A Method to Quantify Uncertainties in Airtightness Measurements: Zero-Flow and Envelope Pressure

Authors: Martin Prignon (UCL), Arnaud Dawans (EJD), Sergio Altomonte (UCL) and Geoffrey van Moeseke (UCL)

Abstract

In airtightness measurements using the fan pressurization method, the calculation of uncertainties does not take into account the zero-flow pressure approximation. It was recently suggested that neglecting the uncertainties related to zero-flow pressure approximation could be an unrealistic hypothesis. In this study, a method for quantifying this source of uncertainty is proposed, illustrated and discussed. The method is applied to a series of 31 zero-flow pressure tests performed on a newly-constructed apartment within a period of 15 days in Brussels, Belgium. For each test, 32 different zero-flow pressure approximations were compared. Since the data had a nested structure, multiple multi-level models were used for their analysis. The results show that, for the tested building, the zero-flow pressure approximation uncertainty is 0.45, 0.91 and 1.52 Pa respectively under low-, medium- and large-wind conditions. These uncertainties can be reduced to 0.42, 0.80 and 1.39 Pa when using longer zero-flow pressure measurement periods. As a comparison, uncertainty in pressure measurement at 50 Pa due to the equipment is 0.25 Pa. The uncertainty of zero-flow approximation makes the envelope pressure uncertainty non-negligible, therefore having an impact on the regression technique used to determine the building airtightness.

Key words: airtightness measurement, fan pressurization test, uncertainties, zero-flow pressure approximation, multi-level modelling, envelope pressure.

1 INTRODUCTION

The measurements of air tightness and infiltration in buildings is an issue of major concern. Kalamees [1] and Jokisalo et al. [2] showed that, in cold climates, air infiltration could be responsible for up to 30% of heating loads. Air infiltration has also been shown to have an impact on the efficiency of ventilation systems, on occupant comfort, and on acoustic insulation [3]. In 2014, Sadauskiene et al. observed that actual building energy use can significantly differ from the value calculated with regulation methods when airtightness is ignored [4]. In this context, however, there is a general consensus that no predictive model can currently replace airtightness measurements [5, 6], this leaving a gap to fill in terms of the availability of reliable and easily implemented modelling methods.

Although fan pressurization tests (also called blower door tests) have been used for almost 50 years [3], there is an important lack of knowledge when dealing with the uncertainties related to these procedures. Several studies have analyzed total uncertainties in repeatability (i.e., successive measurements carried out under the same conditions) and reproducibility (i.e., successive measurements under changing conditions) testing [7-11]. Research has also focused on multiple sources of uncertainty [12-16] among which wind is likely the most challenging, since it can cause precision and bias errors in both envelope pressure and airflow measurements [14].

In envelope pressure measurements, the impact of wind – and more generally of weather conditions – is taken into account through the calculation of zero-flow pressure. The pressure difference measured during a fan pressurization test is the combination of the differential in pressure induced by the fan and that induced by climatic conditions (hence, the zero-flow pressure). One important step in the evaluation of envelope pressure difference is the separation of these two induced pressures. In practice, the measurement of zero-flow pressure is not possible during the fan pressurization test since this requires the fan to be stopped and sealed. The European standard EN ISO 9972:2015 [17] suggests to measure the zero-flow pressure before and after the test, and to assume that the zero-flow pressure during the test is the average of these measurements (Equation 1).

Equation 1

$$\Delta p_0 = \frac{\Delta p_{0,1m} + \Delta p_{0,2m}}{2} = \frac{1}{2}\left(\frac{\sum_{i=1}^{N} \Delta p_{0,1i}}{N} + \frac{\sum_{j=1}^{M} \Delta p_{0,2j}}{M}\right)$$

Where: Δp_0 is the zero-flow pressure approximation; $\Delta p_{0,1m}$ and $\Delta p_{0,2m}$ are the averages of zero-flow pressure measurements during a given measurement period, respectively before and after the fan pressurization test; $\Delta p_{0,1i}$ and $\Delta p_{0,2i}$ are the zero-flow pressures measured at each point during these periods; and, N and M are the number of measurements made within these periods. Standard EN ISO 9972:2015 requires that at least 10 measurements are made during a period of minimum 30 seconds [17].

This method leads to an important source of uncertainty because, in reality, the weather conditions, and therefore the zero-flow pressure, are not constant during the fan pressurization test. Uncertainties in zero-flow pressure have been studied by other authors [15], but Delmotte was the first to raise the issue of uncertainties due to zero-flow pressure approximation, although the sources of such uncertainties were not quantified [13].

In this paper, we propose and illustrate a new method for the quantification of uncertainties in zero-flow and envelope pressure, and we discuss the findings from its application to a series of 31 tests performed on a single apartment in Brussels, Belgium, within a period of 15 days in October 2017. The paper is structured as follows. In the Methodology section, we illustrate the tests performed, drive a comparison between different approximation techniques, and describe the statistical approach behind the newly-proposed quantification method. In the Results section, we provide the outcomes of the tests, the uncertainties quantified by the application of the new method, and the impact of the zero-flow pressure approximation uncertainties on envelope pressure uncertainties. In the Discussion section, we highlight the impact of our findings against the relevant literature and illustrate critically

the limitations of this study. Finally, in the Conclusions, we summarize the outcomes of this work, describing the further research needed in the field of airtightness measurement uncertainties.

2 METHODS

This section is divided in three parts. The first describes the weather data, the equipment used for the series of tests, and the apartment where measurements took place. The second offers a comparison between different zero-flow pressure approximation techniques and the indicator used for the evaluation. The last part focuses on the statistical analysis and on the justification of methodological choices.

2.1 MEASUREMENT METHOD

2.1.1 Zero-Flow Pressure Test

A zero-flow pressure test consists in the measurement of zero-flow pressure every second during three successive periods. In this paper, the initial and last periods have been respectively labelled the first and second “approximation periods”, while the second period (i.e., the middle one) has been called the “fictitious period” (Figure 1). The approximation periods correspond to the times when measurements are used to perform zero-flow pressure approximations. The fictitious period corresponds to the duration that a typical fan pressurization test would take in practice.

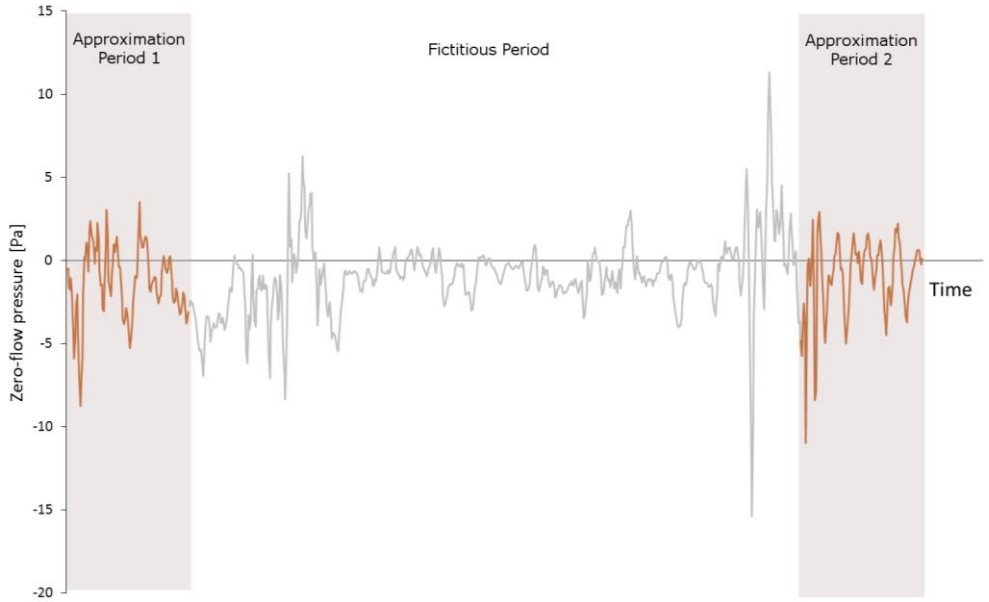


Figure 1 – Zero-flow pressure measured during a period of 840 seconds (one measurement every second), illustrating the three successive periods of a zero-flow pressure test

The paper by Delmotte [13] described for the first time a zero-flow pressure test. To remain consistent with his work, the present study adopted a fictitious period of 600 seconds. The duration of the fictitious period has a strong impact on the uncertainty. Indeed, a longer duration of the fictitious period might imply greater likelihood of weather variations of higher degree. 600 seconds correspond to the measurement of 10 couples with 30-second duration and 30-second delay between measurements. These three values (i.e., number of couples, measurement duration, and delay between measurements) could vary depending on the operator, equipment and weather conditions.

Approximation periods of 120 seconds were selected instead of the 30 seconds used in [13]. This is because Delmotte was interested in the uncertainties related to minimum standard requirements, while the present study focuses on the impact of approximation period duration on zero-flow pressure uncertainty.

2.1.2 Measurement Equipment and Weather Data

A DG-700 pressure gauge (last calibration in June 2017) from The Energy Conservatory (TEC) measured the pressure difference between inside and outside the apartment. According to the manufacturer, the standard uncertainty (i.e., the expected standard deviation for a large number of repeated measures) of the DG-700 is the greatest between $\pm$ 0.5% of the reading and $\pm$ 0.1 Pa, and it has a resolution of 0.1 Pa [18]. For the DG-700 pressure gauge, this uncertainty takes into account the uncertainty of the pressure reference used to calibrate the gauge, the uncertainty of the gauge itself (including hysteresis effect), the temperature effect on the sensor, and the drift of the sensor over time [19].

During the tests, a weather station (Ahlborn FMD 760) placed on the roof above the apartment measured the outside air temperature, the wind speed and the wind direction every 10 seconds. The weather station provided 10-minute averages for the outside air temperature (T_{ext}), the mean wind speed ($\overline{v_w}$) and the mean wind direction. It also provided the maximum wind speed measured during these 10-minute periods. A thermometer (Testo 417) was used to measure the inside air temperature (T_{int}) before each test with an accuracy of $\pm$ 0.5°C and a resolution of 0.1 °C. Table 1 shows, for each weather variable, the minimum, maximum, mean (M) and standard deviation (SD) of all the values recorded by the weather station (10-minute samples) and the thermometer (before each test) during the 15 days. Wind direction was mostly from west and southwest.

	Min	Max	M	SD
T_{ext} [°C]	8.8	24.3	14.2	2.77
T_{int} [°C]	20.5	22.0	21.3	0.57
$\bar{v}_w$ [m/s]	0	3.8	1.3	0.75
$\max(v_w)$ [m/s]	0.3	8.3	3.3	1.57

Table 1 – Minimum, maximum, mean (M) and standard deviation (SD) of weather-related data measured during the 15-day testing period

2.1.3 Measured Apartment

The 31 tests were performed on a newly constructed (2017) apartment within a period of 15 days in October 2017. A short testing period was chosen to avoid impacts of ageing and seasonal variations on airtightness. Since no previous research has defined an appropriate sample size for the statistical analysis of data, a large number of tests were repeated during this period. The apartment was a masonry construction of 228 m^3 with a floor area of 90 m^2 located on the second floor of a 3-storey building in Brussels. Only two perimetral walls were exposed to the outside (Figure 2). Over recent years, masonry apartments have represented a large share of new residential constructions in Brussels.

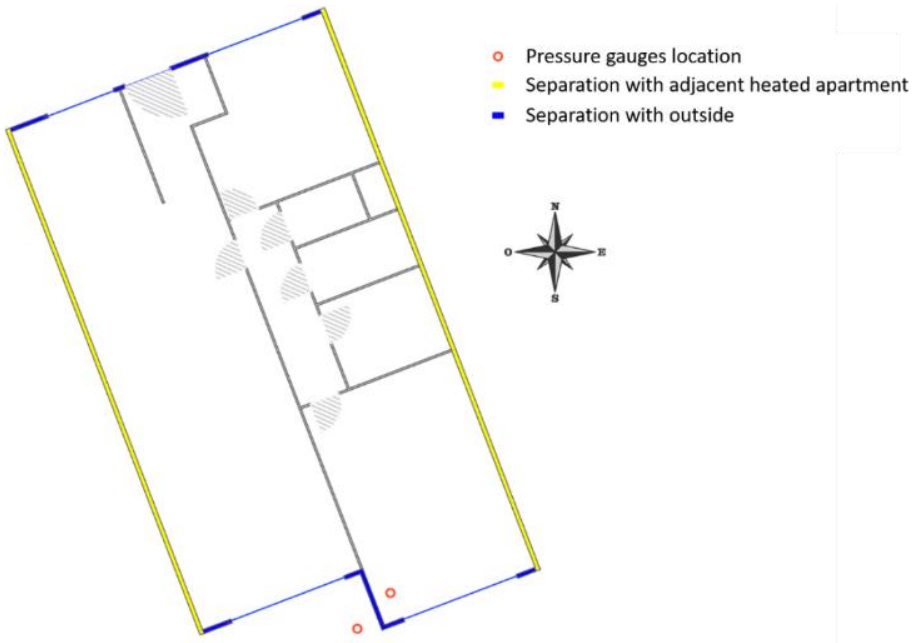


Figure 2 – Schematic plan of the tested apartment, including separation type (with adjacent heated apartment or with outside) and pressure gauges location

In the absence of a fan, if parts of the envelope are in underpressure others are in overpressure [13]. Pressure gauges placed at different locations undergo different zero-flow pressure. Therefore, the pressure gauge location is expected to have an influence on the uncertainty due to zero-flow pressure approximation. In this study, the pressure gauges were placed at the southeast façade because of the

exposed surroundings of the building. Conversely, the northwest façade opens to an enclosed space surrounded by high buildings that could induce an undesirable wind effect.

There were no airtightness requirements for the apartment itself, but the whole building (i.e., all the apartments pressurized simultaneously) had to reach a maximum air change rate of 0.6 h^{-1} at 50 Pa. The preparation of the apartment during the tests was consistent with method 1 described in ISO 9972:2015 [17].

2.2 Approximations Techniques

2.2.1 Different Approximations

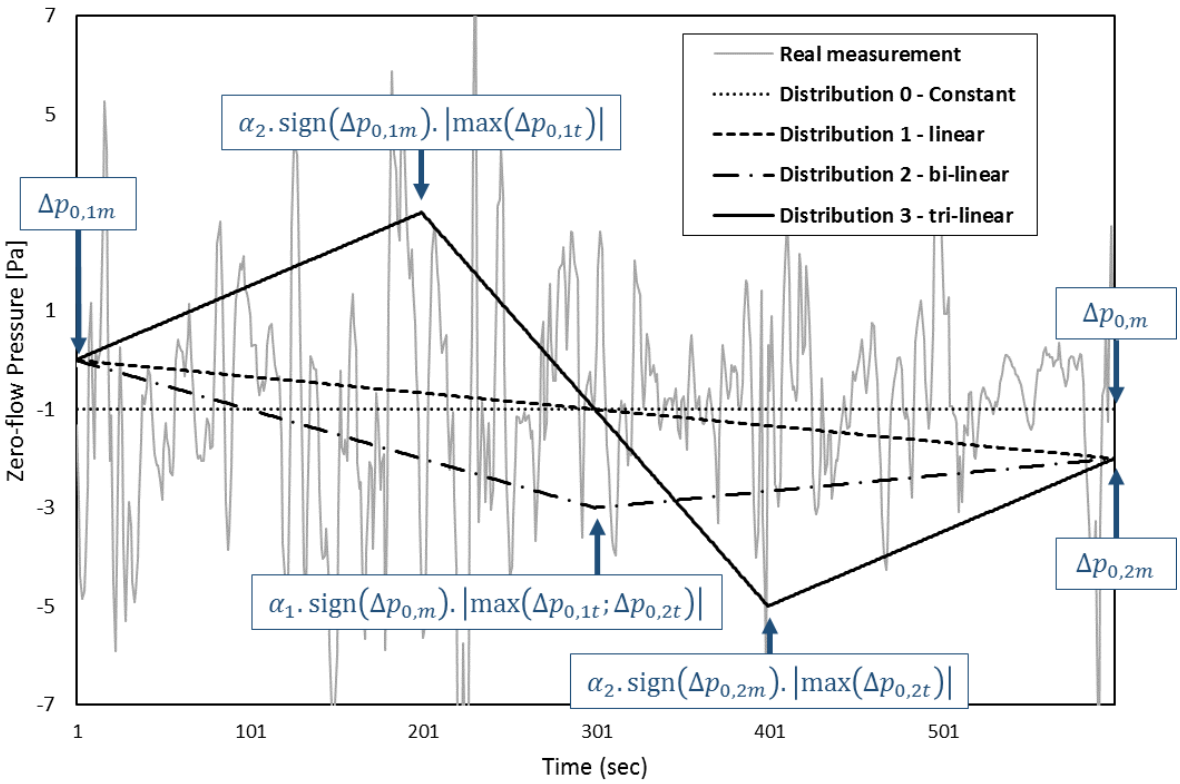
The different approximation techniques compared in this study are obtained by varying three parameters: the duration of approximation periods, the time step between measurements, and the distribution of the approximation value over time. Except for the change in the distribution, whatever the values of the duration and the time step, the approximation always fits within the requirements of the European standard ISO 9972:2015 (i.e., at least 10 measurements made during a minimum of 30 seconds) [17].

The duration of the approximation periods takes four different values: 30, 60, 90 and 120 seconds. The time step between measurements is either 1 or 3 seconds. These values were selected to cover the minimum requirements (i.e., 30 seconds with a time step of 3 seconds). The longest period of 120 seconds was chosen to avoid greater approximation periods that might have had two drawbacks. First, it would have increased the testing period. Second, it would have inflated the probability to record a weather variation during the approximation period, hence not reflecting the conditions during the fictitious period.

For each combination of parameters and for each test, the same conditions (duration period and time step) are applied on approximation periods 1 and 2. The variation of the parameters taken for these

two conditions affects the value of zero-flow pressure before and after the fictitious period, but not during it. The third parameter is the zero-flow pressure distribution, i.e. the variation of the approximation value during the fictitious period.

Figure 3 graphically represents the four different distributions compared in this study.



180 *Figure 3 – Graphical representation of the four distributions (black) for a typical zero-flow pressure test compared to the real zero-flow pressure (light grey)*

The distribution 0 (constant distribution – dotted black line) is the distribution imposed by the standards. In distribution 0, the zero-flow pressure approximation is constant during the fictitious period and its value ($\Delta p_{0,m}$) is the mean of the zero-flow pressure measurements made before ($\Delta p_{01,m}$) and after ($\Delta p_{02,m}$) the fictitious period. In distribution 1 (linear distribution – dashed black line), the approximation value during the fictitious period goes from $\Delta p_{01,m}$ at time 0 to $\Delta p_{02,m}$ at time 600. Distribution 2 (bi-linear distribution – dashed and dotted black line) goes from $\Delta p_{01,m}$ at time 0

to a peak at time 300 and then to $\Delta p_{02,m}$ at time 600. The peak is defined as a percentage (α_1 on the figure) of the maximum (if $\Delta p_{0,m}$ is positive) or the minimum (if it is negative) of the zero-flow pressure measurements made before and after the fictitious period. Distribution 3 (tri-linear distribution – solid line) starts from $\Delta p_{01,m}$ at time 0 and goes to $\Delta p_{02,m}$ at time 600, but it has two peaks at time 200 and time 400. These peaks are also percentages (α_2) of maximum (or minimum, if negative values) of $\Delta p_{01,m}$ at time 200 and $\Delta p_{02,m}$ at time 400, respectively. Even if the two peaks are different, the same α_2 is applied for both. The values of α_1 and α_2 were chosen to provide the best quality indicator (see section 2.2.2) when averaging the 31 tests. No specific distribution relevant to all the zero-flow pressure tests was observed. Therefore, the linear distribution was chosen to take into account possible changes in steady wind pressure. Such change between the beginning and the end of the fictitious period is expected to be better approximated using a linear rather than a constant distribution. The choice of bi-linear and tri-linear distributions follows the same logic, yet considering that changes in steady wind pressure work in steps instead of progressive variations. However, it should be considered that these distributions were chosen to test the research hypotheses, while other distributions might be more suitable to verify other hypotheses.

In this study, each distribution (0, 1, 2 and 3) was applied to each time step (1 and 3), within each measurement period (30, 60, 90 and 120) and for each test (1 to 31), leading to a total of 992 approximations (32 for each test).

2.2.2 Quality Indicator

In his study [13], Delmotte compared the average of zero-flow measurement during the fictitious period with the average of zero-flow approximation during the same period (Δ_{ijkl}), according to Equation 2:

Equation 2

$$\Delta_{ijkl} = \frac{\sum_{t=1}^{600} \Delta p_{0,it}}{600} - \frac{\sum_{t=1}^{600} \Delta \tilde{p}_{0,itjkl}}{600}$$

Where, $\Delta p_{0,it}$ is the real zero-flow pressure measurement at time i of test t, and $\Delta \tilde{p}_{0,itjkl}$ is the approximation of zero-flow pressure at time i of test t using the measurement period j. The parameter k represents the time step and l is the distribution.

The Δ_{ijkl} indicator gives information about the ability of an approximation technique to fit, in average, the real zero-flow pressure. However, no information is provided as to how the approximation fits locally the zero-flow pressure. For example, a linear approximation (distribution 1) with $\Delta p_{0,1m} = -1$ Pa and $\Delta p_{0,2m} = 1$ Pa would have the same Δ than a constant approximation (distribution 0) with $\Delta p_{0,m} = 0$, while the difference between approximation and real measurement at every point would be different.

Since this study focuses on the uncertainties of envelope pressure evaluation at each measurement, there is a strong interest in the ability of an approximation to fit the real zero-flow pressure at each measurement point. This was addressed by calculating the average of the differences between approximation and real measurement at every measurement point during the fictitious period (ε_{ijkl}). From this point, the "approximation quality" refers to this average, based on Equation 3:

Equation 3

$$\varepsilon_{ijkl} = \frac{\sum_{t=1}^{600} \left| \Delta p_{0,it} - \Delta \tilde{p}_{0,itjkl} \right|}{600}$$

It should be noted that the square of the difference could be used instead of the absolute value if the evaluation aimed to penalize huge gaps between approximation and real measurement. However, this

study was interested in ε_{ijkl} values since it could be used for the quantification of uncertainties (see paragraph 3.3.2).

2.3 Statistical Analysis

In this study, four different durations of approximation period are considered within each test, two time steps are tested within each duration, and four distributions are evaluated within each time step. This methodology leads to hierarchically structured data, hence requiring appropriate tools to perform a rigorous statistical analysis.

Traditional statistical tests (e.g., parametric tests such as a Student t-test or ANOVA) assume the data being independent from each other. This is not the case here since the results of different approximation techniques highly depend on the set of data used. When tests assuming independence are used to analyze dependent sets of data, there is a risk to inflate the occurrence of Type I errors. A Type I error occurs when a true null hypothesis is incorrectly rejected (i.e., an effect is statistically observed while in fact there isn't) [20]. This introduces biases in the estimation of parameters [20, 21]. In addition, the use of tests dealing with dependent data, but ignoring their clustered structure, can cause an underestimation of standard errors [22]. Conversely, the statistical tool used for the analysis should consider the dependency of results and the nested structure of data. In this study, this was addressed by using multi-level modelling (MLM) [23].

2.3.1 Multi-Level Modelling

Multi-level modelling is a method of analysis that deals with complex structures of data and recognizes their hierarchical structure by allowing for residuals at multiple levels in the hierarchy [22]. Indeed, MLM partitions the variance of the results into *between-level* (e.g., the variance of approximation quality between different tests) and *within-level* variance (e.g., the variance of approximation quality

between different approximation techniques within the same test). This partition is made at each step of the hierarchy and is mathematically expressed in Equation 4:

Equation 4

$$\varepsilon_{tjkl} = \gamma_0 + e_l + e_k + e_j + e_t$$

Where γ_0 is the grand mean (i.e., the mean when considering all the results) and terms e_l, e_k, e_j and e_t are deviations from the grand mean specific to the type of distribution, the time step, the measurement period and the test, respectively.

Besides its ability to deal with nested data, MLM allows to decide whether a model parameter can be specified as a *fixed* or *random* effect. A fixed effect is a factor that does not change over time and that is applied equally on each unit at every level in the hierarchy, regardless of the level under which it is nested [20, 24]. For example, the possible values of measurement periods remain the same (i.e., 30, 60, 90 and 120 seconds) whatever test is considered. On the contrary, random effects are expected to change within a level (i.e., from test to test) [24, 25]. For example, weather conditions are likely to vary between different tests. One main difference between fixed and random effects is in the calculation of standard errors. In MLM, other sources of uncertainties can be added by the inclusion of further random effects into the model. As a result, fixed-effect models might underestimate the standard error and increases the Type I error rate [24, 26]. When designing an experiment, it is good practice for the experimenter to decide in advance whether an effect is fixed (i.e., experimentally controlled) or random (i.e., occurs by chance).

2.3.2 MLM Comparison

The comparison of models with and without a given effect allows determining if this effect has an impact on the approximation quality. If a model with an effect has a significantly better fit than the same model without it, then it can be assumed that the effect has an impact on the approximation quality. Table 2 shows the four different models that have been compared in this study. Since the

goodness-of-fit of each model depends on the effect chosen, the order of addition of the effects has an impact on the results. In this use of MLM, the logical order is not as obvious as in other domains (e.g., the hierarchy in educational research is straightforward: country, district, school, classroom, students). The order was chosen as follows: duration of the period, number of measurements and distribution of the value computed. However, it was checked that the conclusions (i.e., the effects having a statistical and practical significance) remained the same whatever the order.

Model	Random effect	Fixed effects
Model 1	Test ID	-
Model 2	Test ID	Measurement period
Model 3	Test ID	Measurement period and time step
Model 4	Test ID	Measurement period, time step and distribution

Table 2 – Different models compared in the analysis with fixed and random effects successively added

The simplest model, *model 1*, had only the test number (Test ID) as a random effect and no fixed effect. The other models (2, 3, and 4) were obtained by successively adding three fixed-effects: the measurement period, the time step, and the distribution. Hence: *model 2* had two effects (test ID and measurement period); *model 3* had three effects (test ID, measurement period and time step); *model 4* had four effects (test ID, measurement period, time step and distribution).

The measurement period, time step and distribution were experimentally controlled and were then considered as fixed effects. On the contrary, different tests (i.e., test ID) experienced variable weather conditions and were, therefore, considered as a random effect.

In the literature, researchers have mainly used the chi-square difference (χ_d^2) test to compare the goodness-of-fit of MLM [27]. This test compares the Log Likelihood (LL) values of the restricted (i.e., less parameterized) and unrestricted (i.e., more parametrized) models based on Equation 5:

Equation 5

$$\chi_d^2 = -2LL_{restricted} - 2LL_{unrestricted}$$

Other authors have recommended the use of the Bayesian Information Criterion (BIC) instead of the chi-square difference [27, 28]. This is because BIC adjusts the Log Likelihood value considering the number of parameters and the sample size. However, the interpretation of the BIC in model comparison is very subjective and the BIC should not be used in cases where the models differ only in fixed effects [28]. Therefore, in this study the χ_d^2 test was used to test the statistical significance of the differences between the models.

Since the χ_d^2 test is a null hypothesis significance testing (NHST), other than calculating the statistical significance (*p*-value) of the differences detected, it is important to also estimate the practical relevance of their magnitude (effect size). In fact, one of the main limitations of NHST is that *p*-values depend both on the size of the sample and on that of the influence under investigation [29]. In fact, a result may be found to be statistically significant either if the effect is strong or the sample is large [30]. The effect size is a standardized measure of the difference between models [31]. Due to the nature of the data, in this study the effect size was calculated using the pseudo square partial correlation (r_p^2, Equation 6), where σ_i is the variability explained by model i [21, 25].

Equation 6

$$r_p^2 = 1 - \frac{\sigma_{unrestricted}}{\sigma_{restricted}}$$

When interpreting the size of an effect, authors often refer to the benchmarks proposed by Cohen [32, 33]. As recommended in the literature, these values are appropriate particularly in the absence of previous knowledge in the area [34]. Since this work is, to our knowledge, the first focusing on the

quantification of uncertainties in zero-flow pressure approximation, the benchmarks by Cohen have been used [33], hence defining effect sizes as small, medium or large, respectively for $r_p^2 > 0.01$, $r_p^2 > 0.09$ and $r_p^2 > 0.25$ [32].

2.3.3 Assessing the Need for MLM

One way of testing the need for MLM is by comparing the variance of the approximation quality between different tests (i.e., the between-test variance) with the variance of the approximation quality between different approximations using the same set of data (i.e., the within-test variance). This is done by computing the intra-class correlation (ICC) according to Equation 7:

Equation 7

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

Where, τ_{00} is the between-test variance and σ^2 is the residual variance (i.e., the within-test variance) [21, 25, 31].

The ICC can be defined as the amount of variation occurring between tests. A value of zero would indicate that all the variation occurs between different approximations and none of it occurs between different tests (when the same approximation is compared). In such cases, traditional statistical tools (e.g., ANOVA) can be used and MLM is not needed. A value of ICC higher than zero indicates an increase in the variation between tests, leading to postulate a violation of the assumption of independence, and therefore the need to use statistical analysis tools that are not strictly based on assuming independence in the collection of data [21, 25].

In this paper, the variance of the approximation quality between tests was 0.672 while the residual variance was 0.098 ($ICC = 0.873$). According to Julian's benchmarks [35], ICC values greater than 0.45 correspond to large intra-class correlation. However, ICC values alone are not sufficient to assess the need for MLM. The design effect quantifies the violation of the assumption of independence on standard error estimates. It provides an estimate of the multiplier to apply on standard errors to take into account the bias resulting from a nested structure of the data [21]. The design effect can be calculated based on Equation 8:

Equation 8

$$Design\ Effect = 1 + (n_c - 1)ICC$$

Where, n_c is the number of approximations per test (32, in our data).

The use of single-level modelling instead of MLM might not lead to misleading results when the design effect is smaller than 2 [21, 36, 37]. In this study, the computation of both ICC (0.873) and design effect (27.2) justified the need for MLM.

2.3.4 Covariance Structure and Parameters Estimation

One of the main advantages of MLM is the flexibility of the covariance matrix (i.e., the matrix that defines how the variances associated with each independent group are related to each other). In fact, multi-level modelling makes it is possible to specify direct assumptions regarding its structure [31]. Since the models used in our analysis featured only one random effect – and this was not a time-based study – an independent covariance matrix was used (that is, the variances of random effects are independent and have the same value) [20].

MLM can be fitted using two different methods for parameter estimation: the full information maximum likelihood (FIML) and the restricted maximum likelihood (REML) method. However, REML can only be used to compare models differing in random effects, and not in fixed effects [38, 39]. Therefore, the use of FIML was preferred in this study.

3 RESULTS

The results are presented in three sections. The first section illustrates the issue of zero-flow pressure approximation by comparing the real zero-flow pressure with the reference case (i.e., minimum standard requirements) for the 31 tests. The second section describes and analyses the results of the model comparison. The third section uses these results to quantify the uncertainty of both zero-flow pressure approximation and envelope pressure.

3.1 ZERO-FLOW PRESSURE IN FICTITIOUS TESTS AND REFERENCE CASE APPROXIMATION

Table 3 gives the minimum, maximum, mean (M) and standard deviation (SD) of: the approximation quality for the reference case applied to the 31 tests; the mean; and, the standard deviation of zero-flow pressure ($\Delta p_{0,mi}$) measured during the fictitious periods for the 31 tests.

	Min [Pa]	Max [Pa]	M [Pa]	SD [Pa]
Approximation quality	0.27	2.79	1.26	0.72
Mean average of zero-flow pressure	-2.05	0.50	-0.73	0.53
Standard deviation of zero-flow pressure	0.34	3.76	1.51	0.89

Table 3 – minimum, maximum, mean and standard deviation for (1) the approximation quality using reference case, (2) the average value and (3) the standard deviation of zero-flow pressure measurements

Comparing the mean average of zero-flow pressure and the approximation quality for the reference case illustrates the issue highlighted by Delmotte [13]. Indeed, the mean approximation quality (i.e.,

the mean difference between approximation and real zero-flow pressure measurement) using the reference case (1.26 Pa) is of considerable magnitude. As a comparison, the absolute value of the mean average zero-flow pressure measured during fictitious periods is 0.73 Pa. Clearly, these values are case-related and should not be generalized without further research. However, a large value of approximation quality can be expected whereas wind is known as one of the most important sources of uncertainty [14].

3.2 Comparison of Approximation Techniques

3.2.1 Model Comparison

Table 4 reports the results of the chi-square difference ($\chi_d^{2\,p}$) – with the interpretation of its statistical significance (NHST) – and the estimation of the pseudo-square partial correlation (r_p^2) in the comparison between models. The values of $\chi_d^{2\,p}$ and r_p^2 are computed considering both quality indicators, ε and Δ, as discussed in paragraph 2.2.2.

Model Comparison	$\chi_d^{2\,p}(\varepsilon)$	$r_p^2(\varepsilon)$	$\chi_d^{2\,p}(\Delta)^{\text{NHST}}$	$r_p^2(\Delta)$
1 vs 2	202.60 **	0.10	14.45 **	< 0.01
2 vs 3	0.844 $^{N.S}$	< 0.01	0.614 $^{N.S.}$	< 0.01
3 vs 4	4.835 *	< 0.01	4.295 *	< 0.01

$^{N.S.}$ Not significant (p > 0.05),* Significant (0.001 < $p \leq 0.05$), ** Highly significant ($p \leq 0.001$)
$r_p^2 < 0.01$: effect with non-substantive magnitude, $r_p^2 \geq 0.01$: small effect , $r_p^2 \geq 0.09$: moderate effect, $r_p^2 \geq 0.25$: large effect

Table 4 – Multi-Level Model comparison (statistical and practical significance) using ε and Δ as quality indicators

The comparison using ε as quality indicator shows that the difference between models with and without the measurement period (model 1 vs 2) is statistically significant and practically relevant ($\chi_d^2 = 202.60$, $p < 0.01$ and $r_p^2 = 0.10$). The difference between models with and without the time step (model 2 vs 3) is not statistically significant and its magnitude is non-substantive ($\chi_d^2 = 0.84$, $p = 0.36$ and $r_p^2 < 0.01$), i.e. the influence detected is not practically relevant. The difference between models with and without distribution (model 3 vs 4) is statistically significant and with an effect of non-

substantive size ($\chi_d^2 = 4.84$, $p = 0.03$ and $r_p^2 < 0.01$). The use of the Δ quality indicator leads to the detection of differences between models whose magnitude is consistently not practically relevant ($r_p^2 < 0.01$).

The random effect Test ID was useful to assess the need for MLM, although this parameter has no physical meaning. Since zero-flow pressure is induced by weather factors (wind pressure and temperature difference), which varied during the test period, the Test ID is expected to hide a random effect related to differences in conditions (leading to a strong variation of zero-flow pressure measurements). However, in this study, the focus is on zero-flow pressure and not on weather conditions. Therefore, the standard deviation of zero-flow pressure measurements during the fictitious period was used to replace the Test ID as a random effect. The 31 tests were divided in three deviation classes: class 1 for calm wind ($\sigma_1 \leq 1\ Pa$); class 2 for moderate wind ($1\ Pa < \sigma \leq 2\ Pa$); and, class 3 for strong wind ($\sigma > 2\ Pa$). In the absence of guidelines in the literature for this classification, the boundaries were chosen to have equal pressure steps (1 Pa) and adequate sample sizes in each class (respectively 416, 192 and 348 for classes 1, 2 and 3) between classes. Further work could more thoroughly investigate how the wind affects the standard deviation of zero-flow pressure measurements.

The model comparison allows determining how parameters affect the approximation quality. However, it gives no information about the quantification of their impact in terms of pressure difference. The next paragraph presents an in-depth analysis of the impact of deviation classes and measurement period on the zero-flow pressure approximation.

3.2.2 Pairwise Comparisons

Figure 4 shows the mean and 95% confidence interval of the approximation quality for 12 categories of approximations based on measurement period and grouped under the 3 standard deviation classes, hence providing a graphical overview of the results in terms of pressure difference.

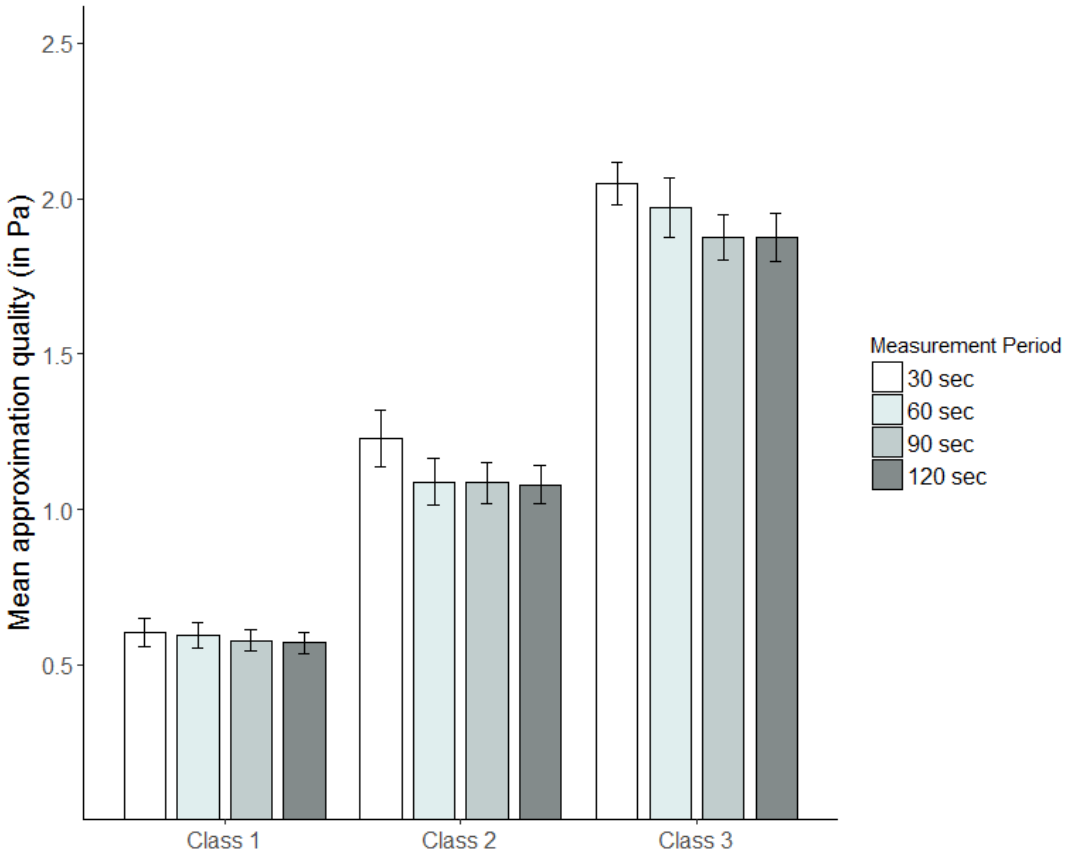


Figure 4 – Mean approximation quality of different approximations grouped in 12 categories

Due to the samples sizes being unequal, preliminary tests were performed on the data to verify eventual violations of the assumptions of normality and homogeneity of variance. In fact, when both the sizes of samples and their variances differ, there is a risk of inflating the occurrence of Type I errors (i.e., the probability of falsely rejecting the null hypothesis) [20, 40, 41]. Since the data were not normally distributed (Kolmogorov-Smirnov test = 0.13, $p < 0.001$) [42], and the variances of the three deviation classes were significantly different (Levene's test $= 40.74$, $p < 0.001$), non-parametric Wilcoxon rank-sum tests were adopted to calculate the statistical significance of the differences detected [20]. In addition, considering that multiple tests were performed on the same data under the same hypothesis, Bonferroni corrections were applied to counterbalance the increase in familywise error rate caused by the significance level inflating across multiple pairwise comparisons. The familywise error is the probability of making at least one Type I error, and is calculated as $1 - (0.95)^n$,

with n being the number of comparisons performed [20]. The probability of making a Type I error was respectively 26% and 14% for the within-class and the between-class comparisons. The Bonferroni correction ensures that the cumulative Type I error rate is kept below 0.05, although it is an adjustment method that is often considered as conservative and potentially vulnerable to Type II errors (i.e., the likelihood of failing to reject a false null hypothesis) [20].

Table 5 presents the results of the Wilcoxon tests with Bonferroni correction. For each comparison, the table provides the sample size of each group (N_1 and N_2), the difference between means and the interpretation of its statistical significance ($(\varepsilon_1 - \varepsilon_2)^{NHST}$), the lower and upper 95% confidence intervals for the mean difference (95% CI_L and 95% CI_U), and the effect size estimated by the Hedge's g coefficient. This was preferred over other indicators of effect size (e.g., Cohen's d and Glass's Δ) since the standard deviation of independent groups were not the same and sample sizes were different. In the calculation of Hedge's g, the standard deviation of each group is weighted with the sample size before pooling, according to Equation 9 [29]:

Equation 9

$$g = \frac{\varepsilon_1 - \varepsilon_2}{\sigma^*_{pooled}}$$

Where, σ_{pooled} is the weighted pooled standard deviation and ε_1 and ε_2 are the mean approximation qualities for both groups. Cohen's benchmarks were used to infer small ($g \geq 0.2$), moderate ($g \geq 0.5$) and large ($g \geq 0.8$) effect sizes [32].

	$(N_1; N_2)$	$(\varepsilon_1 - \varepsilon_2)^{NHST}$	95% CI_L	95% CI_U	Effect size (g)
Between-class comparison					
Class 1 vs Class 2	(416 ; 192)	-0.53 **	-0.57	-0.49	-2.41
Class 1 vs Class 3	(416 ; 384)	-1.36 **	-1.40	-1.31	-4.38
Class 2 vs Class 3	(192 ; 384)	-0.82 **	-0.88	-0.77	-2.32
Within-class comparison					

Class 1 ($N = 416$)	(104 ; 104)				
30 vs 60	-	0.01 N.S	0.00	0.02	0.04
30 vs 90	-	0.03 *	0.01	0.04	0.12
30 vs 120	-	0.03 *	0.02	0.05	0.16
60 vs 90	-	0.02 N.S	0.00	0.03	0.08
60 vs 120	-	0.02 *	0.01	0.04	0.12
90 vs 120	-	0.01 N.S	0.00	0.02	0.05
Class 2 ($N = 192$)	(48 ; 48)				
30 vs 60	-	0.14 **	0.11	0.16	0.50
30 vs 90	-	0.14 **	0.11	0.17	0.52
30 vs 120	-	0.15 **	0.12	0.18	0.56
60 vs 90	-	0.00 N.S	-0.01	0.02	0.01
60 vs 120	-	0.01 N.S	-0.01	0.03	0.04
90 vs 120	-	0.01 N.S	0.00	0.01	0.03
Class 3 ($N = 384$)	(96 ; 96)				
30 vs 60	-	0.08 *	0.03	0.12	0.19
30 vs 90	-	0.17 **	0.14	0.21	0.50
30 vs 120	-	0.17 **	0.13	0.21	0.48
60 vs 90	-	0.10 **	0.06	0.13	0.23
60 vs 120	-	0.10 **	0.06	0.13	0.22
90 vs 120	-	0.00 N.S	-0.01	0.01	0.00

N.S. Not significant ($p > 0.016$), *Significant ($0.00033 < p \leq 0.016$), ** Highly significant ($p \leq 0.00033$), for between-class comparison
N.S. Not significant ($p > 0.008$), *Significant ($0.00017 < p \leq 0.008$), ** Highly significant ($p \leq 0.00017$), for within-class comparison
$g < 0.2$: effect with non-substantive magnitude, $g \geq 0.2$: small effect , $g \geq 0.5$: moderate effect, $g \geq 0.8$: large effect sizes

Table 5 – Between-classes and within-classes comparisons based on Wilcoxon parirwise comparisons with Bonferroni correction (effect sizes were estimated by Hedge's g coefficient)

The results of the pairwise comparisons confirm the graphical observations from Figure 4. The differences in means between deviation classes are statistically significant and practically relevant. The results of the comparisons between measurement periods, however, depend on the deviation class. Within class 1, the comparison 30vs90, 30vs120 and 60vs120 are statistically significant but all the differences detected have non-substantive magnitudes. Within class 2, all the comparisons featuring measurement periods of 30 seconds (i.e., 30vs60, 30vs90 and 30vs120) are highly statistically significant with differences of moderate magnitude. Within class 3, all the comparisons featuring periods of 30 seconds are statistically significant (30vs60) and highly significant (30vs90 and 30vs120) with practically relevant effect sizes (small for 30vs60; moderate for 30vs90 and 30vs120). Within this class, the comparisons 60vs90 and 60vs120 are also highly statistically significant and have a practical effect of small magnitude.

3.2.3 Standard Deviation Measurements

Since, in practice, the real zero-flow pressure cannot be measured during the fictitious period, its mean average and standard deviation cannot be computed. In order to check if the standard deviation of the measurements within approximation periods is a good indicator of the standard deviation within the fictitious period, a Spearman's rho correlation test was performed. This was preferred to the classical Pearson's r coefficient because the standard deviations were not normally distributed (Kolmogorov-Smirnov test = 0.17, $p < 0.001$) [29, 43, 44]. According to Cohen's benchmarks, the correlation between standard deviation measurements during the approximation and fictitious periods had a practical association effect of strong magnitude (Spearman's rho $= 0.829$, $p < 0.001$) [32].

3.3 Quantification of Uncertainties in Envelope Pressure Measurements

3.3.1 Equation for Standard Uncertainty of Envelope Pressure

There are two types of uncertainty evaluation methods: Type A is when uncertainties can be evaluated by statistical analysis of a series of observations; Type B is when an estimate cannot be obtained from repeated measurements and the uncertainty must be evaluated by scientific judgement (e.g., previous measurements or manufacturer's specifications) [45]. In practice, the envelope pressure difference is obtained by averaging multiple measurements (N) of the same pressure difference. According to the literature [45], in case of Type A evaluation methods, the standard uncertainty is the square root of the experimental variance of the arithmetic mean (Equation 10) [12].

Equation 10

$$u_c(\Delta P_{env}) = \sqrt{\frac{s^2(\Delta P_{meas})}{N}}$$

Where, $u_c(\Delta P_{env})$ is the standard uncertainty of the envelope pressure difference, $s^2(\Delta P_{meas})$ is the experimental standard deviation of the envelope pressure measurement, and N is the number of repeated measurements.

When there are large wind fluctuations during a test, however, it might not always be appropriate to consider the envelope pressure measurement as a series of repeated independent measures. The left side of Figure 5 shows this issue by plotting 30 pressure-airflow couples measured within an interval of 30 seconds during a typical fan pressurization test with a target pressure of -40 Pa. The right side of the figure shows the results if envelope pressure measurements were perfectly independent. Delmotte discussed this issue when dealing with standard uncertainty propagation [12, 46].

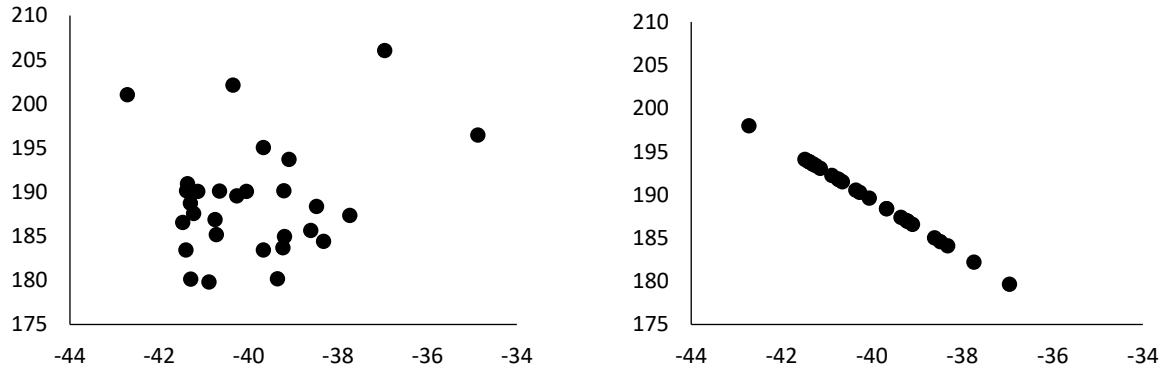


Figure 5 – 30 pressure-airflow couples measurements (one per second): when a -40 Pa target pressure is set during a typical fan pressurization test (left); if measurements were completely uncorrelated (right)

When measurements are no longer independent, the type B evaluation method must be used. In such cases, the standard uncertainty is evaluated by scientific judgement based on the available information on the variability of the results [45]. In the development of the standard uncertainty calculation formula, Delmotte considered each measurement point separately, and therefore did not deal with propagation uncertainty when calculating the means [12]. Similarly, in this study the uncertainty on pressure envelope was considered at each measurement point and not on mean averages.

The envelope pressure is the difference between the pressure measured and the zero-flow pressure. Therefore, the standard uncertainty in envelope pressure difference can be computed with Equation 11 [12, 45]:

Equation 11

$$u_c(\Delta P_{env}) = \sqrt{u_c^2(\Delta P_u) + u_c^2(\Delta p_0)}$$

Where, $u_c(\Delta P_u)$ is the uncertainty related to the pressure measured at gauge location and $u_c(\Delta p_0)$ is the uncertainty related to the zero-flow pressure estimation.

This equation is only valid if ΔP_u and Δp_0 are independent. In the calculation performed in this study, a constant distribution was considered for the zero-flow pressure approximation. Thus, Δp_0 is constant over the test while ΔP_u values are scattered Figure 5; the assumption of independence is therefore satisfied. Since the hypotheses related to envelope pressure are considered acceptable (i.e., negligible uncertainty), the standard uncertainty of pressure measurement was deduced from the equipment described in paragraph 2.1.2 (i.e., the greatest between $\pm$ 0.5% and $\pm$ 0.1 Pa [18, 19]).

3.3.2 Standard Uncertainty in Zero-Flow Pressure

The standard uncertainty in zero-flow pressure is composed of two parts: the zero-flow pressure measurement and its approximation. The uncertainty of zero-flow pressure measurement depends on the approximation technique and on the pressure gauge used. The standard uncertainty of the equipment used to measure zero-flow pressure in this study was assumed constant ($\pm$ 0.1 Pa) because the zero-flow pressure is always lower than 5 Pa (ISO 9972:2015 standard requirement [17]). Since the uncertainty of the measurement is constant, and it is independent of the approximation, Equation 12 can be used to compute the standard uncertainty in zero-flow pressure:

Equation 12

$$u_c(\Delta p_0) = \sqrt{u_{c,m}^2(\Delta p_0) + u_{c,a}^2(\Delta p_0)}$$

Where, $u_{c,m}(\Delta p_0)$ is the standard uncertainty related to the zero-flow pressure measurement, and $u_{c,a}(\Delta p_0)$ is the standard uncertainty related to the zero-flow pressure approximation.

The standard uncertainty related to zero-flow pressure measurement is the standard uncertainty of the equipment divided by two, since the zero-flow pressure is given by the average between pre- and post-test measurements and both have the same standard uncertainty [12].

The standard uncertainty in zero-flow pressure approximation cannot be evaluated by series of observations because the weather conditions cannot be kept constant between tests. Thus, it must be evaluated based on scientific judgement and available information on the variability of results. The indicator ε is defined as the mean of $\left|\Delta p_{0,ti} - \Delta\tilde{p}_{0,ti}\right|$ and both ΔP_0 (Shapiro-Wilk test: $W = 0.94$, $p = 0.08$; Komolgorov-Smirnov test: $D = 0.16$, $p = 0.37$) and ε (Shapiro-Wilk test: $W = 0.94$, $p = 0.06$; Komolgorov-Smirnov test: $D = 0.16$, $p = 0.4$) follow distributions not significantly different from normal. Therefore, it can be assumed that $\Delta P_0 - \varepsilon/2$ and $\Delta P_0 + \varepsilon/2$ are respectively the lower and the upper limit of the interval containing 50% of the zero-flow pressure. Since the Z-scores of the normal distribution are respectively 0.675 and -0675 when considering 75% and 25% of the results, Equation 13 can be used to find the standard deviation of the zero-flow pressure based on approximation quality.

Equation 13

$$\pm 0.675 = \frac{\left(\Delta p_{0,a} \pm \varepsilon/2\right) - \Delta p_0}{\sigma(\Delta p_0)}$$

It is assumed that the zero-flow pressure approximation ($\Delta P_{0,a}$) equals the average of the zero-flow pressure measurements made during the fictitious period (ΔP_0). This hypothesis is acceptable since the difference between both is already considered in the term ε of the equation. Therefore, the standard uncertainty of the approximation method ($u_{c,a}(\Delta p_0)$) is given by the standard deviation of the zero-flow pressure around its mean value ($= \varepsilon/1.35$) [45]. Table 6 provides $u_{c,a}(\Delta p_0)$ values obtained in this study for different deviation classes and measurement periods based on ε averages.

	30 seconds	60 seconds	90 seconds	120 seconds
Class 1	0.45	0.44	0.43	0.42
Class 2	0.91	0.81	0.80	0.80
Class 3	1.52	1.46	1.39	1.39

Table 6 – Values for the uncertainty due to zero-flow pressure approximation for the 12 groups of approximations considered in the in-depth analysis of this study [Pa]

The results presented in Table 6 confirm the observations made in the pairwise comparison (Figure 4, section 3.2.2). Indeed, increasing the approximation period to 60 seconds reduces the uncertainty within Class 2 and 3. Increasing the measurement period to 90 seconds reduces the uncertainty in Class 3.

3.3.3 Standard Uncertainty in Envelope Pressure

Since only a constant distribution has been considered, Equation 14 gives the total standard uncertainty of envelope pressure evaluation as a function of the measurement at the pressure gauge (ΔP_u):

Equation 14

$$u_c(\Delta P_{env}) = \sqrt{\max(0.005 * \Delta P_u; 0.1)^2 + \frac{0.1^2}{2} + \left(\frac{\varepsilon}{1.35}\right)^2}$$

In the literature on uncertainties in airtightness evaluation, results are often given in terms of expended uncertainties [12, 15]. The expanded uncertainty is applied when dealing with a Type B evaluation method and is the equivalent to the confidence interval in a Type A evaluation method. The expanded uncertainty is obtained by multiplying the standard uncertainty by a coverage factor (k). If the probability distribution is assumed normal, coverage factors of 2 and 3 correspond respectively to confidence intervals of ≈ 95% and ≈ 99% [45]. In this study, the envelope pressure was assumed normally distributed and a coverage factor of 2 was used to give an expanded uncertainty equivalent to a 95% confidence interval.

Figure 6 represents the expanded uncertainty ($k = 2$) as a function of the pressure difference measured by pressure gauges (ΔP_u) with (solid lines) and without (dashed lines) consideration of the zero-flow pressure approximation component in the envelope pressure uncertainty. The results are plotted for the three different deviation classes, considering for each class a measurement period of 30 seconds. The left part of the figure is the relative standard uncertainty (expressed in percentage of ΔP_u), while the right part is the absolute standard uncertainty (in Pa).

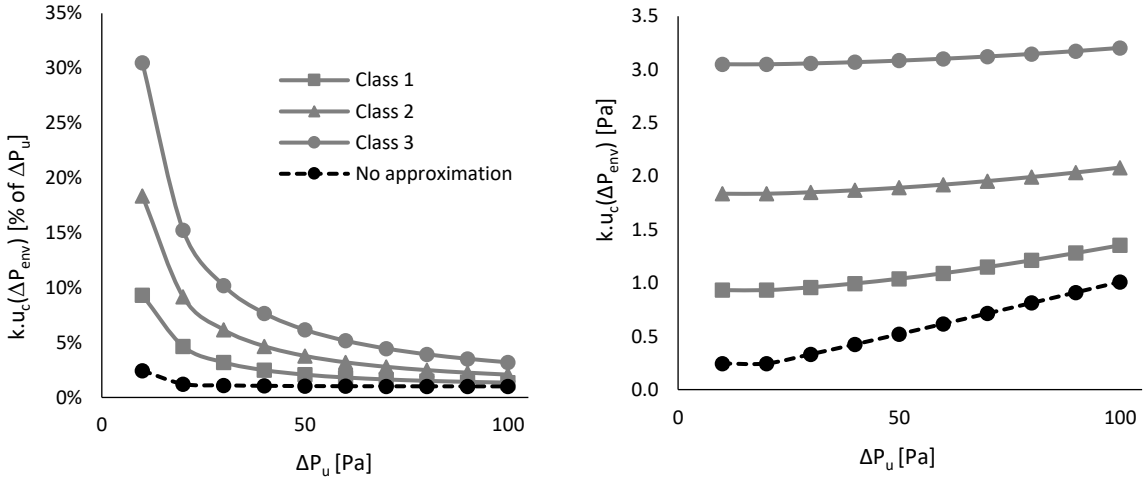


Figure 6 – Results of the calculations for relative (left, in % of ΔP_u) and absolute (right, in Pa) expanded standard uncertainty of the envelope pressure evaluation with (solid lines) and without (dashed line) considration of the approximation uncertainty component

These graphs show that the uncertainty due to zero-flow pressure approximation has a large impact on the envelope pressure uncertainty. Indeed, for a measurement at 50 Pa when using approximation periods of 30 seconds, the uncertainty due to zero-flow pressure approximation is 76%, 93% and 97% of the envelope pressure uncertainty for low-, medium- and high-wind conditions respectively.

Table 7 shows the differences in uncertainty calculations without the uncertainty due to zero-flow pressure approximation and with it for three different deviation classes.

	Low Pressure (10 Pa)		Medium Pressure (50 Pa)		High Pressure (100 Pa)	
	[Pa]	[%]	[Pa]	[%]	[Pa]	[%]
No approximation	0.24	2.4	0.52	1.0	1.01	1.0
Deviation class 1	0.93	9.3	1.04	2.1	1.35	1.4
Deviation class 2	1.84	18.4	1.89	3.8	2.08	2.1
Deviation class 3	3.05	30.5	3.08	6.2	3.20	3.2

Table 7 – Results for the calculations of expanded standard uncertainty (k=2) of the envelope pressure in terms of % of ΔP_u and in terms of Pa. Results are given for low-, medium- and high-pressure

The ISO 9972:2015 European standard allows the first measurement to be made at a pressure of three times the zero-flow pressure [17]. Although that multiplier is applied on the zero-flow pressure value and not on the deviation class, it is reasonable to expect that, for deviation class 3, a measurement at 10 Pa might be sometimes irrelevant.

The value presented in Table 7 and in Figure 6 should be considered carefully in the calculations. Indeed, when performing a test, each airflow-pressure couple recorded is the average of multiple measurements. Furthermore, in practice the operator often tries to take a measurement when wind fluctuations are low. This reduces the impact of wind fluctuations and decreases this component of uncertainty. The values of uncertainties found might overestimate the uncertainty related to zero-flow pressure approximation encountered in practice.

4 DISCUSSION

This study shows that the uncertainty due to zero-flow pressure approximation is substantial and can considerably increase the uncertainty in pressure measurements. Indeed, this uncertainty is 1.52 Pa for deviation class 3 when considering 30-second measurements for the approximation period, while at 50 Pa the uncertainty due to the equipment is 0.25 Pa. Figure 7 represents the uncertainty in the Y variable ($\ln(Q)$) and in the X variable ($\ln(\Delta P)$), with (X1, dashed grey line) and without (X0, dotted grey line) consideration of the uncertainty due to zero-flow pressure approximation, as a function of the pressure measurement in a typical fan pressurization test performed on an apartment in Brussels.

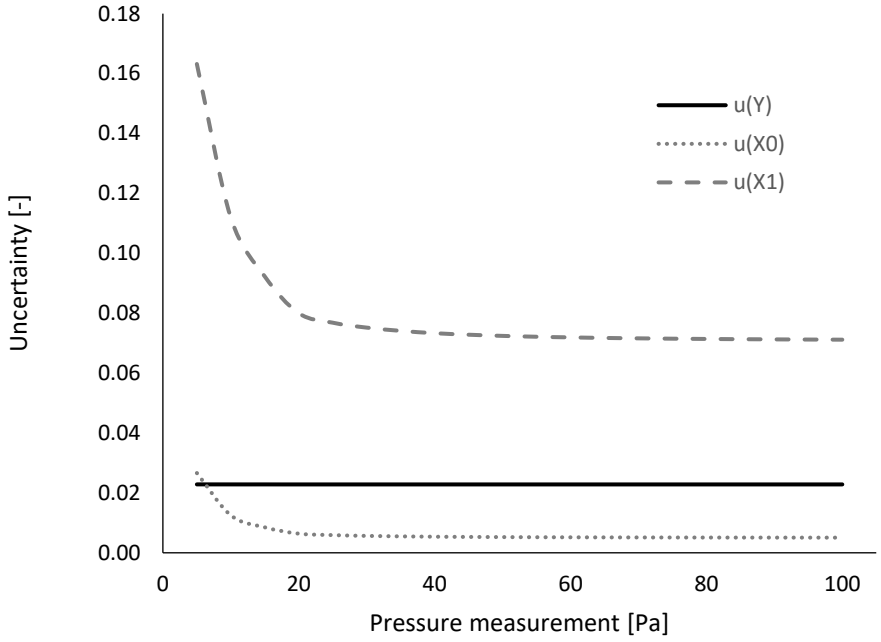


Figure 7 – Uncertainties in Y variable and X variable - with (X1) and without (X0) consideration of the zero-flow pressure approximation - as a function of the pressure measurement for a typical fan pressurization test

This observation has an important impact on the protocol for building airtightness evaluation. A linear regression technique is generally used to determine building flow characteristics (i.e., flow exponent and leakage coefficient) from airflow-pressure couples, but the ISO 9972:2015 standard [17] gives no indication about the choice of the method. The ordinary least square (OLS) method is often utilized in practice, but this method assumes that the uncertainties in the envelope pressure are negligible [47].

Various authors have already questioned the use of the OLS method, even without considering the uncertainty due to zero-flow pressure approximation [12, 48]. The findings of this study strengthen, as illustrated by Figure 7, their point of view in the debate and supports the development of other regression techniques, as for example the iterative weighted least square (IWLS) suggested by Okuyama and Onishi [48] or the weighted line of organic correlation (WLOC) proposed by Delmotte [13]. Recent research has also shown that using the OLS method leads to a poor estimation of uncertainties in building flow characteristics [13, 49].

This study quantifies the impact of uncertainties in zero-flow pressure approximation on envelope pressure uncertainty. Its influence on the global uncertainty is more complex to quantify because the study of different regression techniques is also implied. It must be considered that this component has no impact on the uncertainty calculation using the OLS method since this assumes that the uncertainty in envelope pressure measurement is negligible. However, it has an impact on the calculation of uncertainties in regression parameters and on the estimation of the correlation between both coefficients when using other regression methods (e.g., WLOC or IWLS). In addition, its impact is expected to strongly vary depending on the variable of interest (e.g., airflow at 50 or 10 Pa) due to the weighting of the measurement induced by alternative regression methods. The study of the change in regression techniques and the uncertainty in zero-flow pressure approximation has an influence on the total uncertainty, and the airtightness estimation is an important work that should be conducted in further researches.

This study also finds that none of the three suggested distributions gives better results than the constant one (i.e., the distribution recommended in the ISO 9972:2015 standard [17]). This is not consistent with the bi-linear distribution suggested by Delmotte [13]. However, the results presented in the previous section do not imply that the constant distribution is the most appropriate. Indeed, this study has only tested constant α_i coefficients in the distribution while the standard deviation seems to have a strong impact on the approximation quality. It could be interesting in further research to test

distributions with α_i coefficients depending on the zero-flow pressure variation (i.e., distributions depending on the test results) before claiming that uncertainties cannot be reduced by changing the zero-flow pressure distribution during the test.

Finally, this study shows that an increase of measurement period has a statistically significant and practically relevant positive impact on the approximation quality, especially on windy days. This is important because this modification is easy to implement in practice. When considering travel time, building preparation, equipment installation, and the measurement itself, a fan pressurization test can easily take half-a-day to complete. Therefore, the addition of 90 seconds to both approximation periods is negligible compared to the total duration of the test itself. Results showed that an increase of the approximation period from 30 to 60 seconds for class 2 and from 30 to 90 seconds for class 3 should be considered, respectively, for days with medium and high wind. Unfortunately, to our knowledge, no other author has yet studied the impact of this variable. One interesting further work would be the generalization of these observations on a large sample of zero-flow pressure tests performed on several different buildings.

In generalizing the findings of this study, some methodological limitations should be acknowledged. In fact, even if the results were obtained by rigorous statistical testing, the data on which they were based were specific to the sample analyzed. Other buildings could undergo other zero-flow pressure (e.g., due to a stack effect in high-rise buildings or different surroundings inducing other wind pressure). Since a correlation between the zero-flow pressure and its uncertainty can be expected, it is also reasonable to assume an influence of the building case on the zero-flow pressure approximation uncertainty. However, the method and the trends observed (e.g., the importance of the zero-flow pressure approximation uncertainty regarding the envelope pressure uncertainty or the reduction of uncertainty when taking higher approximation periods) should still be appropriate for other cases. In addition to the generalization of the results, applying this quantification method on a large sample of different buildings would also allow to analyze the impact of building-related parameters (e.g.,

dwelling type, volume or height, etc.) on uncertainty due to zero-flow pressure approximation. Another variable that could be tackled in further research is the location of the pressure gauge. As mentioned in section 2.1.3, the location of the pressure gauge is expected to have an influence on the zero-flow pressure average and standard deviation and, therefore, on the uncertainty due to zero-flow pressure approximation.

5 CONCLUSION

This study provides 4 key-findings:

- Multi-level modelling can, and should, be used to estimate uncertainties due to zero-flow pressure approximation and to study the impact of different variables on it.
- The approximation period should be increased to 60 seconds on medium-wind days (to reduce the uncertainty by 11%), and to 90 seconds on high-wind days (to reduce the uncertainty by 9%).
- A change in the zero-flow pressure distribution or in the time step between measurements does not provide statistically or practically better results.
- Uncertainties in zero-flow pressure represents more than 75% of the envelope pressure uncertainties whatever the wind conditions, and therefore they have an impact on the choice of regression method used.

This study is an important step in the quantification of uncertainties in airtightness measurements, but further research in this field is needed, including: applying the proposed method on a large sample of data from different buildings; studying the impact of uncertainty due to zero-flow pressure on the airtightness estimation for different regression techniques (OLS, IWLS or WLOC); quantifying the other relevant uncertainty components of airtightness measurements uncertainties.

ACKNOWLEDGMENTS

Authors gratefully acknowledge the support of Dr Michael Kent from the University of Nottingham in the statistical analysis of data. Authors also thank Christophe Delmotte from the Belgian Building Research Institute and Felipe Ossio from Pontificia Universidad Católica de Chile for helping with a proofreading of the paper. The tests were performed on a site from Jacques Delens Company and this work was supported by an INNOVIRIS grant (2016-DOCT-9) awarded to the first author for his PhD research at Architecture et Climat, Université catholique de Louvain.

REFERENCES

1. Kalamees, T., *Air tightness and air leakages of new lightweight single-family detached houses in Estonia.* Building and environment, 2007. **42**(6): p. 2369-2377.
2. Jokisalo, J., et al., *Building leakage, infiltration, and energy performance analyses for Finnish detached houses.* Building and Environment, 2009. **44**(2): p. 377-387.
3. Loncour, X. and C. Mees, *L'étanchéité à l'air des bâtiments*. 2015, Centre Scientifique et Technique de la Construction (CSTC).
4. Šadauskienė, J., et al., *Impact of Air Tightness on the Evaluation of Building Energy Performance in Lithuania.* Energies, 2014. **7**(8): p. 4972-4987.
5. Thor-Oskar, R., H. Sverre, and T.J. Vincent, *Airtightness estimation—A state of the art review and an en route upper limit evaluation principle to increase the chances that wood-frame houses with a vapour-and wind-barrier comply with the airtightness requirements.* Energy and Buildings, 2012. **54**: p. 444-452.
6. Prignon, M. and G.V. Moeseke, *Factors Influencing Airtightness and Airtightness Predictive Models: A Literature Review.* Energy and Buildings, 2017. **146**: p. 87-97.
7. Novak, J., *Repeatability and reproductibility of blower door tests - four years' experience of round-robin tests in Czech republic*, in *9th International Buildair Symposium*. 2015: Germany.
8. Delmotte, C. and J. Laverge. *Interlaboratory tests for the determination of repeatability and reproducibility of buildings airtightness measurements*. in *32nd AIVC conference and 1st TightVent Conference: "Towards Optimal Airtightness Performance"*. 2011. Brussels, Belgium.
9. Persily, A. *Repeatability and accuracy of pressurization testing*. in *ASHRAE/DOE Conference Thermal Performance of the Exterior Envelopes of Buildings II*. 1982. Las Vegas, USA.
10. Kim, A.K. and C.Y. Shaw, *Seasonal variation in airtightness of two detached houses*, in *Measured Air Leakage of Buildings: A Symposium*. 1986, ASTM International: Philadelphia, USA. p. 16-32.
11. Bracke, W., et al., *Durability and Measurement Uncertainty of Airtightness in Extremely Airtight Dwellings.* International Journal of Ventilation, 2016. **14**(4): p. 383-394.
12. Delmotte, C. *Airtightness of buildings-Calculation of combined standard uncertainty*. in *34th AIVC Conference "Energy conservation technologies for mitigation and adaptation in the built environment: the role of ventilation strategies and smart materials"*. 2013. Athens, Greece.
13. Delmotte, C. *Airtightness of Buildings - Considerations regarding the Zero-Flow Pressure and the Weighted Line of Organic Correlation*. in *38th AIVC Conference "Ventilating healthy low-energy buildings"*. 2017. Nottingham, UK.
14. Sherman, M. and L. Palmiter, *Uncertainties in fan pressurization measurements*, in *Airflow performance of building envelopes, components, and systems*. 1995, ASTM International: Philadelphia, USA. p. 266-283.
15. Carrié, F.R. and V. Leprince, *Uncertainties in building pressurisation tests due to steady wind.* Energy and Buildings, 2016. **116**: p. 656-665.
16. Walker, I., et al., *Applying Large Datasets to Developing a Better Understanding of Air Leakage Measurement in Homes.* International Journal of Ventilation, 2013. **11**(4): p. 323-338.
17. ISO-9972, *NBN EN ISO 9972:2015 - Performance thermique des bâtiments - Détermination de la perméabilité à l'air des bâtiments - Méthode de pressurisation par ventilateur (ISO 9972:2015) CEN, Brussels, Belgium.* 2015.
18. TEC, *Operating Instruction for the DG-700 Pressure and Flow Gauge*. 2012: The Energy Conservatory.
19. TEC, *An explication of the DG-1000 accuracy specifications*. 2016, The Energy Conservatory.
20. Field, A., J. Miles, and Z. Field, *Discovering Statistics Using R*. 2012, London: SAGE Publications Ltd. 957.

21. Peugh, J.L., *A practical guide to multilevel modeling.* Journal of School Psychology, 2010. **48**(1): p. 85-112.
22. Rasbash, J., et al., *A User's Guide to MLwiN*. 2015, University of Bristol: Centre for Multilevel Modelling.
23. Goldstein, H., *Multilevel Statistical Models*. 4th ed. 2010, New York: John Wiley & Sons Ltd. 1-14.
24. Kent, M.G., *Temporal Effects in Glare Response*, in *Department of Architecture & Built Environment*. 2016, University of Nottingham: United Kingdom.
25. Hayes, A.F., *A Primer on Multilevel Modeling.* Human Communication Research, 2006. **32**(4): p. 385-410.
26. Beretvas, S.N. and D.A. Pastor, *Using Mixed-Effects Models In Reliability Generalization Studies.* Educational and Psychological Measurement, 2003. **63**(1): p. 75-95.
27. Whittaker, T.A. and C.F. Furlow, *The Comparison of Model Selection Criteria When Selecting Among Competing Hierarchical Linear Models.* Journal of Modern Applied Statistical Methods, 2009. **8**(1).
28. Seltman, H.J., *Experimental Design and Analysis*. 2012, Pittsburgh: Carnegie Mellon University 2012. 428.
29. D Ellis, P., *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. 2010, Cambridge: Cambridge University Press. 173.
30. Schiavon, S. and S. Altomonte, *Influence of factors unrelated to environmental quality on occupant satisfaction in LEED and non-LEED certified buildings.* Building and Environment, 2014. **77**: p. 148-159.
31. Kent, M.G., et al., *Temporal effects on glare response from daylight.* Building and Environment, 2017. **113**: p. 49-64.
32. Cohen, J., *A power primer.* Psychological bulletin, 1992. **112**(1): p. 155.
33. Durlak, J.A., *How to select, calculate, and interpret effect sizes.* Journal of pediatric psychology, 2009. **34**(9): p. 917-928.
34. Volker, M.A., *Reporting effect size estimates in school psychology research.* Psychology in the Schools, 2006. **43**(6): p. 653-672.
35. Julian, M.W., *The Consequences of Ignoring Multilevel Data Structures in Nonhierarchical Covariance Modeling.* Structural Equation Modeling: A Multidisciplinary Journal, 2001. **8**(3): p. 325-352.
36. Maas, C.J.M. and J.J. Hox, *Sufficient Sample Sizes for Multilevel Modeling.* Methodology, 2005. **1**(3): p. 86-92.
37. Muthén, B.O. and A. Satorra, *Complex Sample Data in Structural Equation Modeling.* Sociological Methodology, 1995. **25**: p. 267-316.
38. Verbeke, G.M., G, *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. 2000: Springer-Verlag New York. 570.
39. Gurka, M.J., *Selecting the Best Linear Mixed Model Under REML.* The American Statistician, 2006. **60**(1): p. 19-26.
40. De Winter, J.C., *Using the Student's t-test with extremely small sample sizes.* Practical Assessment, Research & Evaluation, 2013. **18**(10).
41. Zimmerman, D.W., *A note on preliminary tests of equality of variances.* British Journal of Mathematical and Statistical Psychology, 2004. **57**(1): p. 173-181.
42. Field, A. and G. Hole, *How to design and report experiments*. 2002: Sage.
43. Ferguson, C.J., *An effect size primer: A guide for clinicians and researchers.* Professional Psychology: Research and Practice, 2009. **40**(5): p. 532-538.
44. Hauke, J. and T. Kossowski, *Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data.* Quaestiones Geographicae, 2011. **30**(2): p. 87.
45. JCGM, *Evaluation of measurement data—guide for the expression of uncertainty in measurement.* 2008, Joint Committee for Guides in Metrology. p. 134.

46. Delmotte, C., *Airtightness of buildings - Calculation of combined standard uncertainty*. 2013, Belgian Building Research Institute - Laboratory Air Quality and Ventilation: Unpublished Report. p. 29.
47. Helsel, D.R. and R.M. Hirsch, *Statistical Methods in Water Resources* in *Techniques of Water Resources Investigations*. 2002, U.S. Geological Survey. p. 522.
48. Okuyama, H. and Y. Onishi, *Reconsideration of parameter estimation and reliability evaluation methods for building airtightness measurement using fan pressurization.* Building and Environment, 2012. **47**: p. 373-384.
49. Prignon, M., A. Dawans, and G. van Moeseke. *Uncertainties in airtightness measurements: regression methods and pressure sequences*. in *39th AIVC-7th TightVent & 5th venticool Conference Smart ventilation for buildings*. 2018.