Evolution of philosophy and description of measurement (preliminary rationale for VIM3)

Charles Ehrlich · René Dybkaer · Wolfgang Wöger

Received: 3 December 2006 / Accepted: 15 January 2007 / Published online: 1 March 2007
© Publication of the United States Government, not subject to copyright 2007

Abstract Different approaches to the philosophy and description of measurement have evolved over time, and they are still evolving. There is not always a clear demarcation between approaches, but rather a blending of concepts and terminologies from one approach to another. This sometimes causes confusion when trying to ascertain the objective of measurement in the different approaches, since the same term may be used to describe different concepts in the different approaches. Important examples include the terms “value,” “true value,” “error,” “probability” and “uncertainty.” Constructing a single vocabulary of metrology that is able to unambiguously encompass and harmonize all of the approaches is therefore

difficult, if not impossible. This paper examines the evolution of common philosophies and ways of describing measurement. Some of the differences between these approaches are highlighted, which provides a rationale for the entries and structure of the August 2006 draft of the 3rd Edition of the *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (VIM3) [1].

Introduction

The concept of measurement covers a wide range of activities and purposes. Different approaches to describing and characterizing measurement have been developed and have evolved to address the various types and uses of measurements, and they are still evolving. Many terms have been used over time in the context of describing measurement, and the evolution of the different approaches to measurement has led to sometimes subtle, but undoubtedly different, uses of some terms.

A “vocabulary” is defined (e.g., ISO 1087-1) as “terminological dictionary that contains designations and definitions from one or more specific subject fields”. Ideally, every term in a vocabulary should designate only one concept, in order to minimize confusion. However, because of the different concepts that are sometimes associated with the same term in the different approaches to measurement, it is virtually impossible to create a vocabulary of measurement that designates only one concept with each term in the vocabulary. This is a major difficulty that has been encountered in developing the 3rd Edition of the *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (VIM3) [1], where

Due to its outstanding importance and in order to reach as many members of the measurement community as possible, this article appears also in NCSL International’s journal ‘Measure’ with the consent of the authors and the editors.

Presented at NCSLI Workshop and Symposium, August, 2006, Nashville, TN, USA

C. Ehrlich (✉)
International Legal Metrology Group,
National Institute of Standards and Technology,
Gaithersburg, MD 20899-2600, USA
e-mail: charles.ehrlich@nist.gov

R. Dybkaer
Department of Standardization in Laboratory Medicine,
Region H: Frederiksberg Hospital,
Copenhagen University Hospital,
Copenhagen, Denmark

W. Wöger
Drosselweg 1, 50735 Cologne, Germany

“metrology” is defined as “field of knowledge concerned with measurement”.

This paper examines the evolution of the more common approaches to describing measurement, highlighting a few of the differences in the use of terms, and providing some of the rationale for how several of the terms are likely to be treated in the final version of VIM3.

In the text, concepts are mostly identified by their full systematic preferred terms of VIM3. In the figures, for convenience, a shortened form, also given in VIM3, is used.

Common elements of most approaches to measurement

There are a few fundamental concepts in most, if not all, approaches to describing measurement. Probably the most fundamental concept pertains to the kinds of things that can be measured, i.e. quantities. Another fundamental concept is the means used to express the magnitude of that which has been measured (in terms of values). Just as fundamental is the concept of measurement itself. The following definitions are taken from the August 2006 draft of the VIM3:

- **Quantity** is “property of a phenomenon, body, or substance, to which a number can be assigned with respect to a reference” (which allows comparison with other quantities of the same kind).
- **Quantity value** (value of a quantity) is “number and reference together expressing magnitude of a quantity”.
- **Measurement** is “process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity”.

In VIM3 the concept **measurand** is defined as “quantity intended to be measured”. This concept has ‘evolved’ from the definition in the *International Vocabulary of Basic and General Terms in Metrology*, 2nd Edition [2], VIM2, which is “particular quantity subject to measurement”, that could be different from the quantity intended to be measured. The distinction must be kept in mind when considering the objective of measurement in the different approaches; this will be discussed further later on.

Figure 1 demonstrates some simple common elements of all approaches to describing measurement. The rectangular box gives the VIM3 definition of “measurand”, and the horizontal scale represents the entire set of values that could possibly be attributed to that type of measurand. Note that there is no measurement unit associated with the horizontal line, because the quantity is an **ordinal quantity**, which is “quantity, defined by a conventional measurement procedure, for which a total ordering relation, according to magnitude, with other quantities of the same kind is defined, but for which no algebraic operations

among those quantities are defined”. Due to the latter characteristic, an average of a set of replicate measurements, illustrated schematically by a histogram, has no meaning.

For those quantities where there are meaningful algebraic operations among the quantities, a **measurement unit** can be defined as “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”. This is indicated in Fig. 2, where the measurement unit is the reference to be associated with the numerical value in the measured quantity value. The concept of measurement unit is common to all approaches to describing measurement (for other than ordinal quantities). The bell curve in Fig. 2 illustrates a ‘Gaussian’ fit to the histogram data. The curve is dashed to indicate that replicate measurements are not always performed in a measurement (that is, sometimes only a single measurement is performed), as will be elaborated below in the discussion of the International Electrotechnical Commission (IEC) Approach.

The two main approaches to describing measurement that will be discussed in this paper are sometimes called the ‘classical’ approach and the ‘uncertainty’ approach. Within each of these approaches are sub-approaches. While the two main approaches are given discrete names, there has in actuality been an evolution of these approaches that makes it difficult to ascribe certain concepts to one approach or another. This evolution of concepts is discussed below. Also, since probability and statistics usually play an important role in most aspects of measurement evaluation, both the ‘frequentist’ and ‘Bayesian’ theories of inference, as used in measurement, will be discussed as appropriate.

Classical approach to measurement

It is generally accepted that the key distinguishing premise of the classical approach to measurement is that, for a specified measurand, there exists a unique value, called the **true value**, that is consistent with the definition of the measurand. This is shown schematically in Fig. 3, where it is indicated that, in the general case, the value being attributed to the measurand based on measurement is different from the true value. This difference could be due to a variety of reasons, including mistakes in formulating the measurement model (such as not taking into consideration all significant factors and influences), and blunders in carrying out the measurement procedure.

Another premise of the classical approach is that it is possible to determine the true value of a measurand through measurement, at least in principle, if a ‘perfect’ measurement were performed. The objective of measurement in the

Fig. 1 Common elements of philosophies and descriptions of measurement 1 **Fig. 2** Common elements of philosophies and descriptions of measurement 2

classical approach is then usually considered to be to determine an estimate for the true value of the measurand as ‘closely’ as possible, or at least as closely as necessary, both by eliminating or correcting for all (known) systematic errors and mistakes, and by performing enough repeated measurements to adequately minimize errors due to random causes.

In the classical approach it is recognized that it is not possible to perform a ‘perfect’ measurement and so there will remain errors, both systematic and random, in the value ultimately being attributed to the measurand based on measurement. This value, frequently referred to as the

‘measurement result’ or sometimes the ‘final measurement result’ in the classical approach, and in other approaches as well, is often obtained as the average measured value, as illustrated in Figure 4. Figure 4 also illustrates the concept of an individual measurement error, defined in the classical approach as the difference between an individual measurement result and the true value. The individual measurement result (‘individual measured value’, denoted by y_i in Fig. 4) is illustrated with respect to the bell-curve, which is now solid to indicate that multiple individual measurements are being considered. Also indicated in Fig. 4 are “systematic error”, defined as the difference between the

Fig. 3 Classical approach to measurement 1 **Fig. 4** Classical approach to measurement 2

unknown mean of the uncorrected measurement results and the true value, and “random error,” defined as the difference between an individual measurement result and the unknown mean of the uncorrected measurement results. Note that the “mean of the uncorrected measurement results” here is meant to be that of a distribution of relative frequencies of measurement results obtained by repeating an experiment infinitely often, always under the same conditions. Thus, in reality, the mean cannot be known exactly. This is illustrated schematically in Fig. 5, where two systematic errors are shown, the lower one (systematic

error_b) with respect to the average of the histogram data, and the upper one (systematic error_a) with respect to the mean of the theoretical frequency distribution for an ‘infinite’ set of data. The bell curve of the theoretical frequency distribution is dashed to indicate that it is not knowable. The systematic error_a line is also dashed to indicate that its length cannot be known, since the mean of the theoretical frequency distribution cannot be known. The question of whether or not the length of the systematic measurement error_b line can be known, as well as the lengths of the three ‘error lines’ in Fig. 4, will be discussed next.

Fig. 5 Classical approach to measurement 3

Knowable error

Two important and related questions that arise in the classical approach are, first, whether it is possible, in principle, to go about identifying and eliminating, or correcting for, all of the errors in a measurement, and, second, if so, how? One possible way of addressing these questions is to hypothesize that it is possible, at least in principle, to determine the true value by carrying out a very large number of different types of measurements of the same measurand, using different measurement procedures, measurement methods or even measurement principles, a large number of times (so that various systematic errors will 'average out'). This would require that a lot of information be obtained through measurement (which may not always be practical, even if the philosophy is sound).

Figure 6 illustrates this idea for just two different measurement principles, and Figure 7 is meant to illustrate the advantage of using multiple measurement principles (indicated by the four different curves). Using this idea in the classical approach, a probability is usually assessed that the true value lies within a stated interval, as could be characterized by the 'width' of the large bell-shaped curve associated with the true value in both Figs. 6 and 7. Since this idea requires that an essentially infinite amount of information be obtained in order to know the true value, it is recognized that, in practice, a true value can never be known exactly using this idea. This is represented schematically in the two figures, where y -double-bar represents the average of the averages of the individual curves in the respective figures.

The questions then remain first, whether it is possible, in principle, in a different way, to identify and correct for all of the errors in a measurement, and, second, if so, how?

Error analysis, frequentist theory in classical approach

One different way of trying to answer these questions is through the application of error analysis, which is based on the frequentist theory of inference as used in measurement. Error analysis is the attempt to estimate the total error using frequency-based statistics. However, the systematic error cannot be estimated in a statistical way, since it is neither observable nor behaves randomly in a measurement series under repeatability conditions. Therefore error analysis, which includes statistical and nonstatistical procedures, leads to inconsistencies in data analysis, especially in error propagation.

Bayesian theory in classical approach

Another way of trying to answer these questions is to apply the Bayesian theory of inference to data analysis. Here systematic and random errors are treated on the same probabilistic basis, where probability is no longer understood as a relative frequency of the occurrence of events, but as an information-based degree of belief about the truth of a proposition, for example, about the true value. Using the Bayesian theory, it is still not possible to determine a true value unless an essentially infinite amount of information is obtained, so that it is again recognized that, in practice, a true value cannot be known.

Difficulties with the classical approach

So far no satisfactory way has been found to identify, let alone correct for, all of the errors in a measurement. The implications are significant, as illustrated in Fig. 8, where a hypothetical three 'known' components of systematic error

Fig. 6 Use of two measurement principles **Fig. 7** Use of multiple procedures, methods and principles

are shown (usually estimated as 'worst-cases'). Since it is virtually impossible to know for sure if there is another component (say, due to a blunder, as indicated by the dashed line), the 'total' systematic error is unknown, as also indicated by a dashed line. If the total systematic error is unknown, then the true value cannot be known. If the true value is not known, then the error cannot be known (as again indicated by a dashed line). The random error, when defined with respect to the average of the histogram data, is calculable, as indicated by the solid line in Fig. 8. However, when random error is defined with respect to the mean of the theoretical frequency distribution, it also becomes unknowable, as illustrated by the dashed line for 'random error' in Fig. 9.

Systematic and random errors can therefore typically only be estimated or guessed. No generally-accepted means for combining them into an 'overall error' exists that would provide some overall indication of how well it is thought that a measurement result corresponds to the true value of the measurand (i.e., to give some indication of how 'accurate' the measurement result is thought to be, or how 'close' the measurement result is thought to be to the true value of the measurand). The difficulty in the classical approach, of the lack of a generally-accepted, good procedure for describing the perceived 'quality' of the measurement result, is one important reason that 'modern' metrology is moving away from the philosophy and language of the classical approach. A solution to this difficulty

number of times
a possible value
that could be
attributed to
measurand is
obtained
(by replicate
measurements)

number of times
a possible value
that could be
attributed to
measurand is
obtained
(by replicate
measurements)

the language and philosophy of the classical approach is a seemingly reasonable thing to do, and many people still do it. This is one of the main reasons that it was decided to keep many of the terms and concepts from the classical approach in the main body of VIM3, and not relegate them to an Annex. Another reason, as mentioned earlier and that will be elaborated further below, is that there is not always a clear demarcation between approaches. As an example, it is not clear to which measurement approach to ascribe the

VIM3 RATIONALE: There are many measurement situations, typically of a relatively simple nature, where it is likely possible to be able to identify and correct for all of the significant systematic errors, as well as to obtain a sufficient number of replicate measurements for the purpose, such that description of the measurement result using

premise of a lack of uniqueness of a true value of a measurand.

Uniqueness of true value

Generally, a measurand cannot be completely specified (except counts with low values), meaning that there will almost always be a set of true values that are consistent with the definition of a measurand. This is illustrated schematically in Fig. 10, where the interval of the set of true values consistent with the definition of the measurand is indicated by a pair of vertical dotted lines. The corresponding range (defined as the difference between the upper and lower limit of the interval) is shown bracketing the average measured quantity value. Even if an infinite series of replicate, arbitrarily precise measurements of (different samples of) the measurand were possible, there would still be a set of measured quantity values having at a minimum that same range, since any individual measurement (sample) could have any value of the set of true values consistent with the definition of the measurand. For a real measurement situation involving random errors, the range would necessarily be greater. The bell curve illustrates such a situation, where a characteristic width of the distribution (e.g., standard deviation) of the measured quantity values would lead to a range that is broader than the range of the set of true values calculated in the same way.

It is often desirable to have a measurement situation where the measurand can be progressively better defined such that the range of the set of true values becomes rel-

atively insignificant with respect to the range of measured quantity values that can be obtained when using the (best) available measuring system, as illustrated in Fig. 11. Under these conditions, the measurand can be regarded as having an ‘essentially unique’ true value (i.e., ‘the’ true value), and the ‘customary’ language and mathematics of measurement can be employed.

However, this situation is not always found. Sometimes the measurand cannot, or needs not, be specified very narrowly. Alternatively, the measurement system is sometimes so precise that it is always capable of producing measured quantity values, illustrated in Fig. 12 by the curve, that are much narrower than the range of the set of true values for that measurand. Under these conditions it is necessary to think differently about the way of describing measurement, irrespective of the measurement approach. For example, in the classical approach, it would no longer be possible to talk about ‘the true value’ of a measurand, or ‘the systematic error’ associated with a measurement result, since such unique values would no longer have meaning. This measurement situation will also be addressed further in the discussion about the uncertainty approach.

Before leaving the discussion of the classical approach, it is worth noting that the classical approach is also sometimes called the ‘traditional approach’ or ‘true value approach.’ However, the latter is a misnomer, since the concept of true value is actually also used in ‘modern’ approaches, such as the ‘uncertainty approach,’ as will be discussed next.

Fig. 10 Non-unique true value 1

Fig. 11 Non-unique true value 2 **Fig. 12** Non-unique true value 3

Uncertainty approach to measurement

The concept of measurement uncertainty had its beginnings in addressing the difficulties described above with the classical approach, namely the questions of 1) whether it is possible, both in principle and in practice, to know the true value and error, 2) whether or not the true value is unique, and 3) how to combine information about random error and systematic error in a generally accepted way that gives

information about the overall perceived ‘quality’ of the measurement. Further, if the true value, or set of true values, is not knowable in principle, then the question arises whether the concept of true value is necessary, useful or even harmful! All of these issues and perspectives will be addressed below.

While different approaches exist within the uncertainty approach, the two most prominent approaches are those put forward in the *Guide to the Expression of Uncertainty in*

Measurement (GUM, 1993 and 1995) [3] and in IEC 60359 *Electrical and Electronic Measurement Equipment – Expression of Performance* [4]. IEC describes its approach as being parallel and complementary to the GUM, but uses a more operational or pragmatic philosophy, focusing primarily on single measurements made with measuring instruments. Both of these approaches, along with their impact on VIM3, will be described.

GUM approach to uncertainty

The GUM approach to uncertainty provides a more refined means than the classical approach for describing the perceived quality of a measurement. One of the main 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, and a means for doing this is provided. Another basic premise of the GUM approach is that it is not possible to know the true value of a measurand (see GUM Section 3.3.1): “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.” A third basic premise of the GUM approach is that it is not possible to know the error of a measurement result (see GUM 3.2.1 Note): “Error is an idealized concept and errors cannot be known exactly.”

In the GUM approach it is explicitly recognized that it is not possible to know, for sure, how ‘close’ a value obtained through measurement is to the true value of a measurand (i.e., to know the error). Instead a methodology for constructing a quantity, called the standard measurement uncertainty, is established that can be used to characterize a set of values that are thought, on a probabilistic basis, to correspond to the true value, based on the information obtained from the measurement. *The objective of measurement in the GUM approach then becomes to establish a probability density function, usually Gaussian (normal) in shape, that can be used to calculate probabilities, based on the belief that no mistakes have been made, that various values obtained through measurement actually correspond to the ‘essentially unique’ (true) value of the measurand.* Note that the GUM does not explicitly state the objective of measurement this way, but it can be inferred through its description of standard uncertainty (see, e.g., GUM 6.1.2). *Another way of viewing the objective of measurement in the GUM approach is that it is to establish an interval within which the ‘essentially unique’ (true) value of the measurand is thought to lie, with a given probability, based on the information used from the measurement.* The modifier “true” has been put in parenthesis here as an alert that the GUM discourages use of the term (but not of the concept)

“true value,” and instead treats “true value” and “value” as equivalent, and thus omits the modifier “true”. This, however, causes terminological difficulties that are treated in VIM3, and are discussed below.

VIM3 RATIONALE for measurement uncertainty. The concept of measurement uncertainty is defined in VIM3 as “parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used”. As stated above, this important concept is introduced in the uncertainty approach to provide a quantitative means of combining information arising from both random and systematic effects (if they can be distinguished at all!) in measurement into a single parameter that can be used to characterize the dispersion of the values being attributed to a measurand, based on the information used from the measurement. The VIM3 definition is modified from the VIM2 [2] (and GUM [3]) definition because of the way that the term “measurement result” has been redefined in VIM3 (see next rationale).

VIM3 RATIONALE for measurement result. The GUM uses the VIM2 definition of “measurement result” (value attributed to a measurand, obtained by measurement), which is the same as the *estimate* mentioned above. However, it was decided by the developers of VIM3 to emphasize the importance of including measurement uncertainty in reporting the outcome of a measurement by incorporating into the definition of measurement result the notion that “a complete statement of a measurement result includes information about the uncertainty of measurement,” as stated in Note 2 of the VIM2 definition of measurement result. Accordingly, measurement result is defined in VIM3 as “set of quantity values being attributed to a measurand together with any other available relevant information,” which requires information not about just a single value, but also about the measurement uncertainty. The “other available relevant information,” when available, pertains to being able to state probabilities.

VIM3 RATIONALE for measured quantity value. Since the term “measurement result” is defined in VIM3 in the more general sense given above, it was decided to introduce a separate concept for the individual quantity values of the set of values being attributed to the measurand based on measurement. Thus, in VIM3, “measured quantity value” is defined as “quantity value representing a measurement result”.

VIM3 RATIONALE for definitional uncertainty. Another basic premise of the GUM approach is that no measurand can be completely specified, as has already been discussed earlier in the context of lack of uniqueness of a true value. In the GUM approach this premise is implemented such that there is always an ‘intrinsic’ uncertainty that is the minimum uncertainty with which an incompletely defined measurand can be determined (GUM

D.3.4). Therefore, in VIM3 the term “definitional uncertainty” was coined for the concept defined by “minimum measurement uncertainty resulting from the inherently finite amount of detail in the definition of a measurand”. The implication of this concept, as discussed above, is that there is no single true value for an incompletely defined measurand. However, a very important point to remember concerning the GUM approach is that it “is primarily concerned with the expression of uncertainty in the measurement of a well-defined physical quantity – the measurand – that can be characterized by an essentially unique value” (GUM 1.2). ‘Essentially unique’ means that the definitional uncertainty can be regarded as negligible when compared with the range of the interval given by the rest of the measurement uncertainty. Therefore, when using the GUM ‘mathematical machinery’ and language, it is important to make sure that this ‘negligibility’ condition applies. If it does not, then use of different approximations and language might be required. This is elaborated further below.

VIM3 RATIONALE for true quantity value. As already noted, in the GUM approach the modifier “true” in “true value” is considered to be redundant (GUM D.3.5), and so a “true value” is just called a “value”. *It is important to recognize that this does not mean that the concept of true value is discouraged or ignored in the GUM.* Rather, the concept of “true quantity value”, defined in VIM3 as “quantity value consistent with the definition of a quantity” has only been renamed “value”, or “the value,” in the GUM. This sometimes causes serious confusion, especially since the same term “value” is also

frequently used in the GUM in the more general, super-ordinate VIM3 sense of “number and reference together expressing magnitude of a quantity”. Another reason for potential confusion is that, if a true value is unknowable, then the need for the concept can be questioned (this will also be discussed later in connection with the IEC approach). However, as discussed earlier, *in the GUM approach, the concept of true value is necessary for describing the objective of measurement. The concept of true value is also necessary for formulating a measurement model.*

The GUM approach to measurement is illustrated schematically in Fig. 13, where the objective(s) of measurement are given at the top. Note that the vertical axis is no longer the number of times that a possible quantity value that could be attributed to a measurand is obtained by replicate measurements. Rather, the vertical axis is now the probability that individual ‘estimates’ of the value of a measurand actually correspond to the (essentially unique true) value of the measurand, where probability here means degree of belief under the assumption that no mistakes have occurred. The curve is now a probability density function (PDF) that is constructed on the basis of both replicate measurements (using so-called Type A evaluation) and other information obtained during measurement, such as values obtained from reference data tables and professional experience (using so-called Type B evaluation).

The combined standard uncertainty, expanded uncertainty and coverage interval are also illustrated in Fig. 13. A **coverage interval** is defined in VIM3 as “interval

Fig. 13 GUM approach to measurement

containing the set of true quantity values of a measurand with a stated probability, based on the information available.” As indicated above, the GUM does not use the word “true” in connection with the concept of true value, and so “(essentially unique true) value” is indicated in Fig. 13. Also shown is the ‘intrinsic’ uncertainty associated with the fact that the (true) value is not unique (but only ‘essentially unique’) in the GUM Approach.

Note in Fig. 13 that the essentially unique true value is not shown to be within the coverage interval. This situation could be due to a variety of reasons, including an unidentified bias (systematic measurement error), inappropriate estimates of the values of influence quantities, or an outright blunder in conducting the measurement.

Incorporation of the terminology explained in the VIM3 rationales discussed above is illustrated schematically in Fig. 14. The objective(s) of measurement are again given at the top of Fig. 14 where the new terminology has also been incorporated. It is important to notice that nothing has changed in going from Fig. 13 to Fig. 14 other than the terminology, which is meant to emphasize that VIM3 is not intended to change the philosophy of the GUM approach, but only to clarify and possibly harmonize some of the terminology.

Figure 15 demonstrates the situation where the definitional uncertainty is not small compared to the rest of the measurement uncertainty, in which case the objective(s) of measurement are stated differently in recognition that probabilities must now be stated with respect to a set of true values, and not to an essentially unique true value.

This measurement regime, and use of probability, is not treated in the GUM. However, the GUM indicates (e.g., GUM Fig. D.2) that definitional uncertainty is to be included in the calculation of measurement uncertainty.

The PDF from Fig. 14 (solid curve) is reproduced as the solid curve in Fig. 15. A broadened PDF (dashed curve) and larger coverage interval are presented in Fig. 15 in order to emphasize the necessity of now incorporating the definitional uncertainty into the probability considerations. Because of the new definition of measurand in VIM3, as “quantity intended to be measured,” if it is thought (but not known) that the quantity actually being measured is different from the measurand, then, using the GUM approach, the corresponding uncertainty associated with a correction is a part of the measurement uncertainty, and similar considerations concerning use of ‘probability’ would apply.

Since they were discussed earlier in connection with the classical approach, it is interesting to consider how the Bayesian and frequentist theories of inference relate to the GUM approach. In a sense, it can be said that the GUM approach, and in fact the uncertainty approach in general, are consequences of the Bayesian theory of describing one’s state of knowledge about a measurand. Using the Bayesian theory in the GUM approach, measurement can be thought to consist of incrementally improving one’s state of knowledge and belief about a true value based on all of the accumulated information that is available through measurement. Using the Bayesian theory, the measurement uncertainty based on probability density functions associ-

Fig. 14 VIM3 terminology for uncertainty approach to measurement 1

Fig. 15 VIM3 terminology for uncertainty approach to measurement 2

ated with a particular measurand will continually change according to additional information obtained through measurement. The frequentist theory of inference can be useful for determining certain Type A components of measurement uncertainty, but is not capable of treating most Type B components. An example of the difficulty of the frequentist theory of inference within the GUM approach is that the frequentist theory is not able to be used to assess the uncertainty of a single measured value when using a measuring instrument, such as a voltmeter. The reason is that the uncertainty here derives from ‘nonstatistical’ information obtained from the instrument’s calibration certificate. This type of single measurement comprises a large fraction of the types of measurements routinely made daily throughout the world.

IEC approach to uncertainty

The other major approach to describing and characterizing measurement that will be discussed here is that used by the International Electrotechnical Commission (IEC), as presented primarily through their IEC 60359 *Electrical and Electronic Measurement Equipment – Expression of Performance* [4]. The IEC philosophy questions the existence, in principle, of a true value of a quantity. The objective of measurement in this view is not to determine a true value of a measurand with a given probability, but concentrates instead on **metrological compatibility of measurement results**, defined by VIM3 as “property of all pairs of measurement results for a specified measurand,

such that the absolute value of the difference of the measured quantity values is smaller than some chosen multiple of the standard measurement uncertainty of that difference”.

The IEC approach is based on a more operational or pragmatic philosophy than the GUM approach. Most notably, the IEC approach treats the concept of true value as both unknowable and unnecessary, discouraging and in fact eliminating at least explicit use of the concept of true value, even in stating the objective of measurement. In the IEC approach, as presented in the Introduction and Annex A of IEC 60359 [4], the stated objective of measurement is to obtain measurement results that are compatible with each other, within their respective measurement uncertainties. The philosophy is that, from an operational perspective, this is all that can really be done in measurement. This is illustrated schematically in Fig. 16, where the four horizontal lines represent sets of measured quantity values for four separate measurements of the same specified quantity being measured (which might be different from the measurand). From the IEC perspective, it could be argued that the concept of true quantity value is potentially harmful, since it leads to thinking about something that is not relevant.

VIM3 RATIONALE. As a result of this key difference in philosophy between the IEC approach and the GUM approach to the uncertainty approach, it is necessary to generalize several of the central concepts and definitions in VIM3 to accommodate both approaches whenever possible. For reasons discussed earlier, the important concept of “true quantity value” is kept in VIM3, but is not explicitly

Fig. 16 IEC approach to measurement 1

used in the context of definitions that also apply to IEC. For example, the definition of “measured quantity value” has been generalized to “quantity value representing a measurement result,” instead of “quantity value representing the set of true values of a quantity ...” so that true value does not need to be explicitly mentioned, but can be still be inferred for the classical and GUM approaches. Similarly, “measurement result”, as mentioned above, has been defined in VIM3 as “set of quantity values being attributed to a measurand together with any other available relevant information”, rather than as, e.g., “set of quantity values estimating the true values of a measurand”. This wording accommodates the IEC view that a measurement result is just a set of values, with every element of the set having equal status. The probabilistic aspect of the GUM approach is left to the end of the definition as “any other available relevant information,” which can be ignored for the IEC approach. A third example is definitional uncertainty, now defined in VIM3 as “minimum measurement uncertainty resulting from the inherently finite amount of detail in the definition of the measurand,” rather than “parameter characterizing the estimated dispersion of the true values of a quantity...,” in order to remove explicit reference to true value.

Another key aspect of the IEC approach is that it focuses on providing guidance for obtaining measurement uncertainty in situations where single measurements are made using measuring systems, and where the measuring system is operating not only under reference conditions, but anywhere within its rated operating conditions. The IEC approach in this regard, as described in IEC 60359 [4], is to construct a calibration diagram applicable under given operating conditions. An interpretation of the IEC cali-

bration diagrams, using a modified terminology that is compatible with the VIM3 terminology, is illustrated in Fig. 17. The horizontal axis, called indication axis (or ‘reading axis’), corresponds to the indication of a measuring system (in unit of indication’). The vertical axis, called measured value axis (or ‘measurement axis’), corresponds to measured values (in ‘unit of measured value’) as obtained using measurement standards. The boundary curves of indication around the calibration curve are obtained during the course of calibration of the measuring system, using measurement standards, and are used to assess the range of indication for a given measurement standard. When subsequently using the measuring system for a measurand with unknown quantity value, a given indication will correspond to a measured quantity value and an assigned range of measured values, which is derived from the boundary curves of indication, as illustrated in the figure. IEC uses this range of measured values in assessing measurement uncertainty.

Returning to the fundamental IEC philosophy that the concept of true quantity value is unnecessary, and that all that really matters is that measurement results are compatible with each other, one might ask what to do when measurement results are not compatible with each other, as illustrated schematically by ‘measurement number 5’ in Fig. 18. In this case it is necessary to investigate whether any mistakes have been made in performing all of the measurements. If no mistakes can be found, then it is assumed that the quantity that was measured was different for some of the measurements. In this case IEC advocates to somehow ‘average all of the measurements’ and create an uncertainty that encompasses all of the measurement results.

Fig. 17 IEC approach to measurement 2 **Fig. 18** IEC approach to measurement 3

Conventional value hybrid approach; knowable measurement error

Before concluding, it is useful here to discuss a hybrid of the classical approach and the uncertainty approach that is frequently employed as a practical solution for handling the conceptual and terminological problems described earlier concerning the inability to know measurement error, without abandoning the concept and term, since they are still so widely used. This hybrid approach, which will be called here the ‘Conventional Value Hybrid Approach’, or CVHA, is typically used in measurement situations where a decision must be made concerning whether a measured quantity conforms to a particular requirement, such as a specified machine tolerance or a legal regulation. The

‘hybrid’ aspect of the CVHA is that, while measurement error is used, measurement uncertainty is also taken into account.

The CVHA is a two-step approach. In the first step a measurement standard is calibrated using a ‘high-level’ measurement procedure and measuring system, and assigned a conventional quantity value. In the second step, a second measurement is performed on the calibrated measurement standard using a ‘lower-level’ measurement procedure and measuring system. Measurement error in the second step is assessed with respect to the conventional quantity value that was assigned to the measurement standard in the first step. This measurement error can be expressed as a rational quantity since it is defined with respect to the conventional quantity value, and not the true

quantity value, of the measurement standard. Figures 19 and 20 schematically illustrate the two-step process of the CVHA.

Figure 19 shows the conventional quantity value being assigned to the measurement standard, through measurement, using a ‘high-level’ measurement procedure and measuring system. In this first step the systematic measurement error, and hence the error, as defined with respect to the true quantity value, cannot be known, and the systematic measurement error is set to zero by convention. The curve represents a fit to a set of histogram data (subscripted ‘1’) that are obtained when calibrating the measurement standard. Note that a measurement uncertainty associated with the conventional value can be determined, but this is not illustrated in this figure.

Figure 20 illustrates the second step of the process, where the quantity associated with the measurement standard (to which a conventional quantity value has been assigned) is now measured with a ‘lower-level’ measuring system. The measured quantity values obtained when using this system are denoted schematically by the “fit to histogram data₂” on the right side, and an individual measured quantity value (y_{2i}) is also indicated. Note that the measurement scale has been shifted in Fig. 20, such that the difference between the conventional quantity value and true quantity value is meant to be the same in Figs. 19 and 20, and the “fit to histogram data₁” in the two figures is also meant to be the same. Figure 20 illustrates that, typically in the CVHA, the measured quantity value using the ‘lower-level’ measuring system is not expected to be as

“close” to the true quantity value as the conventional quantity value is and, further, the width of the “fit to histogram data₂” is not expected to be as narrow as that of the “fit to histogram data₁”. More importantly in Fig. 20, however, is the illustration that systematic measurement error and error can be defined in the second step of the CVHA both with respect to true quantity value (in which case they are unknowable) and with respect to conventional quantity value (in which case they are knowable). Note that systematic measurement error here is also defined with respect to the average of the histogram data₂ and not a mean of the respective theoretical frequency distribution, as discussed earlier (Fig. 5). Figure 20 illustrates a calibration of the lower-level measuring system.

The advantage of the CVHA is that it can be used in measurement situations where the measurement uncertainty associated with the conventional quantity value is small with respect to the typical “knowable measurement error”. Then it is possible to perform relatively straightforward measurements using the lower-level systems, and make equally straightforward conformity assessment decisions, without having to perform a possibly complicated measurement uncertainty analysis. This approach has been used for many years and covers many types of measurement situations where, in fact, a “knowable measurement error” is frequently treated as a measurand.

An example of the CVHA is the use of a standard weight to verify the performance of a balance. The weight is the (calibrated) measurement standard, and the balance is the lower-level measuring instrument used to obtain the mea-

Fig. 19 Conventional value hybrid approach to measurement 1

Fig. 20 Conventional value hybrid approach to measurement 2

sured quantity value in Fig. 20. The knowable measurement error is the difference between the indication and the conventional quantity value of the weight that is placed on the balance. This measured knowable error is then compared to a maximum permissible error (MPE) quoted in a regulation for that type of balance in order to make a decision about whether the balance conforms to the MPE requirement.

As modern measuring equipment used for even routine measurements becomes more sophisticated, it is not always possible to find a measurement standard or measuring instrument that is significantly better than the lower-level measuring system, and so the knowable measurement error is not always significantly larger than the expanded measurement uncertainty associated with the conventional quantity value of the measurement standard. Further, as the pressure increases to become more efficient in every phase of business, including that concerning measurement, there is a need to make better conformity assessment decisions. The irony is that it is then becoming increasingly important, when using the CVHA, to consider the uncertainty of the (knowable) measurement error. It therefore becomes necessary to consider whether there is less terminological and conceptual confusion by calculating the measurement uncertainty associated with the measured quantity value itself (and specifying a maximum permissible uncertainty) [5], than by estimating the knowable measurement error.

VIM3 RATIONALE for measurement error. The dual usage of the term “error”, both in an unknowable sense when a measured quantity value is compared with a true quantity value, and in a knowable (calculable) sense when that same measured quantity value is compared with a conventional quantity value, is another dilemma faced in

the development of VIM3, since two different concepts are being designated by the same term. The solution presented in VIM3 is to slightly re-define “measurement error” in a more general sense, as “difference of measured quantity value and reference quantity value,” where the reference quantity value may or may not be the true quantity value (e.g., it could be a conventional quantity value). This new definition then encompasses both meanings of the term “error”, the unknowable and the knowable “error”.

VIM3 RATIONALE for measurement accuracy. A concept closely related to “measurement error” is that of “measurement accuracy,” mentioned earlier, which even in the classical approach is in common use and is therefore kept in VIM3. The VIM3 definition: “<classical approach> closeness of agreement between a measured quantity value and a true quantity value of a measurand” is similar to the VIM2 definition, which also is based on true quantity value. However, since IEC does not use the concept of true quantity value, and also because a somewhat different usage of “accuracy” has developed in connection with the uncertainty approach, it was decided to include a second definition of measurement accuracy: “<uncertainty approach> closeness of agreement between measured quantity values that are being attributed to the measurand.” This is a situation where a harmonized definition was not considered possible.

Summary

Different philosophies and approaches to measurement still exist and are in common use, most notably the classical approach and the uncertainty approach. Trying

to create a vocabulary of metrology that harmonizes the language of measurement among the different approaches, and that keeps one term designating only one concept, has presented tremendous challenges in developing VIM3. While a principle used for VIM3 has been to harmonize terminology to the extent possible (e.g., “measurement error”), it has in a few cases been necessary to allow two concepts having the same term (e.g., “measurement accuracy”), or different terms for the same concept (e.g., “value”/“true quantity value”), in the different approaches. Several of the decisions and rationales have been presented.

Future

At the time of publication of this paper, the VIM3 has not been finalized. Once the VIM3 has passed the second international comment and review process and has been published, there are plans by the authors to develop an updated and expanded version of this paper for publication and wide distribution.

The plans for publication of VIM3 include its availability, for no charge, on the BIPM web site. Hard copies of VIM3 will likely be available, for a fee, from ISO.

Acknowledgements The authors wish to acknowledge the valuable contributions of P. De Bievre, W.T. Estler, L. Gonella, C. Hockert, H. Imai and S. Morris.

Disclaimer: Material discussed in this paper does not represent the current policy of the National Institute of Standards and Technology (NIST).

References

1. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, 3rd Edn (VIM3), Joint Committee for Guides in Metrology (JCGM), expected 2007
2. International Vocabulary of Basic and General Terms in Metrology (1993) 2nd Edn (VIM2), International Organization for Standardization (ISO)
3. Guide to the Expression of Uncertainty in Measurement (GUM), (1995) 1st edn, International Organization for Standardization (ISO), 1993; corrected and reprinted
4. Electrical and Electronic Measurement Equipment – Expression of Performance, International Electrotechnical Commission (IEC) International Standard 60359, 3rd edn, 2001–2012
5. Kallgren H, Lauwaars M, Magnusson B, Pendrill L, Taylor P (2003) Role of Measurement Uncertainty in Conformity Assessment in Legal Metrology and Trade. *Accred Qual Assur* 8:541–547