

Performance evaluation of the construction products as a research challenge. Small error – big difference in assessment?

E. SZEWCZAK* and A. PIEKARCZUK

Instytut Techniki Budowlanej, 1 Filtrowa St., 00-611 Warszawa, Poland

Abstract. The paper refers to the importance of test results uncertainty in the assessment of the construction product performance. Uncertainty understood as dispersion attributed to any value which is result of the tests, calculations or other evaluations, occurs at every level of assessment (material, product, construction). Authors presented the various approaches to uncertainty estimation, especially in situation of small number of tests results which is characteristic for testing of construction products. Effects of uncertainty on final assessment of construction product was analyzed using the example of bearing capacity of thin-walled structure obtained by numerical calculation. Different values of material tests uncertainty resulting from different approaches to its assessment was taken into account. It was demonstrated that the difference in the results of strength tests of a material, which falls within the limits of uncertainty, may result in a very significant difference in the evaluation of a structure.

Key words: uncertainty, numerical calculation, uncertainty influence on product assessment.

1. Introduction

During decision-making processes, knowledge about uncertainty of information which support the decision is crucial. Uncertainty is present at all levels of civil engineering: from testing of materials and products, material and structures modelling to building process [1–3]. In this paper only part of the process starting from material testing to subsequent assessment of construction product is considered.

Civil engineering uses an enormous variety of construction products which ultimately affect the fulfilment of safety-related requirements by a structure. Consequently, these products need to be evaluated. The evaluation is performed at numerous levels: starting from the microstructure of materials through the properties of component materials and ending with the performance of a finished product. Laboratories perform various tests, including mechanical, physical and functional ones, on finished construction products for all types of resistance to exposure to e.g. temperature, water, UV radiation, pressure, mechanical shocks, fire, electromagnetic radiation and biological impacts. The range of requirements and the required laboratory tests on component materials and finished products is determined individually for each product type (e.g. in harmonised technical specifications) [4].

However, the fundamental question is: to what extent is the assessment of a product based on tests reliable?

In this context, the determination of relationships between the results of laboratory tests based on various conventions and the actual behaviour of a product during the subsequent use thereof may be a scientific challenge. It is not, however, the subject of this paper.

Each value resulting from tests may only be considered in terms of probability. Unfortunately, in a small number of cases it is possible to express this probability using a probability density function (PDF) which is characterised by the expectation value and standard deviation (SD). In all tests which yield qualitative results (ordinal scale with two categories e.g. “fulfilled”, “not fulfilled”) or semi-quantitative results (ordinal scale with more than two categories), which, in the authors’ opinion, are the majority of tests to which construction products are subjected, the determination of a PDF is an issue requiring extensive inter-laboratory tests performed on repeatable items [5–7]. Usually this is not possible for tests on construction products.

Linking a value being measured with a PDF that is assigned to it is much easier for quantitative methods, for which the estimation of the uncertainty of a test result, allowing one to provide a PDF, is subject to numerous guidelines and instructions. The basic principles have been developed by the Joint Committee for Guides in Metrology [8], and are commonly applied by virtually all laboratories. However, as will be demonstrated for destructive (e.g. mechanical) testing, the estimation of uncertainty may be associated with numerous ambiguities.

Presented by Walker et al. [9] typology and terminology of harmonized uncertainty provide uncertainty identification by location, including inter alia parameter uncertainty (associated with the variability of input data) and model uncertainty (associated also, but not only with lack of sufficient understanding of the system). These two kind of uncertainty sources are presented in test results and in estimation of uncertainty of the test results (uncertainty of estimated uncertainty). Variability relating to laboratory testing results has many components. Some of them, e.g. metrological parameters of testing equipment, can be relatively easily expressed numerically

*e-mail: e.szewczak@itb.pl

(of course these parameters are also burdened with parameter uncertainty). There are, however, many unrecognised factors likely to affect the repeatability and reproducibility of tests, which are difficult to be identified precisely. Lack of sufficient knowledge leads to a very simplified and different models adopted by laboratories to estimation of the test result uncertainty (model uncertainty).

The stage that follows laboratory tests is taking the results into account in further evaluations. At this stage, a uniform procedure is also missing. For certain products, statistical methods have been determined for taking the dispersion of tests results into account in the conformity assessment. However, this is usually (for example, for concrete [10]) related to tests repeated as part of the factory production control, where knowing of the variability of the entire population is essential. Methods for the evaluation of construction products' quality based on the assumed producer's risk and consumer's risk have also been described [11]. In such schemes it is not necessary to know the population standard deviation and only the standard deviation of current results is taken into account; however, the minimum number of samples under such control is fifteen. In addition, specifications containing principles for taking variability into account relate only to the random dispersion of current results, expressed by the standard deviation. Most frequently, this is not the only source of potential variability of a result or the only component of uncertainty. In turn, general documents addressing the issue of taking uncertainty into account in the evaluation of conformity [12] are primarily adjusted to metrology, and their application in the assessment of construction products yields non-rational results i.e. there is no possibility for either positive or negative assessment.

As will be demonstrated in this paper, results of laboratory tests performed on the same test item may be characterised by a great variability, which is not easy to estimate. Where a laboratory performing tests on a small (3÷5) number of samples due to necessity (costs, standard-related requirements) has poor knowledge of the possible variability of a result, uniform methods for the determination of an uncertainty are missing, and there are no rational principles of taking uncertainty into account in the evaluation of a product, considerable doubt arises over reliability of this evaluation. This issue is particularly relevant to the characteristics related to safety of the structures in which the product will be used.

This paper, in order to discuss the influence of the variability of a material test result on the assessment of the product, uses the example of strength tests and evaluation of load-bearing capacity.

The authors have not found a publication which addresses this issue.

2. Basic aspects of estimation of uncertainty

In order to determine variability which may characterise a test result, uncertainty is estimated. The philosophy of the notion of uncertainty, arising from the laws of physics, is expressed in the statement contained in [8]: "The uncertainty of the result of

a measurement reflects the lack of exact knowledge of the value of the measurand". A formal definition of uncertainty presented in the VIM Vocabulary of Metrology [13] clearly indicates that this is a numerical value: "non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used".

While in the case of measurements of physical quantities (geometrical dimensions, force, weight) we speak of a measurand, in the case of numerous tests on materials and construction products we rather speak of tested performance or characteristic, the testing of which is based on specific procedures, and the result is defined by the method (e.g. determination of puncture resistance by falling cone method). When applying the philosophy presented in EA Guidelines 4/16 [1], while estimating the uncertainty, a test result must be treated similarly to a measurement result. However, examples of significant constraints relating to the estimation of uncertainty of test results as compared to the estimation of uncertainty of measurement will be demonstrated further on.

These constraints, do not allow one to obtain reliable uncertainty values, and therefore do not allow one to thoroughly evaluate the variability which may characterise a test result. Furthermore, laboratories adopt different approaches to estimating the uncertainty, which – for the same sets of results – may produce divergent uncertainty values. The main differences arise from other methods of taking factors affecting the result into account, and applying other expansion coefficients.

2.1. Evaluation of the influence of external factors. In order to construct a PDF for a test result, a mathematical model is required which takes into account the dependence of the output quantity on all quantities affecting the final result [8, 1]. Where it is not possible to directly include a quantity affecting the result (e.g. a change to environmental conditions or the influence of an operator) into a mathematical model, the basic research method assessing the effect of individual factors on the variability of results is multiple repetitions with an isolated change to particular factors [14]. Such experimental evaluations of component uncertainties are possible where measurements are involved which may be repeated many times on the same item (e.g. measurement of weight or length), or where it is possible to obtain a large amount of homogeneous test material (e.g. testing for a solution's concentration in chemistry). However, it is difficult to imagine experiments involving multiple repetitions which are aimed at evaluating the influence of random effects on the result of strength tests of large-size concrete structural components. Unreasonably high costs are not the only problem in such case. Another aspect ruling out such an experiment is the fact that it would not be possible, even with multiple repetitions, to assign variability of a result to specific factors during the test, because the test item itself is characterised by significant variability (heterogeneity), and the test cannot be repeated on the same sample.

Determination of flexural resistance of lintels according to EN 846–9 [15] is one of the tests examples in which the standard deviation of results is affected by both random effects

associated with the test itself and the heterogeneity of the tested items. Table 1 shows illustrative results of the tests of four different lintels with three specimens of each. This example clearly shows that sample SD taken from results obtained in the same laboratory and similar or the same conditions (like equipment, operator, temperature), differ significantly for different product. In fact, laboratory cannot separate SD contributions coming from product heterogeneity and from laboratory uncertainty.

Obviously, the difference between values of standard deviations is not only due to the differences in homogeneity between individual products; it is also related to the fact that the SD was calculated based on three results, and not the entire population.

However, sample SD is often the only information on variability available to the laboratory, and it is frequently treated as the main component of the uncertainty.

Another extremely different approach in the case of a significant influence of a product's heterogeneity on the SD of the result is to limit oneself to taking into account in uncertainty only the components resulting from the accuracy of the used testing equipment (information obtained from calibration certificates). Both approaches may yield extremely different results. For data presented in Table 1, lintel 3, the standard uncertainty related to the values of a mean from three samples and taking sample SD into account would be 7.3 kN, while taking into account only the accuracy of measuring equipment could yield standard uncertainty of approx. 0.5 kN (for the 1st class of a testing machine).

Table 1
Results of flexural resistance of five different lintels, kN
(Tests performed at the Building Research Institute Laboratory)

specimen	Lintel			
	1	2	3	4
	flexural resistance			
1	89.1	82.1	89.1	82.5
2	90.6	79.6	100.0	95.4
3	95.4	71.4	74.7	89.0
Sample standard deviation: kN	3.3	5.6	12.7	6.5
% of sample mean	3.6	7.2	14.4	7.3

2.2. Coverage factors. Laboratories are required to provide expanded uncertainty of results at a confidence level of 95% [8, 1]. The expanded uncertainty U is obtained by multiplying the standard uncertainty u by a coverage factor k_x :

$$U = k_x \cdot u \quad (1)$$

Selection of the k_x factor results from PDF. If u stands for combined uncertainty which is a resultant of many input quantities with different PDFs, selection of the k_x factor becomes

a significant mathematical problem associated with model uncertainty [9] in estimation of tests uncertainty. Various approaches to this issue may result in significant divergences in estimated uncertainty of the test result.

3. Methods for the evaluation of results variability

Low precision (high SDs of repeatability and reproducibility) of certain research methods may result in a divergence of results obtained by different laboratories. Therefore, the same product can be evaluated differently when it is re-tested [16].

For science branches such as chemistry or biology, tests for proficiency and inter-laboratory tests are of great significance to the evaluation of variability of results. A possibility of external estimation of the assigned value in inter-laboratory tests (e.g. by the use of standard solutions) allows one to draw conclusions about the bias, precision and uncertainty of a particular laboratory [e.g. 17, 18]. For most tests on construction products, the assigned value is unknown. There are no reference materials and it is impossible to repeat a test on the same item because it is destroyed. This reduces possibilities to draw conclusions about the laboratory uncertainty based on the results of inter-laboratory comparisons [19]. Therefore, laboratories are usually condemned to rely on their own, simplified analysis of uncertainty.

3.1. Dispersion of results and standard deviations. In order to evaluate how great the difference between results of tests on the same material can hypothetically be, an example of tensile strength tests in accordance with the EN ISO 6892–1 standard [20] was used. It was based on the results of thirty tests performed in one laboratory (Building Research Institute Laboratory) under repeatability conditions. In this article these results will be referred to as the “population”.

Tensile tests are usually performed on five samples, which can be compared to the sampling of a five-element set from a population of results. Therefore, ten sets of five samples each were randomly selected from the population of stress–strain curves $R = f(e)$. The random selection was performed using “Sampling” function in the Analysis ToolPak package of Microsoft Excel software.

Such sampling may yield significantly different results in terms of both values and their dispersion. This is presented in Fig. 1. For eleven selected values of strain, the following differences were shown:

$$\Delta = \bar{R}_p - \bar{R}_i \text{ and } \Delta s = \sigma - s_i \quad (2)$$

Where:

\bar{R}_p – mean value of stress at a particular strain e for the entire population of results, MPa

\bar{R}_i – mean value of stress for set i consisting of five randomly selected results from the population, MPa

σ – standard deviation for the population

s_i – sample standard deviation for set i

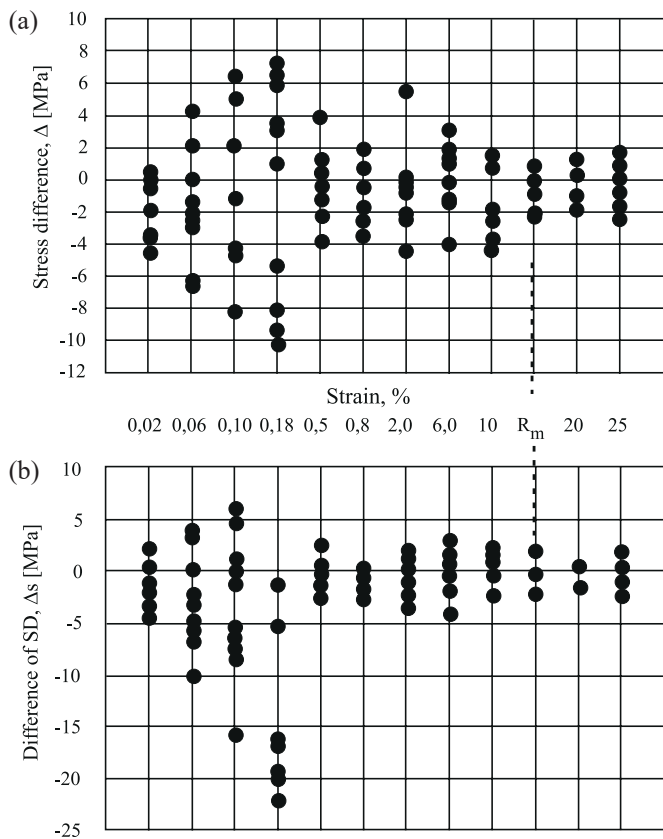


Fig. 1. Differences between the values of the means (a) and standard deviations (b) obtained in accordance with equation (2)

The analogously obtained values Δ and Δs for the maximum stress R_m were also included in the figure.

As can be seen in Fig. 1, the differences between stress results for five-sample sets randomly selected from the population are surprisingly large.

The largest difference between the mean stress for the five-sample set and for the population reached a value of $\Delta_{max} = 10.2$ MPa (approx. 3% of the value being measured). The largest difference between the standard deviation of the population and the five-sample set was $\Delta s_{max} = 22.1$ MPa. These results confirm that where tests on the same material in the same laboratory are performed several times, obtained values can be significantly different in both the value and the uncertainty assigned to that value.

3.2 Examples of methods for the estimation of uncertainties, and the consequences of their use. The following examples (3.2.1–3.2.6) show the differences which may occur in the evaluation of the variability of results. An estimation of uncertainties was performed, making various theoretical assumptions and using various sets of results from the same population. Some of the presented methods for the estimation of uncertainties cannot be considered correct; however, they are presented in order to illustrate the issue, as they occur very frequently in laboratory practice. The aim of this paper, however, is not to conduct full analysis of factors contributing to the uncertainty of tensile tests

results. Therefore, simplifications were used, which involved *inter alia* not taking into account the contribution of the elongation measurement in the uncertainty budget. Therefore, the values of maximum stress R_m , in relation to which a certain independence from elongation can be adopted, were selected as the data for the examples presented below.

Results of the estimated uncertainties and the maximum and minimum results for R_m obtained using the methods described below are presented in Table 2.

Table 2
 Differences in uncertainty values obtained using different methods and different sets of results. u_{max} , u_{min} / U_{max} , U_{min} – maximum and minimum obtained value of standard uncertainty/expanded uncertainty. k_x – coverage factor. All values except for k_x , in MPa

Method described in point:	R_m max R_m min (MPa)	u_{max} u_{min} (MPa)	k_x	U_{max} U_{min} (MPa)
3.2.1	399.6	0.57	2	1.2 1.2
3.2.2, 3.2.3		6.7	2	13,4
3.2.4	391.8	7.1 6.7	2	14 13
3.2.5	388.6	7.10 6.67	2.1 1.9	14.9 12.7
3.2.6.a	391.8	2.4 0.4	2.78	6.6 1.1
3.2.6.b	388.6	2.4 0.4	2.13	5.1 0.8

3.2.1. Uncertainty resulting from standard deviation of the population. In order to obtain information on the dispersion of results within the population, standard deviation $\sigma(R_m)$ for all results of the maximum stress was estimated. The value of the SD was 3.1 MPa.

As the result of the test for maximum stress is a mean value from all samples, the following was adopted as standard uncertainty in this case:

$$u = \frac{\sigma}{\sqrt{N}} \quad (3)$$

Where:

σ – SD for the population

$N = 30$ – number of samples

Having assumed a PDF with a normal distribution, expanded uncertainty was obtained by using $k_x = 2$ i.e. an approximate value of the coverage factor k for normal distribution and the confidence level of 95%.

In this case, R_m max and R_m min are, respectively, the maximum and minimum value of R_m in the entire population.

The presented method for estimating uncertainty does not take into account the variability resulting from the measurement equipment used, which, as will be shown in the next section, yielded an understated value of uncertainty in this case.

3.2.2 Taking parameters of the measurement equipment used into account. A typical method used in laboratory practice for the determination of uncertainty in tensile stress testing, taking into account also the parameters of the equipment used (such an approach was presented by e.g. L. Brunarski et al. [21]), is based on the following model formulated to account for the interrelation of the input quantities that influence the result:

$$R = \frac{F}{a \cdot b} + dR \quad (4)$$

Where:

F – force, N
 a, b – respectively, width and thickness of the sample, mm
 dR – correction related to the random nature of the strength results, MPa

The conversion of standard uncertainties of input quantities $u(x_i)$ into standard uncertainty of the output quantity $u(Y)$ in the case of non-correlated input quantities x_i , is carried out in accordance with the following equation [8]:

$$u(Y) = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)} = \sqrt{\sum_{i=1}^n c_i^2 u^2(x_i)} \quad (5)$$

Where c_i is the sensitivity coefficient.

The PDF of component F is characterised by the expectation value, which is a result of the measurement of force and $u(F)$ – standard uncertainty related to the measurement of force. In simple terms, when assuming the 1st class of a testing machine, a rectangular distribution of the PDF can be assumed, and it can be presented as [8]:

$$u(F) = \frac{1}{\sqrt{3}} \cdot 0,01 \cdot F \quad (6)$$

Similarly, in order to estimate the components resulting from measurement of the width and thickness of the sample: $u(a)$ and $u(b)$, one can use the data resulting from the calibration of measuring instruments. Assuming that the error in a calibration certificate is equal to 0, the following are taken into account: uncertainty of calibration (PDF with normal distribution) and the possible reading error resulting from resolution of the instrument (PDF with rectangular distribution). It is also frequently assumed that the reading error has been taken into account in the uncertainty of calibration. Such an assumption was made in this example. The quantities $a = 19.50$ mm and $b = 0.95$ mm were measured using a slide calliper, for which the expanded uncertainty in the calibration certificate was 0.031 mm (normal distribution of PDF).

The influence of all random factors, during the measurement of both the force and the dimensions of the sample, is taken into account in the spread of results dR with a PDF with normal distribution and the expectation value of 0. In this case, the complete population of samples was still used, therefore the value obtained in section 3.2.1 was adopted as standard uncertainty.

At the level of estimation of expanded uncertainty, one must take into account that the PDF assigned to the output quantity R , may have a distribution other than normal, as it is a combination of four functions with various scales and distributions.

However, when a practical consequence of the Central Limit Theorem (CLT) indicated in the GUM Guide [8] is applied, namely where the combined standard uncertainty “is not dominated by a standard uncertainty component obtained from a Type A evaluation based on just a few observations, or by a standard uncertainty component obtained from a Type B evaluation based on an assumed rectangular distribution”, it can be assumed that the output quantity has a normal distribution, and we obtain for expanded uncertainty of the output value for a level of confidence 95%, $k_x \approx 2$.

The estimated uncertainty is presented in Table 2, section 3.2.2, and, as can be seen, its value is several times higher than the value of the uncertainty based only on the standard deviation of test results (3.2.1) (in contrast to the results described in section 2.1).

3.2.3. Welch-Satterthwaite formula. The dominant influence of the components associated with the measurement may deny the possibility for application of an assumption based on the CLT with a normal distribution of the input quantity. Where the application of CLT becomes unjustified, it is recommended [8] that t-distribution with an effective degree of freedom ν_{eff} obtained from the Welch-Satterthwaite formula be applied.

$$\nu_{eff} = \frac{u_c^4(y)}{\sum \frac{u_i^4(y)}{\nu_i}} \quad (7)$$

Where:

$$u_c^2(y) = \sum u_i^2(y)$$

$$u_i^2(y) = c_i^2 u^2(x_i)$$

ν_i – number of the degrees of freedom for input quantity x_i

y – output quantity

In this case, however, given that for the components estimated using method B (according to [8]: “method of evaluation of uncertainty by means other than the statistical analysis of series of observations”, e.g. calibration certificates, permissible error) the number of the degrees of freedom $\nu = \infty$ is adopted in accordance with [8], and these components had a dominant effect on the uncertainty, the effective number of the degrees of freedom reached values exceeding 300, therefore factor $k_x = t_{95}(\nu_{eff}) = 2$.

As is indirectly apparent from the data shown in Fig. 1, standard deviations of results for values of stresses other than R_m may be significantly greater, and hence the ν_{eff} values may be different as well. This is of no practical significance where dR

was determined based on results for thirty samples; however, for $N = 5$ samples, where dR becomes the dominant component, $v_{eff} \rightarrow 4$ and $t_{95} \rightarrow 2.78$.

3.2.4. Uncertainty for five-sample sets. Examples 3.2.1, 3.2.2 and 3.2.3 concerned a situation where a population of results is at our disposal, which usually does not occur in a third party's testing laboratory. In this example, two sets, with the maximum and minimum value of sample SD of the R_m values, were selected from ten five-sample sets randomly selected in accordance with the above description in section 3.1. The maximum value of SD was $s_{max} = 5.33$ MPa, the minimum value was $s_{min} = 0.84$ MPa. In further calculations, the value of SD for the mean was taken into account, in accordance with formula (3), where $\sigma = s$ and $N = 5$.

For both sets, with the maximum and minimum SD, the uncertainty was estimated based on the model of measurement presented in equation (4).

Two values of standard uncertainty were obtained, namely u_{max} and u_{min} . Table 2 presents values corresponding to them, namely R_{mmax} and R_{min} , which are mean values from five samples set with the maximum and minimum value of SD.

Due to the unknown distribution of the combined uncertainty (the input PDFs have normal and rectangular distributions with various scales), in order to estimate the expanded combined uncertainty, a t-distribution with an effective number of degrees of freedom v_{eff} obtained from the Welch-Satterthwaite formula [8] was applied. Similarly to section 3.2.3, in both cases $k_x = t_{95}(v_{eff}) = 2$ was obtained.

3.2.5. Monte Carlo Method. Where combined uncertainty is affected by several quantities with various distributions, and analytical methods are troublesome, the application of the probabilistic Monte Carlo Method in order to determine the PDF for the output quantity [22] is also allowed. In this study, it was applied using a tool developed by the National Institute of Standards and Technology [23]. Assumptions identical to those in section 3.2.4 were adopted as well as the number of realisations of the output quantity: 1,000,000.

In this case, the value of factor k_x indicated dependence on the value of SD, and for s_{max} it was 2.1, while for s_{min} it was 1.9. The results are presented in Table 2.

3.2.6 Coverage factors in the case of uncertainty based on standard deviation for a small number of samples. When we assume the lack of a significant influence of the measurement equipment on combined uncertainty (as in the case of the results of flexural resistance of lintels – Table 1), the unresolved question, given such a small number of samples, remains: what coverage factor k_x should be applied? Due to the small number of samples, the application of factor $k_x = 2$ resulting from normal distribution is not justified.

In his study, H. Huang [24] compared three models for calculating the expanded uncertainty using the experimental standard deviation: the Student's t model, Craig model, and Bayesian model. The results indicated that the Student's t-model is the least precise and accurate. However, that model is very

often applied practically. Expanded uncertainty is expressed by the following equation:

$$U = t_{95} \cdot \frac{s}{\sqrt{N}} \quad (8)$$

Where:

t_{95} – value from the t-distribution for confidence level of 95% and the number of the degrees of freedom $N-1$

N – number of samples

s – sample SD.

According to H. Huang [24], when prior information on the SD population is not available, the Craig model is preferred. According to the Craig model, for N -samples taken from a population with a normal distribution, expanded uncertainty should be expressed by the following formula [24, 25]:

$$U = \frac{k \cdot s}{c_4 \cdot \sqrt{N}} \quad (9)$$

Where:

s – sample SD

k – coverage factor for normal distribution

$$c_4 = \sqrt{\frac{2}{N-1}} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}, \quad (10)$$

Where:

Γ – Gamma function.

In Table 2, section 3.2.6.a, values are provided which result from the application of the Student's model, in accordance with equation (8), for the number of samples $N = 5$. In section 3.2.6.b, the Craig model was applied in accordance with equations (9) and (10), for the number of samples $N = 5$.

The differences resulting from the application of the Craig and Student's models were discussed in detail by H. Huang [24]. In this study, examples 3.2.6 a and b are treated only demonstratively.

4. Numerical calculations of bearing capacity of thin-walled structure

4.1. Material models. The considerations presented in section 3 show that tests performed on the same material may yield, in the case of a small number of samples, both divergent results (section 3.1) and divergent information on the variability of results (section 3.2).

The authors have thus asked themselves the question: what influence does this have on further evaluation of a construction product manufactured from the test material?

Based on results selected from the population and on the estimated uncertainties, material models were determined, based

on which further calculations of the load-bearing capacity and stability of the structural component of a building used as an example were made. The aim of these calculations was to determine to what extent the variability of a material model based on laboratory tests may affect the evaluation of the actual structure.

For further consideration, to create models A and B, results being extremely different to one another were selected from the population, in accordance with the following formulas:

$$R_A(e) = R_{\min}(e) - U_{\max}(e) \quad (11)$$

$$R_B(e) = R_{\max}(e) + U_{\max}(e) \quad (12)$$

Where:

$R_{\max}(e)$, $R_{\min}(e)$ – values of mean stresses calculated from five defined samples for a particular level of elongation. In order to calculate $R_{\max}(e)$, the five highest values of stresses selected from the entire population were used, and to calculate $R_{\min}(e)$, the five lowest values were used.

$U_{\max}(e)$ – the maximum value of expanded uncertainty at a confidence level of 95% for a particular level of elongation, calculated in accordance with the scheme presented in section 3.2.4 at a coverage factor $k_x = 2$ (the most often used in laboratories).

It was assumed that the laboratory had no knowledge of the material submitted for testing, and the coefficients required for computational evaluation of bearing capacity (including the elasticity module) were only obtained based on tests, therefore the Young's modulus (E) was calculated from curves $R_A(e)$ and $R_B(e)$.

The thus obtained material models, whose values are presented in Table 3, are characterised by the following: at a confidence level of 95% within the entire population, no value of stress lower than $R_A(e)$ and higher than $R_B(e)$ for a particular strain e may occur, even if the uncertainty of the result is taken into account.

For numerical computations using Ansys software, an elastoplastic multilinear material model with isotropic hardening was adopted [28, 29], which presented in coordinates $\sigma_{true} - \varepsilon_{ln}$ based on the following formula:

$$\begin{aligned} \varepsilon_{ln} &= \int_{l_0}^l \frac{dl}{l} = \ln\left(\frac{l}{l_0}\right) = \ln\left(\frac{l_0 + \Delta l}{l_0}\right) = \\ &= \ln\left(1 + \frac{\Delta l}{l_0}\right) = \ln(1 + \varepsilon_{eng}) \end{aligned} \quad (13)$$

$$\sigma_{true} = \sigma_{eng}(1 + \varepsilon_{eng}) \quad (14)$$

Where:

ε_{ln} – relative logarithmic strain, mm/mm

σ_{true} – true stress, MPa

$\sigma_{eng} = R$ – engineering stress (test result), MPa

$\varepsilon_{eng} = e$ – engineering strain (test result), mm/mm

Δl – increase in the sample length, mm

l_0 – initial sample length, mm.

Table 3

Values of stresses adopted for the material models applied in further calculations, based on the values of $R_A(e)$ and $R_B(e)$ calculated in accordance with equations 11 and 12

Strain e , (%)	Stress, R_A (MPa)	Stress, R_B (MPa)
0.02	41.1	73.8
0.06	106.5	159.0
0.1	170.7	240.1
0.14	234.6	312.1
0.18	284.4	380.0
0.5	344.0	384.9
0.8	342.8	382.1
2	361.4	391.2
6	356.4	398.6
10	360.2	400.3
15	373.0	407.7
20	373.8	408.9
25	363.8	403.8
Rm	374.0	410.2
E-Young's modulus	162.0	207.8

Unlike the data listed in Table 3, the strains are expressed in m/m. Such preparation of data is necessary for their proper implementation into the Engineering Data Sources module of the Ansys software. Graphical presentation of material models A and B is provided in Fig. 2.

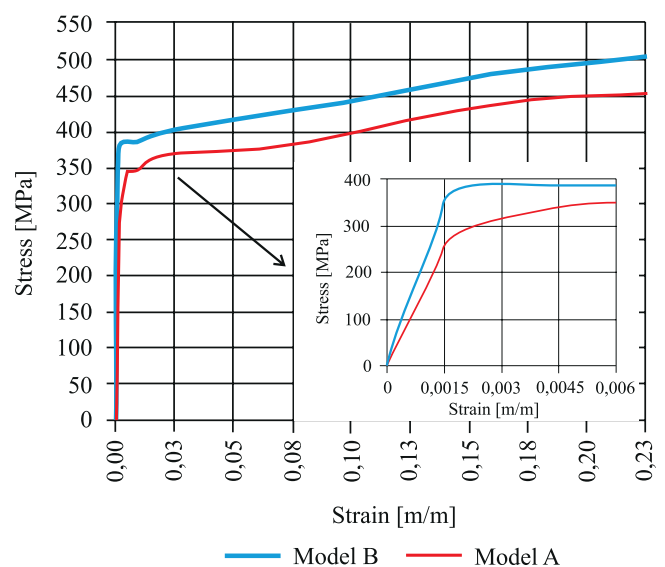


Fig. 2. Stress – strain characteristics of the material model (a) model A; (b) model B. Multilinear isotropic hardening

4.2. Test item. A comparison of the influence of the adopted material models on the was performed based on thin-walled structures used in the construction sector. A good example of such structures are profiled, circular arc-shaped steel sheets, used as roof covering for civil structures.

This type of structure was selected because they are particularly sensitive to changes in strength parameters of the material, as under certain forming conditions (e.g. large spans exceeding 20 m), the structure moves within a non-elastic range, and the loss of load-bearing capacity results from the emergence of local yield hinges.

Thin-walled profiled steel sheets are manufactured through the process of cold rolling, which allows one to obtain a single curved profile with characteristic corrugated web surfaces (Fig. 3).

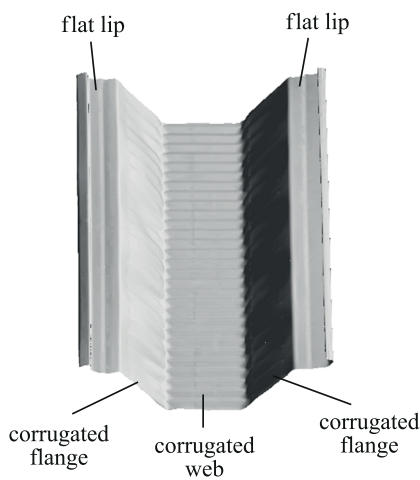


Fig. 3. Profile surface after rolling

Individual profiles are joined together by kneading the free edges in order to form a continuous trapezoid surface. The thus formed arc-shaped surface supported on the extreme edges is a self-supporting component with no supporting substructure.

The corrugating on the middle surfaces of the profile (Fig. 3) change its flexural rigidity as compared to a profile with straight walls, and increase the tendency towards the formation of local instabilities.

The issue of load-bearing capacity and rigidity of cold-rolled profiles has been addressed in studies and experiments [26, 27]. Based on these studies and experiments, it was concluded that the load-bearing capacity and stability of profiles are determined by strong nonlinearities associated with the geometry of the profile and material model.

For comparative purposes, it was assumed that the calculations would be performed on the same structure model and boundary conditions, only the material models would be changed.

4.3. Geometry of the model and boundary conditions. For numerical computations, a model of a roof covering was adopted in the form of a circular arc with the following dimensions:

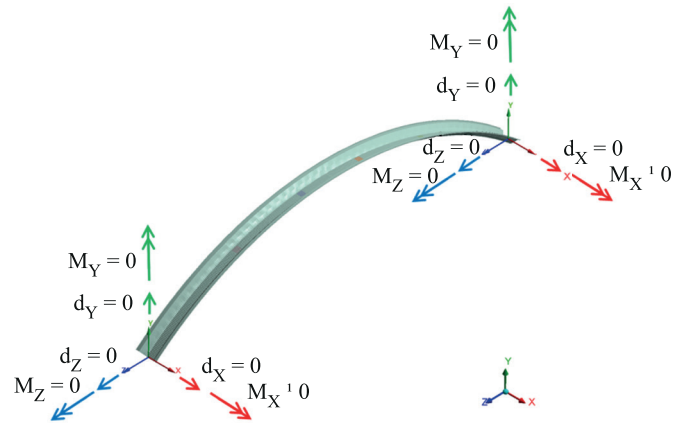


Fig. 4. Method of supporting the arc model

bend radius of 9 m; span of 12 m; elevation of 2.3 m. The model comprises a single segment with corrugated and wavy surfaces, as presented in Fig. 3. Corrugations and waves on the profile surface were adopted based on the measurements of the geometry of the actual component [26]. In order to represent physical conditions of the structure's behaviour, the influence of the adjacent arc segments was modelled by appropriate boundary conditions ("remote point" [28]) sets which allow a segment to be shifted in the direction of the force (direction along the Y and Z axes of the global co-ordinate system, Fig. 4)). The other degrees of freedom are blocked. Support is effected on the arc ends by "remote displacement" type supports, Fig. 4, in which displacements are blocked in three dimensions: d_x , d_y , d_z , rotations: M_y , M_z . Rotations of M_z supports in the direction perpendicular to the plane of the arc are released.

The load was represented using a set of concentrated forces, which is equivalent to the surface area loading, e.g. snow load (Fig. 5). In calculations, the set of forces was replaced by coupled kinematic inputs with gradual displacement up to the loosing of stability Kinematic impuls provides the convergence of nonlinear solutions.

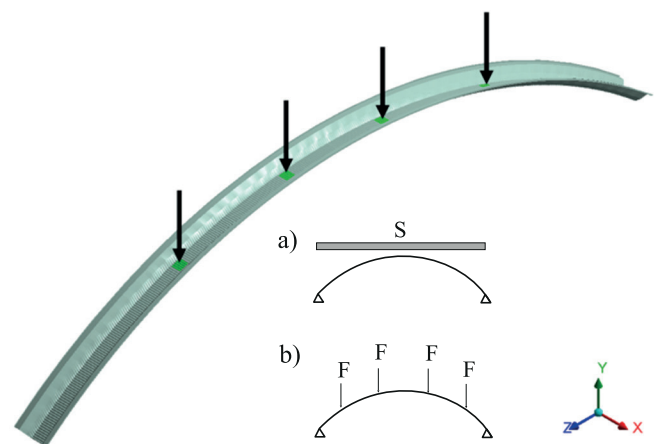


Fig. 5. Method of load application to the model; (a) load evenly distributed, (b) equivalent load

4.4. Selection of finite element mesh. In the model, a Shell 181 type element [28] with linear interpolation function was used, which is a 4-node shell element with six degrees of freedom at each node (three displacements and three rotations). Parameters of the finite element mesh were verified by standard diagnostic tools within the Ansys software. The parameters and evaluation of the quality of the mesh are as follows: number of nodes: 49.291; number of elements: 49.603; parameter element quality: 0.829; parameter aspect ratio: 1.262

Parameter element quality concerns the quality of the finite element mesh. This parameter falls within the range between 0 and 1, where the value of 0 for the surface (2D) is defined as an element of insufficient quality, and the value of 1 denotes a very good mesh.

The parameter Aspect Ratio in the surface arrangement (2D) is defined as the ratio of the longest side to the shortest side of the finite element. The mesh has the best parameters when the Aspect Ratio coefficient achieves the value of 1.0. Detailed rules for the determination of these parameters are presented in paper [28].

It was assumed that the finite element mesh had sufficiently good quality and could be adopted for further calculations.

4.5. Computation results. Fig. 6 presents computation results for the two analysed material models. Full lines present the maximum deformation of the model as a function of force (scale: load – total deformation), broken lines present the maximum von Mises stress as a function of force (scale: load – equivalent stress).

The loss of stability of model A (point A1) occurred at a load of 14.04 kN and deflection of 68.19 mm. The loss of stability of model B (point B1) occurred at a load of 17.63 kN and deflection of 56.03 mm. At boundary values of stress and deflection,

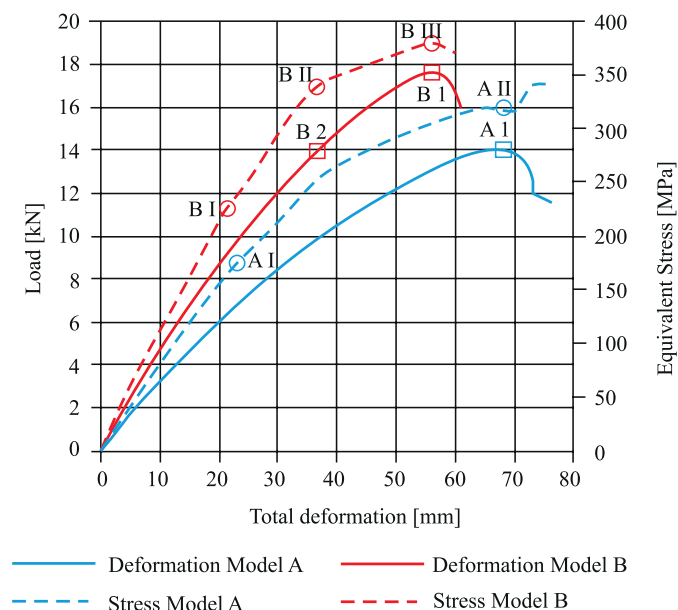


Fig. 6. Results of computations of maximum displacements and stresses

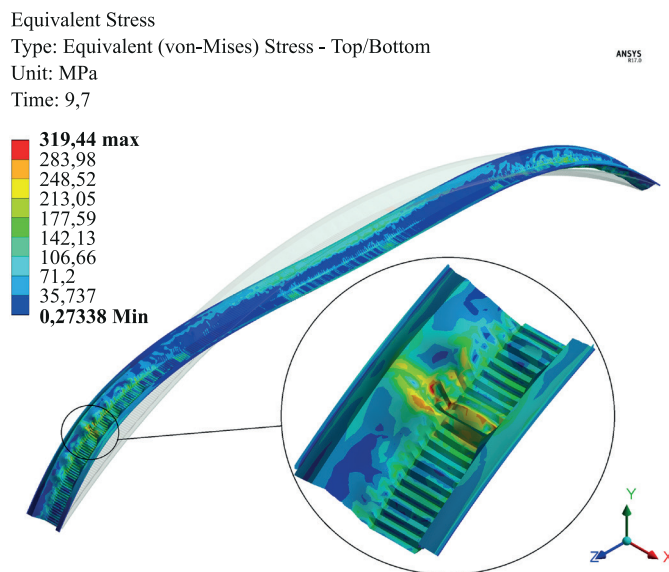


Fig. 7. Map of stresses for model B. Local loss of stability

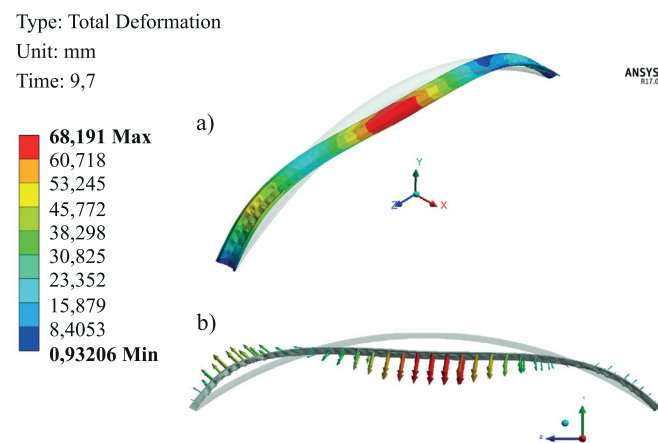


Fig. 8. Deflections of model A (a) map of displacements, (b) displacement vectors

the difference between the obtained results is 20% in relation to the stress and 18% in relation to the deflection. The stresses get into the prebuckling state [29] at the following values, respectively: model A (point AI), deflection of 23 mm, stress of 176 MPa, and model B (point BI), deflection of 21.3 mm, stress of 226 MPa. Having referred the boundary values of the load and deflection obtained in model A (point A1) to model B (point B2), it was concluded that at a load of 14.04 kN in model B, deflection of 36.73 mm was obtained, which is approximately 55% of the deflection of model A. At the moment when model A loses stability due to the exhaustion of load-bearing capacity (Fig. 7), the directional displacement vectors (Fig. 8) show non-uniform displacements, particularly in the lateral areas of the arc.

This is due to the emergence of a local yield hinge, and the structure's transition into the stage of failure.

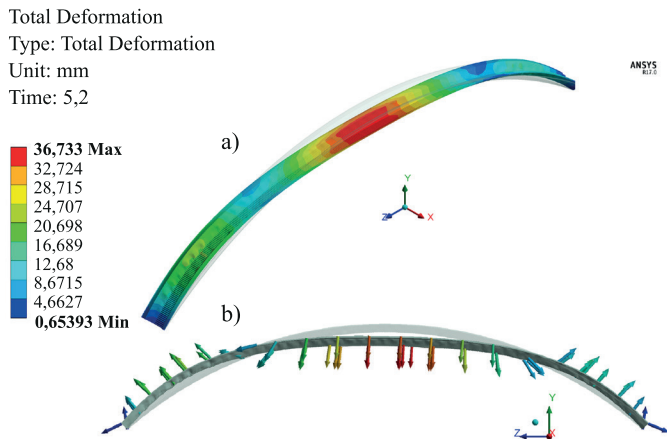


Fig. 9. Deflections of model B (a) map of displacements, (b) displacement vectors

At the same level of load (14.04 kN), in model B (Fig. 9) the displacement vectors are symmetrically distributed on both sides of the arc. This means that at this level of load, no signs of the occurrence of local instabilities become apparent.

It is also worth noting that at the moment when model A entered the stage of failure (Fig. 6, point A II), model B, at the same level of load (Fig. 6, point B II), shows the reserve load-bearing capacity (section B II and B III).

Having analysed an example of a canopy roof of a warehouse building with dimensions as those presented in section 4.3, it can be concluded based on [30] that permissible deflection should not exceed $L/250$, where L is the span of the covering roof. In the case under discussion $L = 12.000$ mm, which yields the critical value of 48 mm. In the analysed models, the permissible load at the deflection of 48 mm is 11.5 kN for model A, and 16.3 kN for model B. When calculated per a unit of the roof surface area, and having deducted the dead load, the permissible surface area loadings due to snow are, respectively, $A = 1.47$ kN/m² and $A = 2.14$ kN/m². Having further adopted the load on the cylindrical roof of a building protected from wind in the 4th snow zone in the first option of load [31], the permissible load of 1.54 kN/m² is obtained. In the case of model A, the serviceability limit state is not satisfied, and in the case of model B, it is satisfied with a reserve of almost 40%. The load-bearing capacity limit state is satisfied in both cases.

Presented different evaluation of the product, results only from the use of different input parameters related to the material (test results) and does not include FEM model uncertainty (different results arising from different geometry, boundary conditions, loads etc.)

5. Discussion

Laboratory testing results are affected by numerous factors which may result in both the value and the variability assigned to it being significantly different in different laboratories. In the other words – both: value and its variability are burdened with uncer-

tainty (consisting of parameter and model uncertainties). Additionally, the performance of laboratory tests on a small number of samples (3÷5) leads to differences between results and their dispersions, resulting from the random selection of samples from the population, which was shown while comparing standard deviations for five-sample sets taken from the same population. The randomly selected set containing a small number of samples may have characteristics which differ significantly (the expectation value, variance) in relation to the population.

During laboratory tests, random factors emerge which affect test repeatability and reproducibility (including, *inter alia*: the operator, type of equipment, environmental conditions, mounting of the sample, and many others), and systematic factors which result from the accuracy of the testing equipment used. All these factors contribute to the uncertainty which characterises variability of the result. As regards tests (e.g. mechanical) in which a sample is destroyed, the fact that, in the variability of results, it is difficult to isolate the proportion of heterogeneity of test items, also plays a significant role. For all of those reasons, a laboratory performing a test on an unknown material usually obtains very rough information on the tested characteristics and its variability. At the same time, due to the lack of uniform methods for the estimation of uncertainties laboratories may obtain different information on variability in relation to the same set of results.

In the case of tensile tests according to EN ISO 6892–1 [20], considered as input data for calculations, the dominant components of uncertainty were due to the measuring equipment. Therefore, the uncertainties estimated using appropriate methods, with the same assumptions adopted in relation to the equipment, did not differ significantly. However, in the case of tests according to EN 846–9 [15], discussed in section 2, the dominant component was the random dispersion of results. In addition, these tests are performed in accordance with the standard on as few as three samples. In such a case, knowledge of variability which can be assigned to a test result is very rudimentary. A typical approach of laboratories involved in testing unknown products using a small number of samples is to estimate the components of uncertainties resulting from random effects by taking into account the current SD obtained in the test. However, a question should be asked here: is such an approach correct? The data presented in Table 1 suggest that for each product, the uncertainty of a test in the same laboratory will be different because sample SD also contains a component associated with variability of the tested product. In addition, differences may appear which result from the method for determination of the coverage factor. For example, $k_x = t_{95}$ for $N = 3$, i.e. for $\nu = 2$ degrees of freedom, is 4.3, while the factor in the Craig model, i.e. $k_x = \frac{k}{c_4}$ for $N = 3$, would be 2.26. With the same SD, this will produce a difference of almost 100% in estimated uncertainty.

This paper demonstrates, based on an example, that the variability of laboratory testing results (in this case, tensile test) may have a very significant effect on the final evaluation of a construction product. For the applied material models A and B, extremely different results were adopted and maximum uncertainties were taken into account. This may arouse suspicion

that the probability of obtaining such great differences is actually low. On the other hand, however, great simplifications were applied: neither the uncertainty associated with elongation nor the variance related to reproducibility were taken into account (data obtained under repeatability conditions were used). The results obtained in two different laboratories could increase the differences between models A and B even more.

The assumed discrepancies in results of tests on the same material appeared to be great enough so that the calculations performed using them indicated a contradictory classification of the same structural component in a civil structure. In this case, it was a covering roof in the form of a sector of a circular arc made of thin-walled profiled cold rolled sheets. As regards data for material model A, the serviceability limit state was exceeded, while for the same conditions (geometry, supports and the method of load application) yet with material model B applied, the serviceability limit state is satisfied with a reserve of almost 40%. The inconsistency of this evaluation brings about far-reaching practical consequences. In the first case (model A), the covering roof design should be rejected or forwarded for correction. In the second case (model B), the design is correct; moreover, one can be tempted to perform dimensional optimisation e.g. by thinning the profile. When designing only based on data concerning materials A or B (without knowing their mutual relationships), each designer can be certain of the correctness of their decision. However, one needs to bear in mind the reasons for drawing such conclusions, particularly in the context of variability of parameters which are of significance in the final evaluation.

There are aspects due to which the issue of taking the variability of laboratory tests results into account in the products evaluation should become the subject of considerably more thorough analyses. Moreover these analyses should be performed at the global level and not only at the level of a single laboratory. These aspects are the lack of uniform principles of test results uncertainty estimation, and the lack of uniform and rational principles -while taking this uncertainty into account in the final evaluations of products and structures. Such principles should be adapted to specific tests and specific products.

Even though Eurocodes related to designing a structure [32] recommend that variability of properties of materials should be taken into account during the design process, they are however recognised as one of the huge number of variables, and little attention is paid to this issue. Designers who work based on Eurocodes apply simplifications which involve adopting tabular values for parameters of the materials. These values have solid foundations in the form of dozens of performed tests; however, these are numbers devoid of readily available and actually used information concerning possible variability. It is hard to believe that the adoption of tabular values based on the type (composition) of a material declared by the manufacturer is burdened with a smaller uncertainty than laboratory tests on the submitted material.

Given the lack of possibility for an analysis of variability of a test result which would be reliable and take all aspects into account, another question arises here: is such an analysis actually necessary? In the authors' opinion, transferring the principles

for estimation of measurement uncertainty onto the actions performed by a testing laboratory is unnecessary and creates the wrong impression of having precise knowledge of the test result and PDF assigned to it. The variability of the test results in the discussed cases is a resultant of the properties of the tested product and the actual uncertainty which can be assigned to the test. What is more, it can be erroneously evaluated where the number of samples is little. Arrangements which involve imposing on a product conditions concerning not only the tested quantity but also its dispersion, and taking into account, in the estimation of uncertainties, only the components associated with the testing equipment, would be much more reasonable.

6. Conclusions

Properly estimated uncertainty of a test result should, at a confidence level of 95%, include a real value, and should be related to the aspects associated with both the repeatability and reproducibility of tests. In most cases of testing construction products, this is not the case. In the absence of the possibility of reliable analysis, simplified methods are adopted, not satisfying these conditions. Laboratories performing tests on construction products evaluate them based on a small number of samples, most often 3÷5. Therefore, this is an evaluation based on poorly known characteristics, which may be burdened with a significant error.

Variability at the level of laboratory tests results may, as has been demonstrated in this paper, have a significant effect on the evaluation of a product or a structure. In extreme cases, this may lead to taking inappropriate decisions on either acceptance or rejection of products, technical implementations or designs directed for implementation.

Despite numerous normative documents and publications on uncertainty, methods for its estimation and taking into account in evaluations, this issue is still imprecise in terms of testing and assessment of construction products. There is a need for development of specific and uniform guidelines concerning individual products and tests. In the authors' opinion, these guidelines should be supported by numerous interlaboratory tests.

Acknowledgements. The authors would like to express their thanks to Professor Lech Czarnecki for his invaluable inspiration, and to their colleagues from the ITB Laboratory, particularly to Przemek Więch, for providing laboratory tests results. The financial support from National Center for Research within the project OPT4BLACH (PBS1/A2/9/2012) is gratefully acknowledged.

REFERENCES

- [1] EA 4/16 G:2003, "EA guidelines on the expression of uncertainty in quantitative testing", <http://www.european-accreditation.org/publication/ea-4-16-g-rev00-december-2003-rev>.
- [2] L. Czarnecki and J.J. Sokołowska, "Material model and revealing the truth", *Bull. Pol. Ac.: Tech.* 63 (1), 7–14 (2015), doi:10.1515/bpasts-2015-0001.

- [3] A.S. Nowak and A.M. Rakoczy, “Uncertainties in the building process”, *Bull. Pol. Ac.: Tech.* 61 (1), 129–135 (2013), doi:10.2478/bpasts-2013-0011.
- [4] EU Regulation No. 305/2011 Of The European Parliament and of the Council of 9 March 2011 laying down harmonised conditions for the marketing of construction products and repealing, *Council Directive 89/106/EEC*, <http://ec.europa.eu/growth/sectors/construction/product-regulation>.
- [5] E. Bashkansky, T. Gadrich, and D. Knani, “Some metrological aspects of the comparison between two ordinal measuring systems”, *Accred. Qual. Assur.* 16, 63–72 (2011), doi:10.1007/s00769-010-0741-2.
- [6] P.Th. Wilrich, “The determination of precision of qualitative measurement methods by interlaboratory experiments”, *Accred. Qual. Assur.* 15, 439–444 (2010), doi:10.1007/s00769-010-0661-1.
- [7] K. Hori, Y. Tsutsumi, Y. Takao, and T. Suzuki, “Calculation of repeatability and reproducibility for qualitative data”, National Sun Yat-sen University, <http://bm.nsysu.edu.tw/tutorial/iylu/ANQ%202008/11.%20Session%20G/G5-01.doc>.
- [8] JCGM 100 (2008), “Evaluation of measurement data: guide to the expression of uncertainty in measurement (GUM)”, http://www.bipm.org/utis/common/documents/jcgm/JCGM_100_2008_E.pdf.
- [9] W.E. Walker, P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, and M.P. Kraye von Krauss, “Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support”, *Integrated Assessment* 4 (1), 5–17 (2003), doi:10.1076/iaij.4.1.5.16466.
- [10] EN 206:2013, “Concrete – specification, performance, production and conformity”.
- [11] ISO 12491:1997, “Statistical methods for quality control of building materials and components”.
- [12] ILAC G8:03/2009, “Guidelines on the reporting of compliance with specification”, <http://ilac.org/publications-and-resources/ilac-documents/guidance-series/>.
- [13] JCGM 200 (2012), *International vocabulary of metrology: basic and general concepts and associated terms (VIM)*, 3rd ed., <http://www.bipm.org/vim>.
- [14] EA-4/02 M:2013, “Evaluation of the uncertainty of measurement In calibration”, <http://www.european-accreditation.org/publication/ea-4-02-m-rev01-september-2013>.
- [15] EN 846-9:2000, “Methods of test for ancillary components for masonry – Part 9. Determination of flexural resistance and shear resistance of lintels”.
- [16] E. Szewczak, “Risks associated with the uncertainty of test results and the assessment of construction products’ conformity”, *Building Materials* 10, 73–75 (2011), [in Polish].
- [17] P. Fiscaro, S. Amarouche, B. Lalere, G. Labarraque, and M. Priel, “Approaches to uncertainty evaluation based on proficiency testing schemes in chemical measurements”, *Accred. Qual. Assur.* 13, 361–366 (2008), doi:10.1007/s00769-008-0402-x.
- [18] S.K. Wong and W.Y. Yao, “Bias analysis of proficiency testing programme results”, *Accred. Qual. Assur.* 21, 185–189 (2016), doi:10.1007/s00769-016-1197-9.
- [19] E. Szewczak and A. Bondarzewski, “Is the assessment of inter-laboratory comparison results for a small number of tests and limited number of participants reliable and rational?”, *Accred. Qual. Assur.* 21, 91–100 (2016), doi:10.1007/s00769-016-1195-y.
- [20] EN ISO 6892-1:2009, “Metallic materials – Tensile testing – Part 1: Method of test at room temperature”.
- [21] L. Brunarski and M. Dohojda, “Diagnostics of concrete strength in structures”, Instytut Techniki Budowlanej, Warsaw, 2014, [in Polish].
- [22] JCGM 101 (2008), “Evaluation of measurement data – Supplement 1 to the Guide to the expression of uncertainty in measurement – Propagation of distributions using a Monte Carlo method”, http://www.bipm.org/utis/common/documents/jcgm/JCGM_101_2008_E.pdf.
- [23] T. Lafarge and A. Possolo, *NIST Uncertainty Machine – User’s Manual*, Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg, Maryland, USA, 2015, <http://uncertainty.nist.gov/>, accessed August 2016.
- [24] H. Huang, “Comparison of uncertainty calculation models”, *Cal. Lab. Int. J. Metrol.* 19 (1), 24–29 (2012).
- [25] M. Roesslein, M. Wolf, B. Wampfler, and W. Wegscheider, “A forgotten fact about the standard deviation”, *Accred. Qual. Assur.* 12, 495–496 (2007).
- [26] A. Piekarczyk, K. Malowany, P. Więch, M. Kujawińska, and P. Sulik, “Stability and bearing capacity of arch-shaped corrugated shell elements: experimental and numerical study”, *Bull. Pol. Ac.: Tech.* 63 (1), 113–123 (2015), doi:10.1515/bpasts-2015-0013.
- [27] A. Piekarczyk and K. Malowany, “Comparative analysis of numerical models of arch-shaped steel sheet sections”, *Archives of Civil and Mechanical Engineering* 16 (4), 645–658 (2016), doi:10.1016/j.acme.2016.04.006.
- [28] *Theory Reference for the Mechanical APDL and Mechanical Applications*, ed. P. Kohnke, Ansys Inc., USA, 2009.
- [29] M. Kotełko, *Load-Bearing Capacity and the Mechanisms of Failure of Thin-Walled Structures*, pp. 18–20, Wydawnictwa Naukowo-Techniczne, Warszawa, 2011, [in Polish].
- [30] EN 1993-1-1:2005, “Eurocode 3: Design of steel structures – Part 1-1: General rules and rules for buildings”.
- [31] EN 1991-1-3:2003, “Eurocode 1: Actions on structures – Part 1-3: General actions – Snow loads”.
- [32] EN 1990:2002, “Eurocode – Basis of structural design”.