24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

MEASUREMENT UNCERTAINTY IN ENVIRONMENTAL NOISE SURVEYS: COMPARISON OF FIELD TEST DATA.

John Fenlon

Department of Statistics, University of Warwick e-mail: j.fenlon@warwick.ac.uk

W. A. Whitfield

Noise.co.uk Ltd, Bretford, Warwickshire, England: e-mail : bill@noise.co.uk

Environmental noise is a problem affecting the quality of everyday life. It is the focus of many planning regulations which affect, building envelope, mitigation design, construction costs and is the cause of numerous legal disputes and issues. Thus it is important to measure environmental noise as accurately as possible and to assess the uncertainty associated with those measurements. In this paper we examine and quantify some of the factors associated with uncertainty in environmental sound measurement for both road and rail traffic, and make some recommendations based on experimental work carried out by Noise.co.uk Ltd. This paper is designed to be read with the corresponding conference presentation which is largely illustrative (graphical).

1. Introduction

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 weightings (estimates) for them. The ISO standard ISO 17025 (2005) 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 can be assessed fairly. We often think of uncertainty in terms of repeatability and reproducibility. Ellison *et al.* (2009) 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. 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. Uncertainty

In an environmental investigation, resources cannot usually be allocated to provide extensive multi-day sampling. A particular problem is that the formal measure of 'day-time' or 24-hour noise using the $L_{Aeq,16h}$ or $L_{Aeq,24h}$ respectively, has no corresponding standard deviation, so that the only way of assessing 'true' day-to-day variation is to take several 24-hour samples, in order to determine the standard deviation (s.d.) of the multiple daily means. Ellison *et al.* (2009) refers to standard uncertainty as the s.d. for quantitative measurements; see also Chapter 4 of Kirkup and Frenkel (2006). To assess uncertainty, some replication of the mean is necessary, though this is both practically and financially very difficult. However, most of the results reported in this document relate to measurements where replication was available.

2.1 Stochastic and Deterministic uncertainties

Generally, in preparing an uncertainty budget, one should consider all the potential sources of variation. The following can be considered as stochastic uncertainties, i.e. of a fairly random nature:

- day-to- day variation
- variation between operators
- variation between types of noise meter
- variation between noise meters of similar type
- 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 can be made for distance, e.g. if the monitoring site is not at the specified position; though 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 involves the statistical analysis of data from several sites where a 'blind' (unattended) monitor was left in position over several consecutive days, the second involves a series of paired (and extended) experiments conducted during 2012 and 2013.

3. Experiments

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. 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, which included several 4-day sets, using two different types of noise meter.

3.1.1 Aims of the project

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 tested the difference between measurement positions chosen by different engineers, although the monitors themselves were set up by a single operator (student). Formally, the effect of different operators setting up equipment was not tested.

3.1.2 Experimental summary

Two basic methods of sampling were employed: extended sampling, when monitors were left for several days adjacent to a road or a railway track. 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. In the first year this was initially done at weekends, but in 2013 monitors were usually set up on Monday and left until Friday. At other times different configurations of monitors were used, e.g. distance (again), paired monitors (both similar and different meters) and opposite sides of the sound source.

For day-time assessments, instruments were generally paired at different distances from the noise source. In addition, a simple 'blind-test' experiment was set up at each of the sites. Six noise.co.uk staff, not all of them experienced engineers, were asked to identify on a map the position of a notional development site. Pairs of monitors were then set up at the identified positions to record noise levels for both road and rail. In general, two sets of monitors were done on a single day together with a Control (near the source) or a Standard (at the nominal position).

3.1.3 Historical data

In the course of environmental monitoring, monitors are sometimes left at a (hidden) location for more than 24 hours. We have identified 50 roadside sites with 2 or more days' worth of data from the last several years. The 50 sites have a total of 198 entries, of which 131 were weekdays, with 34 Saturdays and 33 Sundays. Each entry will have 288 5-minute samples covering a 24-hour period, comprising equalised *LA* samples, together with corresponding percentile data, e.g. 1%, 10%, 50%, 90% and 95%, maxima and minima. These data allow us to examine the relationship between daily means, and determine various measures of variability.

4. Statistical methods and Data monitoring

4.1 'Average' noise level

For most purposes a daytime 'average' noise level (i.e. the logarithm of the mean exponential sound level between the hours of 0700 and 2300) is the standard by which to make an environmental assessment. In the following, for road data in particular, this measure is regarded as the 'gold standard', in the sense that it is the measure by which most regulations are tested. However, from a practical point of view it is not always possible to leave monitors unattended, and measurement is frequently conducted over shorter periods of time. Where comparative experiments are being considered, such as simultaneous measurement at different distances, or paired measurements of monitors at the same distance, then the difference between means provides a useful alternative, although the range of data may be more limited. Nevertheless, even if data are collected over a shorter time period, e.g. 0900 to 1700, this will usually incorporate much of the noisier part of the day, so that any bias in the mean measurement will tend to be positive, i.e. the shorter term mean will be greater than the full day-time mean.

4.2 Preliminaries

Suppose that $y_d = 10.\log\left[\frac{1}{16}\sum_{i=8}^{23}10^{x_i/10}\right]$ is the day-time mean determined from the BS formula,

where the x_i correspond to the hourly LAEq's., so y_d is the log-transformed exponential average of the sixteen equalised hourly noise levels from 0700 to 2300. For 5 minute intervals each hourly average is derived from the 12 × 5min samples making up the hour; it can easily be shown that y_d, the 16-hour estimate, is the log-transformed exponential average of all 192 × 5-minute 'average' values that span the period. The corresponding simple arithmetic mean for day-time is given by $\bar{x}_d = \frac{1}{192}\left[\sum_{i=8}^{23}\sum_{j=1}^{12} x_{ij}\right]$; the standard deviation, given by: $s.d.(x_d) = \sqrt{\frac{1}{191}\left[\sum_{i=8}^{23}\sum_{j=1}^{12}(x_{ij} - \bar{x}_d)^2\right]}$ relates to the overall variability around the mean, rather than the s.d. of the mean itself which is given by $s.d.(\bar{x}_d) = \frac{s.d.(x_d)}{\sqrt{192}}$. However, for y_d there is no corresponding measure of internal variability, and re-sampling using 'bootstrapping' of the data set will only reflect the variability within a single day's data, not the variation from day to day. Hence, the only realistic way of assessing variability / uncertainty for the y_d is to determine its s.d. from multiple-day samples.

4.3 Train event data

The West Coast rail line was sampled with Virgin trains running at speeds up to 120 mph on this section and London Midland commuter trains at a slightly lower speed. At such speeds passage times are relatively short with a triangular profile [Smith, Fenlon & Whitfield (2015)]. Using 1 minute time intervals individual trains can be identified: e.g. there are ~6 express trains per hour during the major part of the day, and ~6 commuter trains; there is also some freight traffic, primarily in the less busy times. For simultaneous data (i.e. measurements made at the same time using two instruments, either adjacent or separated), it is relatively straightforward to identify events.

4.4 Statistical models

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. Ellison *et al.* (2009) or Snedecor and Cochran (1980). 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.

4.4.1 Two instruments

The simplest type of experiment involves two instruments set up side by side (i.e. at the same distance) by the same operator on the same day, and we can write down two simple time series models (linear)

$$Y_{1t} = d_t + \delta_{1t}, \qquad Y_{2t} = d_t + \varepsilon_{2t} \tag{1}$$

where the observations Y_{it} represent the measurements at time *t* on two separate instruments ($i = 1, 2$) and δ_{1t}, ε_{2t} are the corresponding 'uncertainty' errors, and d_t is the 'actual' response at time *t*. Clearly, observations at the same distance should give identical observations, assuming proper calibration. So, we can write:

$$E(Y_{1t} - Y_{2t}) = 0 \quad \text{and} \quad \mathrm{var}(Y_{1t} - Y_{2t}) = \sigma_\delta^2 + \sigma_e^2 .$$

There are some subtle effects here: in principle, there is an operator 'error', or, perhaps more appropriately, a set-up 'error' which should be regarded as a random factor, so that:

$$\mathrm{var}(Y_{1t} - Y_{2t}) = 2\sigma_o^2 + \sigma_\delta^2 + \sigma_e^2$$

where σ_o^2 represents the operator variance, but we could assume that the operator variance is absorbed into the 'instrument' variances σ_δ^2 and σ_ε^2. Clearly, if the instruments are the same, then

$\mathrm{var}(Y_{1t} - Y_{2t}) = 2\sigma_\delta^2$. So, we can consider this type of measurement 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. Altman and Bland (1980) give an exposition of this problem in the medical context 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.

An extension of model (1), where instrument 1 is set at a 'control' position and instrument 2 at a non-zero distance from instrument 1, might be

$$Y_{1t} = d_t + \delta_{1t}, \quad Y_{2t} = d_t + c + \varepsilon_{2t}. \tag{2}$$

In this model, instrument 2 measures the response with a non-zero bias *c*, i.e. by a constant amount above or below the measure of instrument 1, added to which is also a simple additive measurement error as before. Suppose that the two instruments are at distances *k* and *l* from the source, then deterministically, we anticipate that the true relationship should be:

$$Y_{1t} - Y_{2t} = 10\log_{10}(k/l), \text{ for all } t,$$

so that $c = 10\log_{10}(k/l)$ in the Model 2 equation. However, we might anticipate that, as instrument 2 is further displaced from instrument 1, a measure of dispersion may also have to be introduced into the equation to reflect extra deviation due to distance.

In both the above instances (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. These can be examined visually, but an important summarising method is analysis of variance (ANOVA), see, e.g. Snedecor and Cochran (1980), where differences between sets of estimates (e.g. representing different distances from source, or different distances between pairs of measurements) 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 all the sources of variation.

5. Results

5.1 Historical data (road)

Ignoring, for the moment, the distinction between weekdays and weekends, analysis was done on the summary samples, i.e. the BS (logarithmic averaging) means and arithmetic means, for daytime (0700 to 2300). Plotting the data identified three clear outliers, all at sites with low overall means (c.55 dB). Apart from these three, there was a clear decline in the standard deviation as the mean noise increased. Removing these three sites from the 'pool' made a big difference: analysis of variance of all 50 sites estimates the 'average' standard deviation of BS means to be 3.16 dB (NB: assume all statistical figures are in decibels dB), falling to 1.98 when the outliers were removed. Essentially, this latter figure is the best estimate we have of uncertainty, as it is estimated over replicated BS mean records at multiple sites. If we differentiate between weekdays and weekends we can reduce these two figures to 2.98 and 1.86 respectively. In terms of measurement uncertainty, this last figure is a fairly robust measure of overall uncertainty for road traffic, implying an overall uncertainty of c. ± 2dB, incorporating day to day variation, operator and instrument effects.

5.2 Results – road data (A428)

5.2.1 Overall uncertainty

During the course of two summers, the A428 was monitored over 24-hour periods for 19 separate days (measured at 10m from the control). The mean of the L_{Aeq} day-time means is 59.5 with a s.d. of 2.46. Note that this is slightly larger than the 'average' s.d. of ±1.98 for the 47 sites quoted in the previous section.

5.2.2 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. There are two useful methods of visualisation for this type of data: simple paired time-series plots, or scatter-plots, i.e. pairs of data plotted as *x*-*y* plots. 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); 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.

5.2.3 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.

5.2.4 Positional differences

Various experiments were set up in 2012 and 2013 with instruments simultaneously at distances of 10, 20, 40 and 80m. Visual representations of the data both as superimposed time-series plots, and multiple scatter-plots are very helpful in validating (checking) the data. Given the differences between the distances we would (theoretically) anticipate a difference of 3 dB(A) between each of the profiles – since these are set in a geometric progression, and $10 \times \log_{10} (d_2/d_1) \approx 3$, when $d_2/d_1 = 2$. As a general remark, the data for 10, 20 and 40m were usually simply displaced, though at 80m the response was somewhat more diffuse, though still consistent. If the differences between the samples are incorporated into the s.d.’s then the average differences and their s.d.’s are 1.75 ± 1.518 for 3 dB (doubling), 4.18 ± 1.774 (quadrupling) and 5.26 ± 1.913 (eight times), suggesting a considerable diffusion error as well as a potential bias as one moves away from the identified sound source.

5.3 Results – rail data

Histograms of 1-minute measurements for 0700 to 2030 for a typical sample show a clear bi-modal plot. There is almost total separation of background (40 dBA) and train event data (60 dBA+). The distribution centred around 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. The two peaks can be approximated by separate normal distributions thereby discriminating between the distinct nature of the two noise components. Clearly, the model operates on arithmetic averages rather than the exponential averages demanded of acoustic theory. For well-behaved data there is often not much difference between the two estimates – for example, in one sample, the arithmetic estimates of train noise for Control and 20m measures are 71.3 and 68.8 dB, compared to the BS log means of 73.2 and 70.4 dB, a difference of < 2 dB in both instances. For convenience and simplicity of interpretation we shall use the arithmetic means, whilst noting that a translation of 2 dB will not materially affect the results on uncertainty.

For simultaneously collected data, i.e. two instruments at the same or spaced positions relative to the track, identification of train events is straightforward, and the method of paired comparisons (Snedecor & Cochran, 1980, or any standard statistical text) allows a measure of difference between instrument means for train events, and an estimate of the variability. See the section *4.4.1* above. Other data, collected over several days, show very similar signatures, though there can be some slippage due to asynchronicity, late trains, etc., and the timetabling does cause occasional overlap (passing) between trains. Nevertheless, it is quite feasible to collate replications of individual train passages.

5.3.1 Between instrument comparisons

A recent confirmatory experiment set three pairs of 360 monitors (bound together) at 10, 20 and 40m from the track over a weekend. The biases between pairs of observations > 60 dB (mean difference between all samples) were 3.4, 1.9 and 2.6 dB(A), with corresponding variances of 1.128, 1.116 and 0.741. The difference between pairs was surprising – all instruments were calibrated and synchronised. The difference between instruments has to be incorporated into the mean square error (the variance plus the square of the bias), and the square root of half this value (recall that the variance encompasses the uncertainty of both instruments) gives instrument uncertainties of 2.56, 1.56 and 1,91 dB(A) respectively. There were some 350 pairs for each set. In nine other paired trials in 2012 and 2013 using different instruments (usually a 360 and a 118), the uncertainty was between 0.63 and 0.84 for four sets, and between 1.33 and 1.62 for four more; the final result was discarded as an outlier (3.49). Most of these trials comprised about 80 paired samples.

5.3.2 Operator variation

Linked to some of the data in the last section pairs of instruments were set up by two engineers at the same time together with a control instrument set up by a student. Analysis of the comparative data for two realisations of this experiment showed that all results averaged over a 7-hour period differed by less than 1 dB(A).

5.3.3 Train noise estimates at different distances

Data for synchronised estimates of train events at different distances suggested that there was very little difference between the estimates for distances up to 40m. This was confirmed an analysis of variance which showed no effect of group and no significant difference between estimates at Control, 10, 20 and 40m.

5.3.4 Samples on different days

Seven sets of data were compiled by leaving measuring instruments unattended over several days, both during the week and at weekends. Where train events occurred at the same time (or very slightly displaced) on different days, a variance for that set of data was calculated. The variances for these events were pooled (a standard statistical technique), and the resulting standard deviation (s.d.) used as a measure of uncertainty. The individual s.d.'s were combined to give an 'average' estimate of 2.53 dB(A). This is an overall measure of uncertainty as for that determined for road data, although there is essentially only a single instrument and no operator effect given that once the instrument is set up, there is no change.

6. Discussion

Until recently, 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 ($L_{Aeq,24h}$) 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), as has the analysis of historic data from noise.co.uk's database. For the rail data we have focused 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.

In examining the historical data it became apparent that the simple arithmetic mean of a 24-hour sample and its standard deviation are very useful indicators of a site's stability. For example, where an arithmetic mean has a low s.d. (measured over 16 or 24 hours) it simply indicates that the data do not fluctuate much, and therefore the BS mean will be quite close to the arithmetic mean. A higher s.d. tends to indicate a 'spiky' response, which results in a separation of arithmetic and BS means, because of the weighting afforded to larger values of the BS mean.

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 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 train noise occurs within a 1-minute interval. This is clearly not the case on a random basis; however, given a 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.

7. Conclusion

In conclusion, the estimated overall uncertainty associated with an $LA_{Eq,\ 18h}$ 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 10725: 2005 *General requirements for the competence of testing and calibration laboratories*, ISO, Geneva, Switzerland (2005).
2. Ellison S.L.R, Barwick, V.J. and Duguid Farrant T.J. (2009). *Practical Statistics for the Analytical Scientist*, RSC Publishing, Cambridge, UK.
3. Kirkup L and Frenkel R.B. (2006). *An Introduction to Uncertainty in Measurement*, Cambridge University Press, Cambridge, UK.
4. Bootstrapping (statistics) – Wikipedia. https://en.wikipedia.org/wiki/Bootstrapping_(statistics)
5. Smith H.N., Fenlon J.S. and Whitfield W.A. (2015). Environmental Noise – L_{Amax}: Determining measurement uncertainty and identification of outliers. Acoustics 2015, Harrogate. Institute of Acoustics, 15 October 2015.
6. Snedecor G.W. and Cochran W.G. (1980). Statistical Methods, 7th ed., University of Iowa Press, Ames, Iowa, USA.
7. Altman D.G. and Bland J.M. (1980). Measurement in Medicine: the Analysis of Method Comparison Studies, The Statistician 32 (1983) 307-317.