COMMITTEE DRAFT OIML CD1

Date: 27 April 2009

Reference number: OIML/TC 3/SC 5/N1

Supersedes document: Not applicable

OIML TC 3 / SC 5 Project p2 Title: The role of measurement uncertainty in conformity assessment decisions in legal metrology Secretariat: Dr. Charles Ehrlich - USA Ms. Régine Gaucher - BIML	Circulated to P- and O-members and liaison international bodies and external organisations for: [ ] discussion at (date and place of meeting): .... [X] comments by: 15 September 2009 [ ] vote (P-members only) and comments by

TITLE OF THE CD (English):
New OIML Document

The role of measurement uncertainty in conformity assessment decisions in legal metrology

TITLE OF THE CD (French):

Nouveau Document OIML

Le rôle de l’incertitude de mesure dans l’évaluation de la conformité en métrologie légale

Original version in: English

THE ROLE OF MEASUREMENT UNCERTAINTY IN CONFORMITY ASSESSMENT DECISIONS IN LEGAL METROLOGY

0 Table of Contents

1 Scope and Objectives

The scope of this International Document is to provide guidance to OIML Secretariats and to members of OIML Technical Committees and Subcommittees on options to consider for incorporating the concept of "measurement uncertainty" into OIML Recommendations and other OIML publications pertaining to conformity testing of measuring instruments and systems in legal metrology.

The main objective is to provide guidance on options to be considered for incorporating text into OIML publications about how to take measurement uncertainty into account when using measured values, obtained during the testing of a measuring instrument or system, as the basis for making pass-fail decisions.

This includes providing information on how to assess the possible "risks" of erroneous conformity decisions (i.e. probability of erroneous acceptance and probability of erroneous rejection) that arise from the measurement uncertainty associated with the measured values obtained during testing of a measuring instrument or system.

This also includes describing the difference between "error" and "uncertainty" in a way that demonstrates how both concepts (and terms) can coexist in legal metrology, and providing guidelines and examples for the determination and expression of measurement uncertainty in legal metrology applications, consistent with the *Guide to the Expression of Uncertainty in Measurement* (hereinafter denoted by GUM) [1] and its Supplements [6, 7].

The guidance provided in this document is intended to be applicable for both the type evaluation and verification of measuring instruments used in legal metrology.

The guidance provided in this document is also intended to be totally compatible with the straightforward application of ISO/IEC 17025 [11] with respect to requirements concerning the use of measurement uncertainty.

Harmonized methods for evaluating measurement uncertainties and implementing them into decision criteria used for the metrological evaluation of measuring instruments and systems are required in order that test evaluations and metrological judgments may yield comparable results from one national responsible body in legal metrology to another. Such comparability is a necessary element for achieving confidence between bodies in recognizing each other's type approvals, leading to the intended operation and function of the OIML Certificate System [2] and Mutual Acceptance Arrangement (MAA) [3]. Such comparability is generally also necessary for providing confidence in verification processes and certificates.

2 Terminology

2.1
quantity value (VIM3 1.1)
property of a phenomenon, body or substance, where the property has a magnitude that can be expressed as a number and a reference

2.2
true quantity value (VIM3 2.11)
quantity value consistent with the definition of a quantity

NOTE 1 In the Error Approach to describing measurement**,** a true quantity value is considered unique and, in practice, unknowable. The Uncertainty Approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. However, this set of values is, in principle and in practice, unknowable. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity.

NOTE 2 In the special case of a fundamental constant, the quantity is considered to have a single true quantity value.

NOTE 3 When the definitional uncertainty associated with the measurand is considered to be negligible compared to the other components of the measurement uncertainty**,** the measurand may be considered to have an "essentially unique" true quantity value. This is the approach taken by the GUM and associated documents, where the word "true" is considered to be redundant.

NOTE 4 (not in VIM3) Historically, in legal metrology the term "true value" is sometimes used to mean the value associated with a measurement standard that is used in the process of testing a measuring instrument. This is <u>not</u> the meaning of the term in this document.

2.3
error (of measurement) (VIM3 2.16)
measured quantity value minus a reference quantity value

NOTE 1 The concept of 'measurement error' can be used both

a) when there is a single reference quantity value to refer to, which occurs if a calibration is made by means of a measurement standard with a measured quantity value having a negligible measurement uncertainty or if a conventional quantity value is given, in which case the measurement error is known, and
b) if a measurand is supposed to be represented by a unique true quantity value or a set

of true quantity values of negligible range, in which case the measurement error is not known.

NOTE 2 Measurement error should not be confused with production error or mistake.

NOTE 3 (not in VIM3) There has been considerable debate in Working Group 1 of the Joint Committee on Guides for Metrology (JCGM WG1) about whether ‘error’ should be defined as a ‘value,’ as in the above definition, or as a ‘quantity’ that has a value. Both uses of the term ‘error’ can be found in the metrology literature. In this document the definition given above will be used. Note that in reference [7] this is not the case.

2.4
measurement uncertainty (VIM3 2.26)
non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used

NOTE In GUM Supplement JCGM 104 [Ref], measurement uncertainty is described as a measure of how well the essentially unique true value of a measurand is believed to be known.

2.5
measurement result (VIM3 2.9)
set of quantity values being attributed to a measurand together with any other available relevant information

2.6
measured quantity value (VIM3 2.10)
quantity value representing a measurement result

2.7
maximum permissible error MPE (VIM3 4.26)
extreme value of measurement error, with respect to a known reference quantity value, permitted by specifications or regulations for a given measurement, measuring instrument, or measuring system

2.8
measurement model (VIM3 2.48)
mathematical relation among all quantities known to be involved in a measurement

2.9
input quantity in a measurement model (VIM3 2.50)
quantity that must be measured, or a quantity, the value of which can be otherwise obtained, in order to calculate a measured quantity value of a measurand

2.10
measurement unit (VIM3 1.9)
real scalar quantity, defined and adopted by convention, with which any other quantity of the same kind can be compared to express the ratio of the two quantities as a number

2.11
metrological traceability (VIM3 2.41)
property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty

2.12
indication (VIM3 4.1)
quantity value provided by a measuring instrument or a measuring system

NOTE 1 An indication may be presented in visual or acoustic form or may be transferred to another device. An indication is often given by the position of a pointer on the display for analog outputs, a displayed or printed number for digital outputs, a code pattern for code outputs, or an assigned quantity value for material measures.

NOTE 2 An indication and a corresponding value of the quantity being measured are not necessarily values of quantities of the same kind.

NOTE 3 (not in VIM3) There has been considerable debate in Working Group 1 of the Joint Committee on Guides for Metrology (JCGM WG1) about whether 'indication' should be defined as a 'value,' as in the above definition, or as a 'quantity' that has a value. Both uses of the term 'indication' can be found in the metrology literature. In this document the definition given above will be used. Note that in reference [7] this is not the case.

2.13
rated operating condition(VIM3 4.9)
operating condition that must be fulfilled during measurement in order that a measuring instrument or measuring system perform as designed

2.14
reference operating condition (VIM3 4.11)
operating condition prescribed for evaluating the performance of a measuring instrument or measuring system or for comparison of measurement results

2.15
maximum permissible measurement error (VIM3 5.21)
extreme value of measurement error, with respect to a known reference quantity value, permitted by specifications or regulations for a given measurement, measuring

instrument, or measuring system

NOTE 1 Usually, the term “maximum permissible errors” or “limits of error” is used where there are two extreme values.

NOTE 2 The term “tolerance” should not be used to designate ‘maximum permissible error’.

NOTE 3 (not in VIM3) There has been considerable debate in Working Group 1 of the Joint Committee on Guides for Metrology (JCGM WG1) about whether ‘maximum permissible error’ should be defined as a ‘value,’ as in the above definition, or as a ‘quantity’ that has a value. In this document the definition given above will be used. Note that in reference [7] this is not the case.

Suggestions for additional entries?

3 Introduction

The formal, probabilistic concept of "measurement uncertainty" [1] has revolutionized modern metrology. There is a growing literature on methods for calculating and using measurement uncertainty for a variety of types of measurement applications, including decision making in legal metrology testing. Some of these methods are more complex and time consuming than others. However, many legal metrology activities are intended for making relatively 'quick and easy' pass-or-fail decisions, and so choice of method of assessing and using measurement uncertainty can be important in bringing efficiency to the activity. Since the use of formal measurement uncertainty is widely recognized as being essential in both the metrology and laboratory accreditation communities, it has become necessary to consider the different ways that it can be routinely incorporated into legal metrology decision-making processes with a minimum of complexity and disruption for any particular application. This document is intended to provide guidance concerning options to consider for effectively, yet efficiently, incorporating formal measurement uncertainty into conformity decisions on a probabilistic basis when testing measuring instruments and systems in legal metrology.

While the formal concept of measurement uncertainty, in the sense of [1], is relatively new, a less formal but equally important notion of 'uncertainty' in measurement has always existed in legal metrology. One example is the practice of establishing 'expanded' or 'conservative' maximum permissible errors (MPEs) in order to draw 'safe' conclusions concerning whether measured errors of indication are within acceptable limits. The practice of specifying a fraction, such as 1/3 or 1/5, for the maximum allowed ratio of the error ('uncertainty'?) of the standard (reference) instrument to the MPE is another example. However, the probabilistic nature of measurement uncertainty is not explicitly considered in this more classical approach. Clause 4 of this document elaborates on this classical, basic approach to conformity testing decisions in legal metrology, since it serves as the foundation for considering conformity testing that then takes formal measurement uncertainty into account.

With the introduction of the formal concept of measurement uncertainty, making conformity decisions in legal metrology becomes more complex, not only because there is more to consider about making the decisions themselves, but also because the language used to make such decisions can sometimes be confusing, and even appear to be contradictory. Most notably, whereas the concepts of "error" and "uncertainty" share a certain similarity, in that they are both related to the quality of a measurement, they are actually significantly different concepts. Perhaps seemingly ironic, an 'error of indication' is something that can itself be measured, and thus have a value with an associated measurement uncertainty. This difference between "error" and "uncertainty," and how they coexist in legal metrology (and other areas of metrology), is elaborated in Annex A.

When formal measurement uncertainty is taken into account in conformity assessment decisions in legal metrology, the method discussed earlier, of comparing a measured 'error of indication' to a specified MPE, is still used. However, in addition, because of the probabilistic interpretation of measurement that accompanies the concept of formal measurement uncertainty, it becomes necessary to think in terms of the degree of belief (expressed as a probability) that the essentially-unique true value (denoted hereafter as ‘true’ value) of an error of indication actually lies outside of the specified MPE limits, even if the ‘measured’ value lies within the MPE limits, and vice versa (see Annex A). Various “decision rules” can be established for deciding whether or not a particular test is considered to “pass,” based on the expressed probability, and associated “risks” for making incorrect decisions can be calculated. Clause 5 elaborates on these and related topics, and provides options to be considered when developing OIML Recommendations and other OIML documents.

Establishing appropriate MPEs for a given testing scenario is also influenced by measurement uncertainty. The cost to the consumer, vendor or manufacturer associated with the use of MPEs that are unnecessarily large or small can be reduced through taking likely measurement uncertainties into account when first establishing the MPEs. Setting MPEs that are very small can be costly to the instrument manufacturer (who will likely pass the additional cost on to the consumer!), who must build a more costly instrument to meet the tighter requirements. By considering likely levels of measurement uncertainty for different uses of measuring instruments, more optimal MPEs can be set such as to yield acceptable levels of risk. Clause 6 elaborates on options for taking measurement uncertainty into account when prescribing MPEs in OIML Recommendations and other OIML documents.

For convenience, options that should be considered for inclusion in an OIML Recommendation or other OIML document are explicitly provided in Clause 7, including specific language pertaining to the incorporation of formal measurement uncertainty.

4 Basic considerations pertaining to conformity testing decisions and measurement uncertainty

One of the key roles of legal metrology is to evaluate the performance and suitability of designs (or types) of measuring instruments and systems (type evaluation), as well as the performance of individual measuring instruments and systems (initial or subsequent verification), for various regulated applications. The basic kind of test that is used to conduct such evaluations involves comparing a measured ‘error of indication’ with a ‘maximum permissible error’ (MPE) that is specified for the particular application. The value (E_I) of the error of indication is typically defined as the difference between the indicated value of the measuring instrument or system obtained when measuring the measurand, and the ‘true’ value of that measurand. Since it is not possible to perform a ‘perfect’ measurement, and so the ‘true’ value of the measurand cannot be known, the error of indication is usually considered operationally to be the difference between the indicated value (Y_I) of the measuring instrument or system obtained when measuring the measurand, and the value (Y_S) of the same measurand as determined when using a measurement standard. Expressed mathematically:

$$E_I = Y_I - Y_S . \qquad (4.1)$$

(Note that, historically, in legal metrology the term "true value" is usually not used in the sense given here, but rather is used to mean the value associated with a measurement standard that is used in the process of testing a measuring instrument. This latter meaning is not the meaning of the term 'true' value in this document; see Annex A for more detail.)

In general, Y_S can be determined through use of a 'measurement model' [1, 4] that relates the value of the measurand to values (x_i) of 'input quantities in a measurement model' [4] (that is, Y_S depends on, or is a function (f) of, the values x_i):

$$Y_S = f(x_1, x_2, \ldots x_n) . \qquad (4.2)$$

Depending on the category of test being performed (type evaluation, initial verification, or subsequent verification), there can be a wide variation in the details of how to conduct the test. The variation may include the number of individual errors of indication that should be obtained (through repeated measurements), and when and how the operating conditions of the instrument should be controlled (if at all). Common to all of the categories of tests, however, is that conformity decisions are ultimately made based on the results of one or more tests that compare measured errors of indication with MPEs.

The concept of comparing a measured error of indication with a set of MPEs (upper and lower), for purposes of making a conformity decision, is shown schematically in Figure 1. The horizontal axis represents possible values of error of indication E_I. The upper and lower MPEs, denoted MPE_+ and MPE_- , respectively, are shown to be symmetric about 0, but this is not always necessary (e.g., when testing radar guns). If only a single measured error of indication is to be used to make a conformity decision, then if that single measured error of indication lies within the interval defined by the MPEs (denoted as "Conformance Zone" in the figure), the instrument is considered to pass that particular test (as shown in the figure). Otherwise, the instrument is considered to fail that test. Note that formal measurement uncertainty is not being explicitly considered in this discussion or in this figure, however the MPEs are assumed to have been established on the basis of likely levels of measurement uncertainty for the particular type of measurement.

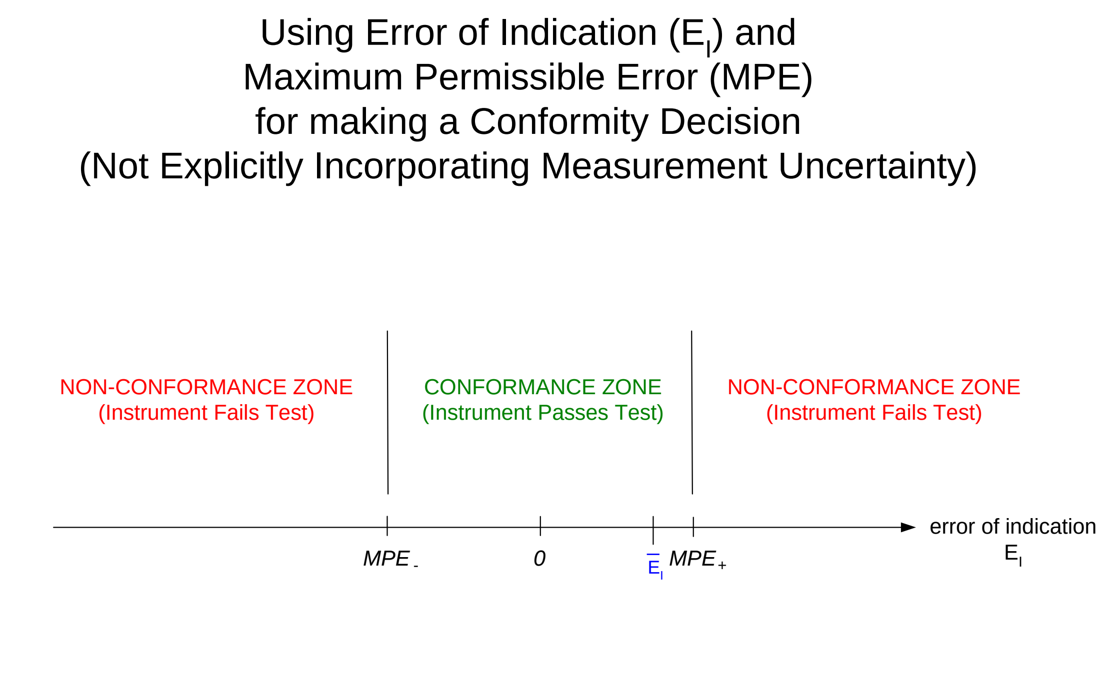


Figure 1

Note that in some OIML Recommendations, in order to account for random variations in measured values, tests are structured such that individual conformity decisions are not based on a single measured error of indication, but rather it is permitted/required to obtain two or more errors of indication and use the average value as the basis of the conformity decision. This is illustrated by the use of the symbol $\bar{E}_I$ in Figure 1, where the test would be considered to pass since $\bar{E}_I$ lies in the conformance zone. Yet another variation is to permit obtaining two or more measured errors of indication, and then require that a certain fraction of them (say, two out of three) lie in the conformance zone. As will be demonstrated in the next clause, when formal measurement uncertainty is taken into account the differences disappear between these ways of making a conformity decision, since measured random variations are incorporated into measurement uncertainty.

5 Conformity testing decisions that formally incorporate measurement uncertainty

As indicated in the Introduction, formally incorporating the concept of measurement uncertainty into conformity testing decisions in legal metrology requires a different way of thinking and talking about such decisions (see Annex A) than is described in Clause 4. Rather than being able to definitively state that a measuring instrument meets specified MPE requirements and so passes a particular conformity test, only a degree of belief (or probability) can be stated that the measuring instrument conforms for each MPE requirement. Inherent in such a probabilistic approach is that certain risks must be considered (e.g., a risk that a decision is incorrect) when ultimately making a pass/fail decision. Measurement uncertainty is used in the process of establishing quantitative values of such probabilities and risks.

It is assumed that the reader of this document has some familiarity with the concept of measurement uncertainty and with the GUM process of calculating it [1]. However, for those who are not familiar, an example is provided in Annex C. Several references [e.g.,] provide more detailed examples. A GUM Supplement [6] is also available that discusses other approaches to calculating measurement uncertainty (e.g., using numerical techniques).

ISO/IEC 17025 [11] has become a widely accepted standard used in the international laboratory accreditation community for assessing the competence of calibration and testing laboratories. This standard states that "testing laboratories shall have and shall apply procedures for estimating uncertainty of measurement," and, further, "When estimating the uncertainty of measurement, all uncertainty components which are of importance in the given situation shall be taken into account using appropriate methods of analysis."

Accordingly, whenever measurement used in the process of testing is specified in an OIML Recommendation (or other OIML document), guidance should be provided on (practical and efficient) methods that can be used to calculate measurement uncertainty for the measurement model(s) appropriate to the type of instrument(s) covered in the Recommendation. In particular, guidance should be provided on how to describe the test apparatus, and on how to identify the input quantities and set up a measurement model (as in Equation 4.2). Further guidance should then be provided on methods that can be used to identify or calculate the associated standard measurement uncertainty (u_S) of the measurement standard or system. Similarly, guidance should be provided on methods that can be used to calculate a standard measurement uncertainty (u_I) associated with the indicated value of the measurand (including components due to indicator resolution, jitter, etc.), and a standard measurement uncertainty associated with repeatability or reproducibility (u_{rep}) of the measuring system and/or procedure. If the indication of the measuring instrument is found to vary over the range of rated operating conditions of the instrument (for a fixed input to the instrument), then a component of measurement uncertainty (u_{roc}) must be included to cover this. Finally, guidance should be provided on how to combine these components of measurement uncertainty to calculate a combined measurement uncertainty (u_{EI}) associated with the error of indication (based on using Equation 4.1). An example of this procedure for establishing a measurement model, identifying and estimating individual components of measurement uncertainty, and finally calculating the measurement uncertainty of the error of indication, is provided in Annex C.

It is important to note that in cases where multiple measurements of a particular error of indication are made for the purpose of assessing the repeatability or reproducibility of the measurement process, it is not necessary to assess the measurement uncertainty associated with each of the individual measured values of error of indication. Rather, the mean value of error of indication ($\bar{E}_I$) can be calculated from the set of individual measured values and used as the 'measured' error of indication, and the standard deviation of the set of individual values can be used as a component of the measurement uncertainty that should be associated with the mean value. OIML Recommendations (and other OIML documents) should emphasize, however, that a random component of measurement uncertainty is not the entire measurement uncertainty, and that systematic components of measurement uncertainty must also be included.

The remainder of this clause discusses ways that the calculated combined standard measurement uncertainty of the error of indication (u_{EI}) can and should be used in order to make conformity decisions for instruments/systems under test.

5.1 Probability density function (PDF)

Inherent in the concept of measurement uncertainty is that the 'true' value of what it is that is intended to be measured cannot be known, since it is impossible to know whether a mistake was made when performing the measurement. And even if it were known that no mistakes had been made in performing a measurement, virtually all measurements have some associated unknown systematic aspects and random variations that are not fully controlled or understood. Accordingly, one must talk in terms of knowing the 'true' value of the measurand on a probabilistic basis, where some values are thought to be more likely than others to correspond to the 'true' value of the measurand. One way of viewing this is that a function can be constructed, known as a probability density function, that gives one's <u>degree of belief</u> about knowing the 'true' value of the measurand.

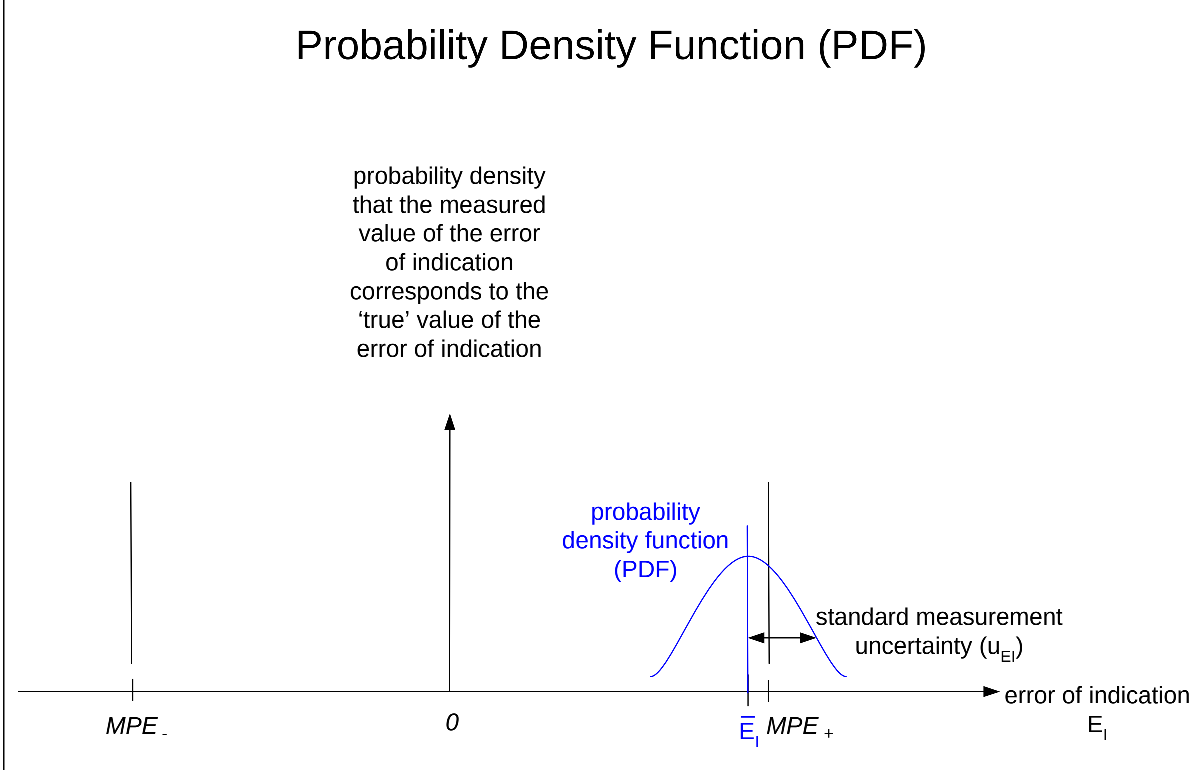


Figure 2

The concept of the probability density function (PDF) is shown schematically in Figure 2. As in Figure 1, the horizontal axis represents possible values of error of indication E_I. In Figure 2, a vertical axis has been added that represents possible probability densities that the 'true' value of the error of indication lies within an infinitesimal region around a particular value of error of indication. The probability (or degree of belief, based on the assumption that no mistakes have been made) that the 'true' value of the error of indication lies between two specified values of error of indication can be obtained by mathematically integrating the area under the probability density function curve bounded by the two specified values.

The PDF curve is shown as Gaussian in shape, which is commonly used (but not always; e.g., see [6]). The mean value ($\bar{E}_I$) of the curve and the standard measurement uncertainty (u_{EI}) are indicated. The curve is normalized such that the total area under the curve is 1, meaning that there is a 100% probability of finding the 'true' value of the error of indication somewhere along the horizontal axis. While this must be the case, it is worth noting that the 'true' value of the error of indication might actually be well outside of the PDF curve, such as if a mistake was made in performing the measurement.

It is also worth reemphasizing that the PDF contains all of the known information about the measurand, including both systematic and random effects. While a curve fit to a histogram of random fluctuations alone frequently has a Gaussian shape, the PDF is not such a fit to a histogram, but rather contains additional information coming from systematic effects in the measurement.

5.2 Probability of conformity

Figure 2 can be used to demonstrate the important differences in making conformity decisions using the classical approach, discussed in Clause 4, and using the GUM uncertainty approach. Using the classical approach, since the mean value ($\bar{E}_I$) of the error of indication is within the conformance zone as defined in Figure 1, the measuring instrument would be considered to pass the particular test shown in Figure 2.

Using the uncertainty approach and taking measurement uncertainty into account for the particular test, it can be seen in Figure 2 that there is a considerable area under the PDF curve that lies outside of the conformance zone (that is, to the right of MPE_+), which means that there is a considerable probability (degree of belief) that the 'true' value of the error of indication lies outside of the conformance zone, even though the mean value ($\bar{E}_I$) of the error of indication is within the conformance zone.

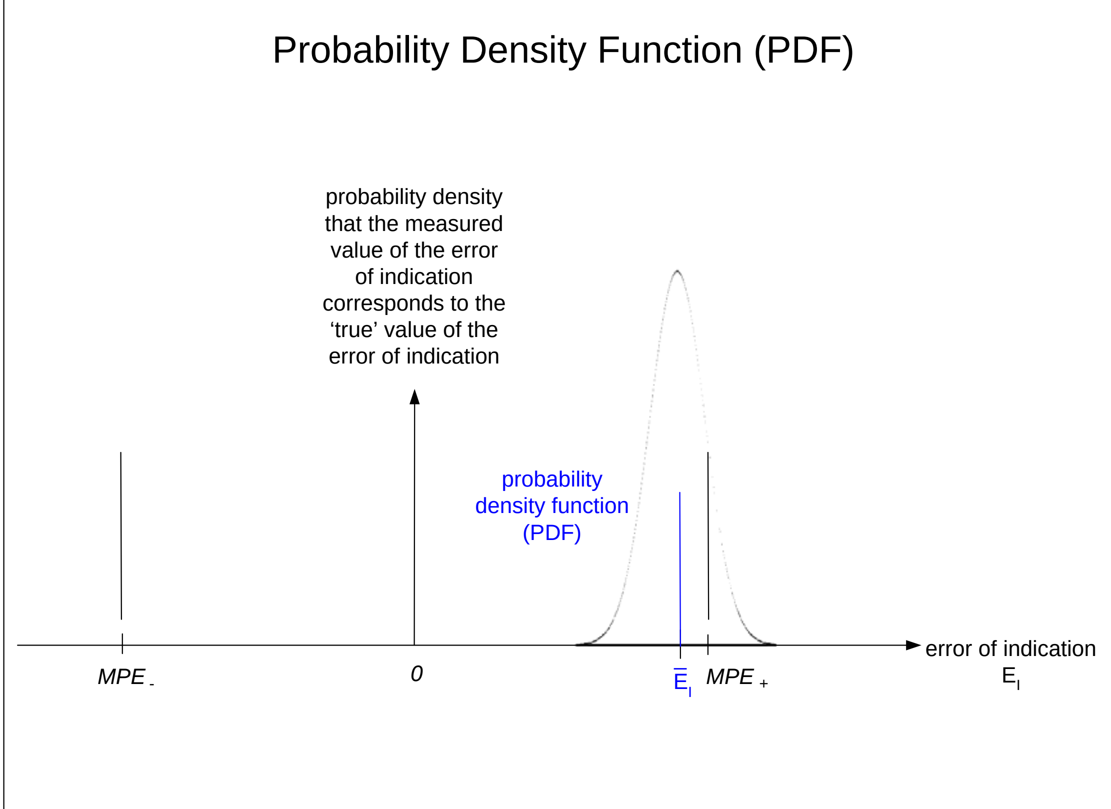


Figure 3

If the area under the PDF curve that lies outside of the conformance zone (as indicated by the un-shaded area under the Gaussian curve in Figure 3) is denoted by A_n, (where "n" stands for 'nonconformance') then the probability p_n that the 'true' value of the error of indication is outside of the conformance zone, and hence that the measuring instrument does not conform to the MPE requirement, is given by $p_n = A_n$ ($= 100 \cdot A_n$ when p_n is expressed in percent (%)). A decision about whether or not the measuring instrument is considered to pass the particular test could then depend upon whether acceptable levels of probability (risk) were met for that kind of test.

Note that if the mean value of the error of indication ($\bar{E}_I$) is just slightly outside of the conformance zone, there can still be a significant probability that the 'true' value of the error of indication lies within the conformance zone. In this case, while the measuring instrument would be considered to fail the particular test in the classical approach, a decision could still be made using the uncertainty approach that the measuring instrument passes the particular test, again depending upon acceptable levels of probability (risk) for that kind of test, and who is considered to be taking the risk. The issue of risk assessment, along with rules for deciding whether a particular test is considered to pass or fail, will be addressed in the next clause.

Constructing PDFs and calculating areas under a PDF curve is in general a nontrivial matter, and so OIML Secretariats and TC/SC members should carefully consider what advice and assistance to provide in this regard in their Recommendation(s). [References?] When the PDF can be treated as Gaussian, there is a convenient method that incorporates what is known as the 'standard normal distribution table' for calculating the area under the curve for a specified $\bar{E}_I$, MPE_+ and u_{EI} [12]. Annex B provides information about the standard normal distribution table, along with an example of how to use it.

5.3 "Risks" and "decision rules" associated with conformity decisions

As already discussed, because of the probabilistic nature of the GUM uncertainty approach to measurement, making a pass-fail decision based on whether or not the measured value of the error of indication lies within the region bounded by the MPEs carries with it the possibility (or risk) that an incorrect decision has been made (that is, the 'true' value of the error of indication may actually lie in a region bounded by the MPEs that is different than the region where the measured value lies). This clause discusses the types of risks associated with the uncertainty approach, and the rules that can be applied to making conformity decisions for testing in legal metrology. These rules should be considered by OIML Secretariats for possible incorporation into OIML Recommendations and other OIML documents.

Various treatments and names have been given to the different types of risks associated with making conformity decisions for tests that are based on meeting tolerance interval requirements such as MPEs [7, 10]. As summary, there are three fundamental types of risks: 1) risk of false acceptance of a test, 2) risk of false rejection of a test, and 3) shared risk.

5.3.1 Risk and decision rule for false acceptance

Risk of false acceptance means that the test is considered to pass, but in reality the MPE requirement might not have been met. In this case, the measured value of the error of indication lies within the region bounded by the MPEs, but the PDF extends into the region outside of the region bounded by the MPEs, as shown in Figure 3, meaning that the 'true' value of the error of indication is believed to possibly lie outside of the region bounded by the MPEs. Note that the risk of false acceptance is taken by the evaluator or user of the measuring instrument or system. The risk is that the instrument or system is not performing 'within specification' even though the test result says it is. The value of the risk of false acceptance is calculated as the area A_n under the PDF curve that is outside of the region bounded by the MPEs, which is the un-shaded area under the curve in Figure 3.

A possible decision rule that can be associated with a legal metrology test is that the probability or risk of false acceptance (p_{fa}) be less than some stated value (for example, 5%). This risk would favor the evaluator or user of the instrument/system, to the detriment of the manufacturer or seller of the instrument/system, since the value of the error of indication $\bar{E}_I$ would lie within the region bounded by the MPEs, and, further, could usually not even lie very close to the relevant MPE boundary if the decision rule is to be met (e.g., see example in Annex D).

5.3.2 Risk and decision rule for false rejection

Conversely, risk of false rejection means that the test is considered to fail, but in reality the MPE requirement might have been met. In this case, the measured value of the error of indication lies outside the region bounded by the MPEs, but the PDF extends into the region inside of the region bounded by the MPEs. Note that the risk of false rejection is taken by the manufacturer or seller of the measuring instrument or system. The risk is that the instrument/system is performing 'within specification,' even though the test result says it is not. The value of the risk of false rejection is calculated as the area under the PDF that is inside of the region bounded by the MPEs when the measured value of the error of indication lies outside the region bounded by the MPEs.

A possible decision rule that can be associated with a legal metrology test is that the risk of false rejection (p_{fr}) be less than some stated value (for example, 2%). This risk would favor the manufacturer or seller of the instrument/system, to the detriment of the evaluator or user of the instrument/system, since the value of the error of indication $\bar{E}_I$ would lie outside of the region bounded by the MPEs, and, further, could usually not even lie very close to the relevant MPE boundary if the decision rule is to be met.

It is important to note that it is not possible to have a decision rule for a given test that incorporates both risk of false acceptance and risk of false rejection. That is, the 'advantage' can go to either the evaluator/user or the manufacturer/seller, but not to both at the same time! It is also important to note that knowledge of the measurement uncertainty (and preferably of the PDF) must be known in order to calculate the risk of false acceptance or false rejection.

5.3.3 Shared risk

Shared risk, on the other hand, is an agreement between the parties concerned with the outcome of the testing that neither will be given an advantage or disadvantage with respect to the consideration of measurement uncertainty for measured values of the error of indication $\bar{E}_I$ that are near enough to the MPE boundaries that risk of false acceptance or rejection would be significant. Implicit in such an agreement is that the measurement uncertainty u_{EI} is 'small' with respect to the MPE (i.e., the ratio (u_{EI}/MPE) is 'small') so that the significant risk of an erroneous decision exists for values of $\bar{E}_I$ that are only very close to the MPE boundaries. This is illustrated in Figure 4 for two possible different PDFs for a given measurement. The uncertainty u_{EI} associated with the leftmost (red) Gaussian curve is probably too large for a shared risk arrangement, whereas the uncertainty u_{EI} associated with the rightmost (green) Gaussian curve would probably be acceptable for most applications.

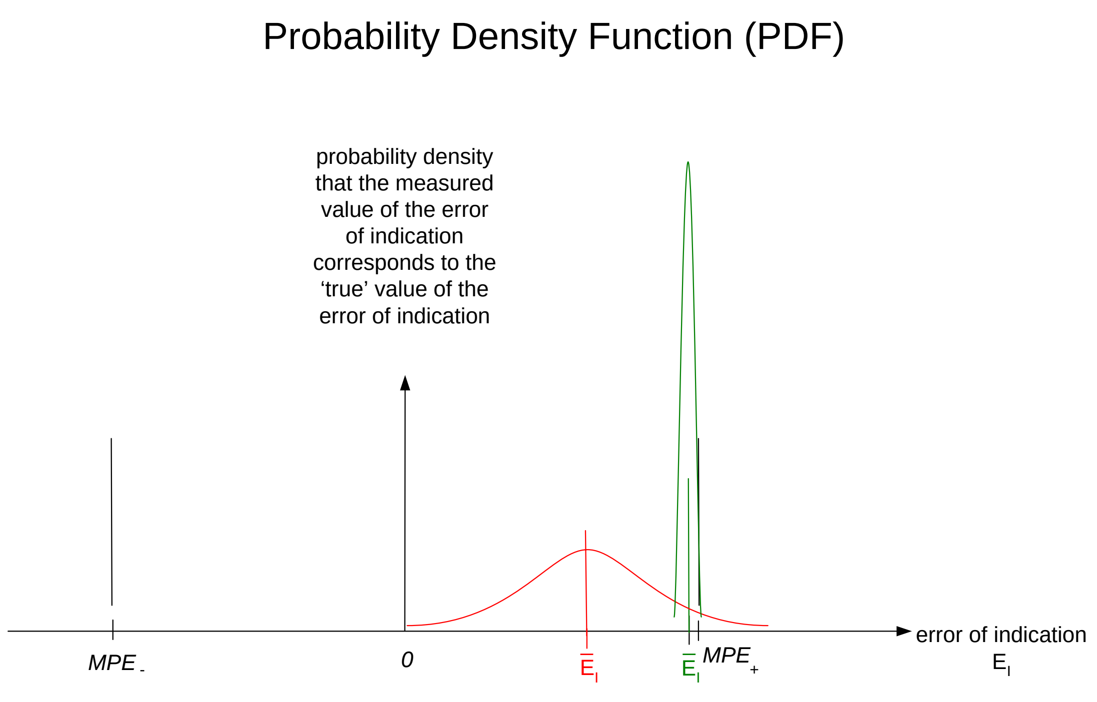


Figure 4

An advantage of the shared risk approach is that it is not necessary to know the PDF for the error of indication, since the risk is shared equally and so no risk calculations are necessary. This advantage makes use of the shared risk approach highly desirable when considering what decision rule to propose in an OIML Recommendation or other OIML document, since it at least partially simplifies the decision making process.

In fact, while not explicitly stated, many OIML Recommendations are currently, at least implicitly, using the shared risk approach. In order to meet the requirements in ISO/IEC 17025 [11] that measurement uncertainty be taken into account, at least at some level of rigor, for all measurements, it is highly recommended that OIML Secretariats explicitly include text in their Recommendations that elaborates that the shared risk principle is being used, when this is the case.

Note that with the shared risk approach it is still necessary to calculate the measurement uncertainty u_{EI} so that the ratio (u_{EI}/MPE) can be examined to see if it is 'small enough,' as discussed in the next clause.

5.3.4 Maximum permissible uncertainty (of error of indication)

It is becoming common (e.g., [13]) to refer to the maximum value that the ratio (u_{EI}/MPE) is allowed to have in terms of a "maximum permissible uncertainty" (denoted symbolically by MPU_{EI}) of the error of indication, defined by:

$$MPU_{EI} \equiv f_{EI} \cdot MPE \qquad (5.1)$$

where f_{EI} is a specified number less than one, usually on the order of 1/3 or 1/5 (0.33 or 0.2) [8]. The maximum permissible uncertainty (MPU_{EI}) is typically thought of as the largest value that u_{EI} can have for a given measurement of the error of indication $\bar{E}_I$ for which the shared risk approach can be used. The decision rule to be applied concerning MPU_{EI} is that if u_{EI} is greater than MPU_{EI} then the test is considered to fail, and means for reducing u_{EI} (or for incorporating an increased MPE) will need to be developed.

Another way of thinking about the need for specifying an MPU_{EI} is that if u_{EI} is comparable to the MPE, then for values of $\bar{E}_I$ that are, say, around halfway between 0 and MPE_+, as shown by the leftmost curve in Figure 4, there can be a relatively large probability that the 'true' value of the error of indication lies far to the right of MPE_+ (i.e., when E_I lies very close to MPE_+), which is an unacceptable risk in many cases. By having an MPU_{EI}, such a possibility is eliminated.

5.3.5 Maximum permissible uncertainty (of measurement standard)

Besides the need for specification of a 'maximum permissible uncertainty (of error of indication),' for the reasons given above, another decision rule that is frequently used is to specify a 'maximum permissible uncertainty (of the measurement standard)' (denoted symbolically by MPU_S), defined by:

$$MPU_S \equiv f_S \cdot MPE \qquad (5.2)$$

where f_S is a specified number less than one, also usually on the order of 1/3 or 1/5 (0.33 or 0.2). Then the maximum permissible uncertainty (MPU_S) is the largest value that u_S is allowed have for a given measurement of the error of indication $\bar{E}_I$.

The rationale for this requirement is that if MPU_S is too large, then the pass-fail decision based on MPU_{EI} above can become dominated by the quality of the measurement standard and/or testing laboratory, rather than on the quality of the instrument/system being tested (note that u_{EI} contains u_S as well as other components of uncertainty). It could be considered unfair to test the instrument manufacturer's instrument with a measurement standard that has an uncertainty that comprises most of u_{EI}, since then the uncertainty of the indicated value (u_I), as well as other possible components of uncertainty associated with the instrument/system, would need to be relatively small in order that the uncertainty of the error of indication remains acceptably small for the particular test (i.e., less than MPU_{EI}). By requiring that f_S be relatively small (say, less than 1/5), then any significant differences or discrepancies among testing laboratories can be avoided. Individual OIML Recommendations should therefore specify an acceptable f_S (or MPU_S) that is appropriate to each particular kind of test.

5.3.6 Summary of considerations for decision rules

When considering what decision rules should be incorporated into the OIML Recommendations and other OIML documents that they are responsible for, OIML Secretariats should take into account the consequences of an incorrect decision when proposing acceptable levels of risk. If the consequences of false acceptance are not considered to be too severe, incorporating the shared risk approach should be promoted, since it is a relatively efficient means of deciding conformity while still taking measurement uncertainty into account. It is usually the case in legal metrology that the shared risk approach can be used successfully for a test, as long as the corresponding MPE for that kind of test does not need to be too 'small' (see Clause 6 below) and that the MPU can be kept acceptably 'large.'

If the shared risk approach cannot be used, and it is instead necessary to use the risk of false acceptance for making a conformity decision, there is a convenient means of doing this, that can minimize the time and effort required by the test evaluator, utilizing the concept of the "measurement capability index" [7], defined for purposes of legal metrology as $C_m = MPE/(2 \cdot u_{EI})$. Annex E provides a discussion and example of how the measurement capability index can be used to make a relatively 'quick' decision on a test when the MPE, risk of false acceptance (p_{fa}), measured E_I and calculated u_{EI} are all known.

For those special cases of using risk of false acceptance (or false rejection) where the uncertainty of the error of indication (u_{EI}) can be considered to be constant, then a particularly convenient method can be used for making conformity decisions, known as "guard banding." Under such conditions, the MPE boundaries are simply 'shifted' inward (for false acceptance) or outward (for false rejection) by an amount corresponding to the respective risks, and conformity decisions are then made on the basis of whether the measured error of indication (E_I) lies within or outside of the shifted conformity boundaries. Reference [7] provides a very useful discussion of the guard band principle.

While decision rules and associated risks, along with their consequences, should be considered and discussed in OIML Recommendations, OIML Secretariats and TC/SC members should consider carefully whether specified levels of acceptable probability for various types of tests should be required or even suggested. If so, this should be done only in the context of regulatory matters. Risks to a manufacturer may have serious economic consequences that are typically outside the scope of a Recommendation.

6 Taking measurement uncertainty into account when establishing MPEs and accuracy classes

Many OIML Recommendations, and some other OIML documents, specify MPEs that are to be used for particular tests. Establishing what values the MPEs should have usually involves a balance of considerations, including adequately protecting the consumer or user of the measuring instrument/system for reasons of cost and sometimes safety, but also protecting the manufacturer or distributor, again for reasons of cost. What is sometimes overlooked is consideration of the lowest level of measurement uncertainty that can be physically attained for the particular test, which sets a lower limit on the MPE that can be used. OIML Secretariats should take this into account when specifying an MPE for a particular test, or when establishing accuracy classes for a type of instrument, especially in cases where MPUs are specified.

For example, in testing cases where the uncertainty u_{EI} is known to typically be of a certain amount (and cannot easily be reduced), then the MPE corresponding to that test should be appropriately specified such that the ratio ($f_{EI} = u_{EI}$/MPE) discussed in clause 5.3.4 can be kept acceptably low. In this case, since u_{EI} cannot be reduced, it may become necessary to increase the MPE such that the condition illustrated by the rightmost curve in Figure 4 can be obtained.

Similarly for the measurement standard, if f_S (= u_S /MPE) is typically too large for a given type of test, then the MPE might not be appropriate and so, if possible, specifying a larger MPE in the Recommendation might be necessary. If the MPE cannot be reduced for other reasons, then it might be necessary to specify a type of measurement standard/system that has a lower measurement uncertainty (u_S).

While outside of the scope of this Guide, OIML Secretariats are encouraged to consult existing literature (e.g., [14]) when considering advice to include in their Recommendations concerning specification of appropriate MPEs and Accuracy Classes.

7 Options pertaining to "measurement uncertainty" that should be considered for inclusion in OIML Recommendations and other OIML documents

When discussing how to incorporate measurement uncertainty into the Recommendations and other OIML documents for which they are responsible, Secretariats and TC/SC members should consider the following:

1) Provide a clause in each OIML Recommendation that emphasizes how measurement uncertainty can and should be incorporated into conformity decisions that are associated with the Recommendation. Suggested text (in *italics*):

"*XX Measurement uncertainty*

The use of measurement uncertainty has become an important and essential element in all aspects of metrology, including legal metrology. The OIML Guide(?) G YY on "The Role of Measurement Uncertainty in Conformity Assessment Decisions in Legal Metrology" should be consulted for a general understanding of the terminology and concepts related to measurement uncertainty, and for guidance on how to assess and use measurement uncertainty.

Measurement uncertainty shall be considered in all aspects of measurement and conformity assessment decisions associated with this OIML Recommendation. Guidance is provided on how to do this (see Clause xxx).

Every measurement result that is reported during testing of a measuring instrument/system when using this Recommendation shall include a measured value along with its associated measurement uncertainty. Exceptions include those cases where individual measured values are obtained for the purpose of assessing a component of measurement uncertainty associated with the repeatability or reproducibility of the measuring instrument/system and/or testing procedure, or where it is determined that a component of measurement uncertainty is not significant in a particular measurement application (this should be so noted)."

2) Individual OIML Recommendations should provide guidance on calculating measurement uncertainty for the measurement model(s) appropriate to the type of instrument(s), testing systems and processes covered in the Recommendation. Examples of such guidance are given in the seven steps below. In general, guidance should be provided on how to:

- (Step 1) Describe the instrument under test (IUT), along with the measuring system that will be used for performing the test(s). Include in the description all quantities that can effect the measuring instrument, all influence quantities that can effect the measuring instrument/system, and specify the conditions (if any) at which the (influence) quantities will be maintained during the testing, or the range(s) that the (influence) quantities must remain within during the testing (e.g., rated operating conditions and/or reference operating conditions of both the measuring instrument/system and IUT);

- (Step 2) Identify all of the different kinds of tests that will need to be performed for the type evaluation and/or verification. Based on the description in Step 1, develop a mathematical model of the measurement (as in Equation 4.2) to be used for performing each kind of test. Each model must ultimately provide an expression for the 'error of indication,' and also include an expression for the standard measurement uncertainty to be associated with each measured error of indication (unless repeated measurements of error of indication are to be obtained, in which case the mean value of the error of indication is to be presented, along with an associated standard measurement uncertainty that incorporates a component obtained from the repeated measurements; see Step 5 below);.

- (Step 3) Calculate the associated standard measurement uncertainty (u_S) of the measurement standard or system;

- (Step 4) Calculate a standard measurement uncertainty (u_I) associated with the indicated value of the measurand (including components due to indicator resolution and/or random fluctuation);

- (Step 5) Calculate a standard measurement uncertainty (u_{rep}) associated with the repeatability or reproducibility of the measuring instrument/system and/or testing procedure;

- (Step 6) Calculate a standard measurement uncertainty (u_{roc}) if the indication of the measuring instrument is found to vary when the instrument is operated over its range of rated operating conditions for a fixed input to the instrument;

- (Step 7) Combine all of these components of measurement uncertainty in order to calculate a combined standard measurement uncertainty (u_{EI}) associated with the error of indication.

OIML Recommendations (and other OIML documents) should emphasize that a random (Type A) component of measurement uncertainty is not the 'entire' measurement uncertainty, and that systematic (Type B) components must also be included.

If they exist, include discussion of special or unusual aspects of assessing the components of measurement uncertainty.

3) For each kind of test identified above in 2), Step 2, the OIML Recommendation should discuss and specify what the appropriate MPE is for that kind of test. For example, for a type evaluation test, the MPE that is specified could correspond to one of several possible accuracy classes that the instrument is being tested for. For a verification test, the specified MPE could be based on a variety of considerations, as discussed in Clause 6.

There should also be discussion of what the likely values of u_{EI} and u_S will be during the test, in order to decide whether values of MPU_{EI} and MPU_S should be specified and, if so, what those values should be (or, rather, what f_{EI} and f_S should be; See Clauses 5.3.4, 5.3.5 and 6.)

4) OIML Secretariats and TC/SC members should consider whether 'acceptable' levels of risk for various types of tests should be suggested in their OIML Recommendations. Decision rules and associated risks, along with their consequences, should be considered and discussed in OIML Recommendations. However, this should be done only in the context of regulatory matters. Risks to a manufacturer may have serious economic consequences that are typically outside the scope of a Recommendation.

Depending on the values of MPU_{EI} and MPU_S specified in the prior step (if any), discussion should be provided on whether the 'shared risk' principle is to be used (see Clause 5.3.3), or whether there is a specified risk (probability) that is to be used and, if so, whether it is a Risk of False Acceptance (see Clause 5.3.1) or a Risk of False Rejection (see Clause 5.3.2). Note that if the 'shared risk' approach is used in an OIML

Recommendation (or in other OIML documents), it should not be used in an implicit manner but, rather, an explicit statement of its use should be provided in the Recommendation.

5) If Risk of False Acceptance or Risk of False Rejection is used, it is further necessary to specify whether u_{EI} is to be considered as fixed for each measurement, in which case a guard band can be used for deciding conformity, or whether u_{EI} is to be calculated separately for each measurement of error of indication, in which case the Standard Normal Distribution Table or Measurement Capability Index can be used each time. Reference to Annex B and Annex E of this *OIML Guide(?) G YY on "The Role of Measurement Uncertainty in Conformity Assessment Decisions in Legal Metrology"* should be provided, along with possible additional discussion of how to use the Standard Normal Distribution Table and/or Measurement Capability Index for the particular Recommendation.

Constructing PDFs and calculating areas under a PDF curve is in general a nontrivial matter, and so OIML Secretariats and TC/SC members should consider what advice and assistance to provide in this regard in their Recommendation(s) (e.g., use of the standard normal distribution, or numerical techniques).

6) While assessing the measurement uncertainty of the error of indication for an individual measurement for a specified type of measuring instrument may be somewhat complex, it is important to note that, once all of the derivation has been performed, and values and associated measurement uncertainties are obtained for typical measurement conditions, the process of obtaining a value of u_{EI} for each subsequent individual measurement performed during a given type evaluation or verification test should become relatively straightforward, since most components of measurement uncertainty will not change from one individual measurement to another. This aspect of the treatment of measurement uncertainty should be included in the discussion in each OIML Recommendation where measurement uncertainty is relevant. Suggested text (in *italics*):

"*While assessing the measurement uncertainty of the error of indication for an individual measurement for a specified type of measuring instrument may be somewhat complex, it is important to note that, once all of the derivation has been performed, and values and associated measurement uncertainties are obtained for typical measurement conditions, the process of obtaining a value of u_{EI} for each subsequent individual measurement performed during a given type evaluation test should become relatively straightforward, since most components of measurement uncertainty will not change from one individual measurement to another. This can simplify the process of incorporating measurement uncertainty in 'field' situations, since guardbands or straightforward Measurement Capability Index tables can be used (e.g., see Annex E in the OIML Guide(?) G YY on "The Role of Measurement Uncertainty in Conformity Assessment Decisions in Legal Metrology.""*

7) OIML Recommendations should provide for explicit entries in the Format of the Test Report document for recording measurement uncertainty, to accompany every measured

value that is recorded (except when measurements for repeatability and/or reproducibility are being obtained). In those cases where measurement uncertainty can be assumed to be negligible, this should be documented with an appropriate notation, rather than leaving a blank entry. Also, if the 'Measurement Capability Index' ($C_{M)}$ method or the 'Guardband' method is to be used, this should also be noted in the Format of the Test Report, along with spaces for recording values of the appropriate parameters, along with the outcome of the test. A space for reference to where to find the C_M chart should that was used also be provided.

8) OIML Recommendations should provide guidance on how to treat measurement uncertainty at the stage of verification testing, emphasizing any differences, precautions and/or special considerations from the guidance provided for type evaluation testing.

8 References

[1] *Guide to the Expression of Uncertainty in Measurement*, ISO, 1993 ; Corrected and Reprinted, 1995.

[2] *OIML Certificate System for Measuring Instruments*, OIML B 3, 2003 (Amended 2006)

[3] *Framework for a Mutual Acceptance Arrangement on OIML Type Evaluations (MAA)*, OIML B 10, 2004 (Amended 2006)

[4] *International Vocabulary of Metrology - Basic and General Concepts and Associated Terms*, Third Edition, ISO, 2007.

[5] Ehrlich, C., Dybkaer, R., and Woger, W., *Evolution of philosophy and description of measurement (preliminary rationale for VIM3)*, OIML Bulletin, April, 2007, p.23-35

[6] *Evaluation of measurement data – Supplement 1 (Monte Carlo Techniques)* …2007

[7] *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment*, Joint Committee for Guides in Metrology (JCGM), Working Group 1 (WG1), Document 106, under development

[8] Sommer, K.-D. and Kochsiek, M., *Role of measurement uncertainty in deciding conformance in legal metrology*, OIML Bulletin, Volume XLIII, Number 2, April 2002, p.19-24

[9] Ehrlich, C. and Rasberry, S., *Metrological Timelines in Traceability*, OIML Bulletin,

[10] *Guidelines on assessment and reporting of compliance with specification*. ILAC-G8:1996

[11] *General requirements for the competence of testing and calibration laboratories* ISO/IEC 17025, 2005.

[12] http://www.itl.nist.gov/div898/handbook/eda/section3/eda3671.htm

[13] Kallgren, H. and Pendrill, L., *Uncertainty in conformity assessment in legal metrology (related to the MID)*, OIML Bulletin, Volume XLVII, Number 3, July 2006, p. 15-21

[14] WELMEC Guide 4.2, Elements for deciding the appropriate level of confidence in regulated measurements, Issue 1, June 2006

[15] *The Expression of Uncertainty in EMC Testing*, United Kingdom Accreditation Service (LAB34), Edition 1, August 2002

Annex A Coexistence of "measurement error" and "measurement uncertainty" in legal metrology (relationship between measurement and testing)

The introduction in 1993 of the *Guide to the Expression of Uncertainty in Measurement* [1] (also referred to as the GUM) opened a new way of thinking about measurement, in particular about how to express the perceived quality of the result of a measurement. Rather than express the result of a measurement by providing a best-estimate of the true value of the quantity being measured, along with information about known systematic and random errors, the GUM provided an alternative approach whereby the result of a measurement is expressed as a best-estimate of the essentially-unique true value (denoted hereafter as 'true' value) of the quantity intended to be measured (the 'measurand'), along with an associated 'measurement uncertainty.' (Note that, historically, in legal metrology the term "true value" is sometimes used to mean the value associated with a measurement standard that is used in the process of testing a measuring instrument. This is not the meaning of the term in this document.)

The concept of measurement uncertainty can be described as a measure of how well the 'true' value of the measurand is believed to be known. (Note that according to the GUM approach it is not possible to know how well the 'true' value of the measurand is known, but only how well it is believed to be known.) The notion of 'belief' is an important one, since it moves metrology (and legal metrology) into a realm where results of measurements must be thought about and expressed in terms of probabilities or degrees of belief. When making decisions in legal metrology about whether measuring systems are performing according to specified requirements, if the GUM approach is to be followed it becomes necessary to make such decisions on a probabilistic basis. This OIML Document provides guidance on how to incorporate the GUM approach and take into account the concepts of measurement uncertainty and probability when making such conformity assessment decisions.

Legal metrology is the process and the practice of applying regulatory structure and enforcement to metrology, which is the science and application of measurement. Much of legal metrology involves testing measuring instrument/system design and use, in both laboratory and field environments, to assure that credible measurements can be, and are being, made when using the instrument/system in regulated situations. Testing in this context means that a decision is being made about whether the measuring system under test is providing indicated values of a quantity being measured that are believed to be 'close-enough' to the 'true' value, as determined by using measurement standards, for the regulatory purpose at hand. The close-enough conditions are specified in regulations, usually in terms of 'maximum permissible errors' (MPEs) or 'accuracy classes.' Using the GUM approach, the objective of testing then becomes to determine the degree of belief that the 'true' value of the 'error of indication' lies within the maximum permissible errors when taking measurement uncertainty (of the measured 'error of indication'!) into account.

Using the concepts of 'measurement error' and 'measurement uncertainty' at the same time like this may at first glance seem inconsistent or otherwise confusing. The GUM seemingly discourages use of the concept of measurement error in favor of measurement uncertainty. However it must be kept in mind that the focus of the GUM is on using calibrated measuring instruments to perform measurements, and not on testing measuring instruments themselves. From the GUM perspective, known measurement errors that arise when using a measuring instrument are to be ‘corrected for,’ so that no known (systematic) measurement error remains. By contrast, in the context of testing in legal metrology (as well as in some other areas of metrology), the ‘measurement’ of error is used to assess the performance of a measuring instrument (and is not corrected for), and error (or, actually, error of indication) can in fact be considered to be a perfectly reasonable measurand, for which its ‘true’ value cannot be known but can be stated on a probabilistic basis. This approach to use of the term ‘error’ is the approach that is taken in this document.

As already indicated, conformity testing in legal metrology typically involves comparing the measured error of indication of a measuring instrument or system to an MPE that is specified in a legal regulation. The measured value of the error of indication is typically calculated in legal metrology as the difference between the indicated value and a value as given by a measurement standard. It is known that the value as given by the measurement standard is very likely not the ‘true’ value of the quantity being measured, but it is typically thought to be very close for a given situation. However, since the ‘error of indication’ is usually meant to be the difference between the indicated value and the ‘true’ value of the measurement standard, the uncertainty associated with the value given by the measurement standard (such as is stated in its calibration certificate) must be taken into consideration when making a conformity assessment decision. This will be elaborated on below.

By utilizing a first-principles approach that incorporates a simple example involving a mass standard and a weighing instrument to be tested, this Annex will now elaborate on how measurement error and measurement uncertainty can coexist when considering measurement in the context of testing.

As in Clause 3 of the GUM, the initial focus of this Annex will be to consider measurement error and measurement uncertainty from the perspective of describing the objective of measurement. The terminology used to do this will be that of the VIM3 [4], which in some cases is somewhat different than that of the GUM, for reasons that will be explained when necessary. Several relevant definitions from the VIM3 are provided in Clause 2 of this document.

The objective of a measurement can be thought of as developing, through some type of ‘experiment,’ a quantitative expression about the ‘measurand’. The expression usually involves the concept and term ‘value’ (‘quantity value’ in VIM3), which is a number and reference that together express magnitude of a ‘quantity.’ The reference is typically a measurement ‘unit,’ which is adopted by convention such that other quantities of the same kind can be compared to it.

Prior to the concept of measurement uncertainty, the objective of measurement was to obtain a measurement result that was typically expressed as a best-estimate of the ‘true’ value of the measurand and was sometimes accompanied by an ‘error analysis’ that contained any systematic errors (that were to be ‘corrected’ for when calculating the best-estimate) and a description of the ‘spread’ of the random errors (if more than one observation was made) that occurred during the measurement. The concept of metrological traceability was used for expressing the measurement result in terms of an appropriate measurement unit by establishing a chain of comparisons or calibrations back to a realization of the measurement unit. Besides stating possible systematic errors associated with the traceability chain, nothing further was typically stated about other possible sources of systematic error.

As discussed earlier, the concept of measurement uncertainty fundamentally changed the way that metrologists think about the objective of measurement. Most notably, one of the basic premises of the GUM approach is that it is possible to characterize the quality of a measurement by accounting for both random and systematic ‘effects’ on an equal footing, thus refining the information previously provided in an error analysis, and putting it on a probabilistic basis. Rather than express a measurement result as a best-estimate of the ‘true’ value of the measurand, along with an error analysis, a measurement result is instead to be expressed as a best-estimate of the ‘true’ value of the measurand along with a measurement uncertainty, which is a measure of how well the stated best-estimate is believed to be known (based on the experimental data and the assumption that no mistakes were made when performing the measurement).

The probabilistic basis of the GUM approach derives primarily from another basic premise of the GUM (Clause 3.3.1), which is that it is not possible to know the true value of a measurand: “The result of a measurement after correction for recognized systematic effects is still only an estimate of the value of the measurand because of the uncertainty arising from random effects and from imperfect correction of the result for systematic effects.” This is a very fundamental and important point to keep in mind. Another related consideration, discussed in D.3.4 of the GUM, is that there is no such thing as a unique true value of a measurand, since at some level there is always an ‘intrinsic’ uncertainty due to the necessarily incomplete definition of the measurand (VIM3 refers to this as “definitional uncertainty”). Clause 1.2 of the GUM elaborates that, therefore, it is not possible to have a unique, true value of a measurand, but rather that it is only possible to have an “essentially unique” true value, which, as mentioned earlier, for shorthand has been referred to in this document as a ‘true’ value.

Note that the Note in Clause 3.1.1 of the GUM explains why the GUM views the terms “value of a measurand” and “true value of a measurand” to be “equivalent,” and so uses only the term “value” when what is meant is the concept of ‘true’ value (as it is defined in B.2.3 of the GUM), namely, a value consistent with the definition of the measurand. The VIM3 [4] and this document do not adopt this GUM convention, and utilize the term “true value” when that concept is what is intended, since the term “value” is already used in the more general sense given above. It is otherwise confusing to use the single term “value” for two different concepts [5].

Besides the concept of "error," another concept (and term) that is discouraged in the GUM, at least in a quantitative sense, is "accuracy." This is because "accuracy" is typically thought of in the inverse sense as "error," in that the larger the error, the lower the accuracy. Since "error" cannot be known in the GUM sense, neither can "accuracy." Therefore, care should be taken in OIML Recommendations to be sensitive to how the term "accuracy" is used, both in connection with "accuracy classes" as well as in the general sense.

Metrological traceability continues to be a very important concept in the uncertainty (GUM) approach to measurement, and in fact takes on an additional aspect that links it very closely to the concept of measurement uncertainty. Besides serving as the basis for establishing a chain of comparisons or calibrations back to the measurement unit so as to be able to express the 'measured value' in terms of a measurement unit, the concept of metrological traceability is also used to be able to track the progression of measurement uncertainty along the traceability chain. In this regard, metrological traceability and measurement uncertainty are inextricably linked [9], as explicitly evidenced in the VIM3 (and VIM2) definition of metrological traceability.

A.1 Measuring

The concepts of 'measurement unit', 'true' value, 'measurement error' and 'standard measurement uncertainty' are illustrated in Figure A1, in the context of measuring (calibrating) a standard weight, which is shown schematically at the top right. It is assumed that the weight is calibrated using a high quality measuring system that is not otherwise mentioned or shown. The calibration certificate of the standard weight contains the measured mass value ($M_{calibrated}$) of the standard weight, along with the associated standard measurement uncertainty ($u_{calibrated}$). The standard measurement uncertainty (or the expanded uncertainty, $U_{calibrated}$) is obtained during the calibration of the standard weight, through the use of the traceability principle, back to the measurement unit shown on the horizontal axis of the figure. The 'true' value of the mass of the standard weight is also indicated in the figure, both at the top right and on the horizontal axis, where it is indicated that it exists, but is unknowable in principle. The small vertical bars around the 'true' value of the mass of the standard weight on the horizontal axis are meant to denote the definitional uncertainty associated with the 'true' value.

Also shown in Figure A1 is a probability density function (PDF) which, as described in Clause 5.1, provides probability densities that the 'true' value of the mass of the standard weight lies within an infinitesimal region around a particular possible 'true' value of the mass of the standard weight. The standard measurement uncertainty ($u_{calibrated}$) is obtained from the PDF, usually as the standard deviation of the (assumed) Gaussian curve, as indicated.

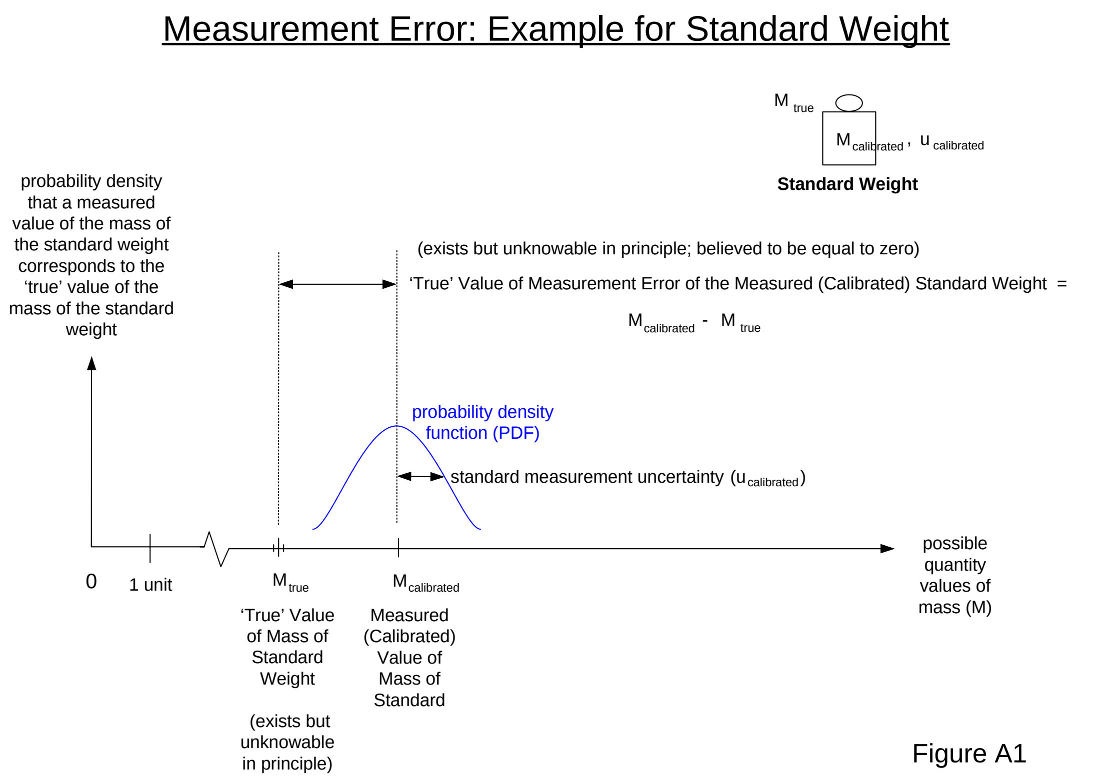


Figure A1

Figure A1 also illustrates the 'true' value of the 'measurement error' of the mass of the standard weight, defined as the difference between the measured (calibrated) value of the mass of the standard weight and the 'true' value of the mass of the standard weight. An important point to note in Figure A1 is that this error is considered as unknowable, since the 'true' value of the mass of the standard weight is unknowable. The GUM discourages use of the concept of error since it is 'unknowable' in this measurement context, and instead favors use of measurement uncertainty, since measurement uncertainty can be calculated, and gives a measure of how well one believes one knows the 'true' value of the mass of the standard weight. It is very important to keep in mind that, in the context of measurement, despite the possible reality illustrated in Figure A1, the 'true' value of the error of the measured (calibrated) mass of the measurement standard is believed to be zero, based on all of the available information from the measurement (calibration), since corrections are to be applied for all known systematic errors.

A.2 Testing

Now consider the situation where the calibrated standard weight is used for the purpose of testing, not calibrating, a weighing instrument, as illustrated in Figure A2. In a testing scenario, indicated values of a quantity being measured when using a measuring instrument under test are compared with measured values (of the same quantity) as obtained when using a measurement standard.

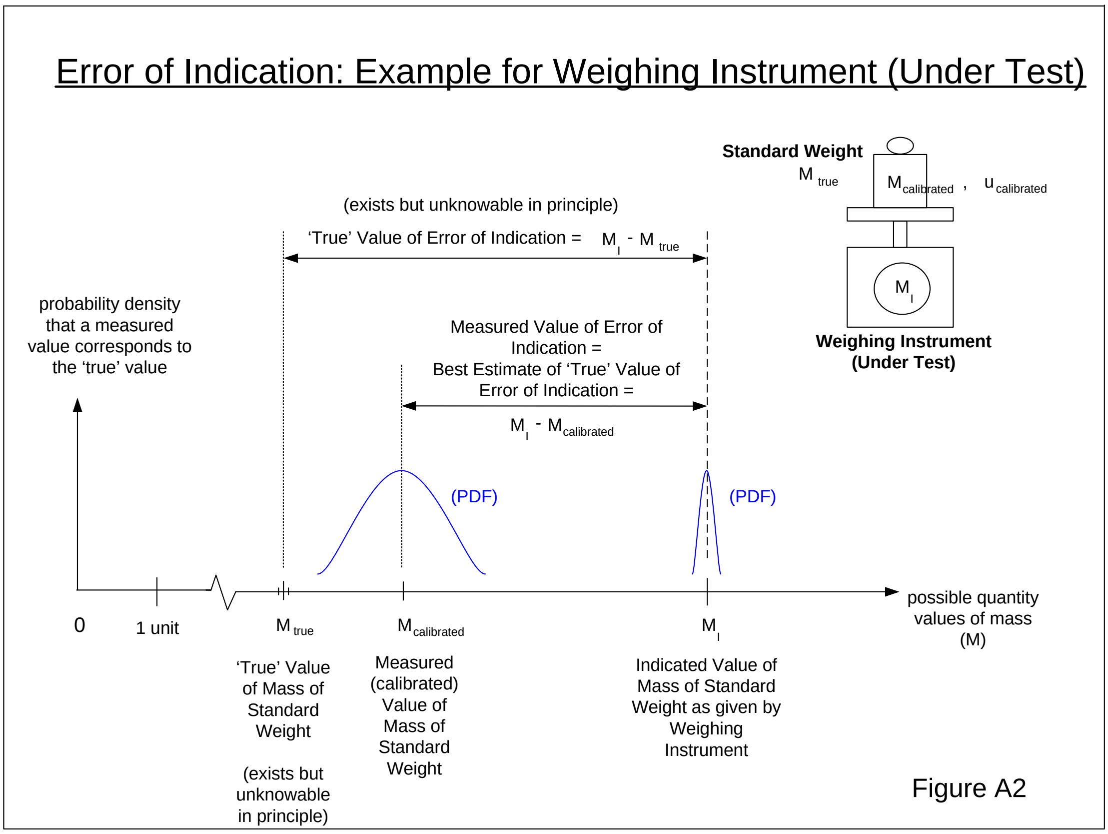


Figure A2

Figure A2 contains much of the same information as Figure A1, but in addition shows the value (M_I) of the indication of the mass of the standard weight as obtained from the weighing instrument under test. Two 'errors of indication' are also shown, one with respect to the 'true' value of the mass of the standard weight (which is still unknowable), and another with respect to the measured (calibrated) value of the mass of the standard weight (which is knowable and, in fact, known). As noted in Figure A2, the measured value of the error of indication is taken as the 'best-estimate' of the 'true' value of the error of indication since, as discussed above, the 'true' value of the error of the measured (calibrated) mass of the measurement standard (standard weight) is believed to be zero.

Testing is typically a quicker and less time consuming process than calibration, and is therefore frequently performed in both 'laboratory' and 'field' environments, especially when controlled laboratory conditions are not available or practical. In a testing scenario, the objective is not to 'correct' or 'adjust' the indicated value to the measured (calibrated) value of the mass standard, but rather to assess whether the measured difference (error of indication) between the indicated value and the calibrated value of the mass standard is within acceptable limits of maximum permissible errors (MPEs, see Clause 4), as expressed in regulation (e.g., in an OIML Recommendation). While it is highly desirable that the error of indication be small (and even zero), this is typically not the case in testing.

Note that use of the term "error" in the context of a testing situation (as 'error of indication') is different than in the context of a measurement (calibration) situation (as 'measurement error'), sometimes leading to confusion since, as mentioned earlier, the GUM discourages use of the term and concept of "error."

Also shown in Figure A2 are two PDFs, one for the measured (calibrated) value of mass of the standard weight (this is the same PDF as shown in Figure A1), and the other for the indicated value of the mass of the standard weight (sources of this uncertainty could come from instability (jitter) of the indicated value, and finite resolution of the indicator). What is desired is to use the information in these two PDFs to be able to make a statement about how well the 'true' value of the error of indication is believed to be known. This is illustrated in Figure A3.

Note that the horizontal axis in Figure A3 is now changed from that in Figures A1 and A2, and is labeled 'possible quantity values of error of indication.' The magnitude of the measured value of the error of indication is the same as is given in Figure A2 and, as discussed earlier, is the best estimate of the 'true' value of the error of indication. As for any measurand, a PDF can be constructed giving the probability density that the 'true' value of the error of indication (the measurand in this case) lies within an infinitesimal region around a particular possible 'true' value of the error of indication. Such a PDF is illustrated in Figure A3, along with the associated standard measurement uncertainty (u_{EI}). This PDF is obtained by combining (sometimes called convoluting) the two PDFs in Figure A2 [6]. It is interesting to note that u_{EI} is the 'uncertainty of the error (of indication),' which explicitly demonstrates the coexistence of the terms and concepts 'uncertainty' and 'error' in a testing scenario.

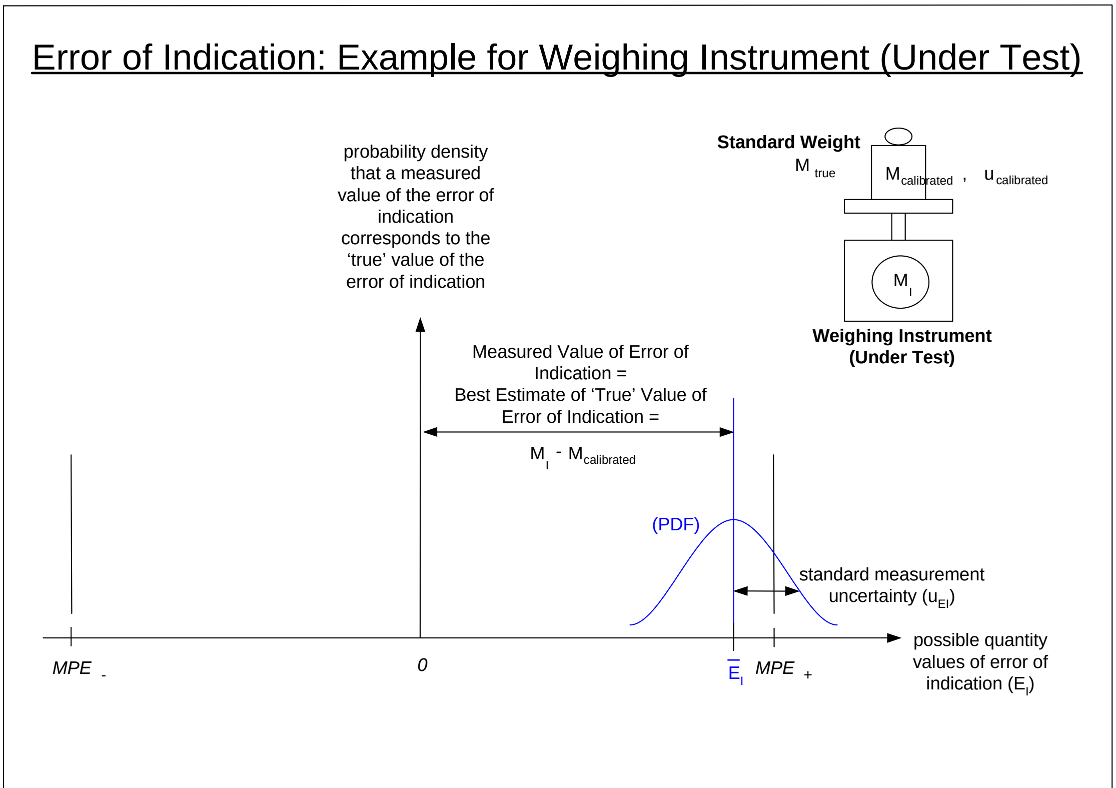


Figure A3

A.3 Brief summary

In summary, while the concept of 'measurement uncertainty' was developed to replace the need for the concept of 'measurement error' and 'error analysis' in the context of performing measurements, the term and concept of 'error' remains useful in the context of testing measuring instruments and systems. In fact, it makes sense to talk about the uncertainty of a measured error of indication! The measurement uncertainty associated with the measurement standard(s) used when performing the testing must be taken into account when making (probabilistic) conformity assessment decisions, since they contribute to the standard measurement uncertainty of the error of indication (u_{EI}).

Annex B Use of the Standard Normal Distribution Table

[Adapted from: NIST Web Site]
http://www.itl.nist.gov/div898/handbook/eda/section3/eda3671.htm

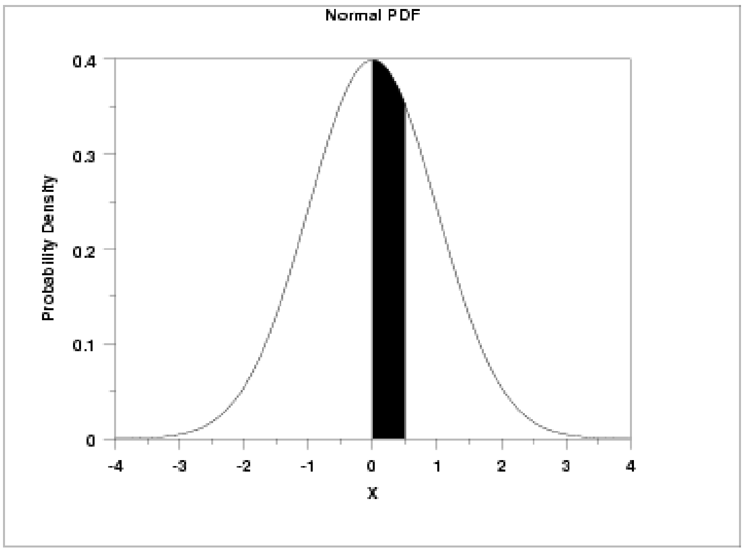
The general formula for the probability density function of the normal distribution is

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

where μ is the location parameter and σ is the scale parameter. The case where $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**. The equation for the standard normal distribution is

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

The figure below illustrates the standard normal distribution (sometimes also referred to as a normalized Gaussian distribution). The shaded area under the curve represents the probability that the parameter *x* is between 0 and α (α = 0.5 in the figure).



Values of areas under the curve for discrete values of α can be obtained from the standard normal distribution table:

Standard Normal Distribution Table

The table below contains the area under the standard normal curve from x = 0 to a specified value x = α.

Area under the Normal Curve from X = 0 to X = α

α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997

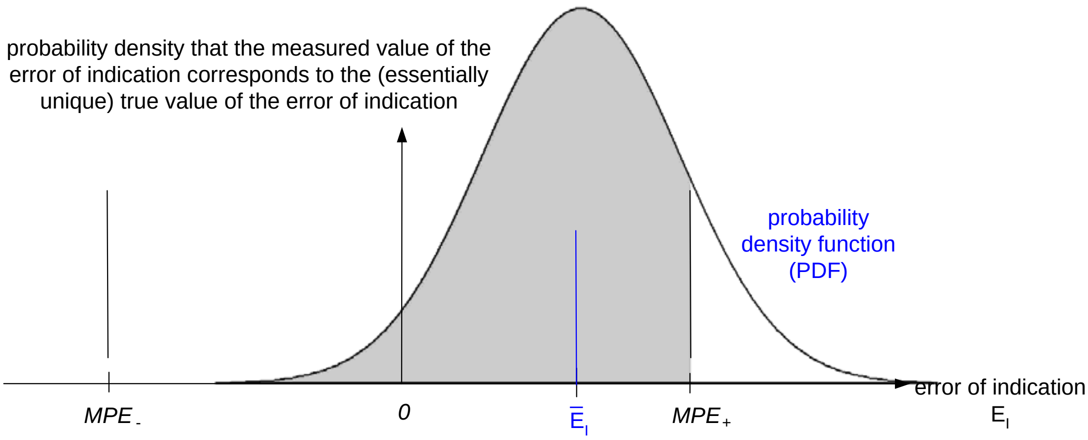
In this document, α is defined as:

$$\alpha = [(MPE_{+} - \bar{E}_{I})/\ u_{EI}] \qquad \text{B.1}$$

for the case where $\bar{E}_I$ is greater than 0. The case where $\bar{E}_I$ is less than 0 is discussed later.

Figure B1 illustrates the relevant parameters:

Making Conformity Decision Based on Combined Standard Measurement Uncertainty of Error of Indication



Possible Conformity Criterion: What is the probability that the true value of the error of indication (E_I) lies inside of the Conformance Zone?

Determine Area under PDF, using Standard Normal Distribution Table.

Figure B1

EXAMPLE

Consider an individual test of a length measuring instrument, such as a line measure, where the indicated value of length (L_I) is 1.0006 m when the value of the reference length of a high-precision line measure (L_R), as obtained from its calibration certificate, is 1.0003 m. The measured value of the error of indication is then:

$$E_I = L_I - L_R = 0.0003 \text{ m} = 300 \text{ µm} \qquad \text{B.2}$$

Say that a calculation of the standard uncertainty of the error of indication gives:

$$u_{EI} = 180 \text{ µm} \qquad \text{B.3}$$

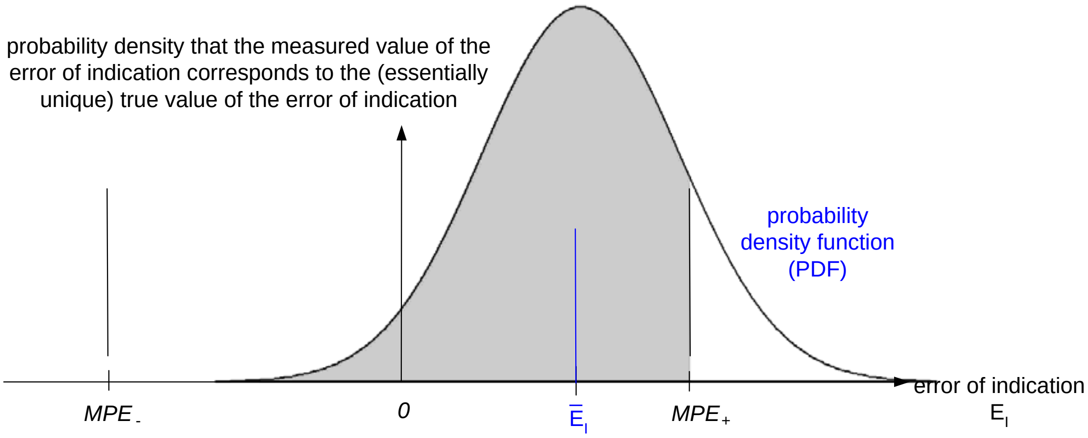
If the MPE for this particular test is given as 500 µm, then α is calculated as:

$$\alpha = [(MPE_+ - \bar{E}_I)/ u_{EI}] = [(500\text{-}300)/180] = 1.11 \qquad \text{B.4}$$

From the Standard Normal Distribution Table above, find the entry for 1.11 by reading down the left column under α to "1.1", then read across the top row to the heading ".01", then read the entry in the table where the column and row intersect, which is .3665.

This means that the area under the curve between $\bar{E}_I$ and MPE_+ in Figure B1 is 0.3665. Therefore, since the area under the curve to the left of $\bar{E}_I$ is 0.5000, the probability (assuming that no mistakes were made in the measurement) that the true value of the error of indication is within the conformance zone is 0.3665 + 0.5000, or 0.8665 (86.7%). Thus, the risk of false acceptance is $p_{fa} = 1 - 0.8665 = 0.133 = 13.3\%$. Note that $f_{EI} = u_{EI}/MPE = 0.36$, so that the maximum permissible uncertainty test would fail if the maximum value of f_{EI} was specified as 1/3 for this test.

Making Conformity Decision Based on Combined Standard Measurement Uncertainty of Error of Indication



Possible Conformity Criterion: What is the probability that the true value of the error of indication (E_I) lies inside of the Conformance Zone?

Area under PDF, using Standard Normal Distribution Table, is 86.7%.

Figure B2

In the case where $\bar{E}_I$ is less than 0 the standard normal distribution table can again be used, taking advantage of the symmetry of the Gaussian curve, but it is then necessary to define α according to:

$$\alpha = [(\bar{E}_I - MPE_-)/\ u_{EI}]\ . \qquad B.5$$

Annex C Example of assessing measurement uncertainty of error of indication

Consider the case of incorporating measurement uncertainty into the decision process (pass/fail) for type evaluation testing of a pressure measuring instrument that utilizes a pressure transducer.

Following the steps given in Clause 9:

(Step 1) Describe the instrument under test (IUT), along with the measuring system that will be used for performing the test(s). Include in the description all influence quantities that can effect the measuring instrument, all influence quantities that can effect the measuring system, and specify the conditions (if any) at which the influence quantities will be maintained during the testing, or the range(s) that the influence quantities must remain within during the testing (e.g., rated operating conditions and/or reference operating conditions of both the measuring system and IUT).

Type Evaluation Testing of a Pressure Measuring Instrument that utilizes a Pressure Transducer

Description of Test System

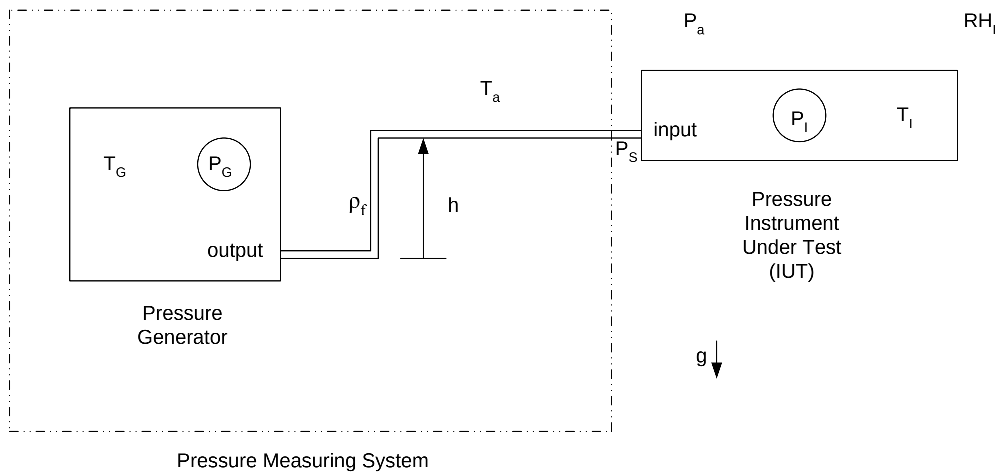


Figure C1

The instrument under test (IUT) is a pressure measuring instrument that utilizes a pressure transducer that, for sake of illustration, will be considered to be configured in the so-called 'gage mode,' meaning that one side of the transducer is open to ambient (atmospheric) pressure (denoted in Figure C_1 by P_a).

The IUT is located such that it either sits on a bench that is open to the atmosphere (as in Figure C1), or is placed in a chamber where the temperature and relative humidity can be controlled. The temperature of the IUT is indicated as T_I, and the relative humidity is indicated as RH_I. The input to the IUT is indicated in the figure, and this establishes the reference level of the IUT with respect to the indicated gage-mode pressure P_I.

The pressure measuring system is indicated by the dotted rectangle, and consists of a pressure generator and rigid tubing that connects the output of the pressure generator to the input of the IUT. The operating fluid (which must be specified) is known to have a mass density denoted by ρ_f, and the height of the reference level of the IUT above the height of the pressure generator is denoted by h (even if the pressure generator sits on the same bench, the two reference levels will likely be different). The gage mode pressure generated by the pressure generator at its reference level is denoted by P_G, and the temperature of the pressure generator is denoted by T_G, which might be different than the temperature of the ambient air, denoted by T_a. The local acceleration of gravity at the test location is given as g.

The influence quantities that can affect the outcome of the test are then P_a, T_a, T_G, RH_I and T_I. The first three will not be controlled during any of the tests, but rather will only be measured (T_G will be monitored to make sure that the pressure generator is always operating within its rated operating conditions). On the other hand, some of the tests will involve changing (and measuring) the temperature of the IUT (T_I) and the relative humidity of the air surrounding the IUT (RH_I).

The other test parameters h, g and ρ_f are not considered as influence quantities since they do not affect the IUT (or pressure generator).

(Step 2) Identify all of the different kinds of tests that will need to be performed for the type evaluation. Based on the description in Step 1, develop a mathematical model of the measurement to be used for performing each kind of test. Each model must ultimately provide an expression for the 'error of indication,' and also include an expression for the standard measurement uncertainty to be associated with each measured error of indication (unless repeated measurements of error of indication are to be obtained, in which case the mean value of the error of indication is to be presented, along with an associated standard measurement uncertainty that incorporates a component obtained from the repeated measurements). Account should also be taken in the uncertainty analysis of the range of values of error of indication that could be obtained when the IUT is operating anywhere within its rated operating conditions.

The different kinds of tests that will need to be performed are given in OIML Recommendations R 101 and R 109. Included are temperature tests, humidity tests, and hysteresis tests.

The basic mathematical model (for error of indication) for all of these types of tests is based on first generating a mathematical expression for the best-estimate of the 'true' value of the hydrostatic gage pressure delivered by the pressure measuring system to the input of the IUT (this pressure is denoted by P_S in Figure C_1):

$$P_S = P_G + \left(\rho_f - \rho_a\right)\cdot g \cdot h \qquad , \quad \text{C.1}$$

where ρ_a is the density of the ambient air. The mathematical model for the error of indication (E_I) of the measuring instrument is then taken as the difference between the indicated value of the measuring instrument (P_I) and the best-estimate of the 'true' value of the hydrostatic gage pressure (P_S) delivered by the pressure measuring system to the input of the IUT:

$$E_I = P_I - P_S \qquad \text{C.2}$$

The combined standard uncertainty of an individually measured value of the error of indication is then obtained from the use of equation 10 of the GUM [1]:

$$u_{EI}^2 = u_{PI}^2 + u_{PS}^2, \qquad \text{C.3}$$

where u_{PI} incorporates only resolution limitation and 'jitter' of the indication of the IUT, and

$$u_{PS}^2 = \sum_i \left(\frac{\partial P_S}{\partial x_i} \right)^2 u_{x_i}^2. \qquad \text{C.4}$$

The summation over the index i covers all of the quantities upon which P_S depends. (Note that equations C.3 and C.4 are based on the assumption that there is no correlation among the quantities. If such correlation exists, equation 13 of the GUM must be used.) From equations C.1 and C.4:

$$u_{PS}^2 = u_{PG}^2 + (g \cdot h)^2 u_{\rho_f}^2 + (g \cdot h)^2 u_{\rho_a}^2 + \left[(\rho_f - \rho_a) \cdot h \right]^2 u_g^2 + \left[(\rho_f - \rho_a) \cdot g \right]^2 u_h^2 \ , \qquad \text{C.5}$$

where the individual components of measurement uncertainty must be obtained from various sources, such as tables or calibration certificates. (Note that u_{ρ_a} itself depends on the temperature and relative humidity of the air.) Equation C.5 can then be combined with equation 3 to obtain an expression for the combined standard uncertainty to associate with an individually measured value of the error of indication.

However, for each type of test for the type evaluation, it is necessary to also incorporate a component of measurement uncertainty for the repeatability of the test (denoted u_{rep}). This can be obtained by performing a series of repeated 'identical' measurements and calculating the standard deviation of the measured values, or by obtaining such information from measurements that were performed earlier (the method used should be specified).

Also, the IUT should be evaluated to determine how the indication changes (for a fixed input) as the instrument is subjected to likely simultaneous changes in its operating conditions during field use. A component of uncertainty (denoted u_{roc}), perhaps obtained as the standard deviation of a set of values obtained as the operating conditions of the

IUT are randomly varied over the range of rated operating conditions, should also be considered for inclusion in the final expression for u_{EI}:

$$u_{EI}^2 = u_{PI}^2 + u_{PS}^2 + u_{rep}^2 + u_{roc}^2, \qquad \text{C.6}$$

where u_{PS}^2 is obtained from equation C.5.

Consider a particular type evaluation test where the IUT is operated at its nominal maximum rated operating pressure of 1.01 MPa (10 atmospheres). Let the pressure generator be set to generate a pressure (P_G) of 1.0000 MPa, with an uncertainty (u_{PG}), as given from its calibration certificate, of 0.0001 MPa (or 100 Pa).

The operating fluid is a liquid with a mass density (as given by the manufacturer) of 900 kg/m^3 and a corresponding stated measurement uncertainty (u_{ρ_f}) of 10%, or 90 kg/m^3.

The ambient air density (ρ_a) depends on the air temperature (T_a) [measured to be 23 ºC, with an uncertainty of 0.01 ºC], the atmospheric pressure (P_a) [measured to be 0.10147 MPa, with an uncertainty of 0.00010 MPa], and the relative humidity (RH_I) [measured to be 60%, with an uncertainty of 5%]. Using known equations for calculating air density, ρ_a is calculated to be 1.194 kg/m^3, with an uncertainty of 0.005 kg/m^3.

As the total variation in the local acceleration of gravity (g) over the surface of the Earth can be as much as 0.5%, the value of the local gravity needs to be established with an uncertainty appropriate for this use. Tables accounting for latitude and height above sea level are available. For this particular test, g is obtained from such a table to be 9.79560 m/s^2, with an uncertainty (u_g) of 0.00005 m/s^2.

The height (h) of the reference level of the IUT above the reference level of the pressure generator is measured to be 0.0213 m, with a measurement uncertainty (u_h) of 0.0001 m.

- *(Step 3) calculate the associated standard measurement uncertainty (u_{PS}) of the measurement standard or system;*

The standard measurement uncertainty (u_{PS}) of the pressure delivered by the measurement system to the input of the IUT can be calculated using equation C.5 as

$$u_{PS}^2 = (100)^2 + (9.79560 \cdot 0.0213)^2 (90)^2 + (9.79560 \cdot 0.0213)^2 (0.005)^2 + \left[(900 - 1.194) \cdot 0.0213\right]^2 (.00005)^2 + \left[(900 - 1.194) \cdot 9.79560\right]^2 .0001^2, \qquad \text{C.7}$$

$$= [10^4 + 352.62 + 1.09\text{x}10^{-6} + 9.16\text{x}10^{-7} + 0.775]\ \text{Pa}^2$$

$$\approx 10{,}353\ \text{Pa}^2, \qquad \text{or}$$

$u_{PS} \approx 102$ Pa

It can be seen immediately from this analysis that the uncertainty in the value of the generated pressure dominates the total uncertainty of the pressure delivered to the input of the IUT, followed next by the uncertainty of the density of the operating fluid. Such an analysis helps to identify where efforts could be best spent, if necessary, trying to reduce the uncertainty of the pressure delivered to the input of the IUT.

- *(Step 4) calculate a standard measurement uncertainty (u_{PI}) associated with the indicated value of the measurand (including components due to indicator resolution and/or random fluctuations);*

Observed random fluctuations (jitter) in the indicated pressure (P_I) of the IUT for a fixed input pressure of 1.01 MPa, and for the operating conditions of the IUT maintained under specified reference conditions, are found to be ± 15 Pa, which translates into a component of uncertainty of u_{PI} of $15/\sqrt{3} = 8.7$ Pa.

The resolution of the indication is found to be ± 5 Pa, which yields a component of uncertainty of u_{PI} of $5/\sqrt{3} = 2.9$ Pa.

The combined standard uncertainty associated with the indication of the IUT is then

$$u_{PI} = \sqrt{[(15/\sqrt{3})^2 + (5/\sqrt{3})^2]} = 9.13 \text{ Pa.}$$

- *(Step 5) calculate a standard measurement uncertainty (u_{rep}) associated with the repeatability or reproducibility of the measuring instrument/system and/or testing procedure;*

A series of repeatability tests is performed on the IUT, where the repeatability condition is that the pressure from the pressure generator is alternately applied and then removed fifty times, everything else remaining constant. Sufficient time is left between pressurizations to allow for thermal equilibrium to be established. Effects due to possible hysteresis are also analyzed. The calculated standard deviation of the fifty values (u_{SD}) is then taken as component of measurement uncertainty to attribute to repeatability/reproducibility for this particular type of test. For purposes of example, assume that a value of $u_{SD} = u_{rep} = 20$ Pa is calculated.

- *(Step 6) calculate a standard measurement uncertainty (u_{roc}) if the indication of the measuring instrument is found to vary when the instrument is operated over its range of rated operating conditions for a fixed input to the instrument;*

Returning to the test conditions in Step 4, now systematically vary (if possible) the operating conditions of the IUT over its range of rated operating conditions, and observe the corresponding variation in the indicated pressure P_I. Again if possible, vary the operating conditions both individually and also all at once, to simulate possible conditions that the IUT could experience in a field environment. Say that for such a test the indicated pressure is found to vary by ± 30 Pa. The corresponding component of measurement uncertainty (u_{roc}) due to (possible) variation in operating conditions over the range of rated operating conditions is then:

$$u_{roc} = 30/\sqrt{3} = 17.3 \text{ Pa}.$$

- *(Step 7) combine these components of measurement uncertainty in order to calculate a combined standard measurement uncertainty (u_{EI}) associated with the error of indication.*

It is now possible to calculate the combined standard measurement uncertainty of the error of indication (u_{EI}) for the particular type evaluation test where the IUT is operated at its nominal maximum rated operating pressure of 1.01 MPa (10 atmospheres), as described above. Using equation C.6:

$$u_{EI}^2 = u_{PI}^2 + u_{PS}^2 + u_{rep}^2 + u_{roc}^2,$$

$$= (9.13)^2 + (102)^2 + (20)^2 + (17.3)^2 = 11{,}187 \text{ Pa}^2 \text{ , or}$$

$$u_{EI} = 105.8 \text{ Pa} \qquad \text{C.8}$$

While this example shows virtually everything that is necessary to consider in order to assess the measurement uncertainty of the error of indication for an individual measurement for this type of measuring instrument (and so may appear somewhat complex) , it is important to note that, once all of this derivation has been performed, and values and associated measurement uncertainties are obtained for typical measurement conditions, the process of obtaining a value of u_{EI} for each subsequent individual pressure measurement performed during a given type evaluation test should become relatively straightforward, since most components of measurement uncertainty will not change from one individual measurement to another.

It is in fact interesting to note that the uncertainty of the error of indication presented in equation C.8 was obtained without ever obtaining an explicit value for an individual error of indication, but rather only a nominal (maximum) pressure value was specified for the test. While some of the components of measurement uncertainty may decrease at lower pressures, it is sometimes convenient to just (conservatively) use what is believed to be the maximum uncertainty throughout the testing for a particular type of test.

It is also interesting to note that, in this case, almost all of the uncertainty in the error of indication comes from the measurement standard (i.e., the pressure generator). This is not always the case.

Now that how to assess the measurement uncertainty of the error of indication for the test arrangement in this example has been presented, it is now possible to extend the example to consider how to establish requirements on MPEs, maximum permissible uncertainties and risk options to be considered for making conformity decisions. This will be done in Annex D.

Annex D Example of risk assessment incorporating measurement uncertainty

For each kind of test identified, the OIML Recommendation should discuss and specify what the appropriate MPE is for that kind of test. For example, for a type evaluation test, the MPE that is specified could correspond to one of several possible accuracy classes that the instrument is being tested for. For a verification test, the specified MPE should be based on a variety of considerations, as discussed in Clause 6.

There should also be discussion of what the likely values of u_{EI} and u_S will be during the test, in order to decide whether values of MPU_{EI} and MPU_S should be specified and, if so, what those values should be (or, rather, what f_{EI} and f_S should be). See Clauses 5.3.4, 5.3.5 and 6.

Continuing with the example from Annex C, consider the case where the IUT is to be tested to determine if that type of instrument can be classified as belonging to a specified accuracy class (say, class 0.06 as specified in OIML R 109), that has a corresponding MPE that is designated as $MPE_{.06}$ = 0.06% (1 MPa) = 0.0006 MPa = 600 Pa.

An analysis must be performed of whether, for the type of test covered in Annex C on this type of instrument, it is most appropriate to use consumer's risk, producer's risk or shared risk. Things to be considered in the analysis are what the consequences would be (safety, economic and otherwise) to the instrument user and instrument manufacturer of an incorrect pass-fail decision (for either the specified, or likeliest, use of the type of instrument), and what likely values are of u_{EI} during the test.

For example, if the type of IUT is typically used to monitor atmospheric pressure for weather forecasting, it might be decided that the shared risk approach is adequate, as long as a specified value of f_{EI} (such as 1/3) is adhered to. On the other hand, for a type of pressure measuring instrument being used to monitor critical vessel pressure in a nuclear power plant, or being used for aviation altimetry, the consumer's risk approach should probably be used, with a relatively conservative (smaller) f_{EI}.

Before deciding on which risk approach to use, it might be necessary (or at least useful) to first perform some preliminary measurements to determine typical values of u_{EI} (which has already been established in equation C.8 in Annex C as being around 105 Pa). These measurements could also be used to help establish an appropriate specified value of f_{EI} such that there would be a very small probability of an incorrect pass-fail decision.

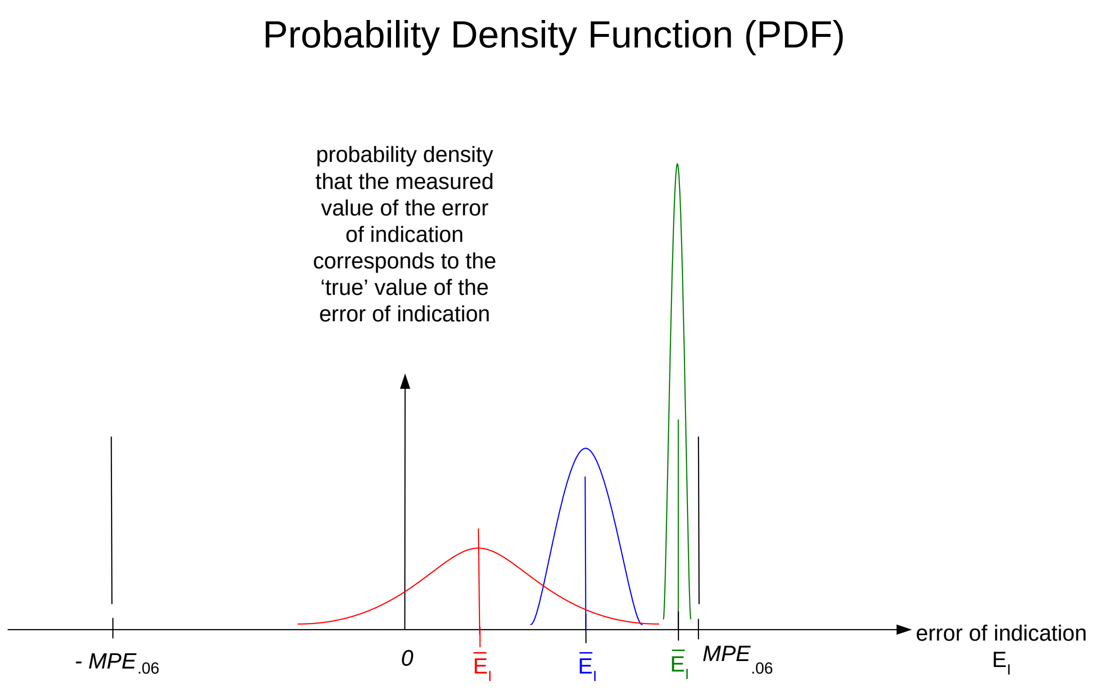


Figure D1

Figure D1 illustrates the situation for the example being discussed. The middle (blue) curve represents a Gaussian PDF where the uncertainty (standard deviation of the curve) is about 1/6 of the MPE (u_{EI} /$MPE_{.06}$ = 105/600). The leftmost (red) curve represents a Gaussian PDF where the uncertainty is about 1/3 of the MPE. By examining these two curves it is then possible, on just a visual basis, to decide a level of comfort with which either the ratios (corresponding to f_{EI}, as discussed in clause 5.3.4), or a different ratio, should be specified as a requirement in the Recommendation. For the particular example being discussed, assume that the type of IUT will be used in a non-critical application, and so an f_{EI} of 1/3 is considered to be acceptable. For a critical application, an f_{EI} of 1/20, as indicated schematically by the rightmost (green) curve in Figure D1, might be more appropriate. In this latter case, in order to achieve this smaller value of f_{EI}, it would be necessary to either reduce u_{EI}, or choose a larger MPE (accuracy class) for this type of instrument to belong to.

Turning to requirements on the measurement standard, values of u_{PS} can be obtained from an analysis of the measuring system, including incorporating information contained in the calibration certificate of the measurement standard, in order to help decide whether the measurement standard and measuring system are appropriate to be used for the particular kind of test. This aspect of the test should be discussed and specified in the Recommendation (e.g., an appropriate value of f_S should possibly be specified, as discussed in Clause 5.3.5). For the pressure example being discussed, the uncertainty associated with the pressure delivered and measured by the 'standard' is given in equation C.7 as u_{PS} = 102 Pa, which is only slightly less than u_{EI}, and so the middle curve in Figure D1 can again be used for deciding (on a visual basis) whether the uncertainty of the standard is acceptable. In this case the decision to be made is whether the uncertainty due to the standard unfairly affects the pass-fail test decision from the manufacturer's point of view, in that most of the uncertainty is due to the measurement standard and not the IUT. For the particular example being discussed, a required value of $f_S = 1/3$ would be acceptable (since the measured value is 1/6).

OIML Secretariats and TC/SC members should consider whether 'acceptable' levels of risk for various types of tests should be suggested in their OIML Recommendations. Decision rules and associated risks, along with their consequences, should be considered and discussed in OIML Recommendations. However, this should be done only in the context of regulatory matters. Risks to a manufacturer may have serious economic consequences that are typically outside the scope of a Recommendation.

Depending on the values of MPU_{EI} and MPU_S (or f_{EI} and f_{PS}) specified in the prior step (if any), discussion should be provided on whether the 'shared risk' principle is to be used, or whether there is a specified risk (probability) that is to be used and, if so, whether it is a Risk of False Acceptance or a Risk of False Rejection. Note that if the 'shared risk' approach is used in an OIML Recommendation (or in other OIML documents), it should not be used in an implicit manner but, rather, an explicit statement of its use should be provided in the Recommendation.

Continuing further with the example from Annex C, next consider the case where the IUT is to be tested for initial verification requirements. In this case, an MPE for initial verification (MPE_{iv}) is to be specified in the Recommendation, and so the Recommendation should discuss the various considerations that go into choosing an appropriate MPE_{iv}, such as needs of the regulator and consumer, and achievable levels of operation of the instrument in a 'field' environment.

As was the case for the type evaluation test, the question of what type of risk and decision rules to use for initial verification must be analyzed, only with now a (typically) larger MPE (it is frequently the case that the MPE_{iv} is chosen to be twice the MPE, however this is not always necessary), and so the answer to the question might be different. For example, for the type evaluation test it might be decided that using consumer's risk is appropriate, along with a specified value of f_{EI}, whereas for the initial (or subsequent) verification test, the use of shared risk (which is easier to handle in the field) is adequate, since, with a larger MPE, the PDF might now look more like the rightmost curve in Figure D1, rather than like the middle curve. In such a case it makes sense to avoid computational complication and share the risk, since an 'incorrect' decision could be made only over the relative width (which is very small) of the rightmost curve.

If Risk of False Acceptance or Risk of False Rejection is used, it is further necessary to specify whether u_{EI} is to be considered as fixed for each measurement, in which case a guard band can be used for deciding conformity, or whether u_{EI} is to be calculated separately for each measurement of error of indication, in which case the z-statistic or Measurement Capability Index can be used. Reference to the OIML Guide(?) G YY on "The Role of Measurement Uncertainty in Conformity Assessment Decisions in Legal Metrology" should be provided, along with discussion of how to use the z-statistic and/or Measurement Capability Index for the particular Recommendation.

Constructing PDFs and calculating areas under a PDF curve is in general a nontrivial matter, and so OIML Secretariats and TC/SC members should consider what advice and assistance to provide in this regards in their Recommendation(s) (e.g., use of the z-statistic or numerical techniques).

For the situation where it is decided that the risk of false acceptance approach is to be used, a decision must be made concerning what is the acceptable level of risk for false acceptance (p_{ca}, see Clause 5.3.1), and a further analysis must be performed about whether the uncertainty of the error of indication can be taken as constant for each measurement, or whether it is necessary to recalculate it each time.

If u_{EI} needs to be calculated each time, then it is necessary to either use the z-table each time (e.g., see Annex B in the OIML Guide(?) G YY on "The Role of Measurement Uncertainty in Conformity Assessment Decisions in Legal Metrology"), or to calculate the Measurement Capability Index (C_M) each time (e.g., see Annex E in the OIML Guide(?) G YY on "The Role of Measurement Uncertainty in Conformity Assessment Decisions in Legal Metrology") and use the corresponding C_M table each time.

If u_{EI} can be considered as a constant for a given type of measurement, and so does not need to be calculated each time, then a guard band can be constructed by shifting the MPE boundaries inward by a fixed amount (so as to keep the probability of false acceptance less than a specified value; see [7]). Pass-fail decisions are then made on the basis of whether the measured E_I lies within the new (reduced) MPE boundaries.

Returning to the type evaluation test for the example in Annex C (and above), assume that it is decided that a 5% level of risk of false acceptance (consumer's risk) will be used for this application of the IUT (i.e., $p_{ca} = .05$, and thus the probability of conformance is $p_c = 0.95 = 95\%$). Since for this example it has been determined that $MPE_{.06} = 600$ Pa and u_{EI} (at maximum pressure of 1 MPa) = 105 Pa, the standard normal distribution table in Annex B can be used to determine the maximum value of the error of indication. Begin by locating the entry in that table for 0.9500 (or actually for 0.4500, since 0.5000 needs to be subtracted from 0.9500 in this case for the table in Annex B), which is between the entries 0.4495 ($\alpha = 1.64$) and .4505 ($\alpha = 1.65$). Using interpolation, the value of α that will be used is then 1.645. Equation B.1 can then be used, in a slightly rearranged form:

$$\bar{E}_I = MPE_{.06} - (u_{EI} \cdot \alpha) \qquad D.1$$

to obtain $\bar{E}_I$ = 425 Pa , which is the maximum value that $\bar{E}_I$ can have where there is no greater than a 5% risk that test should have been considered to fail even though it is considered to pass. This situation is demonstrated graphically in Figure D2.

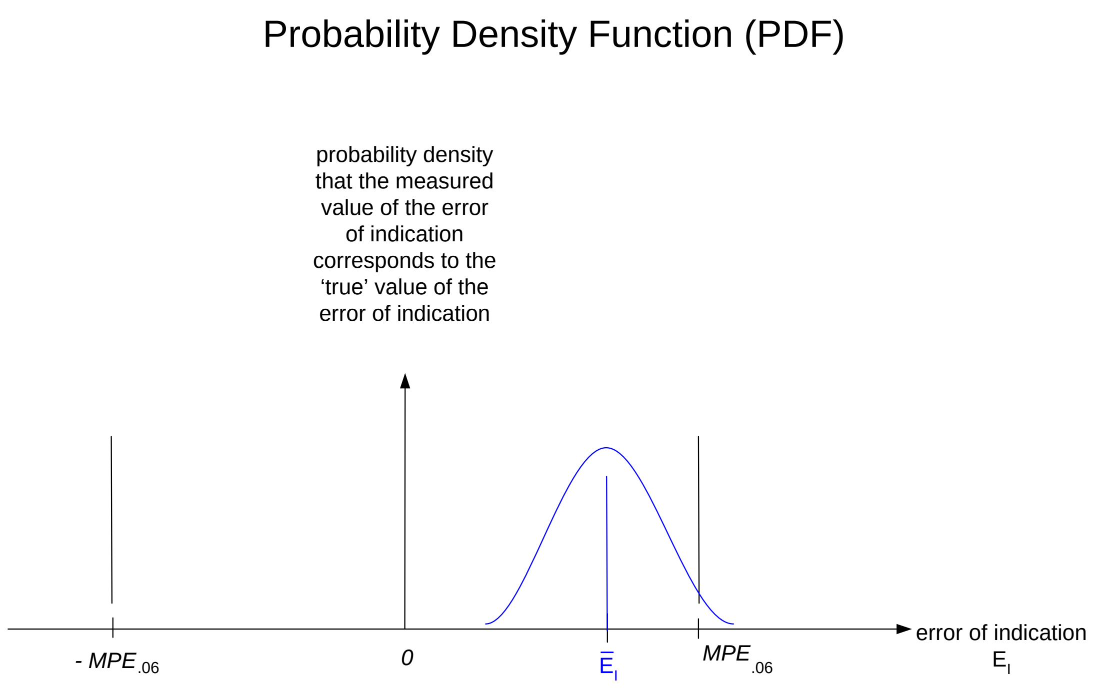


Figure D2

Rather than using the z-table, it may be more convenient to use the measurement capability index chart to arrive at this same conclusion (see Annex E). In this case, the measurement capability index is first calculated using equation E.1 as $C_M = MPE/[2 \cdot u_{EI}]$ = 600/[2·105] = 2.86. Using the 95% chart in Annex E, the corresponding value of $\hat{E}$ is about 0.85. Rearranging equation E.2, $E_I = MPE\,(2 \cdot \hat{E} - 1) = 600\,(1.7 - 1) = 420$ Pa, which is close to the 425 Pa obtained when using the more precise z-table.

While assessing the measurement uncertainty of the error of indication for an individual measurement for a specified type of measuring instrument may be somewhat complex, it is important to note that, once all of the derivation has been performed, and values and associated measurement uncertainties are obtained for typical measurement conditions, the process of obtaining a value of u_{EI} *for each subsequent individual measurement performed during a given type evaluation test should become relatively straightforward, since most components of measurement uncertainty will not change from one individual* *measurement to another. This aspect of the treatment of measurement uncertainty should be included in the discussion in each OIML Recommendation where measurement uncertainty is relevant.*

If it is determined experimentally that there is significant variation in u_{EI} from one measurement to the next, then it will be necessary to use either the z-table or measurement capability index for each measurement of E_I. However, as indicated earlier, it is unlikely that u_{EI} will vary appreciably for each measurement and, besides, it is sometimes more convenient to take a conservative approach and treat the u_{EI} determined in Annex C as the likely upper bound of all of the u_{EI}'s, and so treat it as a constant. In this case, a guard band can be created (where the new MPE is moved inward from 600 Pa to 425 Pa) and the decision making becomes much simpler, where tests involving measured values of E_I less than 425 Pa are accepted, and those greater are rejected. This guard band approach is illustrated in Figure D3.

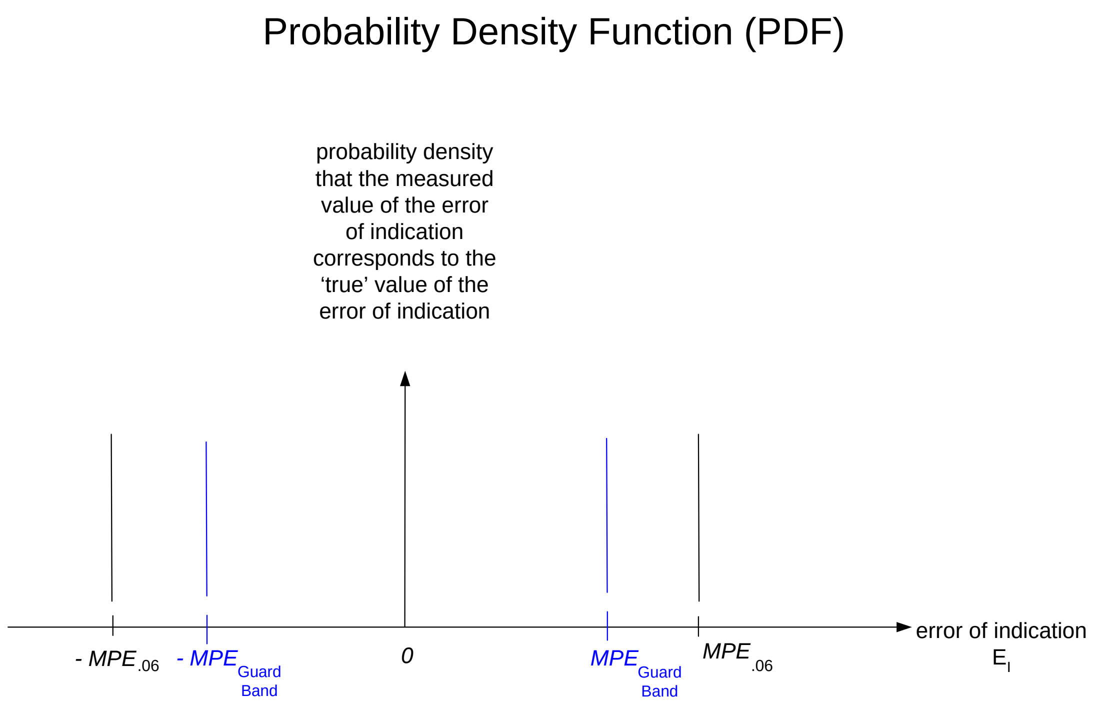


Figure D3

Annex E Measurement Capability Index (C_M)

The "measurement capability index," defined and discussed in [7], is a useful tool for quickly assessing whether a measured error of indication (E_I), with associated combined standard uncertainty (u_{EI}), is considered to conform to the maximum permissible error (MPE) requirement within a specified conformance probability (p_c).

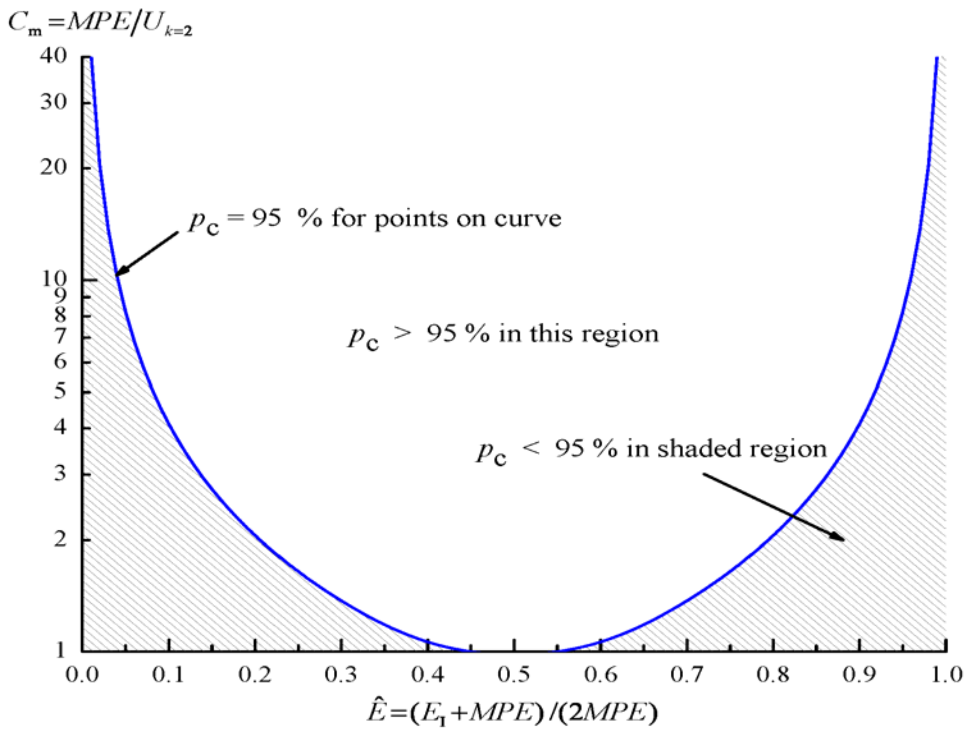
The measurement capability index is dimensionless, and defined for legal metrology as:

$$C_m = MPE/[2 \cdot u_{EI}] = MPE/U_{k=2} \qquad \text{E.1}$$

In order to use the measurement capability index, it is first necessary to calculate another dimensionless parameter Ê, defined as:

$$\hat{E} = [E_I + MPE]/[2\ MPE] \qquad \text{E.2}$$

(Note that for $-MPE<E_I<MPE$, then $0<\hat{E}<1$.) A chart such as the one below can then be constructed for a given p_c (shown here for $p_c = 95\%$), where the intersection of Ê and C_M can be found to see if it lies in the shaded region (test fails) or un-shaded region (test passes). (Figure courtesy of W. Tyler Estler).



Annex F Establishing measurement uncertainty to use with conformity tested measuring instruments/systems

Once a measuring instrument has passed an initial or subsequent verification test, it is sometimes used to perform a measurement where it is required that the measured value is accompanied by its associated measurement uncertainty. In such a situation, unless the instrument was not only tested but also calibrated, all that can be said about any measured value obtained when using the instrument is that the ‘true’ value of the measurand is believed to be best represented by the measured value (as given by the indication of the measuring instrument), but that the ‘true’ value could lie anywhere (with equal probability) in the range given by the measured value, plus or minus the MPE. This is the so-called ‘rectangular probability distribution,’ and is treated in 4.4.5 of the GUM [1].

According to that analysis, the measurement uncertainty that should be associated with the measured (indicated) value is

$$u = MPE / \sqrt{3} \quad , \qquad \text{F.1}$$

where MPE is the maximum permissible error that was used when the measuring instrument was tested.