12th IMEKO TC1 & TC7 Joint Symposium on
Man Science & Measurement
September, 3 – 5, 2008, Annecy, France

SUBJECTIVELY- VS. OBJECTIVELY-BASED UNCERTAINTY EVALUATION IN METROLOGY AND TESTING

Franco Pavese [1]

[1] INRIM, Torino, Italy, f.pavese@inrim.it

Abstract: The general problem is wide and extensively treated in the literature. The paper intends to limit the discussion from the viewpoint of the metrologist to some aspects of the problems related to the concept of 'true value', abandoned by GUM but still present in other international reference documents, and about the use of non-quantitative information or of *prior* knowledge in connection with the uncertainty evaluation in measurement.

Keywords: subjective information, objective information, uncertainty evaluation.

1. INTRODUCTION

In 1980-81 the first outcomes appeared of a process initiated in the frame of BIPM [1–3], reforming the way the uncertainty was expressed and based on the distinction of errors in random and systematic components, which was becoming considered unsatisfactory. That approach, eventually formalized in the *Guide to the Expression of Uncertainty in Measurements* issued in 1993 [4], was also adopted by ISO and other major international organizations. Evidence of the terms of the debate in that period of time is well summarized by a couple of other publications: a NBS one [5] and a Colclough article also appeared on the NBS Research Journal [6].

In the following 15 years, despite the strong support provided to the new approach by a wide metrological literature, it is questionable whether it has effectively replaced the previous in metrology. In fact, (i) the BIPM companion document, aiming at providing the nomenclature of the basic metrological concepts, the *International Vocabulary of Basic and General Terms in Metrology* (VIM), in its changes from the 2nd [7] to the 3rd edition [8] does not appear to be full consistent with GUM but shows a further evolution of concepts, as discussed in [9], and, (ii) the corresponding basic document in the contiguous frame of testing, ISO 5725 [10], remained unchanged and only in 2007 the relevant ISO committee (TC69) formally decided to start undergoing its revision [11].

The adoption by the GUM of a new approach in the expression of uncertainty was prompted, ultimately, by the inherent contradiction of the concept of 'true value' (of a quantity) in metrology, which, on one side is said to be "in practice, unknowable" [8], on the other side is necessary to obtain a measure of the "systematic" components of 'error'. In fact, GUM Annex D is devoted to this issue, where it states (D.3.5) "The term 'true value of a measurand' … is not used in this Guide because the word 'true' is viewed as redundant", because "the 'true' value of the measurand (or quantity) is simply the value of the measurand (or quantity)".

However, does GUM solve the conceptual issue, i.e., is it actually implementing its stated goal: "The concept of uncertainty adopted in this Guide is based on the measurement result and its evaluated uncertainty rather than on the unknowable quantities 'true' value and 'error'", i.e. on "observable quantities"?[1] Consequently, did GUM condition metrological terminology and practice?

The reader is directed to [9] for a comprehensive set of citations of the relevant documents and for their comparison concerning both metrology and testing issues. Therefore, in the following the documents will sometimes be referred to omitting their relevant paragraph number.

2. PRELIMINARY ISSUES

As also GUM and other documents are forced to do, it is not possible to reason about uncertainty in measurement without also preliminarily reasoning about the value of the quantity intended to be measured.

The goal of GUM can be considered consistent with Wittgenstein's "verification principle" [13]: "The meaning of a question is the method of answering it", a wording that in other translations from German reads "The meaning of a statement is the method of its verification". Without entering the issue of the general validity of this principle –by itself not verifiable– that was quite contrasted in other fields, namely the theological and transcendental philosophy ones, there is little doubt on its validity in science in modern physics, where experimental evidence is generally required to substantiate theoretical statements and models.

[1] This does not mean in itself that the GUM is explicitly replacing "uncertainty" to "error" as said in [12]. The text is a bit tortuous, but, e.g., it reads (D.5.1) "the uncertainties associated with the random and systematic effects that give rise to the error can be evaluated. But even if the evaluated uncertainties are small, there is still no guarantee that the error in the measurement result is small …". One should instead note that the GUM, in assuming in (3.2.3) that the error expectation is zero after correction (see Section 2 that follows), is contradicting its Appendix D.

Consequently, the concept of 'true value' would be meaningless according to this principle, since it is said to be unverifiable. The expected consequence should be that one can deal only with measured –or otherwise 'verified'– values[2].

These measured values arise from observations, i.e., are measures of "observable quantities", hence assumed to be or approximate values of a measurand.

However, the GUM, while dropping the term "true", is not adding to "value" the basic specification 'measured'. In fact, GUM makes a distinction between "the measurand" (D.1), defined as "the quantity to be measured" (the 1993 VIM definition, changed by VIM in 2007 into "the quantity intended to be measured", a conceptual change), and "the realized quantity" (D.2), "an approximation of the measurand".

The implication of this distinction is that GUM prescribes that "the result of the measurement of the realized quantity is corrected for the difference between that quantity and the measurand". In doing so, GUM seems to fail in its goal of being only based on "observable quantities", being both terms of the difference, "realized quantity" and "measurand", not observable: the first because its value is as unknowable as that of the "quantity", the second because the knowledge of the measurand can only be approximated. In fact, the GUM specifies that "neither the value of the realized quantity nor the value of the measurand can ever be known exactly".

To illustrate this issue, a very useful classification will be used (Fig. 1) introduced by Colclough [6] and as follows:

1) "each result may differ from the true value by the same amount and with the same sign,
2) each error may vary randomly realizing a stable random distribution with a non-zero mean,
3) each error may vary randomly realizing a stable distribution with a zero mean,
4) each error may vary non-randomly (e.g. cyclically or by failing to produce convergent frequencies)".

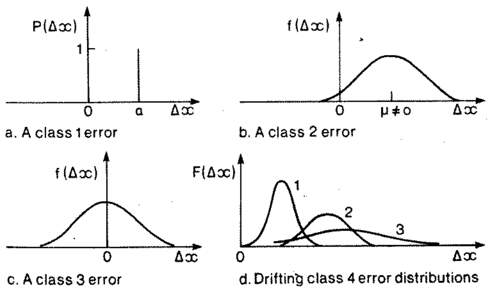

Fig. 1. The four classes of error (from [6]).

In the figure the abscissa zero value is the "true value" of the quantity, but for the reasoning in this paper it can be seen as coincident with GUM's "value of the measurand". Being it referred to "errors", a concept not used in GUM, is not relevant to the comments that follow. Let us also omit to discuss here the implications of Class 4.

First of all, the term "correction" in GUM explicitly does not apply to Class 1, but only to Class 2, as specified in (3.2.3), where an uncertainty of the correction is assumed existing.

Then, since "after the correction, the expectation or expected value of the error arising from systematic effect is zero" (3.2.3), after correction GUM states the error to reduce to Class 3. However, since the same limitation of the "true value" –its being unknowable– apply to the "value of the quantity", no experimental –nor conceptual– verification of the difference between the expectation of the realized quantity and the measurand (or quantity) can be construed. This means that the effect of the GUM correction is a lack of evidence that reduction of Class 2 error to Class 3 is achieved[3]. One may only empirically assess a usually-expected 'reduction' of the 'bias' [4,5].

In all instances, despite the fact that in a 'well-designed experiment' one should assume to have identified all (meaningful) sources of systematic effects, one should adopt for the systematic effects DIN distinction of the "total systematic error" into two components [14]:

(a) one covering the *known* systematic measurement error components;

(b) the other covering the *unknown* systematic measurement error components.

GUM corrections only concern "recognized systematic effects", another reason for a possible failure of reduction of Class 2 error into Class 3.

Therefore, in addition to a possible failure in avoiding to resort to the quantity (true) value, it seems that that GUM approach would also fail in its claim of 'randomising' the effect of systematic error, while it would be "essential to the randomatic theory that any distribution used to calculate uncertainties is of Class 3", since "a standard deviation … cannot reflect uncertainty arising from an unknown and unobservable non-zero distribution mean" [6].

3. IS A VERIFIABLE –OBJECTIVELY-ONLY BASED– DEFINITION OF QUANTITY VALUE POSSIBLE ?

VIM 2nd edition [7] (1.19), reported in GUM, was providing a definition of 'true value' that essentially remained unaltered in the 3rd edition [8]: (2.11) "quantity value consistent with the definition of a quantity'', in practice a tautology useless from an operational viewpoint.

From the operational viewpoint, one should ask whether the metrologist is ever dealing with the 'true value' of a quantity, or, instead, more likely with a "conventional quantity value", according to VIM definition [8] (2.12)

[2] Values associated to an observation can also be non-numerical, e.g., lexical. In this paper only numerical data are considered.

[3] The GUM actually admits this in (D.4): "A corrected measurement result is not the value of the measurand …".

[4] Otherwise, the correction itself would be useless.

[5] With the exception of the cases where one make a correction with respect to a verifiable or unique condition of the quantity definition: e.g., when the use of an "ideally pure substance" is stated.

"quantity value attributed by agreement to a quantity for a given purpose".

A value of any "realized quantity" results from complex operations, being obtained from:

(i) the values of the measured quantities,
(ii) the realised values of the units of the measured quantities,
(iii) a mathematical model of the relationship between the above quantities and also all the identified *uninvestigated*[6] influence quantities, involving a certain number of parameters,
(iv) the estimate of the values of the model parameters,
(v) the conventional values of 'constants' (when used).

Causes for the measured value of the "realized quantity" to differ from the 'true value' of the 'quantity' are:

(i) statistical reasons, for the values of all measured quantities;
(ii) inaccuracy of the realized sizes of the units of the measured quantities;
(iii) inaccuracy of the mathematical model used for the relationship between the above measured quantities and for the identified *uninvestigated*[6] influence quantities (e.g., higher-order terms omitted, incomplete knowledge about variability), and model incompleteness for unidentified influence quantities;
(iv) inaccuracy of the numerical values assigned to the parameters and statistical reasons when they arise from experimental investigation,
(v) trueness of the conventional values of 'constants' and, when an uncertainty is associated to the value, probabilistic reasons.

Some of these causes arise from the practical difference between a "realized quantity" and the "quantity", others from the imprecision of the investigations concerning the "realized quantity". It is difficult, in a general case, to separate the two parts.

GUM is referring to the 'realized quantity'. In fact, the effect of an "incomplete definition of the measurand", which certainly is one of the causes for the conceptual difference with respect to the "quantity", is excluded from the "results of measurement". VIM [8], by indicating the measurand as the "quantity *intended* to be measured", is making basically the opposite choice of indicating the "quantity" as the aim of the measurement.

According to GUM approach, item (i) above concerns Type A component of uncertainty. Item (ii), in the frame of an investigation on a quantity value, should be considered to concern *a priori* information, i.e., Type B component of uncertainty, basically in the form of the uncertainty of the supporting calibration. Also the uncertainty associated to item (v), when attributed to the conventional values used in the model, should concern Type B component of uncertainty.

Items (iii) and (iv) are only partially taken into account by the GUM, Appendix D. The "final result of measurement" is out of its scope, omitting estimation of "unknowable" contributions (remaining error, difference to the value of the measurand, measurand incomplete definition). This omission appears to be somewhat inconsistent with the use of a model, a mandatory need for the GUM approach. In fact, the value of the "result of measurement" according to the GUM is the corrected expectation of only the observations and the variance is not augmented by the unknown amount due to the effect of the measurand incomplete definition, either in the sense indicated by the GUM for the latter and by item (iii). Yet, the model should also include, in addition to the measured quantities and those relevant to the "recognized systematic effects"[7], also an estimate of its imperfection for not being able to identify all systematic effects. Finally, there is basic difference, not specifically indicated in the GUM, between models based on theoretical considerations, where the parameter values are to be obtained from *prior* information (originating Type B uncertainty components), and empirical models, where the parameter values need a specific experimental determination, a 'calibration', a *reproducibility* study, etc.

The above illustration concerned so far one set of *repeated* observations performed in a single laboratory[8]. There is already evidence that some components of knowledge are objectively- and others are subjectively-based and that some arise from *prior* knowledge. The latter can, in turn, involve a first level of possible ambiguity in respect to being subjectively based, for arising from an expert judgement or also from the result of a measurement process not entirely objectively-based.

However, the question that has been placed at the top of this Section cannot be answered without taking into account *more than one set of measurements*. This need, which is generally necessary to the analysis in testing (see, e.g.,

[6] 'Uninvestigated' means here that experimental evidence is not available, at all or in a degree considered sufficient, or that conventional values are used without further investigation, including the 'constants' in (iii).

[7] "The mathematical model of the measurement procedure that transforms the set of repeated observations into the measurement result is of critical importance since, in addition to the observations, it generally includes various influence quantities that are inexactly known". Should these quantities be uninvestigated, the knowledge on which the correction is based, is not consisting of statistical data, so the associated uncertainty is part of the Type B components, estimated by other means (*subjective* like *a priori* probability distributions, bounds based on expert judgment, stipulated values, …; or *objective* like certificates, literature information, ...).

[8] They must be repeated measurements (not in the sense of GUM definition 3.1.5, but in the sense of general statistics, homoscedastic data) because GUM is stating that all detectable non-zero effects due to a probability (or to an empirical) distribution are not attributed to the set of "*uncorrected* observations".

[15,16]), is apparently omitted in the GUM, with some peculiar consequences.

In fact, this is the general case also in metrology, where, possibly except when providing data to a hierarchically lower-level user, no statement is possible about the accuracy of the measured quantity value and no definite confidence level can be attributed to the associated uncertainty, until one can compare two or more sets of results of measurements performed on the same quantity, in the same laboratory or in different laboratories.

The very same GUM concept of “realized quantity” is generally a relative concept, at least partially, meaning ‘quantity as realized at a given time by a specific laboratory’. If no more than one realization exist, the concept itself may vanish for lack of evidence: if only one copy of a standard is considered –or exists– and is circulated for measurement among all possible laboratories, practical differences between the concepts of ‘realized quantity’ and ‘quantity’ may become irrelevant; only those arising from the imprecision of the investigations concerning the ‘realized quantity’ remain. If only one realization of a physical state –e.g., of the triple point of a substance– is available, no experimental evidence can be brought for the temperature value of the pure substance, unless the substance is so close to the zero-impurity condition that the effect of correction for this systematic effect can demonstrated to be irrelevant for the stated level of uncertainty [17]. Also the size of the unit of a quantity (item (ii) above), is no more an issue for measurements performed with the same standard. The factor (in a multiplicative model) due to the realization of the unit affects the measurements results, but the effect has to be taken into account only when comparing observations involving more than one independent realizations of the units.

Let us now limit the analysis to the most common case concerning the analysis of independent measurements originating from different experimenters in different laboratories. They need, almost invariably, a specific analysis for their *combination*, resulting in specific decisions (i.e., lacking generality) to be taken to obtain a summary statistics.

There are excellent examples: one are the fundamental constants, the value of which is obtained, in general, by adjustment of several sets of measured values. Another are the key comparisons in the frame of the CIPM MRA, where a “reference value” is generally computed, involving specific decisions since no more than 50% of the key comparison data in the BIPM KCDB were found to outcome a consistent set of data [18, 19].

In all instances, the corrected result being the “best estimate of the ‘true’ value” as stated in [4] (D.3.4), or the key comparison reference value being “in most cases, … a close, but not necessarily the best, approximation to the SI value” as stated in [20], remains an unverified and unverifiable statement. The best one can say is that the ‘degree of confirmation’ [21] of the meaning of the stated value depends on the efficacy of the experimental evidence. It seems uncontroversial that this efficacy depends on the robustness of the experimental evidence and that this is stronger when more independent (basically uncorrelated) determinations of the same value are available, and to the degree to which these values are consistent with each other.

In conclusion, as far the *quantity value* is concerned, in the vast majority of cases, if not all –difficult to verify,– starting from measured values, one eventually assigns the result of measurement basically the character of a ‘*consensus value’*, though only some of them are explicitly declared as such. The consequence is that, *in metrology*, with some exceptions, the ‘true value’ is not a relevant issue, even when the phenomenon is known to be objectively single-valued –e.g., as for the fundamental constants.

In testing the situation is, at least for the field-level of most laboratories, quite different: a “trueness” is defined as ‘‘the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value’’ [10] and the relationship between the latter and the ‘true value’ of the quantity is generally out of the field of the interest of testing procedures. The ‘true value’ is replaced by a “conventional true value” and “in practice, the accepted reference value is substituted for the true value” [22] .

In this frame, the difference between a ‘consensus value’ and a ‘reference value’ is that, the latter is an information *a priori* derived from characterisation measurements of a batch of material, the former is directly obtained with a *consensus statistical procedure* from the participants’ results [23].

4. TO WHICH DEGREE CAN NON-QUANTITATIVE INFORMATION AND *PRIOR* KNOWLEDGE CONTRIBUTE TO GENERATE AN OBJECTIVELY-BASED UNCERTAINTY EVALUATION ?

From the general principles, the fact that statistical data are objectively-based, at least in physical sciences when they are the result of measurement, is little controversial.
Different is the case of non-quantitative information and of *prior* knowledge. The field is extremely vast, so let us restrict a tentative reply to the question limited to what is concerning the process bringing to the ‘*consensus value*’ and to the overall uncertainty statement associated to it.

Non-quantitative information is often called ‘qualitative’ in the sense that the source information is not consisting of numerical data (even if it can then be transformed into a numerical ‘scale’ or in a binary (e.g., yes/no) or fuzzy form). Traditionally involving a broad range of testing activities, it is more and more used also in the field of metrology, namely for its relatively recent consideration of fields like medecine and biology, earth sciences, chemistry, environment.

Closer to the preceding discussion, it also comprises expert judgment: taking a decision concerning a ‘consensus value’ necessarily involves also a shared decision about combining data, not directly or fully stemming from the data themselves. For example, the well-known issue of selecting a summary statistics for a set of comparison data generated in a key comparison (KC, an MRA exercise [20]), called KC reference value and related uncertainty, has been in many cases the source of a great deal of discussion and of controversy in the past 10 years. This is an indication of the fact that often a unique objective evidence cannot be drawn even from statistical data.

An even more evident case is when, like in the assessment of acceptance of a calibration and measurement capability (CMC) [20], the expert judgment is itself concerning non-quantitative information, like the scoring of a 'visit' to a laboratory to assess its 'quality', and this information have to be combined with the quantitative results obtained by the same laboratory in a KC. It is a typical decision-taking problem (see e.g., [24]).

The consequence should be that the uncertainty associated with the lack of firm evidence that the method or decision chosen (e.g., mean or weighted mean or median) is the 'best' (whatever is the sense one may give to this term) or attached to the non-quantitative information, should *increase* the overall uncertainty of the 'consensus' numerical value (in the case of a KC) or limit (in the case of a CMC). Author's impression is just the opposite, of consensus being considered as a method to minimise uncertainty.

Considering now *prior* knowledge, clearly it is, in principle, advantageous to use the full knowledge available, especially in metrology, since this is likely to decrease the uncertainty associated with the overall knowledge. For example, for a standard having proved to be stable in time (see item (d) in the list below), one can increase by this way the number of the *repeated* observations available for that specific standard over the years, which can then be statistically analysed all together, and the uncertainty be considered as part of Type A components. Therefore, Type A components of uncertainty can also embed *prior* information. In addition, in these cases, the same objectivity level of the last-produced observations can be maintained.

The VIM [8] provides the following list for items whose uncertainty that should be considered as part of the Type B components:

a) "associated with authoritative published quantity values;
b) associated with the quantity value of a certified reference material;
c) obtained from a calibration certificate;
d) about drift;
e) obtained from the accuracy class of a verified measuring instrument;
f) obtained from limits deduced through personal experience."

Thus, *prior* information is considered in (c) the calibration certificate of a device, in (b) the reference value of a batch of material, or in stating the degree of equivalence originated by a MRA exercise [20]. To these values also an uncertainty is associated and possibly, in addition, a probability distribution (empirical, if assigned on the basis of the results of the original studies), most often Gaussian. For the user, they act as being assigned (or stipulated), so loosing their possible original content of subjectivity that could have contributed to their determination when the original studies were performed.

Similarly one should consider item (a)[9] and (e). On the contrary, item (f) should be considered as an example of an 'expert judgement'. Where the uncertainty is expressed as bounded in a confining interval [25] or as a *prior* probability distribution (uniform one included), in both cases it should be considered as subjectively-based.

The VIM list is not exhaustive but it would be impossible to list all possible cases of *prior* knowledge: for each of them a thoroughly analysis is necessary to identify its right character.

4. CONCLUSIONS

The paper is intended mainly to places questions, and only some examples of possible replies are provided from a metrologist viewpoint.

Basically, the concept of *accuracy* of a quantity value was found to generally lack the necessary "verification" or, at least, a sufficient "degree of verification". As a consequence, the values used at the top of the metrological traceability chain should be considered as 'consensus values' or 'reference values' also in metrology and treated as such.

Yet, of the three traditional steps in the metrological knowledge-gaining process about a standard –repeatability, reproducibility and accuracy– also the first one, *repeatability*, actually suffers of uncertain objectivity. In fact, the necessary specification "over a short period of time" is indefinite and essentially means 'over a period of time sufficiently short for reproducibility to be true', an evident tautology. If repeatability cannot be independently verified, its concepts is vanishing and merging into the one of reproducibility [9].

It seems that only *reproducibility* be objectively based, but with a limitation about its "degree of confirmation" indicated in the GUM "If all the quantities on which the result of a measurement depends are varied, its uncertainty can be evaluated by statistical means". The less this applies, the more reproducibility becomes a wishful thinking.

However, the GUM remedy, of using instead a model, is apparent, at least if used as the only resource: as said at the beginning of this paper, no model is meaningful unless it is validated by experimental verification, with specific additional uncertainty components arising from model incompleteness or inadequacy; and, in all instances, then, uncertain numerical values must fill up parameters in the model. After all, a statistical study, especially if conducted with a non-parametric method, is a model-less approach in many cases with less restrictions.

A possibility to escape from some of the above limitations, which are sometimes unacceptable, would be to recognise that our knowledge is only relative and to use only differences of values, simpler in a hierarchical scale, less easy *inter pares*: this issue will be treated in a subsequent paper.

[9] If critically used.

REFERENCES

[1] “Report on the BIPM Enquiry on Error Statements”, Rapport BIPM 80/3, BIPM (in French) (1980)

[2] Kaarls, R., “Proces Verbaux des seances du CIPM”, vol 49, pp A1–A12 (1981)

[3] Giacomo, P., Metrologia 18 43–44 (1981)

[4] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML “Guide to the expression of uncertainty in measurement (GUM)”, 2nd edn. ISO, Geneva, Switzerland (1993)

[5] NIST “Expression of the Uncertainty of Final Measurement Results: Reprints”, NBS Special Publication 644 (1983)

[6] Colclough, A R., “Two theories of experimental error” J Res Natl Bur Stand 92 167-185 (1987)

[7] BIPM “International Vocabulary of Basic and General Terms in Metrology (VIM)”, 2nd edn BIPM/ISO (1993)

[8] BIPM “International Vocabulary of Basic and General Terms in Metrology (VIM)”, 3rd edn BIPM/ISO (2007)

[9] Pavese, F., “Replicated observations in metrology and testing: modelling of repeated and non-repeated measurements” ACQUAL 12 525-534 (2007)

[10] ISO 5725 (1994) Accuracy (trueness and precision) of measurement methods and results. International Organization for Standardization, Geneva, Switzerland

[11] ISO TC69 “Minutes of the 2007 General Meeting”, Lo Skolen, Denmark, ISO, Geneva (2007)

[12] Rabinovich, S., “Towards a new edition of the GUM” Accred. Qual. Assur. 12 603-608 (2007)

[13] Wittgenstein, L., “Philosophical Remarks”, Basil Blackwell, Oxford, 1975 (a collation of writings from Wittgenstein)

[14] DIN 1319 “Fundamentals of metrology –Part I: basic terminology”, Berlin, Beuth (1995)

[15] Bremser, W., in “Data Modeling for Measurement in Metrology and Testing”, F. Pavese and A.B. Forbes eds., Birkhauser, Boston, October 2008

[16] De Bièvre, P., in “Data Modeling for Measurement in Metrology and Testing”, F. Pavese and A.B. Forbes eds., Birkhauser, Boston, October 2008

[17] Pavese, F., McConville, G.T.: "The triple-point temperature of pure equilibrium deuterium", Metrologia 24 107-120 (1987).

[18] BIPM KCDB, http://bipm.kcdb.org

[19] Cox, M.G., IT in Metrology, BIPM-PTB Workshop, Berlin 2007, www.bipm.org

[20] CIPM (1995) “Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes”, Bureau International des Poids et Mesures, Sèvres

[21] Carnap, R., “Testability and meaning”, in Philosophy of Science, III (1936) and IV (1937)

[22] ISO 3534-2 Vocabulary and symbols —Part 2: Applied statistics (2006) International Organization for Standardization, Geneva, Switzerland

[23] Baldan, A., van der Veen, A.M.H., Prauß, D., Recknagel, A., Boley, N., Evans S., Woods, D. “Econimy of proficiency testing: reference value versus consensus value”, Accred Qual Assur 6 164-167 (2001)

[24] Ayyub, B.M., Klir, G.J. “Uncertainty Modeling and Analysis in Engineering and the Sciences” Chapman & Hall/CRC, Boca Raton, 2006

[25] Kreinovich, V., in “Data Modeling for Measurement in Metrology and Testing”, F. Pavese and A.B. Forbes eds., Birkhauser, Boston, October 2008