

## CHECKING THE ACCURACY OF LABORATORY TESTS

*Otto Dvořák, František Wald*

*University Centre for Energy Efficient Buildings of Czech Technical University in Prague, Třínecká 1024, 273 43 Buštěhrad, Czech Republic; otto.dvorak@uceeb.cvut.cz*

### ABSTRACT

In fire engineering, the quality of results of crucial test methods is key for validation of advanced numerical models. Therefore, it is important that the correct functioning of the test equipment and test procedures is achieved. This contribution is an assessment of the accuracy respectively the precision of the test methods based on the knowledge of the repeatability  $r$  and the reproducibility  $R$  determined by the round robin test (inter laboratory comparison) with reference testing materials.

### KEY WORDS

Repeatability; Reproducibility; Repeatability limit; Reproducibility limit; Test method; Test result; Accuracy; Precision; Test apparatus; Reference material

### INTRODUCTION

Thermal technical and fire characteristics of materials are important input data of CFD fire models that determine the quality of their predictions. These characteristics are usually determined by ISO, IEC, CEN or national test standards/procedures. These standards are validated. Standards for the determination of thermal technical characteristics of materials/products sometimes specify the conformity of measurements under repeatability and reproducibility conditions using the repeatability limits  $r$  and the reproducibility limits  $R$ . They also contain procedures for verifying the correct function of the test apparatus by comparing the difference between the test results measured by the interlaboratory or intralaboratory comparison on CRM (Certified Reference Materials) or SWS (Secondary Working Standards) substances/materials and their nominal values, with the tolerance expressed in the relevant standards with the  $r$  and/or  $R$  inequality.

**The repeatability limit,  $r$**  is the difference between the two test results obtained by the same operator with the same instrument under constant operating conditions on the same test material over the long term under the normal and correct execution of the test method exceeded the value given in the standard only in one case from twenty. (e. g.  $r = 1.4 \text{ }^\circ\text{C}$  (see EN ISO 13736 [1])).

**The reproducibility limit,  $R$**  is the difference between the two individual and independent test results obtained by different operators working in different laboratories with the same test material over the long term under the normal and correct execution of the test method exceeded the value in formula (3) only in one case of twenty. (e. g.  $R = 3.2 \text{ }^\circ\text{C}$  (see ISO 13736 [1])).

The above quoted standard requires verification after purchase experimental apparatus and subsequently periodically to verify its trueness/ not damage either:

a/ with one test performed on CRM and SWS, with the difference between a single result and a CRM or SWS reference value being within the following tolerance:

$$|x - \mu| \leq R/\sqrt{2} \quad (1)$$

where:

$x...$  is the test result,

$\mu....$  is the certified CRM value or SWS reference value,

$R...$  is the limit of reproducibility of the test method,

b) or n-times repeated tests performed on CRM or SWS; the difference between the average of n results and the CRM certified value or the SWS reference value should be within the following tolerance:

$$|\bar{x} - \mu| \leq R_1/\sqrt{2} \quad (2)$$

where:

$\bar{x} ...$  is the arithmetic mean of repeated test results,

$\mu ...$  is the CRM certified value or SWS reference value,

$R_1...$  is equal to the formula given by formula (3):

$$R_1 = \sqrt{R^2 - r^2(1 - 1/n)} \quad (3)$$

where:

$R...$  is the limit of reproducibility of the test method,

$r...$  is the limit of repeatability of the test method,

$n...$  is the number of retests performed on CRM or SWS.

The question is, whether the above mentioned inequalities are statistically correct.

### Statistical model

The error of the determination result can be defined as the difference between the result and the actual / reference value of the measured quantity, i.e. according to equation (4)

$$x - \mu = \xi + e \quad (4)$$

where:

$x...$  is the test result,

$\mu ...$  is the actual / reference value of the measured quantity,

$\xi...$  is the error caused by deviation from defined experiments conditions,

$e...$  is the random error that occurs even if the repeatability conditions are met,

The variance of the total error ( $X - \mu$ ) denote  $\sigma^2$ , equals the sum of the variance  $\xi$  and the variance  $\epsilon$  according to the Equation:

$$\sigma^2 = \sigma_\xi^2 + \sigma_e^2 \quad (5)$$

where:

$\sigma^2...$  is the total variance under reproducibility conditions,

$\sigma_\xi^2...$  is the component of the whole variance representing the variability of uncontrollable factors,

$\sigma_e^2...$  is the variance under repeatability conditions.

These components of the total variance correspond to the respective reproducibility limits  $R$  value and repeatability  $r$ . If  $R$  and  $r$  are known from the round tests (e.g. they were determined during validation of the test method), the test concerned standard requires that two measurements  $x_1$  and  $x_2$ , measured under the following conditions:

1. the repeatability met the following criterion with a probability of 0.95:

$$|x_1 - x_2| \leq r \quad (6)$$

If the difference is greater than  $r$ , i.e. according to Equation (7), we would consider the measured values as being overloaded by an excessive error (as "unrepeatable").

$$|x_1 - x_2| > r \quad (7)$$

2. the reproducibility met the following criterion with a probability of 0.95:

$$P((x_1 - x_2) \leq R) = 0.95 \quad (8)$$

If the difference is greater than  $R$ , i.e. according to Equation (8), we would consider the measured values as being overloaded by an excessive error (as "unreproducible").

$$|x_1 - x_2| > R \quad (9)$$

If the reference / nominal values  $\mu$  of measured variables, such as CRM or SWS materials, are available, then a pair of  $x_1$  and  $x_2$  values is not necessary to verify e.g. the "reproducibility" of the results; one measurement/one result would be sufficient, but the following criterion would have to be met:

$$P((x) \leq R/\sqrt{2}) = 0.95 \quad (10)$$

If the reference value  $\mu$  of the measured quantity is known and a possibility of repeating this determination is possible, the determination can be realized of either:

a/ **n- times under repeatability conditions** (same day, in the same laboratory, on the same apparatus, with the same operator),

b / or **n- times under reproducibility conditions** (e.g. in  $n$  different laboratories, other times, on a third-party apparatus, of course, with other staff).

#### Ad a /

By selecting a particular laboratory, the measurement date, a specific operator, and a test equipment is chosen a specific system of repeatability conditions. Then a total error  $(x - \mu)$  is equal to the sum of the error of this particular system  $\xi$  and the arithmetic error mean  $\bar{e}$  under repeatability conditions, that is

$$|\bar{x} - \mu| = \xi + \bar{e} \quad (11)$$

This corresponds to variable:

$$\text{Var } |\bar{x} - \mu| = \text{Var} (\xi) + \text{Var} (\bar{e}) \quad (12)$$

From the Equations (5) ( $\sigma_{\xi}^2 = \sigma^2 - \sigma_e^2$ ) and (12) result that:

$$\text{Var } |\bar{x} - \mu| = \sigma^2 \cdot \sigma_e^2 + \sigma_e^2 / n = \sigma^2 - \sigma_e^2 (1 - 1/n) \quad (13)$$

Since  $\sigma^2$  and  $\sigma_e^2$  are proportional to the limits of reproducibility  $R$  and repeatability  $r$  (in the same order), equation (13) can be rewritten as follows:

$$|\bar{x} - \mu| \leq \frac{1}{\sqrt{2}} \cdot \sqrt{R^2 - r^2(1 - 1/n)} \quad (14)$$

The validity of the above criteria according to Equations (2) and (3) is confirmed.

Furthermore, it is clear that failure to meet the criterion  $|\bar{x} - \mu| \leq R_1 / \sqrt{2}$  indicates that:

- a) a test device/method does not meet the required reproducibility limit or,
- b) a laboratory in question is working with a large systematic error, or that the environment conditions are significantly different from normal conditions, etc.

#### Ad b /

*The standard test method* could further recommend in its Annex to the Testing Laboratory another option to verify the trueness of its test equipment and the trueness of the test by inter-laboratory comparison with another laboratory and with the same test sample, e.g. under reproducibility conditions.

Then, according to the laws of probability and mathematical statistics:

$$\text{Var } |\bar{x} - \mu| = \text{Var } (\xi + e) \cdot 1/n = \sigma^2 \cdot 1/n^1) \quad (15)$$

With a following assessment criterion:

$$|\bar{x} - \mu| \leq R / \sqrt{2}n \quad (16)$$

To complete:

- If the values  $x$  and  $\mu$  on a specific experimental device are known, then on average of many repetitions should apply that  $|\bar{x} - \mu| = 0$ ,
- if  $|\bar{x} - \mu| \neq 0$ , it signals a systematic error on the device,
- if  $|\bar{x} - \mu| > \xi$ , and  $P(|\bar{x} - \mu|) > \xi = \alpha$

- balance sheet can be supplemented by:

- if  $|\bar{x} - \mu| > \xi$ , and  $P(|\bar{x} - \mu|) > \xi = \alpha$   
when  $\alpha$  is the significance level,  
 $\xi = 1,96.s$  when  $\alpha = 0,05$   
 $\xi = 1,64.s$  when  $\alpha = 0,1$ .

etc, when  $s$  (or  $\sigma$ ) is the standard deviation of random errors,

- if from repeated measurements  $|\bar{x} - \mu| > \xi$  (systematic error exists) then  $\xi = 1,96. s / \sqrt{n}$

<sup>1)</sup> The variance of a random variable  $X$  is often referred in statistics either as  $\text{Var}(X)$ , (*variance*  $X$ ),  $\sigma^2(X)$ ,  $s^2(X)$  or  $D(X)$

and  $\xi = t \cdot s / \sqrt{n}$ . The selective standard deviation  $s$  can be estimated from the known formula

- $s = \frac{1}{n-1} \cdot \sum_1^n (x_i - \bar{x}^2)^{1/2}$ , when the coefficient  $t$  depends on  $n$  and  $\alpha$ . Related standards, see [2-7].

## CONCLUSION

The practical applicability of the above mentioned criteria for the correct function of a test device has been confirmed by the facts according to the relations (1), (2) and (3).

A more precise justification of a test equipment provides inter laboratory comparison (round – robin test) with other laboratories and evaluating the determination results according to the above derived relationships (15) and (16).

This option could be recommended during the revision of test standards in the form of a supplement to an already established Annex for verification of test equipment. It could also be used by research laboratories developing new test methods and / or accredited testing laboratories or laboratories with a good laboratory practice.

When working with specific  $r$  and  $R$  values, they should not be labelled as repeatability and reproducibility, as is commonly in the standards, but correctly as the limits of repeatability and reproducibility [7].

## ACKNOWLEDGEMENT

This work has been supported by the Ministry of Education, Youth and Sports within National Sustainability Programme I (NPU I), project No. LO1605 and the Ministry of Interior, project No. VI20162019034.

## REFERENCES

- [1] EN ISO 13736: 2013. Determination of flash point- Abel closed cup method.
- [2] ISO 5725-1: 1994/Cor. 1:1998-02. Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions.
- [3] ISO 5725-2: 1994/Cor. 1:2001-10. Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.
- [4] ISO 5725-3: 1994/Cor. 1:2001-10. Accuracy (trueness and precision) of measurement methods and results – Part 3: Basic method for the determination of the precision of a standard measurement method.
- [5] ISO 5725-4: 1994. Accuracy (trueness and precision) of measurement methods and results – Part 4: Basic method for the determination of the trueness of a standard measurement method.
- [6] ISO 5725-5: 1998/Cor.:2005-08. Accuracy (trueness and precision) of measurement methods and results – Part 5: Basic method for the determination of the precision of a standard measurement method.
- [7] ISO 5725-6: 1994/Cor.1:2001-10. Accuracy (trueness and precision) of measurement methods and results – Part 6: Use in practice of accuracy values.
- [8] ISO 3534-1:2006. Statistics – Vocabulary and symbols – Part 1: General statistical terms used in probability.
- [9] ISO 3534-2:2006. Statistics – Vocabulary and symbols – Part 2: Applied statistics.