



HAL
open science

Data comparisons and uncertainty: a roadmap for gaining in competence and improving the reliability of results

Abdérafi Charki, Franco Pavese

► To cite this version:

Abdérafi Charki, Franco Pavese. Data comparisons and uncertainty: a roadmap for gaining in competence and improving the reliability of results. *International Journal of Metrology and Quality Engineering*, 2019, 10, pp.1. 10.1051/ijmqe/2018016 . hal-02531103

HAL Id: hal-02531103

<https://univ-angers.hal.science/hal-02531103>

Submitted on 7 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Data comparisons and uncertainty: a roadmap for gaining in competence and improving the reliability of results

Abdérafi Charki^{1,*} and Franco Pavese²

¹ University of Angers, LARIS-ISTIA, 62 avenue Notre Dame du Lac, 49000 Angers, France

² INRIM, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce, 91, 10135 Torino, Italy

Received: 24 September 2018 / Accepted: 19 November 2018

Abstract. This paper traces a roadmap for gaining in competence and for improving the reliability of results in a laboratory. The roadmap was built from the requirements concerning the results quality and measurement uncertainty, which accreditation bodies use for the accreditation of testing and calibration laboratories. In industry, accreditation is the accepted proof of a laboratory's assigned level of competence. The level of performance of a laboratory is demonstrated through the quality of its management of test and calibration results. Inter-laboratory comparisons and the evaluation of measurement uncertainties are recommended as the most appropriate methods for demonstrating continuous improvement in laboratories. The common methods used for data comparisons and for the evaluation of measurement uncertainties are highlighted. An overview of the main indicators used in data comparisons is presented. Some recommendations are made that are useful to the design of a roadmap for gaining in competence and for improving the quality of results obtained by a laboratory.

Keywords: ISO/IEC 17025 Standard / laboratory / comparisons / proficiency testing / uncertainty / metrology / reliability of results / competence

1 Introduction

Performance can only be assessed if it can be measured by means of an indicator. An organization, a product, a person, etc., can only be appropriately assessed if one or more relevant indicators are available with which to make the assessment.

An evaluation of the performance of an entity inevitably looks at the way in which results are obtained, and expects to find proof of appropriate levels of aptitude and competence.

Within the industry, more and more laboratories are opting for accreditation in order to ensure recognition of their competence.

In the field of testing and calibration, for example, the Standard used for the accreditation of laboratories is ISO/IEC 17025 [1], entitled "General requirements for the Competence of Testing and Calibration Laboratories".

This Standard describes the organizational and technical instructions laboratories must follow in order to be able to demonstrate their competence via a certified level of quality of laboratory management. It is important to note that simply adapting and applying the

instructions is not sufficient. Each laboratory must also be evaluated by a recognized, independent, impartial and reliable authority.

In this article, the main technical requirements of the Standard are examined, with focus primarily on the notions of making the correct choices and following the correct procedures in estimating uncertainties and reliability of measurement results.

These requirements are absolutely essential when assessing the competence of a laboratory and therefore evaluating its performance.

Among the technical requirements of the ISO/IEC 17025 Standard is the requirement to estimate the uncertainties and to provide proof of the quality of the results supplied to customers by means, for example, of inter-laboratory comparisons. A few principles regarding the organization of an inter-laboratory comparison are also set out.

The requirements of this Standard encourage laboratories to examine the key areas where improvements may be achieved, namely, in the method, the means, the environment, the materials, the manpower, as well as the management.

It is impossible to satisfy the technical criteria of the ISO/IEC 17025 Standard without attention to organization. The requirements concerning quality management of the Standard must therefore be met.

* Corresponding author: abderafi.charki@univ-angers.fr

One of the recommended methods that is increasingly used by testing and calibration laboratories for evaluating measurement uncertainties is the ISO/IEC Guide 98-3 (GUM) document [2]. It is very straightforward to apply it to calibration laboratories, though less for testing laboratories, and is therefore less commonly used by them.

Today, there is little doubt about the need to evaluate uncertainties of measurement, thanks to the ISO/IEC 17025 Standard. In fact the performance and competence of accredited laboratories is largely expressed in terms of their mastery of measurement uncertainty and the trueness of the results obtained. Companion standards ISO 13528 [3] and ISO 5725 [4], mainly used by testing laboratories, are complementary to it in the case of proficiency testing and inter-comparisons.

The concept of comparison has become widely used in many sectors, and many new Proficiency Test (PT) schemes are currently launched each year worldwide [5,6]. A detailed study of comparison testing has been conducted [6]. The International Organization for Standardization (ISO) has published a reference document on general requirements for proficiency testing [7] and a standard on statistical methods for use in proficiency testing [3]. The International Laboratory Accreditation Corporation (ILAC) has also published a document on requirements for proficiency testing [8].

Pavese et al. [9] go over the steps that allow knowledge to be developed, presenting the aims and characteristics of the various inter-laboratory comparison methods, notably referring to the tools established by documents such as the BIPM MRA (the Mutual Recognition Arrangement) and the ISO 13528 [3] and ISO 5725 [4] standards.

This article presents a roadmap for gaining in competence and for improving the reliability of results obtained by a laboratory. Section 2 summarizes the main requirements regarding uncertainty measurement, validation methods and inter-laboratory comparisons. Section 3 deals with the different types of data comparisons (intra- and inter-laboratory comparisons). Section 4 highlights the different methods used for uncertainty evaluation in calibration and testing laboratories. Section 5 gives the main key indicators for monitoring the performance of a laboratory, and highlights certain recommendations to improve the methodology for data comparisons and the accreditation process of a laboratory.

2 Ensuring reliability of results

The accreditation standard for testing and calibration laboratories is ISO/CEI 17025 [1], which contains the requirements relating to management aspects and the technical requirements to ensure the accuracy and reliability of results obtained by a laboratory.

A great many factors can influence the accuracy and reliability of testing and/or calibration carried out by a laboratory [1]. Some of these factors may include the following:

- level of competence of personnel;
- equipment and ambient conditions;

- handling and storage of objects for test and calibration;
- measurement traceability;
- sampling;
- collection;
- testing and calibration methods (whether developed or adopted, standardized or non-standardized).

The means used to ensure the quality of testing and calibration results:

- regular use of certified reference materials and/or internal quality control with use of secondary reference materials;
- participation in comparison programs between laboratories or proficiency tests;
- tests or calibration repeated using identical or different methods;
- renewed tests or renewed calibration of stored objects;
- correlation of results for the different characteristics of an object.

The section 7.7 of the ISO/IEC 17025 Standard also raises other points:

- the laboratory shall have a procedure for monitoring the validity of results. The resulting data shall be recorded in such a way that trends are detectable and, where practicable, statistical techniques shall be applied to review the results;
- the data quality must be analyzed, a programmed action must be implemented in order to correct the problem and avoid the recording of invalid results;
- this implies that the laboratory has established predefined means of monitoring, particularly as regards to measurement or test equipment, the use of certified reference materials (reference substances), the relevance of methods, etc.

The monitoring of equipment is, moreover, specified in separate sections:

- section 6.4.10: When intermediate checks are necessary to maintain confidence in the performance of the equipment, these checks shall be carried out according to a procedure;
- section 7.7.1 c): Functional check(s) of measuring and testing equipment;
- section 7.7.1 e): Intermediate checks on measuring equipment.

The checks necessary to maintain confidence in the calibration status of the primary reference, transfer, or work measurement standards as well as of reference materials, must be conducted in accordance with a clearly defined procedure and time frame.

These paragraphs highlight in particular the fact that any deviation in equipment that impacts on the quality of the results must be monitored. A small deviation will not necessarily be insignificant when it comes to the estimation of uncertainties.

When a piece of equipment has a material effect on the accuracy of a result, it is important to ensure that it is calibrated prior to use. This requirement implies good management and monitoring of equipment, and therefore adherence to sound metrological principles within the company.

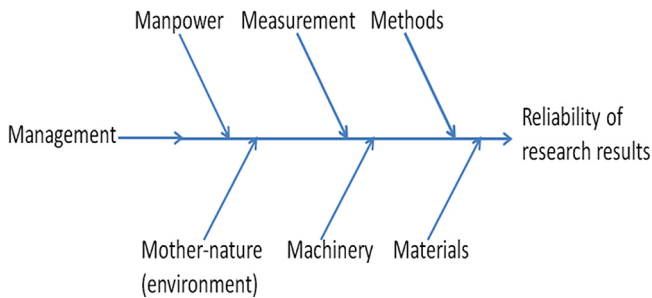


Fig. 1. Ishikawa "Fishbone" diagram (principle of 7 M's).

The following should be recorded as a minimum, if only to ensure measurement traceability:

- identification details of the object being tested or calibrated;
- identification details of the customer;
- date of testing or calibration;
- name of the operator or operators;
- detailed list of documents used;
- raw measurement data;
- reference documentation for measurement-standard indexing;
- calibration uncertainty of the measurement means;
- report detailing any adjustments made;
- maximum error tolerance, if appropriate.

The ILAC-G2 document [10] details the basic principles of measurement traceability.

The organization of the metrological function [11] plays a prominent role in the measurement process, whatever the magnitude being measured. When looking to achieve reliable results, it is evident that one cannot do without the sixth and the seventh M (Management and Measurement) in addition to the traditional 5 M's (Method, Manpower, Machinery, Mother-Nature (environment) & Materials) – see Figure 1. It is also very important, therefore, to adhere to the requirements concerning quality management in the ISO/IEC 17025 Standard.

Good organization and a high level of performance in a laboratory are dependent on a sound policy for implementing its more specific competencies, as well as upon recognition of these competencies.

The section 7.7 of the ISO/IEC 17025 Standard [1] on quality assurance of testing and calibration results stipulates that measurement results must be recorded and that, whenever possible, interpretation of results should include a statistical analysis in order to highlight deviations and trends, as specified by the Standard.

In order to monitor trends, the laboratory may, "but not be limited":

- take part in inter-laboratory comparison programs;
- regularly check the equipment used in order to ensure its stability and integrity;
- compare the results obtained by the method used with those of another method;
- assess the correlation of the results obtained for the object tested or calibrated with the results for an object with different characteristics.

Underlying all these requirements are the fundamental concepts of statistics and metrology: repeatability and reproducibility between and within laboratories, deviation, estimation of uncertainty, etc.

In order to ensure the reliability of measurement results obtained by a laboratory prior to reporting them, it is advisable to take note of section 7.7 of the Standard. This section can be found just before the section 7.8 detailing the requirements for the drawing up of a report for a customer, and just after the requirements regarding the following aspects:

- personnel;
- accommodation and environmental conditions;
- test and calibration methods and method validation;
- equipment;
- measurement traceability;
- sampling;
- handling of test and calibration items.

It goes without saying that all these factors vary from one laboratory to another, and that it is essential to have sound information regarding variations.

For these reasons:

- uncertainties of measurement should be identified and recorded in the report whenever required;
- calibration results must be linked to national and international measurement standards;
- it is advisable to use the results of inter-laboratory comparisons to confirm the precision and trueness of testing or calibration results.

Depending on the circumstances, data comparisons, whether carried out internally or among multiple organizations, provide an undeniable and readily recognized opportunity to any laboratory, organization or individual to put data collected to good use, to progress or profit from it and gain in terms of performance and competence.

A laboratory wants to demonstrate that a method of quantitative analysis is fit for its intended purpose following an initial assessment procedure of the performance of that method, comprising the following factors [12]:

- repeatability;
- intermediate precision;
- trueness and accuracy;
- calibration function (linearity);
- detection limit and quantification.

This means allowing internal method validation of methods [1] that are:

- non-standardized;
- designed/developed by the laboratory;
- standardized but used outside their intended field of application;
- amplified or modified with the respect to the standardized method;
- provided by suppliers or manufacturers, frequently involving the use of logic controllers or reagent kits,
- compared by an alternative method with respect to the reference method;
- able to confirm or assess the performance of a standardized method used in the laboratory.

Internal quality control and interpretation of inter-laboratory comparisons must also be in place to enable confirmation of the validation of the testing method [1].

3 Establishing data comparisons

3.1 Within-laboratory

This between-series within-laboratory analysis can be considered as the first level of data comparison, characterized as follows [13]:

3.1.1 Repeatability

The set can be formed by several series of data taken over a period of time that is much longer than the “short period” indicated for “repeatability” of data to apply.

In metrology, this is the typical case of a standard constructed with the aim of preserving a stable value with time.

In testing, this is the typical aim of a laboratory issuing test results over time under assumed repeatability conditions. The testing case is different, because the test material changes each time, but applies to checks made using a “reference material”: however, the latter additional information is external to the within-laboratory knowledge. The repeatability condition is assumed to be obtained by correctly performing the test according to an approved procedure, and this is basically why the result of a test can be obtained as a single value associated with an acceptance limit (tolerance interval), in contrast with the situation in metrology.

3.1.2 Reproducibility

A reproducibility study consists in preparing a “Design of Experiment” that can obtain sensitivity coefficients for the different influence quantities.

It consists in varying by known amounts each influence factor separately, and checking the overall effect.

These coefficients can also be computed without experimentation by differentiating the model expressed in closed form (analytically): this method may suffer from model imperfections.

The results do not directly inform about the actual variability of an experimental setup in each specific real condition.

The set of results of a run forms a single series of non-repeated data [9].

A variability level of the setup should be obtained by performing a specific experimental condition, called “reproducibility condition” (or “intermediate condition” when focused on only specific effects).

It is assumed that an evaluation of reproducibility is achieved, but the truth of this assumption is not particularly easy to check.

3.1.3 Drift with time

Replicated measurements, not falling under the category of “reproducibility” studies, do not always fall either, for their full period of time, under the “repeatability” condition,

irrespective to the length of the time span of the series of measurements, because “short time” condition does not indicate precisely a period of time. Replicated measurements are repeatable only until each specific experiment keeps the relevant necessary conditions unchanged during that time. These conditions may instead change even without any bug in the experimental procedure. In fact, a change can occur due to an instability in the characteristics of an instrument, or to a change of the measurand itself.

Drift can be either, an “instrumental drift” (*International Vocabulary of Metrology* – VIM – clause 4.21 [9]) or, a measurand drift (not defined in VIM or GUM [2]). The instrumental drift is said in VIM being “related neither to a change in a quantity being measured nor to a change of any recognized influence quantity”. However, being due to an instrumental bias whose value is changing with time, it should be considered as a possible influence quantity in the design of the experiment, because it induces a “continuous or incremental change over time in indication (of the instrument)”. If detected, it will induce a non-negligible effect on the result.

Effects like that of drift can better be detected and treated, rather than at this stage, at the end of the knowledge-gaining process. Drift in time of the measurand value [9] is a typical case when several series of data are obtained over extended periods of time.

3.2 Inter-laboratory

The inter-laboratory comparison allows a laboratory to position its results with respect to other laboratories and “occasionally” to check if its practices lead to consistent results. It must be remembered that a laboratory must be able to guarantee at all times that its results are incontestable.

The aim of an inter-laboratory comparison is also to check the competence of laboratories by comparing their results, qualified by the obtained values and their uncertainty.

The measurement results are compared with the reference values obtained by a reference laboratory or a consensus mean value.

The different levels of inter-laboratory comparisons are:

- comparisons organized by the BIPM (International Bureau of Weights and Measures) as part of the MRA (Mutual Recognition Arrangement) [14], generally limited to national laboratories;
- comparisons organized by the RMOs (Regional Metrology Organizations; EURAMET [15], for example), also as part of the MRA;
- proficiency tests performed according to ISO 13528 [3];
- comparisons organized by accredited laboratories. In Europe, these comparisons were organized by the EA (European Cooperation for Accreditation) until 2005. Since then, an agreement between the EA and EURAMET has entrusted them to EURAMET [15].

For the field of testing, COFRAC (the *COmité FRANçais d’Accréditation* – the French accreditation body) has issued a reference document, LAB CIL REF 02 [16] which can be downloaded from the <http://www.cofrac.fr> internet site.

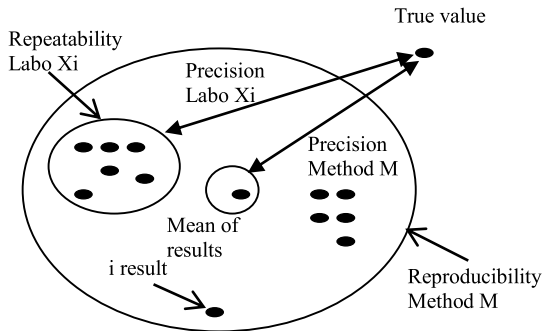


Fig. 2. Repeatability and reproducibility.

Pavese et al. [9] give a detailed overview of the different types of data comparisons (intra- and inter-laboratory comparisons).

3.3 Principle of the ISO 5725 standard

The accuracy of a result signifies the combination of both precision and trueness. The ISO 5725 series of Standards [4] serve as a reference in this area.

The estimation of accuracy is expressed by the relation concerning each test measurement result:

$$A = m + B + \varepsilon, \quad (1)$$

where m is the general mean (expectation), B is the laboratory component of the bias under repeatability conditions, and ε is the random error present in all measurements under repeatability conditions. B and ε are assumed to be normally distributed.

The general term m can be written as follows:

$$m = \mu + \delta, \quad (2)$$

where μ is the accepted reference (or certified) value and δ is the bias of measurement method.

Figure 2 illustrates the various components that result from an inter-laboratory comparison: the laboratory estimate of repeatability standard deviation relative to intra-laboratory variance, as well as of standard deviation reproducibility, including intra-laboratory variance and inter-laboratory variance.

To improve the accuracy of a result, a laboratory may need to ask questions about the possibility of further influence factors, such as the environment, the manpower, the measurement means, etc., all of which logically hinges on the estimation of the uncertainty related to the influence these factors exert.

3.4 Principle of the ISO 21748 and ISO 13528 Standards

The statistical model on which the ISO 21748 guidance is based is formulated as:

$$y = \mu + \delta + B + \sum c_i x'_i + \varepsilon, \quad (3)$$

where y is a test result, assumed to be calculated from the equation: $y = f(x_1, x_2, \dots, x_n)$ where the variables are independent; x_i is the deviation from the nominal value of x_i ; c_i is the sensitivity coefficient, equal to $\partial y / \partial x_i$. The other terms are the same as those used in the model (1) and the expression (2) defined in the ISO 5725 Standard.

The ISO 13528 Standard [3] deals with proficiency tests, which consist in testing the capability of a laboratory via measurements carried out on a reference sample.

Generally speaking, estimation of measurement uncertainties and inter-laboratory comparisons can only be carried out in a sufficiently rigorous manner if the personnel involved are sufficiently competent. To apply a method correctly, a certain amount of expertise in the field of metrology is required (i.e. in measurement and calibration), as well as training and awareness actions. These also ensure recognition of the laboratory's competence within its field, as well as ensuring that customers have competent persons dealing with their questions or requests. The accreditation of laboratories by means of the ISO/IEC 17025 Standard ensures the recognition of competence and, therefore, of performance.

There is no doubt that the estimation of uncertainties, carried out in conjunction with an inter-laboratory comparison [9], provides a wealth of information about the measuring means used, and ensures an expertly-managed measurement process; however, it is essential to go further and detect all the various influence factors at play to ensure the correctness of the testing and that the analysis or calibration conditions are optimal by highlighting any flaws that might impinge on the quality and reliability of measurements.

Furthermore, when a call for tenders is issued along with a Technical Requirements Specification (TRS), it is never certain that the requirements regarding the MPE (maximum permissible errors) of the various measurements requested are perfectly understood by the customer [17]. We have all known cases where a TRS has been drawn up with requirements that are simply too difficult, if not impossible, to fulfill due to the exorbitant cost. Perfect knowledge of the relevant methods, therefore, in particular via estimation of uncertainties by inter-laboratory comparisons, safeguards against the signing of an unworkable contract, thus helping avoid a potential conflict with the customer.

4 Assessing uncertainties

4.1 Basic principle

Various methods and tools to evaluate uncertainty exist [18–20]. All methods comply with the concepts and recommendations of the Guide to the Expression of Uncertainty in Measurement (GUM). [21] introduces several possibilities for laboratories, notably those based on within-laboratory and inter-laboratory approaches.

Figure 3 shows a general roadmap for evaluating uncertainties. Two approaches are generally used: the first one is based on a physical model and the second one is based

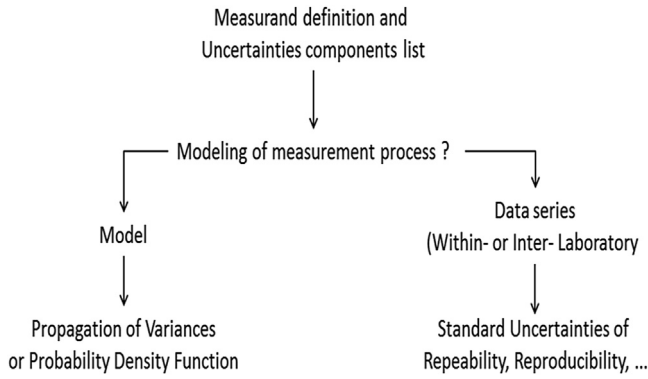


Fig. 3. Roamap for evaluating uncertainties.

on a statistical model (in terms of the ISO 5725 Standard). The two approaches may be used in conjunction with each other.

Many factors can affect the quality of a measurement (see Fig. 1). Among these factors, the following may be cited: the operator, the instrumentation used, the method, the environment, etc. The accuracy of a measurement is thus characterized by a number of factors.

Measurement results are dependent, among other things, on the measurement apparatus used, the characteristics of which may vary during its lifetime. For this reason, it is essential that it be regularly calibrated and/or verified (NB: the concepts of calibration and of verification are quite distinct – cf. [22]). These steps allow the apparatus to be linked to national measurement standards and the user to detect any corrections that may need to be made to the instrumentation (linearity, offset, hysteresis, etc.), to make any necessary adjustments, and to take into account the uncertainty of the corrections.

Moreover, where it is necessary to validate (or otherwise) a test measurement result, and subsequently pronounce on the acceptance (or not) of, for example, a manufactured part, it is vital that the apparatus being used be approved and appropriate for the measurements being carried out. If the measurement result is to be positioned in a tolerance range and is, moreover, close to either the upper or lower limit, there must be as little doubt as possible concerning said measurement result, which must be qualified by means of a measurement uncertainty that takes into account all the factors that have an influence on the result.

Due to the various errors affecting a measurement [23] the result is expressed as a range, which represents the measurement uncertainty within which it is assumed the true value should be comprised within a given confidence level; a measurement result is therefore written as:

$$Y = \hat{y} \pm U, \quad (4)$$

where \hat{y} is the estimated expected value (mean) of Y and U is the expanded uncertainty of measurement.

Determining \hat{y} and U may be achieved by means of an uncertainty propagation method in the common case where more than one influence quantity is involved, and more

specifically by the algebra of variables. Reference works [2,13,18–20] detail a comprehensive approach to the estimation of measurement uncertainty for testing and calibration.

The ISO/IEC Guide 98-3 document [2] differentiates between two types of methods for determining uncertainty components. The type A method makes use of statistical analysis. The series of observations pertains to a probability distribution whose standard deviation can be computed, known as the standard uncertainty. With the type B method, the uncertainty components are instead determined using other techniques that do not rely on statistics. Together they make use of all the information available.

A comparable uncertainty evaluation is achieved using a Bayesian approach [18], the principle of which is to use all prior information, typically including expert opinions, experiment feedback, etc. (in statistics, it is a matter of assessing likelihood).

The expanded uncertainty is written as $U = k u_c(y)$, where u_c is the combined standard uncertainty, which is expressed by the following relation:

$$u_c(y) = \sqrt{\sum_{i=1}^{i=n_A} (u_{Ai})^2 + \sum_{i=1}^{i=n_B} (u_{Bi})^2} \quad (5)$$

where u_{Ai} and u_{Bi} are the standard uncertainty estimated by type A and type B methods, respectively [2], and n_A and n_B are the component numbers that represent the influence factors taken into account for the estimation of the combined uncertainty u_c .

In general, expression (5) is used under the hypothesis of the central limit theorem and all variables are assumed to be uncorrelated. The result is generally given with a coverage factor k associated with a confidence level that can be attributed. When dealing with a normal distribution, $k=2$ corresponds to a 95.45% confidence level.

Consequently, in order to establish the uncertainty of a measurement result composed of the results of several input magnitudes, each with its own corresponding measurement uncertainties, the various measurement uncertainty quotients must be added geometrically.

To give an example for the uncertainty computing methods, if we were to look at an influence quantity based on environmental conditions (temperature, pressure, humidity, etc.), there are already-existing physical models that relate the resulting measured quantity to environmental parameters such as the law of dilatation of the materials measured. In other instances, the interactions between different factors should be determined using experimental design models.

To estimate the uncertainty of a quantity value with only a tolerance range and no distribution law, a rectangular distribution is often used, as in Figure 4. That is to say that, when re-measuring the quantity, it is considered that valid values fall within the tolerance range and that the probability of a measurement value outside the range is zero or is discarded.

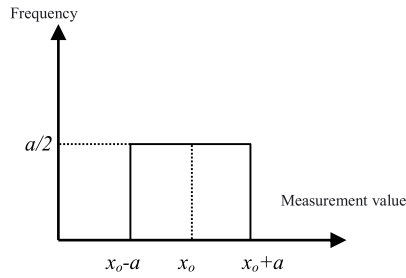


Fig. 4. Rectangular distribution.

Within the range $(x_o - a; x_o + a)$, the probability of occurrence is considered constant. To calculate the standard uncertainty of measurement, the tolerance range must be divided by $\sqrt{3}$ [2].

Otherwise, other probability density functions exist (triangular, arc sinus, etc.) that are used in different measurement fields.

4.2 Evaluating uncertainties: main methods

4.2.1 Using a physical model

The method commonly used is based on the law of propagation of uncertainty. Various concrete examples in metrology can be found in the ISO/IEC Guide 98-3 [2] – Guide to the Expression of Uncertainty in Measurement (GUM). The principle of this guide has already been presented in section 4.1. In a non-linear form of the physical model, other methods of uncertainty evaluation are more appropriate [18].

A supplement to the GUM, GUM S1 [23], specifies uncertainty evaluation in terms of a probability density function. It describes the numerical propagation of the probability density function through the physical model. The Monte Carlo methodology as presented by the GUM S1 involves the propagation of the distributions of the input sources of uncertainty by using the model to provide the distribution of the output. When all the input PDFs are defined, a number of Monte Carlo trials is selected. Generally, the higher the number of simulation trials, the better the results. The Monte Carlo Simulation is easy to apply and gives the required information about the model distribution. However, it also has some limitations: the model simulation runtime may be long in some complex cases. Furthermore, the choice of the proper probability distribution functions for the model parameters may be difficult. The accuracy of the numerical simulation also depends on the quality of the random number generator.

Bayesian statistics based on the Bayes theorem can also be used for the uncertainty evaluation [18]. Bayesian inference is made through the probability density function that describes the information acquired from measurement and the knowledge about the quantity before the measurement is performed. Bayes formula is a tool that combines the a priori information on the parameters and the information provided by the data. Regarding the Bayesian method, it is often difficult to assign prior-information distribution to the quantity being measured. To construct the prior density function, one should make

use of all available prior information, such as calibration data, data from other similar experiments, and other available evidence pertaining to the quantity. Since that information involves personal judgment, this can result in significant uncertainties.

4.2.2 Using a statistical model

The ISO/TS 21748 Standard [13] presents an approach for evaluating uncertainty of measurement based on an estimation of repeatability, reproducibility and trueness. It is related to the ISO 5725 Standard [4].

Given the relationship described in equations (1)–(3), the uncertainty $u(y)$ associated with an observation can be estimated using the following expression:

$$u_c^2(y) = u^2(\delta) + s_B^2 + \sum c_i^2 u^2(x_i) + s_\varepsilon^2, \quad (6)$$

where s_B^2 is the variance of B , s_ε^2 is the variance of ε , $u(\delta)$ is the uncertainty associated with δ , and $u(x_i)$ is the uncertainty associated with x_i (see Eq. (3)), the other terms are the same as those used in the model (3). B and ε are assumed to be normally distributed and uncorrelated.

In order to study the effects of variables, it is possible to incorporate other terms in equation (6), or higher order terms or correlation terms exactly as described by the GUM [2,13]

The reproducibility standard deviation s_{Repro} is given by:

$$s_{\text{Repro}}^2 = s_B^2 + s_\varepsilon^2.$$

Given expression (2), the uncertainty associated with δ is estimated as:

$$u^2(\delta) = s_\delta^2 + u^2(\mu),$$

where $u(\mu)$ is the uncertainty associated with the reference (or certified) value μ . The standard deviation s_δ is calculated as:

$$s_\delta^2 = \frac{s_{\text{Repro}}^2 - (1 - \frac{1}{n})s_\varepsilon^2}{p},$$

where p is the number of laboratories; n is the number of replicates in each laboratory.

The ISO 11352 [24] Standard describes the procedures for the assessment of measurement uncertainty based on random errors and systematic errors. The evaluation of the measurement uncertainty is based on the testing results and validation data, which represent the reproducibility in the laboratory and the bias of the laboratory method.

An approximate estimate of the combined uncertainty using the reproducibility standard deviation (s_{Repro}) from an inter-laboratory trial (with p laboratories), can be expressed as follows [24]:

$$u_c^2(y) = \sqrt{\frac{s_{\text{Repro}}^2}{p}}.$$

5 Key performance indicators for monitoring

5.1 Internal data comparisons

5.1.1 Control charts

Control charts can be constructed and be useful for all activities of a laboratory. All data is checked frequently taking into account the following limits:

Upper (LU) and Lower Limits (LL):

$$L_L^U = \mu_t \pm 3 \times \sigma_{\text{Lab}},$$

where μ_t is the target value and σ_{Lab} is the standard deviation (can be combined at least with the repeatability and/or reproducibility standard deviations) of the laboratory,

Upper (CLU) and Lower Control (CLL) Limits:

$$CL_L^U = \mu_t \pm 2 \times \sigma_{\text{Lab}}$$

The control chart will be effective if the type of graph selected is appropriate to the requirements of the user. Several configurations can be observed as follows:

The chart displays all the results in succession. But the trend rules could lead to anomalous actions because of confusion between repeatability and change of conditions. This type of chart may be justified if the repeatability is not negligible compared with change of conditions.

The chart displays daily averages. But with application of trend rules to averages without paying attention to the dispersion of data corresponding to those averages. This type of chart may be justified if the repeatability is not negligible compared with change of conditions and if its purpose is only to display trend.

5.1.2 Internal Normalized Mean Square Error

The common key indicator used for internal data comparisons is the Internal Normalized Mean Square Error (*INMSE*) that is expressed as:

$$INMSE = \frac{|\mu_x - \mu_{\text{IR}}|}{\sqrt{u_x^2 + u_{\text{IR}}^2}},$$

where μ_x and u_x are respectively the mean and standard uncertainty estimated values of results obtained with one instrument and/or with a method used and/or by one person (or group of persons) and/or with an intermediate material (produced in the laboratory).

μ_{IR} and u_{IR} are respectively the mean and standard estimated uncertainty of results measured by a qualified person and/or obtained with an internal reference material (produced in the laboratory).

If $INMSE < 1$, the results are considered satisfactory.

If $INMSE$ is lower than or equal to 2, then the observed bias is judged to be statistically negligible compared with

$$\sqrt{u_x^2 + u_{\text{IR}}^2}.$$

The user may consider μ_x the target value but must take into account u_{IR} in the uncertainty estimation.

Otherwise, the user must include the bias in the calculation of the uncertainty if μ_x is the target value of the control chart.

For the evaluation of u_x and u_{IR} standard uncertainties, it is possible to use only the estimate of repeatability of each component. The estimate of repeatability with several reference materials (p samples) depends on the stability and repeatability in the field studied. If the standard deviations ($s_{\text{sample},i}$ is the standard deviation for each sample i) are judged to be significantly constant, u_{repeat} the repeatability standard uncertainty is estimated as:

$$u_{\text{repeat}} = \sqrt{\frac{\sum_{i=1}^p s_{\text{sample},i}^2}{p}}$$

5.2 External data comparisons

5.2.1 External Normalized Mean Square Error

The common key indicator used for external data comparisons is the External Normalized Mean Square Error (*ENMSE*) that is expressed as:

$$ENMSE = \frac{|\mu_x - \mu_{\text{ER}}|}{\sqrt{u_x^2 + u_{\text{ER}}^2}},$$

where μ_x and u_x are respectively the mean and standard uncertainty estimated values of results, obtained with one instrument and/or with a method used and/or by one person (or group of persons) and/or an internal material (produced in the laboratory).

μ_{ER} and u_{ER} are respectively the mean and standard uncertainty values of the reference measurement standard or results obtained with an inter-laboratory comparison or with a method known in literature (a reference method) or by another laboratory (accredited or with a recognized competency level) or a reference material (certified).

If $ENMSE < 1$, the results are considered satisfactory.

In the case of inter-laboratory comparison, if s_{repro} is the reproducibility standard deviation, n being the number of laboratories, the standard uncertainty u_{ER} is calculated as follows:

$$u_{\text{ER}} = \frac{s_{\text{repro}}}{\sqrt{n}}.$$

In the case of a reference method, if s_{method} is the standard deviation of m results obtained with a reference method, the standard uncertainty u_{ER} is expressed as:

$$u_{\text{ER}} = \frac{s_{\text{method}}}{\sqrt{m}}.$$

5.2.2 Z-score

For each inter-laboratory comparison, the report gives the assigned value as the expected or estimated reference value μ_{ER} with its standard uncertainty. If the reproducibility standard deviation of the laboratories is

s_{repro} , then, the Z-score of a laboratory is expressed in the following way:

$$Z_{\text{-score}} = \frac{\mu_x - \mu_{\text{ER}}}{s_{\text{repro}}}$$

If $|Z_{\text{-score}}| < 2$ then the laboratory has come up with “good” results.

If $2 < |Z_{\text{-score}}| < 3$ then the laboratory should be asking questions.

If $|Z_{\text{-score}}| > 3$ then the laboratory must take action.

Based on the experiment design implemented when the accuracy study was carried out, the trueness of a method can be verified when the bias $\text{Bias} = \mu_x - \mu_{\text{ER}}$ is lower than an acceptability limit which has been set in accordance with either a regulatory or standard requirement, or fixed by the customer, or by the laboratory itself [25].

5.3 Feedback and recommendations

However, organizing comparisons can be problematic in various technical and scientific fields for a number of reasons:

- the tests are too long or costly;
- the reference has not been established or assigned;
- there is no organization either willing or able to conduct the comparison;
- the data is confidential.

For research laboratories, it is nevertheless essential to make use of this type of tool for the validation of techniques or methods that have been developed.

Organizations have become involved in data comparison either voluntarily, for regulatory reasons, or because of requirements stipulated in an official Standard.

A number of different situations have been observed:

- a laboratory uses its data solely to monitor the quality of its results, either to assess staff and compare highly qualified individuals (who have recognized experience and expertise) with other members of staff in the laboratory, or to compare methods, equipment, etc.;
- a laboratory uses reference material or reference gauges to compare results obtained, in which case it is necessary to ensure conformity of the reference used or the quality of traceability to the reference standard;
- a laboratory compares itself with another laboratory, in which case it is necessary to ensure that the other laboratory has at least the same levels of requirements (in terms of acquisition of competences), such as those mentioned in ISO/CEI 17025 Standard [1]. If this is the case in the field of calibration, then the uncertainty level of the other laboratory needs to be known.

Experience shows that unfortunate situations occur, such as a laboratory comparing itself with another that has obtained a very different quantity value.

There are more and more laboratories (with or without accreditation) that organize inter-laboratory comparisons, which is in fact a good thing. But, it is important to be sure that all the organizers (in particular in very sensitive fields, for example those linked to health) implement things correctly and that they follow and know how to use

correctly the tools recommended in the ISO 13528 [3] or the ISO 5725 [4] Standards.

The ISO 17043 [7] Standard is a very important document that allows the organizers to maintain or achieve recognition for their level of competence and to state their independence, impartiality and integrity.

The organizers of comparisons in certain cases do not demonstrate at all their independence, and organizations that conduct comparisons are sometimes open both to external laboratories and to laboratories that are part of the organizer arrangement itself.

The risk of not allowing the participation of an external laboratory (without a valid reason) therefore increases.

Among other less than perfect situations that occur are the following:

- the preparation of the reference materials or samples is not carried out by an organization with sufficient recognized competency;
- participants constitute an overly heterogeneous group (for example, a portion has 17025 accreditation but not the rest);
- participants compare results before these results have been provided to the organizer;
- participants do not take the time to carry out tests in a proper manner and hand in a result that is approximately the same as a neighbor’s or colleague’s.

When inter-laboratory comparisons are conducted, and when laboratories recognize the undeniable advantages of such comparisons, these laboratories must also ensure that the necessary time and money are dedicated to this activity. It seems that considerable effort is still needed to increase awareness among those laboratories that do not yet participate in these comparisons.

For accredited laboratories, inter-laboratory comparison programs do involve substantial cost, especially if the laboratory takes part in inter-laboratory comparison campaigns for all parameters (or matrices) or for all the methods for which it has accreditation (a debatable practice for which there is no actual consensus).

In many countries, laboratories have not yet become involved in inter-laboratory comparisons, firstly because of a lack of understanding of the advantages to be had from inter-laboratory comparison, and secondly because of the associated cost. Although not everyone is helped by the quality and metrology structures of their country.

One more point worth mentioning is that the use to which comparison results are put is on occasions seriously disappointing. There are times when the results are not used at all. There are other times when a laboratory sits on its results (even if, for example, the Z-score is outside acceptable limits), failing to ask questions, and filing the report away in a drawer.

Finally, though staff may have training in statistical tools, the effectiveness of that training can fall short of the real needs of laboratories.

6 Conclusion

This paper gives a methodology based on data comparisons and uncertainty evaluation principles. It allows

laboratories to gain in competence and to improve the reliability of results.

The methodology is built from the main requirements and methods used to ensure the quality of measurement results in accordance with ISO standards.

The authors have focused on the accreditation of testing and calibration laboratories, which is a crucial proof of the laboratories' competence and consequently of their performance. When thoroughly understood and appropriately applied, the requirements of the ISO/IEC 17025 Standard ensure that accredited laboratories ask themselves appropriate questions and take action accordingly, which ultimately results in the satisfaction of their customers.

When seeking to limit doubts concerning measurement results to a minimum, the estimation of measurement uncertainty and participation in inter-laboratory comparison is a must, since it offers, among other things, the most pertinent answers.

References

1. ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories* (ISO, Geneva (Switzerland), 2017)
2. ISO/IEC Guide 98-3, *Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement* (GUM: 1995) (ISO, Geneva (Switzerland), 2008)
3. ISO 13528, *Statistical methods for use in proficiency testing by interlaboratory comparison* (ISO, Geneva (Switzerland), 2015)
4. ISO 5725-Part 1 to 6, *Accuracy (trueness and precision) of measurement methods and results* (ISO, Geneva (Switzerland), 1994)
5. M. Thompson, S. Ellison, R. Wood, *The International Harmonized Protocol for the proficiency testing of analytical chemistries laboratories*, Pure Appl. Chem. **78**, (2006), DOI: [10.1351/pac200678010145](https://doi.org/10.1351/pac200678010145)
6. R. Lawn, M. Thompson, R. Walker, *Proficiency testing in analytical chemistry* (The Royal Society of Chemistry, Cambridge, 1997)
7. ISO 17043, *Conformity assessment – General requirements for proficiency testing* (ISO, Geneva (Switzerland), 2010)
8. ILAC-G13, *Guidelines for the requirements for the competence of providers of proficiency testing schemes* (ILAC, Silverwater (Australia), 2000), Available from <http://www.ilac.org/>
9. F. Pavese, A. Charki, Data inter-comparisons in the context of the knowledge-gaining process: an overview, Acta IMEKO **2**, 73–83 (2018)
10. ILAC-G2, *Traceability of measurement* (ILAC, Silverwater (Australia), 1994)
11. ISO 10012, *Measurement management systems – Requirements for measurement processes and measuring equipment* (ISO, Geneva (Switzerland), 2003)
12. S. Belouafa, F. Habti, S. Benhar, B. Belafkih, S. Tayane, S. Hamdouch, A. Bennamara, A. Abourriche, Statistical tools and approaches to validate analytical methods: methodology and practical examples, Int. J. Metrol. Qual. Eng. **8**, (2017), DOI: <https://doi.org/10.1051/ijmqe/2016030>
13. ISO 21748, *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation* (ISO, Geneva (Switzerland), 2010)
14. CIPMMRA-D-05, *Measurement comparisons in the context of the CIPM MRA, BIPM Document* (CIPM, Sèvres (France), 2016), Available from <https://www.bipm.org/utis/common/documents/CIPM-MRA/CIPM-MRA-D-05.pdf>
15. EURAMET Guide n° 4, *EURAMET Guide on comparisons, version 1.1* (EURAMET, 2016)
16. LAB CIL REF 02, *Exigences pour l'accréditation des organisateurs de comparaisons interlaboratoires selon la norme NF EN ISO/CEI 17043*, Cofrac document, Available from <https://www.cofrac.fr/documentation/LAB-CIL-REF-02>
17. ISO 14253-1, *Geometrical product specifications (GPS) – Inspection by measurement of workpieces and measuring equipment – Part 1: Decision rules for verifying conformity or nonconformity with specifications* (ISO, Geneva (Switzerland), 2017)
18. A.B. Forbes, Approaches to evaluating measurement uncertainty, Int. J. Metrol. Qual. Eng. **3**, 71–77 (2012)
19. EA-4/02, *Expression of the uncertainty of measurement in calibration*, December 1999
20. EA-4/16, *EA Guidelines on the expression of uncertainty in quantitative testing*, December 2003
21. M. Desenfant, M. Priel, Road map for measurement uncertainty evaluation, Measurement **39**(9), 841–848 (2006)
22. NF X 07-001, *Basic Standards – International vocabulary of basic and general terms in metrology (normes fondamentales – vocabulaire international des termes fondamentaux et généraux de métrologie)*, December 1994
23. ISO/IEC Guide 98-3/Suppl 1, *Propagation of distributions using a Monte Carlo method* (ISO, Geneva (Switzerland), 2008)
24. ISO/DIS 1135 2, *Water quality – Estimation of measurement uncertainty based on validation and quality control data* (ISO, Geneva (Switzerland), 2012)
25. A.G. Chunovkina, A.V. Elster, I. Lira, W. Wöger, Analysis of key comparison data and laboratory biases, Metrologia **45**, 211–216 (2008)

Cite this article as: Abdérafi Charki, Franco Pavese, Data comparisons and uncertainty: a roadmap for gaining in competence and improving the reliability of results, Int. J. Metrol. Qual. Eng. **10**, 1 (2019)