various sources of variation.

Although the sample standard deviation of independent measures of $LA_{Eq,T}$ should be regarded as the ‘gold standard’ for estimating uncertainty, it is also possible to examine the paired time series plots of 1-minute (rail) or 5-minute (road traffic) samples. For example, Figure 6 shows simultaneous plots of the 5-minute output from two monitors placed side by side close to the A428. Note how one meter reads slightly higher than the other in the first half of the recording period, but then trails it later in the day. Summing the squares of the $LA_{Eq,5min}$ difference between the monitors provides an estimate of the variance (square of the standard deviation), and therefore of the uncertainty associated with a pair of monitors. Although this estimate is based on the set of individual 5-minute samples, the fact that the two profiles move almost in unison implies that the uncertainty measure will be very close to that for the integrated measure.

4. RESULTS

4.3. Road

Figure 1 illustrates the full range of sound over a 24-hour period for the A428. This is generally only achievable when a measuring instrument is left at a site continuously, which is expensive regardless of whether the instrument is monitored (attended) or not. If an instrument can be left at a monitoring site for several days, it will provide some replication (and therefore a direct measure of uncertainty) for $LA_{Eq,16h}$ and $LA_{Eq,24h}$ measurements. More commonly, in routine surveys, an instrument only samples sound during the working day (say, 0800 to 1700) when it can be attended. The integrated measure $LA_{Eq,0800\text{-}1700}$ may approximate the 16h day measurement, though there will be a bias which is not consistent from site to site. During the day-time on a moderately busy road the data are essentially continuous, with the noise level hardly dropping below 50 dB(A). The 16h LA_{Eq} is 55.9 dB(A), compared to 56.3 dB(A) for the shorter working day. Note the volatile behaviour of the sample during the night-time hours of 2300 to 0700, where the response is much more event-orientated, with the base sound level falling to around 35 dB(A) in the early hours of the morning; the 8h (night-time) LA_{Eq} is 49.6. Note that in this paper we are considering uncertainty in terms of L_{AEq} so that the concept of a baseline (or background) level refers to L_{AEq} and not L_{A90}.

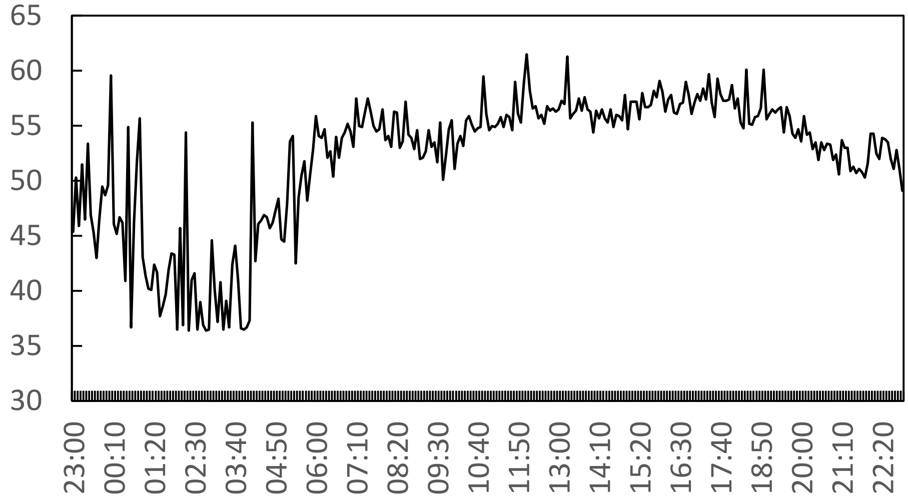

Figure 1: $LA_{Eq,5\ min}$ observations on the A428 mid-way between Coventry and Rugby (Warwickshire) on Friday 20th July 2012.

This particular site was measured over a 24-hour period on 19 separate days (during the months of July and August in both 2013 and 2013, at 10m from the road edge. The data are presented graphically in Figure 2. Several things are worthy of note with regard to these data: virtually all of the 'shooting' values during the 0700 to 2300 period occurred on consecutive days in August 2013. The mean value of the other 17 $LA_{Eq,16h}$ is 59.2

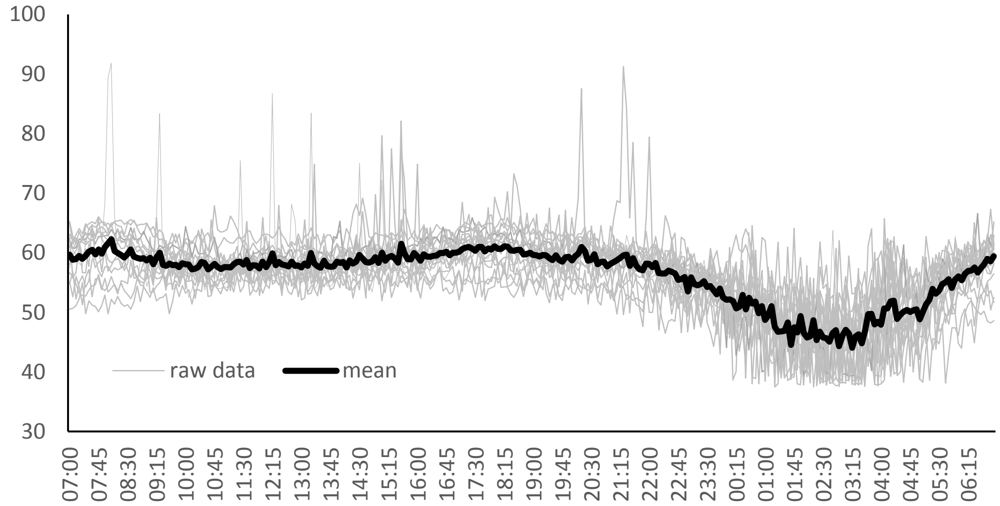

Figure 2: Combined sound pressure level data from 19 x 24 hour samples on the A428 during July and August of 2012 and 2013. The variability of each sample around the overall mean is shown.

dB(A) with a standard deviation (s.d.) of 2.44 dB(A), which suggests that the majority of observations will fall within approximately ±5 dB(A) of the fitted mean. This is certainly suggested by the 'shadow' created by all the other observations. The actual range of the individual daily $LA_{Eq,16h}$ values was 53.8 to 62.8 dB(A), so that a single observation set has a significant uncertainty associated with it. The s.d. gives a statement of that uncertainty [8]. Given that these observations were taken over quite a long period (different days) with different operators, positions and instruments, the variability (or uncertainty) will contain elements due to all these factors, though the day-to-day variation will dominate.

4.3. Rail

Figure 3 shows a typical plot of integrated 1-minute sound pressure of train noise (LA_{Eq}) over a 24h period. The day is actually a Saturday, but the profiles are similar for other days inasmuch as there is a background noise level (ambient, when no trains are passing) and peaks (where a train passes). Clearly, there is very limited 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) in day-time. There are clearly troughs in the afternoon where the sound level is lower. If one looks at the peaks there appear to be three different levels. Although there are different train types: inter-city and commuter/local trains, their sound profiles are not hugely different. The ⅓-octave sound profiles were compared for several samples of background noise, commuter trains and inter-city trains, and the latter two were found to be very similar. It has therefore been inferred that the differences in sound levels are 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.

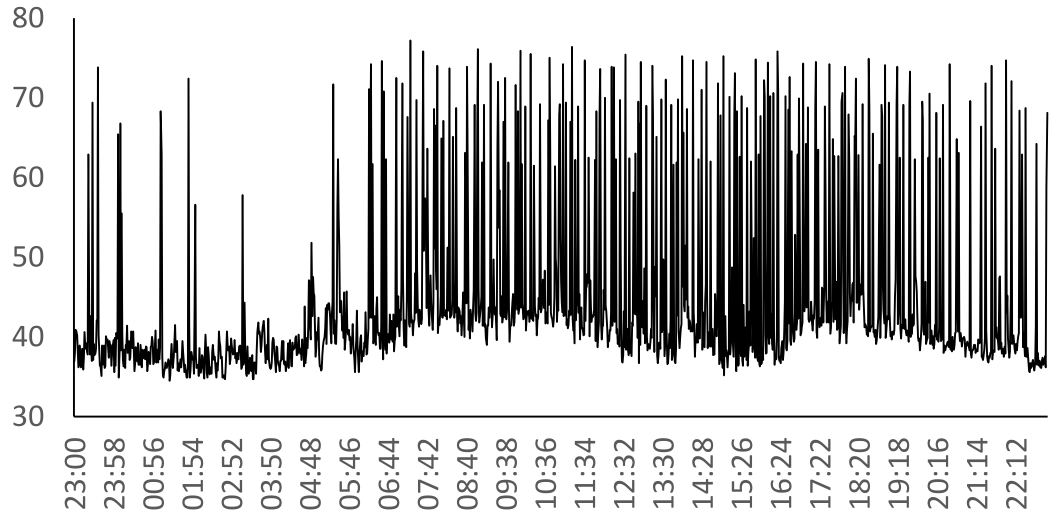

Figure 3: $LA_{Eq,1min}$ observations on the West-coast main line between Coventry and Rugby (Warwickshire) on Sunday 15 July 2012.

There is clearly some periodicity in the data in Figure 3, 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, sub-sampling may be appropriate as a proxy for full 16-hour sampling.

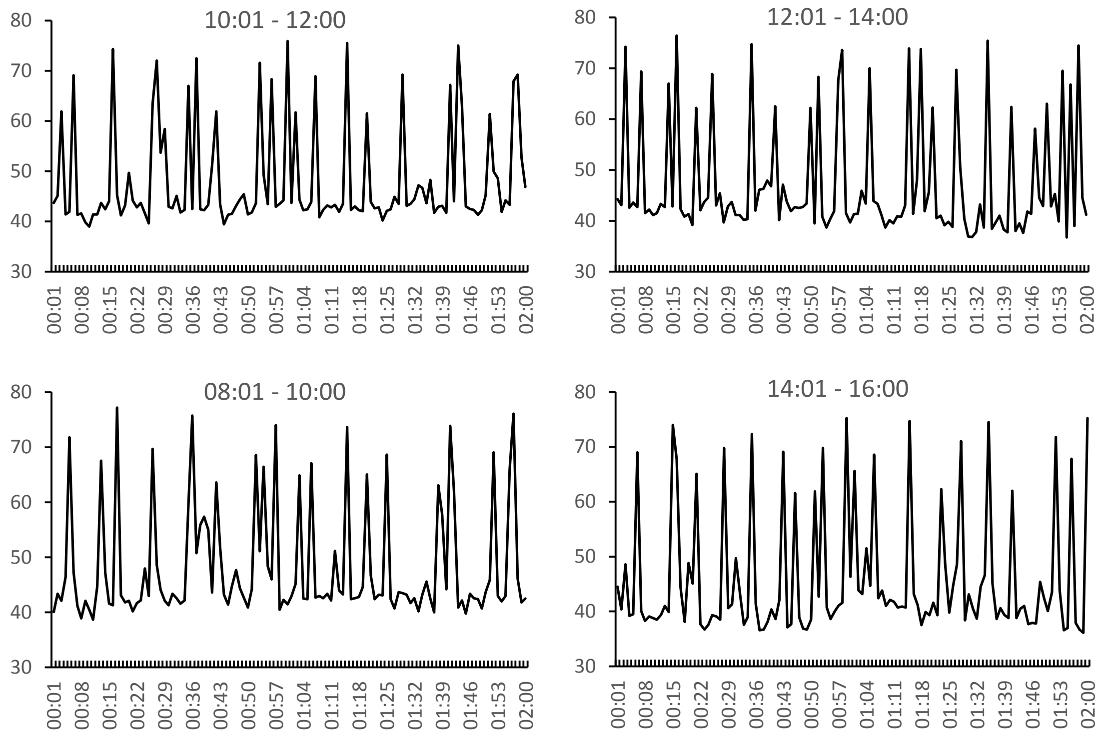

Figure 4: Four consecutive 2-hour profiles from Figure 3.

Figure 4 illustrates four consecutive 2-hour profiles from Figure 3, which clearly demonstrate the similarities between the different samples, but also indicates the quite distinct differentiation between the noise of passing trains and the background ambient noise. The site is in a rural location, some distance from the nearest road, and it is clear that the background day-time noise level is of the order of 40 dB(A). It was commented earlier that 1-minute train noise samples can be considered as events, which 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 5 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. 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. 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 73.2, 72.7, 73.1 and 72.9, demonstrating the potential effectiveness of subsampling. Clearly, this 'neat' separation might not be expected in suburban or urban locations where other noise interference might be anticipated.

4.3. Paired experiments

Although the sample standard deviation of independent measures of $LA_{Eq,T}$ should be regarded as the 'gold standard' for estimating uncertainty, it is also possible to examine the paired time series plots of 1-minute (rail) or 5-minute (road traffic) samples. For example, Figure 6 shows simultaneous plots of the 5-minute output from two monitors placed side by side close to the A428. Note how one meter reads slightly higher than the

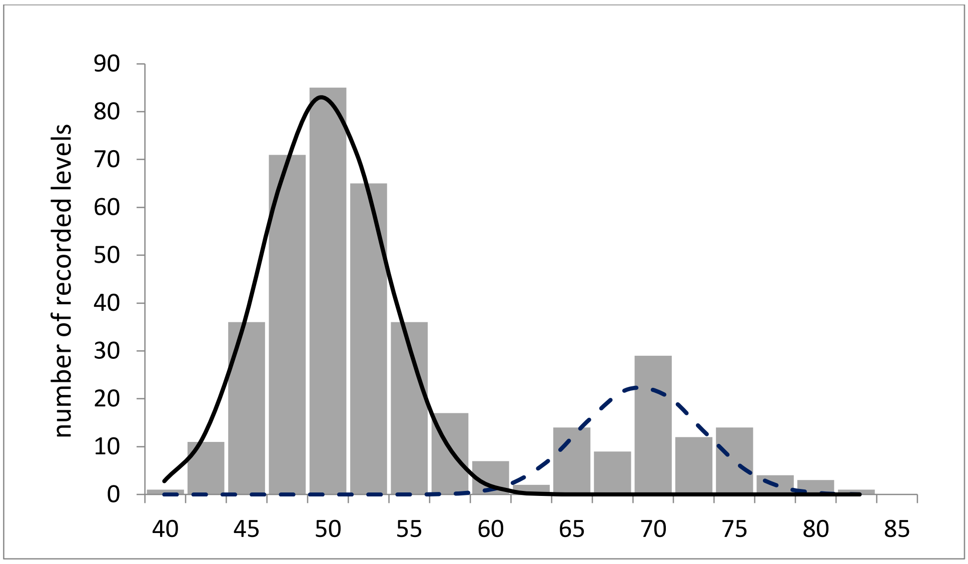

Figure 5: Histogram of 1-minute data for 15-July 2012 @ 10m

other in the first half of the recording period, but then trails it later in the day. Summing the squares of the $LA_{Eq,5min}$ difference between the monitors provides an estimate of the variance (square of the standard deviation), and therefore of the uncertainty associated with a pair of monitors (Fenlon & Whitfield, 2017). Alternatively, one can simply integrate the sample over the full measurement period. For the data presented here, the integrated $LA_{Eq,T}$ (0925 – 1700) values are 58.8 dB(A) for Instrument 1 and 59.3 dB(A) for Instrument 2, a difference of 0.5 dB(A). This compares quite favourably with a difference of 0.4 for the arithmetic mean or a similar value for analysis of differences.

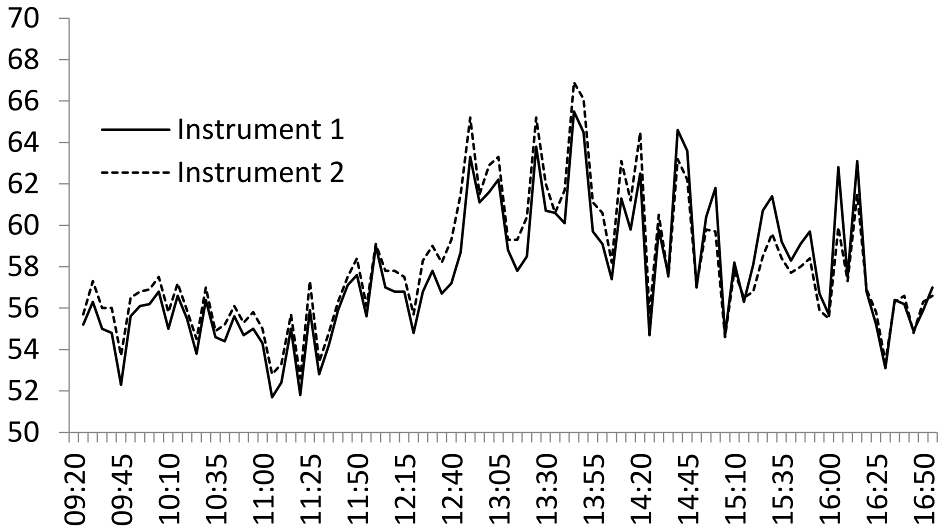

Figure 6: $LA_{Eq,5min}$ observations for two identical monitors situated side-by-side close to the A428 mid-way between Coventry and Rugby (Warwickshire) on 25 July 2013.

Comparison of instrument measurement data is a key feature of the Design of Experiments (DoE). Observing the time histories for several sets of data, the generic estimate of measurement error due to the instrument was c. 0.5dB(A) irrespective of instrument make and model. The estimate is based on the statistical methods alluded to above in the **Statistical methods** section, and set out in more detail in [3].

In one series of experiments (not detailed here) pairs of meters were installed at 10m, 20m, 40m and 80m from the control point and the data compared between pairs on the same day generally between 0930-1700hrs. Visualisation is essential to eliminate incompatible data, e.g. where a meter malfunction suggests that data do not match. Such visualisation can be effected with either a simple *x-y* scatterplot, or overlaid time-series, as in Figure 6.

A visual representation of a set of data from three monitors at 20, 40 and 80m from the sound source data is shown in Figure 7. The data show the 40m and 80m data plotted against the 20m data, which is simply represented by the $y = x$ line. In this way the decline in sound level over distance can be assessed for different distances, but also any interaction between distance and sound level can be seen. In this particular example, the 80m sample appears to be parallel to the 20m sample but some 3 dB lower. Note that the 40m sample appears closer to the 80m sample below about 50 dB but migrates towards the 20m sample at higher sound levels, though there is a considerable amount of overlay between the 40 and 80m samples. Clearly, the empirical rule of 3 dB diminishing with doubling distance does not appear to hold!

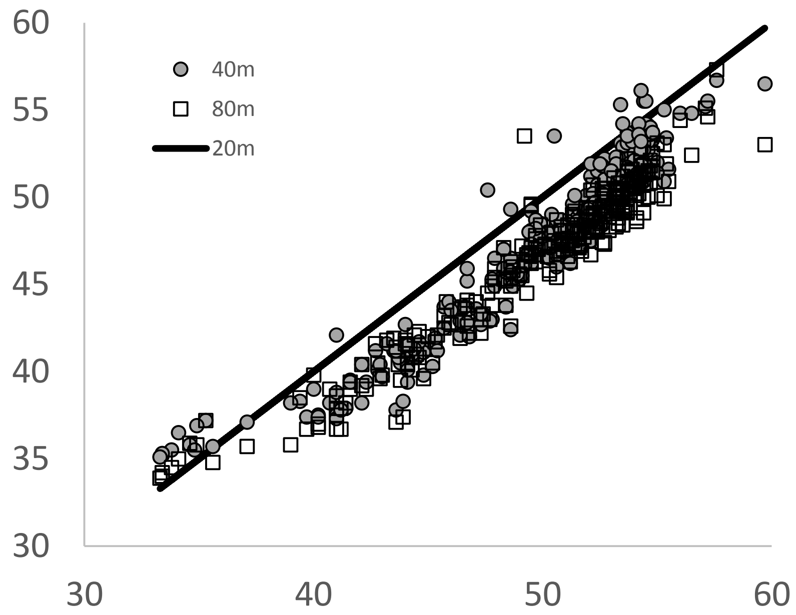

Figure 7: x-y scatterplot of sound-pressure level data: 40m and 80m plotted against 20m sample for road traffic noise data on the A428 – Saturday 11 August 2012.

5. DISCUSSION

Measurement uncertainty has tended to focus on instrument accuracy and precision, and not really considered the stochastic variation inherent in the process. 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 no two days' profiles would be the same, and we would expect the compound means ($L_{Aeq,16h}$ or $L_{Aeq,24h}$) 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.

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). For the rail data we have focused on events rather than an L_{Aeq} estimate, but it is clear that subsampling offers a method of measuring uncertainty, though day to day variation is likely to be greater than within day variation, simply because of weather / atmospheric conditions . 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 a reasonable substitute.

With regard to the practical aspects of environmental sound recording, it is clear that some sets of observations are incorrect, e.g. the sound profile does not 'look' appropriate, or there appears to be a drift in some or all of the data. The most important thing with any set of data is to visualise it as a first check of its validity. If downloading to a spreadsheet such as Excel, this is done simply by capturing the time and LA_{Eq} and producing a simple time-plot. A 'spiky' response suggests sudden loud noise, which may be real, but can often be an artefact, such as a bird singing close to a monitor – monitors are often sited close to hedges! The extended (percentile) data can be used to draw out a true exceedance budget which may be helpful in censoring the data. A particular problem is that outliers seriously affect the BS log mean because of its reliance on exponentiation.

With respect to train noise, there is a general assumption that a train event occurs within a 1-minute interval. This is clearly not the case on a random basis; however, given a triangular trace of a passenger train's sound profile, the apex of the triangle is critical, it can be easily shown that even if the apex is within a couple of seconds of the start or end of the sampling interval, the LA_{Eq} will only fall by about 2 dB(A) well within the sort of sampling variation encountered in this exercise.

The statistical discipline of Design of Experiments is an important methodology for organising experiments or surveys to estimate components of variance, i.e. breaking down the uncertainty associated with the measurement system to enable a statement to be made about the veracity of individual measurements, but also enabling the engineer to focus on the significant factors affecting the variability. Empirical studies suggest that day-to-day variation is the dominant component in most environmental studies.

6. REFERENCES

1. BS EN ISO/IEC 17025 2017: "General requirements for the competence of testing and calibration laboratories"
2. Lawson, J (2015). *Design and Analysis of Experiments with R*, CRC Press, Boca Raton, Fl, USA.
3. Fenlon, J.S., Whitfield W.A. (2017). Measurement uncertainty in environmental noise surveys: comparison of field test data, in *ICSV 24*, 2017.
4. Ellison, S.L.R., Barwick, V.J, Farrant, T.J, *Practical Statistics for the Analytical Scientist*. 2009: RSC Publishing, Cambridge UK.
5. Scrosati, C., Scamoni, F., Zambon, G. (2015). Uncertainty of façade measurement in building by a Round Robin Test. *Applied Acoustics*, 96, 27-38.
6. Snedecor, G.W., Cochran, W.G., *Statistical Methods*. 7th Edition ed. 1980: University of Iowa Press, Ames, Iowa, USA.
7. Dean A, Voss D, Draguljić D (2017). *Design and Analysis of Experiments, 2nd ed.*, Springer International Publishing AG, Switzerland.
8. Hughes, I.G., Hase, T.P.A. (2010). Measurements and their Uncertainties. Oxford.