Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Expected mean in an environmental noise measurement and its related uncertainty

Marco Paviotti and Stylianos Kephalopoulos

European Commission, via e. fermi, 1, 21020 Ispra, Italy
marco.paviotti@jrc.it

In the context of the implementation of the Environmental Noise Directive 2002/49/EC a study on noise measurement uncertainty was performed. Averaging over different samples of noise measurements , there might be assumptions over the distribution and independency of the samples. In the context of environmental noise, this might be the case of a series of measurements of a constant noise source like an industrial plant or a fluctuating noise source like road traffic. Using a series of 15 minutes L_{Aeq} samples, the average of these values is usually considered as the expected mean, however, the error caused by the specific selection of the samples is not usually evaluated. Statistically speaking, before establishing an average value, at least the distribution of the samples and the effect of adding-up several uncertainties should be evaluated. This article focuses on the mathematical formulas which could be used and discusses the differences in assessing the expected mean for normally distributed values, or for log-normally distributed, and finally suggests an approach to properly adding-up all uncertainties related to a long-term environmental noise measurement campaign.

1 Introduction

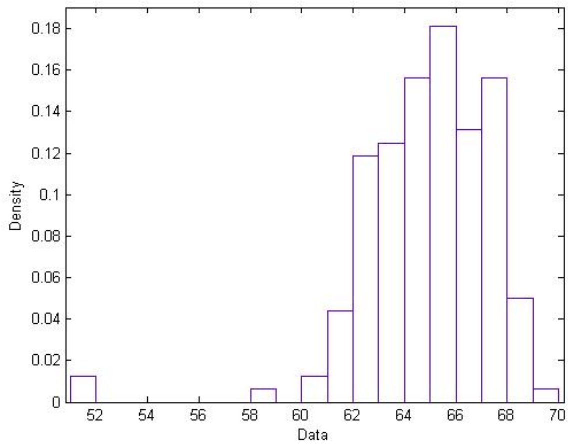
A still unsolved problem in environmental noise measurements is how to correctly measure a long-term noise indicator such as the European L_{den} level. A good statistical analysis should be performed on the measured values to allow for the best estimation of the L_{den}. L_{den} is a noise indicator that requires averaging the source variations over one year and the meteorology that affects the propagation of the sound over several years. .

This article describes possible approaches to be followed for assigning correct numbers to the aforementioned noise indicators, although no definitive solution is given since some problems remain still to be solved.

2 Description of the data characteristics

The present study was performed using data recorded during a long-term environmental noise measurement campaign in an urban situation and close to a road with 25000 vehicles pass-bys on average per day. Data were recorded next to the road, to assess the road noise levels, and then at approximately 200m away from the road, next to a house, where some other local noise sources were simultaneously present. Given the large number of values recorded, it was possible to see the differences in the prediction of the long term noise levels, when the calculations were based on a large set of recordings or just subsets of these recordings. All values used are 15 minutes long L_{Aeq} samples. In Fig.1 the distribution of the samples is given for a selected subset, that is a subset during the same night-time hours (from 01:00 to 05:00) during working days and along several weeks. A few 15 minutes L_{Aeq} samples recorded during the night reported no pass-bys at all, and therefore they were out of the average group of L_{Aeq}. Nevertheless, it was decided to include them, otherwise a bias would have been introduced in the evaluation.

The same evaluation is performed on the same recorded time periods at the house location, far from the major source. The results are plotted in Fig.2.



.Fig.1 Distribution of 15 min L_{Aeq}.

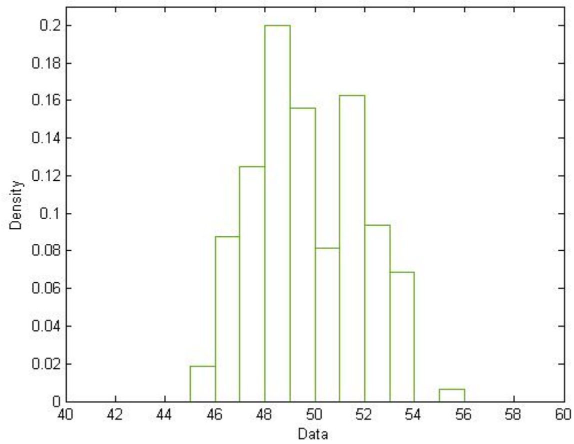


Fig.2 Distribution of 15 min L_{Aeq}.

Both data distributions have a quite common shape. These sets of data were therefore selected as a representative sample of data for analysing how to attribute uncertainties.

3 Analysis of a single set of L_{Aeq}

The first step of the analysis consists in identifying the correct approach for evaluating the true mean. For clarity, this is the mean that would have been obtained if an infinite time would have been allowed for measurements under the same conditions (e.g.: using only data over night-times, next to the road, and during working days only). Two randomly selected subsets were extracted and the means were compared (see Table 1).

Set of samples	Mean L_{Aeq}
All	65.4
1/20th subset 1	66.7
1/20th subset 2	65.3

,Table 1. Example of how averaging over a random subset of L_{Aeq} values can differ from the mean obtained from a number of samples 20 times larger.

Looking at the normality of the distribution, it can be ensured that the set of values chosen did not indeed include 'strange' samples and it seems that a single class of events was captured. In other words, it means that the samples were correctly selected from the same population. It would not make certainly sense to mix a road noise measurement during mid-day with a corresponding measurement at midnight, since it is expected that the traffic will differ a lot between day and night, therefore the samples would come from two statistically different populations.

4 Analysis of the best approach to the true value

Before proceeding in evaluating the uncertainties related to a long-term environmental noise measurement campaign, it is necessary to understand the properties of the mean to be calculated. In environmental acoustics, the mean value of noise levels is commonly given by:

$$\overline{L} = 10 * \log_{10}\left(\frac{1}{n}\sum_{i=1}^{n} 10^{0.1*L_i}\right) \tag{1}$$

and not by the geometrical mean of the L_i samples although from a purely statistical point of view the quantity under assessment is the noise level. Concerning the noise source under assessment, all L_i values should be equivalent levels recorded for the same time interval and under the same operating and meteorological conditions. Also, the measurements should be independent from each other, and this might be possibly checked.

However, the quantity that the instrument "feels" is the sound pressure p, and not the level L_i. Although the pressure is averaged keeping the root mean square (rms) of the p values along a certain time, for each of the sampled periods it will be the p and not the L_i the quantity measured.

It is possible to use the same data set and see how the p values are distributed (Fig.3). Looking at Fig. 3 it can be noticed that even the skewed values of the original distribution showed in Fig.1 now seems to be properly described (the one on the left side of the distribution) by a normal distribution. This can be possibly attributed to the fact that in real phenomena, the sound pressure varies evenly around a certain value.

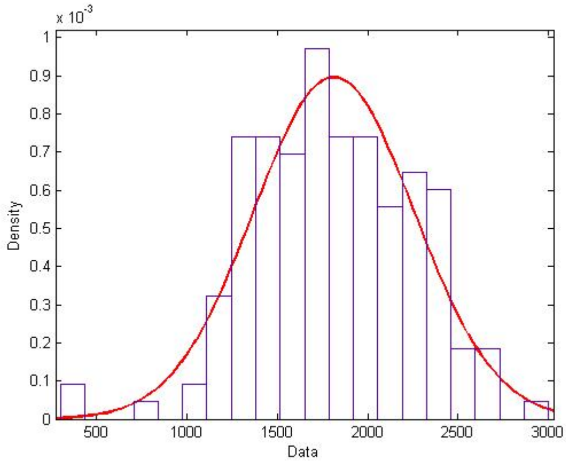


Fig.3 Distribution of 15 min sound pressure (p) values relative to the data in Fig.1.

One would conclude that it is therefore possible to calculate the mean of the samples as:

$$\overline{p} = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{2}$$

however, this will produce a different result compared to the one obtained using the standard way to get the average sound pressure, i.e. by the rms of the pressure levels.

Another question to be answered is whether the average obtained by means of Eq. (1) or Eq. (2) is the correct one for estimating the evolution and fluctuation of the environmental noise phenomenon that is impossible to practically measure thoroughly over a very long time length (i.e., one year). The environmental noise measurements over short periods of time are always performed with the aim to assess what the noise level would be on average over an infinite time period. Therefore the average noise level calculated as described above is never the true mean, however, it should be as close as possible to the true mean. Since the best way to be close to the true mean is to have the largest amount of as much as possible independent and unbiased samples, and given that Eq. (1) is the only correct approach that considers the rms, there is no other solution than using Eq. (1).

Using this approach, nevertheless, there is one conflicting hypothesis: if the mean of the population is defined by Eq. (1), this means that the averaged values are $10^{(0.1*L_i)}$ instead of L_i (i.e., squared sound pressures instead of levels). Consequently, the uncertainty should also be expressed in terms of squared sound pressure instead of levels. Unfortunately, for calculating the standard deviation of the noise levels, it is commonly used the following Eq. (3):

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(L_i - \overline{L}\right)^2} \tag{3}$$

The reason why Eq. (3) is used is that for most of the measurements the levels are considered to be "normally" distributed, not the squared sound pressures, nor the sound pressures. To follow a consistent statistical approach the question is therefore whether to accept Eq. (2) and modify Eq. (1) or vice versa.

Given the computational reason expressed (rms of *p*) and since the regulations (e.g., ISO 1996-2) require performing a squared sound pressure averaging of the different time periods (and this reflects the physical phenomena) Eq. (1) cannot be modified.

Therefore, statistically speaking, the population of values must be the population of squared sound pressures, and this seems to be log-normally distributed.

In the case of environmental noise, since the $10*\log_{10}(p^2/p_0^2)$ is used everywhere, it should be better to assume that the $10*\log_{10}(x)$ is normally distributed (here $x = p^2/p_0^2$ and not *p* because the rms of the x values is chosen). Given this assumption, the corresponding probability distribution function (PDF) can be defined as:

$$PDF(x) = \frac{10}{\sqrt{2\pi}\cdot x\cdot \ln(10)\cdot \sigma}\, e^{\left(-\frac{[10\cdot\log_{10}(x)-\mu]^2}{2\cdot\sigma^2}\right)} \quad (4)$$

where, μ and σ are respectively the mean and the standard deviation of the population of values expressed in levels (L=10*log10(x)), or:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} L_i \quad (5)$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(L_i - \mu\right)^2} \quad (6)$$

Considering the 1^{st} moment of x, it can be shown that the expected mean can therefore be expressed as:

$$\left[E(x)\right] = \int_0^{+\infty} x\cdot PDF(x)\cdot dx = e^{\left(\frac{\ln(10)}{10}\mu + \left(\frac{\ln(10)}{10}\right)^2\frac{\sigma^2}{2}\right)} \quad (7)$$

It is not the intention to go deeper in the analysis concerning ways for calculating the possible means, however, at least for the data set used as an example in this article, the mean values obtained by using the three different approaches discussed above are presented in Table 2.

Mean type	Mean L_{Aeq}
Following Eq. (1)	65.4
Following Eq. (2)	65.2
Following Eq. (7)	65.6

.Table 2. Means of the set of L_{Aeq} values presented in Fig.1 and Fig. 3 calculated by three different approaches.

The slight differences in the calculated means can be explained as follows: the log-normal approach leads to slightly higher values since with respect to the standard mean calculated by Eq. (1) the one calculated by Eq. (4) describes a slightly higher probability that high values will be encountered than when a normal PDF is used. This is illustrated in Fig.4.

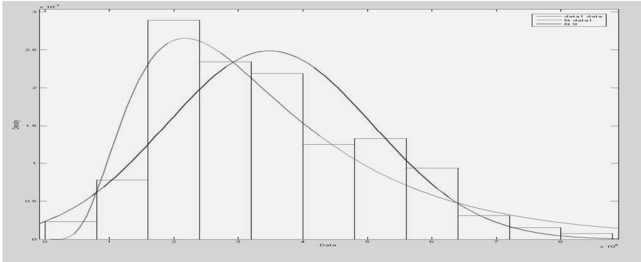


Fig.4 Differences between a normal and a lognormal fitting of the $(p/p_o)^2$ values.

Although the differences between the means are not great, and one can therefore conclude that this does not constitute a major problem, the implications when performing environmental noise measurements are relevant.

5 Standard deviation of a single set of values

When a mean value is calculated, the standard deviation is commonly used to know how the measured values are distributed around the true mean value (the calculated mean would converge to the true mean if noise would have been measured over an infinite time period).

It is common, for normally distributed data (and not for log-normally distributed), to define the uncertainty related to a given set of measurements by addressing the standard deviation σ. In the case of a normal distribution of data, the interval μ±σ should statistically include the 65% of the measured values (only for that specific set of data recorded, not for the entire "population"). Often, this result is used for improperly addressing the true mean and the uncertainty with which this is known. For example, if measuring road traffic noise 100m away from the road, technicians usually perform measurements over one or two weeks and "infer" the yearly average just from the mean calculated from the short term measurements. Also, the standard deviation is sometimes used in environmental noise studies to "infer" how good that mean is, regardless of the fact that this gives information only about the distribution of the measured values and not on the probability these values being close to the true mean.

Indeed, measurements aim at revealing a good approximation of the true mean and provide information on the reliability or distance of that evaluation from the true mean. Given only a small number of samples from the entire population (e.g.: in case of road traffic noise the entire population are daily levels all over the year) it should be possible to use statistics to address the true mean (e.g.: the yearly average level).

The incorrect approach is therefore taking σ to represent the uncertainty with which the true mean could be known. A correct approach is given by the DIN 55 303 part 2 [1,2] where the exact approach for infering the true mean is described using the Student distribution for normally distributed data. This norm implies the use of the squared sound pressure measured values (e.g.: one sample of p^2/p_0^2 each day during a few days along the year in our case) as if the samples would have been taken from the entire population of values (365 daily samples over one year), and using these values as they would have been normally distributed. The only error associated to this approach is that the population is not normally but log-normally distributed. Based on this assumption, and considering Eq. (1) for the mean μ as well, the following expressions for σ and the confidence interval for the true mean ($\mu_{min} < \mu_{true} < \mu_{max}$) are used:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(10^{0.1*L_i} - \mu\right)^2} \quad (8)$$

$$\mu_{max} = \mu + \frac{\sigma * t_{n-1}}{\sqrt{n}} \quad (9)$$

$$\mu_{min} = \mu - \frac{\sigma * t_{n-1}}{\sqrt{n}} \quad (10)$$

The objection for using Eq. (8), (9) and (10) is that the 10^(L/10) values used are log-normally and not normally distributed. So, the μ_{min} and μ_{max} are not strictly applicable (indeed, in case of a large spread of values, negative μ_{min} values could be obtained, which are not possible since the log function is only defined for positive real values).

Mean type	μ-σ	Mean L_{Aeq}	μ+σ
Following Eq. (1)	65.1	65.4	65.7
Following Eq. (2)	62.7	65.2	67.1
Following Eq. (7)	19.5	65.6	129.2

6 Adding up standard deviations

If the formula for normal distribution of data is anyway used, it is then possible to add up the several standard deviations that are calculated for a real noise measurement campaign. Basically, two conditions may exist:

$$L_{average} = 10 * \log_{10}\left(\frac{1}{\sum a_i}\sum_{i=1}^{n} a_i 10^{0.1*L_i}\right) = 10 * \log_{10}\left(\frac{1}{\sum a_i}\sum_{i=1}^{n} a_i x_i\right) \quad (11)$$

$$L_{sum} = \sum_{i=1}^{n} L_i = 10 * \log_{10}\left(\prod_{i=1}^{n} 10^{0.1*L_i}\right) = 10 * \log_{10}\left(\prod_{i=1}^{n} x_i\right) \quad (12)$$

Eq. (11) is used each time a weighted average level is needed, for example a weekly average, where it should be considered that in a given week from Monday to Friday the traffic is high, whereas for Saturday and Sunday is lower (e.g.: $L_{Mo\text{-}Fr}$ is weighted for 5/7 and $L_{Sa\text{-}Su}$ is weighted 2/7 and the two values are added up to have the weekly average). Another example is the L_{den} indicator.

Eq. (12) is used when the level at the receiver is obtained as the sum of the source level plus a transfer function (e.g.: L_{source}=70 dB, $L_{transfer}$=20 dB, then $L_{receiver}=L_{source}-L_{transfer}$).

A matrix that contains the a_i coefficients is built up and solved, to have a single expression of the L_{den} as a function only of the input L_i values. This is essential for the correct estimation of the uncertainty. In the past, the authors themselves erroneously added more times the same standard deviation, since for example a level was recorded for nighttime road traffic noise, and then this level was combined with different transfer functions from the road to the house (receiver point).

7 Calculation of the overall standard deviation in an environmental noise measurement

A procedure to measure the average L_{den} was developed in the context of the European IMAGINE project based on the GUM and on the ISO 1996-2 [3, 4, 5]. In this procedure, general techniques to measure long-term noise values are described. The basic idea is that for each assessment position noise levels are recorded during periods (short-term or long-term) which do not coincide with the entire year. Statistics is then used to describe the mean and the uncertainty associated to the measurement. The mean is obtained as a combination of different short term measurements under different source conditions (different traffic or operating conditions) and mixed with the occurrences of the different meteorological propagation conditions. In general, the source is measured close enough to avoid any meteorological effects, then the propagating part is measured, and finally corrections are applied, like the correction for the residual noise and for the position in front of the façade, or correction for the atmospheric absorption [6, 7]. Classes of emission and propagation are filled in with the corresponding log-normal means and mean boundaries and afterwards added up. A very simplified example of a combination of these measurements is given in figure 1. In a real noise measurement like the road traffic noise measurement campaigns performed within the IMAGINE project, the number of classes was as follows:

Source class: 5 classes from Monday to Friday, 5 classes Saturday, 5 classes Sunday;

Propagation: 4 meteorological classes in summer time and 4 in winter time;

a_i coefficients (% time) – 120 percentages (15 traffic X 8 propagation)

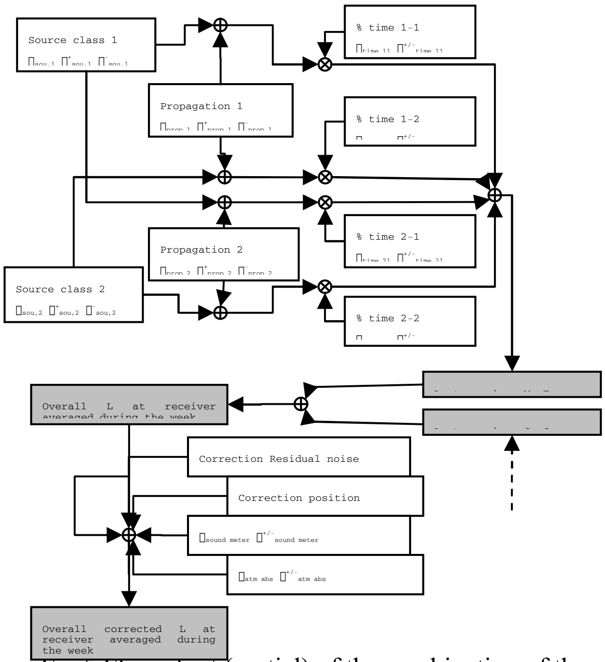


Fig.5 Flow chart (partial) of the combination of the different source and propagation classes.

The calculation chart illustrated in Fig. 5 considered the following uncertainties:

- source daily variation;
- source seasonal variation;
- propagation meteorological day/evening/night variations;
- propagation meteorological seasonal variations;
- meteorological parameters uncertainty;
- residual noise at the receiver;
- residual noise at the source;
- position of the microphone in front of the façade;
- microphone class 1 uncertainty.

8 Conclusion

This article explored the different approaches to use for assessing a long-term environmental noise level. Different means are discussed as a function of the distribution of values and the requirements of the Environmental Noise Directive (2002/49/EC). A methodology was then presented to address the overall standard deviation. Some questions still remain open, namely concerning the correct approach for calculating the mean, the approach for calculating the single standard deviation and for adding up the several standard deviations, therefore no definitive solutions are given. This work is still under development and it is expected to be concluded in the next months after having reached wide scientific consensus on the methodology to be finally adopted.

References

[1] DIN 55 303 -2 “Statistical interpretation of data; tests and confidence intervals relating to expectations and variances”, May 1984.

[2] VDI 3723 “Anwendung statisticher Methoden bei der Kennzeichnung schwankender Gerauschimmissionen”, May 1993.

[3] GUM, Guide for the expression of uncertainty in measurement (GUMBIPM, IEC, IFCC, ISO, IUPAC, OIML)

[4] H. Jonasson et al., IMAGINE report, IMA32TR-040510-SP10 “Determination of L_{den} and L_{night} using measurements”, 2006

[5] M. Paviotti, S. Kephalopoulos, IMAGINE report IMA30TR-061016-JRC01, “Test of the measurement standard concerning the determination of L_{den} and L_{night}”, 2006

[6] S. Kephalopoulos, M. Paviotti, D. Knauss and M. Bérengier, “Uncertainties in long-term road noise monitoring including meteorological influences”, Noise Control Eng. J. 55 (1) 2007

[7] M. Paviotti, S. Kephalopoulos, A. Iacoponi, “Techniques to measure the European L_{den} noise indicator for major sources in complex urban environments”, ICSV13 Vienna (Austria), 2-6 July 2006