MEASUREMENT UNCERTAINTY OF FAÇADE SOUND INSULATION: A COMPARISON BETWEEN ANOVA AND RRT UNCERTAINTY EVALUATIONS.

C. Scrosati and F. Scamoni

Construction Technologies Institute, National Research Council of Italy, San Giuliano Milanese, Milan, Italy email: c.scrosati@itc.cnr.it

W. A. Whitfield

Noise.co.uk Ltd, Bretford, Warwickshire, England: email: bill@noise.co.uk

B. M. Gibbs

Acoustic Research Unit, University of Liverpool, England: email: bmg@liverpool.ac.uk

In recent years the attention given to the measurements in the low frequency range has considerably increased, as well as their uncertainty evaluation; nevertheless, the uncertainty of field measurements, in particular façade sound insulation, has not been comprehensively investigated. ISO 12999-1 gives the uncertainty for airborne and impact sound insulation. This specific standard, however, is inaccurate as far as the façade sound insulation is concerned, because its uncertainty is considered equal to the airborne sound insulation uncertainty; indeed, the façade sound insulation measurement method is extremely different from the airborne sound insulation measurement method for partition walls and floors. This study analysed the uncertainty of the measurement method of façade sound insulation for field measurements, with the global loudspeaker method. The uncertainty evaluations were analysed using both advanced analysis of variance (ANOVA) techniques and Round Robin Test (RRT) uncertainty evaluation. Also, a comparison between these two uncertainty evaluation methods was done.

Keywords: façade sound insulation, uncertainty, round robin tests, analysis of variance

1. Introduction

When reporting the result of the measurement of a physical quantity, it is compulsory that some quantitative indications of the quality of the result be given so that those who use it can assess its reliability. Without such indications, measurement results cannot be compared, either with one another or with reference values given in a specification or standard. It is therefore necessary, in order to characterize the quality of the result of a measurement, to evaluate and to express its uncertainty. In general, uncertainties should preferably be determined following the principles laid down in ISO/IEC Guide 98-3 [1], the Guide to the expression of uncertainty in measurement (GUM:1995). According to current knowledge, it seems impossible to formulate these models for the different quantities in building acoustics. Therefore, the concepts of repeatability and reproducibility stated in ISO 5725 [2] are necessary to determine the uncertainty of building acoustics measurements. Beside these standards, for the identification of the different contributions to the uncertainty of building acoustics measurements, the advanced analysis of variance (ANOVA) could be used [3].

This paper looks at the uncertainty associated to the field measurements of façade sound insulation, analysing and comparing the uncertainty of a façade Round Robin Test (RRT) [4] and a façade

Gauge Repeatability and Reproducibility study (GRR) using both ISO 5725 and ANOVA calculation methods.

2. Methods of calculating uncertainty

The two dominant current methods for calculating uncertainty are ISO 5725 [2] and GUM [1], and a good comparison of these methods and their strengths and weaknesses is offered by Deldossi and Zappa [5]. Lyn et al. [6] proved, by a case study, that the estimate of sampling uncertainty made using the modelling approach (GUM) resulted to be six times larger than that found using the empirical approach (based on an experimental design as ISO 5725). The difficulty in establishing reliable estimates for the input variable for the modelling approach is thought to be the main cause of the discrepancy and the empirical approach to uncertainty estimation was recognized to be generally the one providing the more reliable estimates.

The other empirical method is the ANOVA approach. The main advantages of ANOVA are listed by Deldossi and Zappa [5] and include the ability to determine the contribution of the operator and part and operator by part interaction. The ANOVA method used for Measurement System Analysis (MSA) is also known as a Gauge Repeatability and Reproducibility study (GRR) with the term “gauge” referring to the measurement instrument. The appropriate GRR special application for the purpose of this research, is described by Burdick et al [7] as a Balanced Two Factor Crossed random model with interaction and it informs this research on achieving an accurate and reliable estimate of the variability in the measurement process due to the part, operator and instrument.

The scope of this research is to compare the uncertainty results obtained from ISO 5725 and ANOVA approaches. One of the major differences between these two approaches is the Design Of Experiment (DOE). The ISO 5725 approach needs a Round Robin Test (RRT) design, while the ANOVA approach needs a GRR design. These two DOEs are shown in the following subsections.

2.1 RRT DOE

Generally speaking an RRT is a test consisting of independent measurements executed several times by different operators. The variability between tests performed by different operators and/or with different equipment will usually be greater than the variability between tests carried out within short interval of time by single operator using the same equipment. The general term for variability between repeated measurement is precision. Repeatability and reproducibility are the two extremes of precision, the first describing the minimum and the second the maximum variability in results.

The reproducibility standard deviation is defined as [2]:

$$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2} \tag{1}$$

where σ_L^2 is the between-laboratory variance and includes the between-operator and between-equipment variabilities, and σ_r^2 is the repeatability variance, which is the arithmetic mean of the within-laboratory variances. In this definition reproducibility contains repeatability and therefore must be greater in magnitude.

The Doe of a RRT is based on the following [2]:

1. the test method under investigation has to be one that has been standardised;
2. the samples for the precision experiment have to be identical or – in case of discrete objects that are not altered by testing – the samples could be exactly the same in the different laboratories;
3. the statistical model for estimating the accuracy of a measurement method assumes that every test result, *y*, is the sum of three components:

$$y = m + B + e \tag{2}$$

where, for the particular material tested, m is the general mean (expectation), B is between-laboratory variation and e is the random error occurring in every test;

4. the laboratories should be chosen at random from all laboratories using the measurement method;
5. the choice of number of laboratories is a compromise between availability of resources and a desire to reduce the uncertainty of the estimates to a satisfactory level. It is common to choose a value of p (number of laboratories) between 8 and 15;
6. the choice of replicate number n, depends on the fact that, if the between-laboratory standard deviation (σ_L) is larger than the repeatability standard deviation (σ_r), as is often the case, little is to be gained by obtaining more than n=2 tests per laboratory per level.

An RRT assesses the uncertainty of measurement methods with a reference value. One of the main aspects is the determination of this reference and its uncertainty. To minimize the uncertainty of an inter-laboratory test (ILT) a reliable – with low uncertainty – reference value is necessary. Due to the typology of the sample test in acoustic measurements, a reference value does not exist; therefore an estimated value is used. The best measuring reference is the mean value.

In building acoustics, the choice of number of laboratories and number of replicates is laid down in ISO 12999-1 [8], the standard on determination of measurement uncertainty in building acoustics. The number of laboratories should be at least p = 8 and the number of test results from each laboratory should be at least n = 5; the combination of p and n should be chosen so that:

$$p(n-1) \succ 35 \tag{3}$$

Concerning the field measurements, in ISO 12999-1 [8] was introduced, for the first time, the *in-situ* standard deviation, s_{situ}, which is an intermediate condition between repeatability and reproducibility standard deviations. The *in-situ* standard deviation condition refers to tests performed on exactly the same object, in the same location, by different operators using different equipment.

For façade RRT, being the same façade in the same location, the *in-situ* standard deviation is calculated.

2.2 GRR DOE

The GRR applies ANOVA to a design of experiment that is described as a balanced two-factor crossed random model with interaction. That is, every level of one factor is run with every level of another factor (crossed) and each measurement is repeated the same number of times (balanced), e.g. every part in the test sample is measured by every operator the same number of times. This DOE is used to draw out the ‘factors’ influence on the measured results. Like a RRT it attempts to assess the same uncertainty due to repeatability and reproducibility but the experimental design also allows the user to draw out the contributions of components of variance due to the instrument, operator and part being measured as well as any interaction that may have occurred in the experiment between the operator and the part. As more than one part is measured it is particularly suited to the field measurement of sound insulation, indeed if the parts selected for measurement are identical in construction as well as the shape and size of the room, the measurement uncertainty of the construction itself can be determined and there is evidence to suggest that this may have its own uncertainty signature particular to the construction [3].

For our experiment, we can model the measurement X by operator i on part j at replication k by:

$$X_{ijk} = \mu + O_i + P_j + (OP)_{ij} + R_{k(ij)} \quad \begin{cases} i = 1, \ldots, 5 \\ j = 1, \ldots, 6 \\ k = 1, 2 \end{cases}$$

, where O_i, P_j, $(OP)_{ij}$ and $R_{k(ij)}$ are random variables corresponding to the operator, the part, the operator by part interaction, and the measurement replications. We assume these variables are independent of each other, and normally distributed with mean 0 and constant variances σ^2_O, σ^2_P, σ^2_{OP} and σ^2_R respectively. μ refers to the overall mean of all the observations X_{ijk} .

We can use the values of S^2 (sample variance), in order to estimate for expected mean square error. Using this estimate we can then further estimate the variance components σ^2_O, σ^2_P, σ^2_{OP} and σ^2_R. After these variance components have been calculated, we can estimate the values of the variance due to repeatability and reproducibility as:

$$\hat{\sigma}^2_{repeatability} = \hat{\sigma}^2_R$$

$$\hat{\sigma}^2_{reproducibility} = \hat{\sigma}^2_O + \hat{\sigma}^2_{OP}$$

NB: Note that for a GRR study the reproducibility does not include repeatability within the definition. The operator by part interaction (OP) can arise from differences in operator's measurements for some parts, but not for others. We can use analysis of variance to determine whether this interaction is significant.

3. Façade RRT

In the RRT [4] experiment, nine teams, coordinated by ITC-CNR – Construction Technologies Institute of the Italian National Research Council – were involved, each of them operating with its own equipment. The building element tested was a prefabricated concrete façade with a 4 mm single glazing wood-aluminium frame window with a MDF (Medium Density Fibreboard) shutter box. The façade is situated at the first floor. In a first RRT analysis approach [9], the low-frequency (LF) bands (50, 63 and 80 Hz) were not included in the uncertainty calculation. As the interest in the LF has grown in recent years, in the expanded and revised version [4] of the paper [9], the LF uncertainties were reported and analysed. The quantity analysed in the RRT is the standardized level difference of façade $D_{ls,2m,nT}$. The subscript *ls* indicates that a loudspeaker was used instead of the real traffic noise (*tr*). In fact the aim of the RRT is the determination of the uncertainty of the measurement method of façade sound insulation. Even if ISO 16283-3 [10] suggests to use the real traffic for whole façade measurements because it is the most accurate method to estimate the outdoor/indoor difference under actual traffic conditions, in this study the global method with loudspeaker was used. In fact, the traffic noise may not be constant during a day or a week and so its repeatability is not known, and it can not be used for the purpose of the study.

3.1 ISO 5725 results

In fig. 1, the uncertainty of façade sound insulation, calculated following the principles laid down in ISO 5725 [2], in terms of repeatability (s_r) and in-situ (s_{situ}) standard deviations are shown [4].

The data of this RRT [4] were also analysed with the Functional Data Analysis (FDA), and it was found [11] that more than half of the between-laboratory variability is explained by the first principal component. This component is negative over the whole frequency range, which indicates a shift of the laboratory-specific mean, with respect to the general mean, in the same direction throughout. It reaches a (negative) peak on the lowest frequencies whose magnitude is four times higher than any other peak.

This means that the greatest variability between laboratories will be found by heavily weighting the lowest frequencies, with only a light contribution from the other frequencies. In short, the quantity $D_{2m,nT}$ is more variable across laboratories on the lowest frequencies. This is shown in Figure **1**.

Regarding low frequencies, a RRT [12] for the comparison of the standard measurement procedure and the low frequency procedure stated in ISO 16283-3 [10] was performed and the results of this RRT will be the subject of comparison between ANOVA and ISO 5725 uncertainty calculation method of the expanded version of the present study, that is being currently drafted .

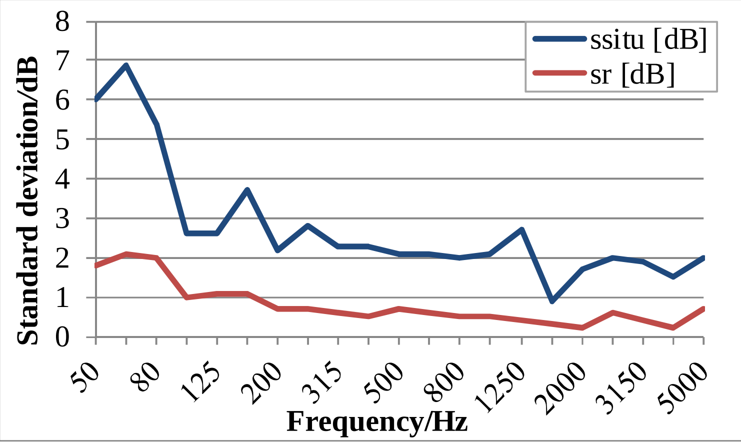

Figure 1: s_{situ} (blue) and s_r (red) of $D_{ls,2m,nT}$ of RRT of façade [4].

4. Façade GRR

The Gauge R&R study contains two repeat tests on 6 different facade elements by 5 different operators. The residential home facades measured had different room sizes, internal finishes and areas of glazing so the 'part' or facade test element, was expected to vary. The DOE attempts to assess the same uncertainty due to repeatability and reproducibility but the experimental design also allows the user to draw out the contributions of components of variance due to the instrument, operator and part being measured as well as any interaction that may have occurred in the experiment between the operator and the part.

4.1 ANOVA results

Figure **2** shows the uncertainty of façade sound insulation measured in the field, calculated using GRR and ANOVA. In this case the variance of the part being measured is clearly the major influence on the measured results and this was expected given the significantly different test situations.

The red line is the total repeatability and reproducibility variance combined and this can be broken down further into its component parts and is shown in Figure **3**.

In this case the reproducibility variance component is split into operator variance and that associated with interaction. In this case the experiment appears not to suffer too badly from interaction between the operator and the part being measured although it does feature to some extent in the 250Hz band where the interaction variance component is higher than the variance of the operator. Figure **2** & Figure **3** are a good illustration of how the measurement uncertainty associated with field testing facade sound insulation can be broken down into component parts without no additional testing.

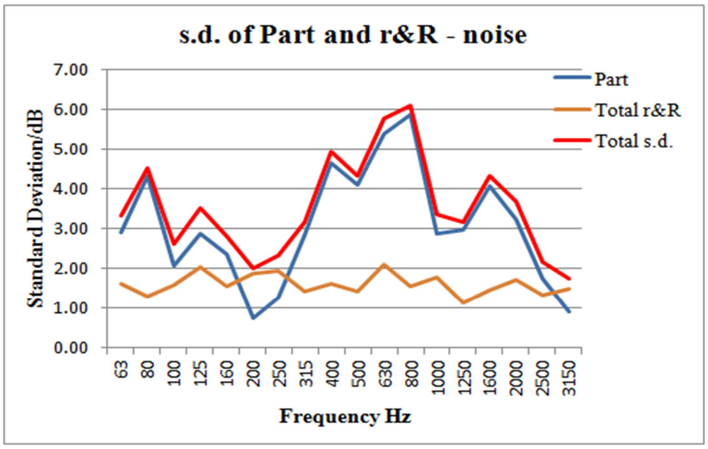

Figure 2: Graph of part, Total r&R and sum of both for each third-octave band

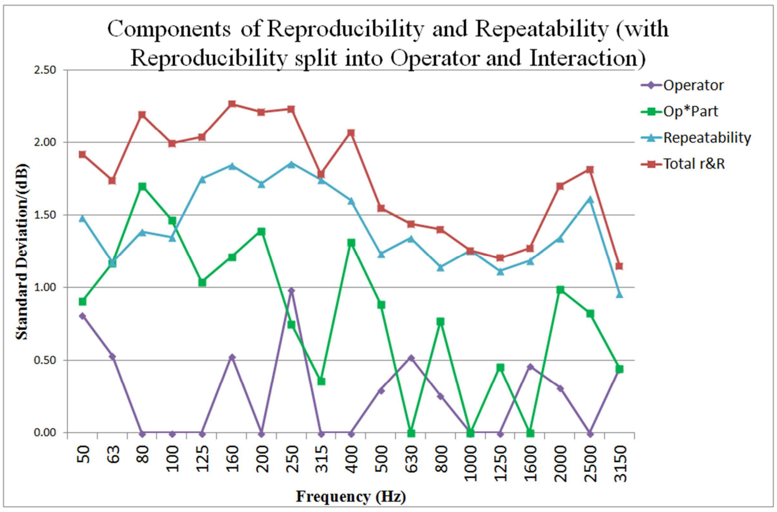

Figure 3: Graph of terms that contribute to Total r&R for each third-octave band, after removal of outliers

5. Discussion

It is possible to overlay the repeatability and reproducibility variance components for both the RRT and GRR global test data. For ease of comparison the variance components have been analysed using the same definitions for r&R where R is the between –lab variance. Taking the repeatability variance component first.

It is immediately apparent that the repeatability of each experiment follows different patterns with change in frequency. The repeatability of the RRT (Figure **4**) is much lower than the GRR experiment across all higher frequencies, whereas in the lower frequencies it is much higher than that of the GRR. The way outliers are dealt with could be behind these differences however. In the noise.co.uk GRR the method was more subjective, meaning data points in the lower frequencies could be classified as outliers even if Cochran's test did not suggest so. As such, repeatability in the GRR is lower in these low frequencies. However, when all operators showed large standard deviations in the middle frequencies for the GRR, then outliers could not be identified and the repeatability remained high.

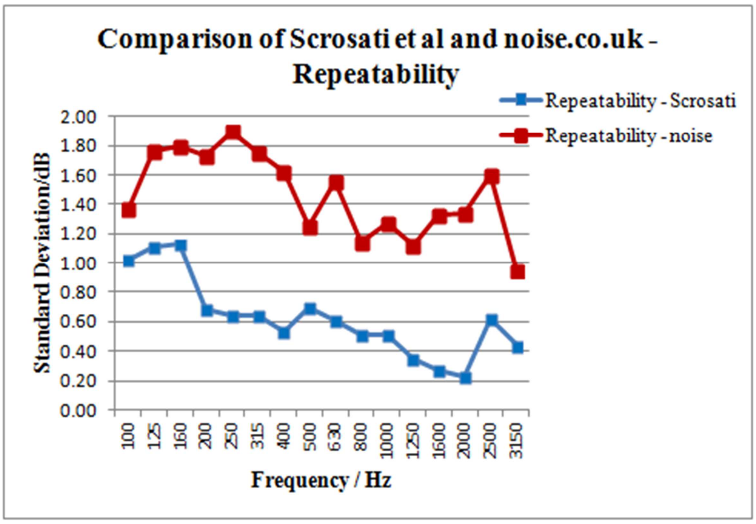

Figure 4: Comparison of repeatability in each experiment, using ISO 5725

This repeatability being higher than that of the RRT could be partly due to operator experience and the fact the experiment was undertaken on a working construction site with imperfect background noise conditions that possibly contributed to the variability for each test. It was noted that the repetitions did not necessarily take place consecutively on site due to time constraints and on site working and as repeatability definitions usually rely on the repetitions taking place over a relatively short time period this may be different to the sampling in the laboratory experiment.

In the case of reproducibility there is a complete contrast in results where GRR field test data has a lower value for the reproducibility component of variance then the RRT.

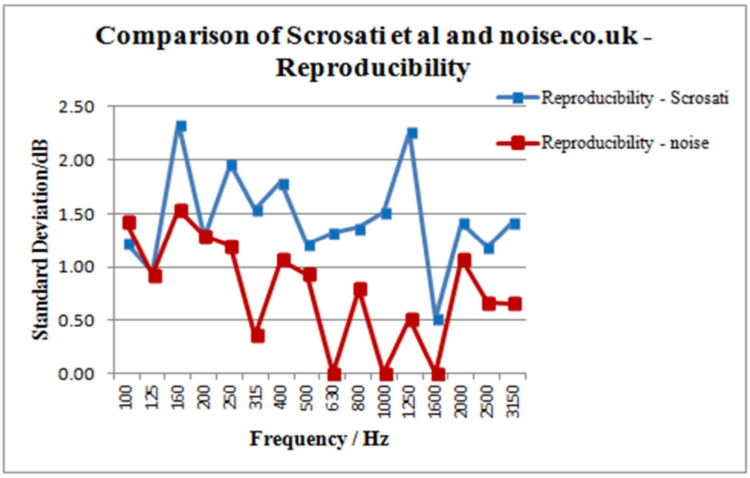

Figure 5: Comparison of reproducibility in each experiment, using ISO 5725

This contrast between variance associated with r & R offers a closer comparison between the experimental data results when the measurement system uncertainties are combined in Figure **6**.

We see the comparison of the total repeatability and reproducibility (total r&R) in the frequency bands between 100Hz and 3150Hz. NB: The low frequency (50-80Hz) bands have been ignored as the RRT showed much higher variance than the GRR and therefore the scale would have caused the graph to be unable to show the patterns in the higher frequencies.

What is clear is that total r&R is similar between the two experiments, something that couldn't be said by just looking at repeatability or reproducibility. This may be due to the fact that sometimes variance is counted as interaction variance in the GRR, mainly because only one replication is left for a particular operator when an outlier has been removed. It is clear that the two experiments follow very similar patterns across this range (except at 1250Hz). This variation may point to facade measurement having a signature and as such we gain an idea of the r&R uncertainty profile for the façade measurements. It is clear the r&R variance in the measurements reduces slightly overall with frequency increases. The factors causing this variation appear consistent across each experiment, likely being related to the variance associated with different meters and different measurement positions within the rooms rather than on site conditions and other factors that contribute to 'outlier' data.

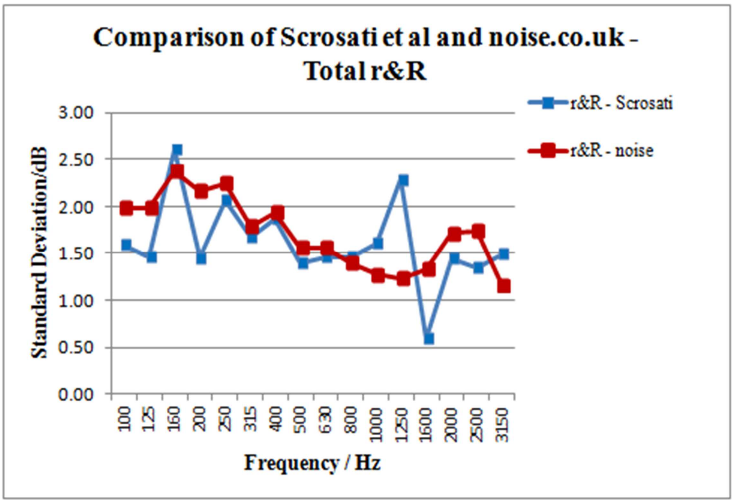

Figure 6: Comparison of total r&R in each experiment, using ISO 5725

6. Conclusion

Two methods for determining uncertainty in the measurement of facade insulation have been undertaken separately, one in Italy and one in England. One focussed on ISO 5725 DOE and was carried out on a single 1st floor facade element on a full scale experimental building at ITC-CNR. The other involved the field testing of a residential housing site during construction where six different room facades were measured. Both experiments involved multiple operators undertaking repeat measurement of the facade/s. The total uncertainty associated with the field testing GRR DOE was clearly due to the variability of the performance of the part and as only one part was involved in the laboratory RRT is seemed sensible to limit the comparison of test data to that due to the measurement system alone. Outliers were identified and removed from both studies using a practical assessment rather than a deterministic statistical outlier identification test. ISO 5725 was used to calculate the uncertainty in each case to reduce dissimilarities and the definitions of repeatability and reproducibility aligned to allow this comparison.

The comparison of variance components associated with repeatability and reproducibility yielded contrasting results mainly due to the repeatability component being significantly and unexpectedly higher than the reproducibility component in the GRR field experiment, this apparent anomaly is due in part to the fact that reproducibility does not contain repeatability in the ANOVA analysis. However, the combined uncertainty of the measurement 'system' (Total r&R) proved to exhibit comparable levels of variance across the frequency range and apart from one or two exceptions a similar shape. These are early reported observations and further work will be carried out on this data.

REFERENCES

1 ISO/IEC Guide 98-3:2008, *Uncertainty of Measurement - Part 3: Guide to the Expression of Uncertainty in Measurement (GUM:1995).*

2 ISO 5725 part 1 and 2:1994 *Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions and Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*.

3 Whitfield W.A. and Gibbs B.M. Measurement Uncertainty in Airborne Sound Insulation: Uncertainty Components in Field Measurement, *Proceedings of the 22nd International Congress on Sound and Vibration*, Florence, Italy, 12–16 July, (2015).

4 Scrosati C., Scamoni F., Zambon G., Uncertainty of facade sound insulation in buildings by a round robin test, *Appl. Acoust.* **96**, 27–38, (2015).

5 Deldossi, L., Zappa, D., ISO 5725 and GUM: comparison and comments. *Accreditation and Quality Assurance*, **14**, 159–166, (2009).

6 Lyn J.A., Ramsey M.H., Damant A.P., Wood R. Empirical versus modelling approaches to the estimation of measurement uncertainty caused by primary sampling, *Analyst* **132**, 1231–1237, (2007).

7 Burdick R.K., Borror C.M, Montgomery D.C., *Design & Analysis of Gauge R&R Studies*. ASA-SIAM Series on Statistics and Applied Mathematics (2005).

8 ISO 12999-1:2014, *Acoustics - Determination and Application of Measurement Uncertainties in Building Acoustics - Part 1: Sound Insulation.*

9 Scamoni F., Scrosati C., Mussin M., Galbusera E., Bassanino M., Zambon G. and Radaelli S. Repeatability and reproducibility of field measurements in buildings, *Proceedings of 8th European Conference on Noise Control 2009, EuroNoise09*, Edinburgh, Scotland, UK, 26-28 October, (2009).

10 ISO 16283-3:2016, *Acoustics - Field measurement of sound insulation in buildings and of building elements - Part 3: Façade sound insulation,* ISO, Geneva, Switzerland (2016).

11 Argiento R., Bissiri P.G., Pievatolo A., Scrosati C., Multilevel functional principal component analysis of facade sound insulation data, *Qual. Reliab. Eng. Int.* **31**, 1239–1253,(2015).

12 Scrosati C., Scamoni F., Prato A., Secchi S., Fausti P., Astolfi A., Barbaresi L., D'Alessandro F., Di Bella A., Schenone C., Zambon G., Uncertainty of Façade Sound Insulation by a Round Robin Test. Evaluations of Low Frequencies and Single Numbers, *Building and Environment* **105**, 253–266, (2016).