Thermochimica Acta 382 (2002) 1–16

thermochimica acta

www.elsevier.com/locate/tca

Measurement uncertainty according to ISO/BIPM-GUM

Wolfgang Kessel*

Physikalisch-Technische Bundesanstalt, Proj. 2.002 "Messunsicherheit im industriellen und gesetzlichen Messwesen", Bundesallee 100, D-38 116 Braunschweig, Germany

Received 25 June 2001; received in revised form 25 June 2001; accepted 27 June 2001

Abstract

In modern life, to make measurements comparable, stating the quality of a measurement in terms of the measurement uncertainty is an absolute necessity. The ISO/BIPM "Guide to the Expression of Uncertainty in Measurement"—usually referred to as the GUM or the Guide—, which was published in 1993, introduces a method to unify the evaluation and the statement of measurement uncertainties. This method has been accepted by almost all calibration services all over the world and has become a quasi-standard in the field of metrology. This paper deals with the background of the GUM, the knowledge of the respective measurement and other fundamental aspects which have been included in the EA-4/02 requirements document published by the European co-operation for Accreditation. It shows how the basic concepts may be converted in a straight forward procedure to evaluate the uncertainty in a measurement. It is an expanded version of a lecture with the same title, held by the author at the workshop "GUM uncertainty budget in calorimetry—questions and ways to realistic determinations of measurement uncertainty" at the 14th Ulm-Freiberg Kalorimetrietage in Freiberg (Saxony) March 21/23, 2001.

Keywords: Standard measurement uncertainty; Expanded measurement uncertainty; Type A and B evaluation; Model of evaluation

1. Introduction

Since their beginnings more than 300 years ago, the calculating natural sciences and metrology, which is based on these, have been accompanied by the question as to how the accuracy of measurement results is to be assessed. For a long time, systematic and random measurement errors were regarded as the cause for the more or less exact determination. This opinion may in the final analysis be ascribed to the measuring means of former times which were well-conceived from the point of view of instrument technology but were not so perfect in terms of modern process technology. In the past few decades, the fast development of measuring technology, and particularly the use of measuring transducers and electrical data acquisition, have led to the measuring means being highly refined. This was the reason for a new approach which abandons an inherent inconsistency of former views.

According to the opinion formerly held, the systematic and random measurement errors, which in principle cannot be completely determined, describe deviations from the so-called "true value" of a quantity. As the determination of the "true value" of the measurand[1] is the objective of any measurement, this value is not known before the measurement. On the other hand, since the measurement errors relate to the "true value" it should really be known already before

* Tel.: +49-531-592-2131; fax: +49-531-592-2130.
E-mail address: wolfgang.kessel@ptb.de (W. Kessel).

[1] For the definitions of metrological terms such as measurand, measurement deviation, etc. see VIM [1], the international dictionary for metrology, or DIN 1319-1 [2].

0040-6031/02/$ – see front matter
PII: S0040-6031(01)00729-8

the respective measurement. We have to do with a circular argument.

The ISO/BIPM "Guide to the Expression of Uncertainty in Measurement" [3], published in the early 1990s and briefly referred to as GUM (or Guide), is based on a different operational approach relying on a description of the more or less complete knowledge of the respective measurement and its conditions [4,5]. The wide acceptance of the GUM may be explained besides the unifying of ideas by the fact that this Guide is supported by seven renowned international institutions dealing with the fundamentals of metrology:

BIPM International Bureau of Weights and Measures, http//:www.bipm.org.
IEC International Electrotechnical Commission, http//:www.iec.ch.
IFCC International Federation of Clinical Chemistry, http//:www.ifcc.org.
ISO International Organization for Standardization, http//:www.iso.ch.
IUPAC International Union of Pure and Applied Chemistry, http//:www.iupac.org.
IUPAP International Union of Pure and Applied Physics, http//:www.iupap.org.
OIML International Organization of Legal Metrology, http//:www.oiml.org.

The approach adheres to the view that the quantity to be measured as a physical object exists before the measurement. It is regarded as the cause for specific effects in the form of indications of the measuring apparatus. The value of the measurand, however, does not pre-exist, it is established only by the measurement—it actually is assigned by the measurement to the quantity to be measured [6]. The assignment of the measurement value to the object measured is made by a comparison with a quantity whose value is known. Two aspects are affected by the question of accuracy: "How well is the physical equality of the measurand with the reference quantity estimated?" and "With what accuracy is the value of the reference quantity known with regard to the generally accepted system of units?". The inadequate knowledge of these two items leads to an uncertainty in the assignment. Thus more or less large ranges of values are obtained, which are compatible with the more or less exact knowledge of the conditions and influences. Each of its values can rightly be considered as the result of the measurement.

In its Section 2 this paper briefly describes some reasons why the determination of the measurement uncertainty is so important in modern business. In Section 3 the operational and metrological aspects of a measurement are presented by the example of the calibration of a mercury-in-glass thermometer. Section 4 deals with the naive approach to the measurement uncertainty, the so-called uncertainty intervals. Section 5 is an introduction to the GUM approach the centre of which is formed by the *standard measurement uncertainty*. Section 6 summarises the method of the GUM. Section 7 explains the so-called industrial view which via the term of *expanded measurement uncertainty* reverts to the uncertainty intervals but in a new and extended way. Finally, following the European calibration guideline EA-4/02 and its supplements [7,8,9], Section 8 gives instructions as to how an uncertainty analysis on the basis of the GUM should suitably be carried out.

As the assessment of the accuracy of a measurement or its result, respectively, is a rational description of more or less exact knowledge, it is important for the terms used for the description to be unambiguous and as precise as possible. This is necessary because inaccurate terms further increase the incompleteness of the knowledge. In this sense, the GUM makes a distinction between the uncertainty as a general expression of a doubt about a statement, and the uncertainty of measurement as a numerical value quantifying the doubt about a measured value. So the statement "I am unsure whether tomorrow it will be as hot as today", expresses only the general doubt, whereas the statement "I have determined the temperature of 978 °C in the annealing furnace with a measurement uncertainty of 1.5 °C", expresses the doubt about the measurement result reported in terms of its value, so that it becomes accessible to certain quantitative comparisons. The author will follow this principle strictly hoping that the reader will not be tired by some lengthy terms.

2. Measurement uncertainty—why?

In [10], the author has already described in a general way what is to be understood by measurement uncertainty within the meaning of the GUM, and what importance it has in the modern context of quality assurance. Here, only some additional metrological aspects will be depicted.

In 1992, Kaarls, the then president of the Western European calibration co-operation (WECC) and later vice-president of the European co-operation for accreditation (EA) pointed out in his lecture at the Eurolab workshop [11]:

> "...Within the modern economic structure it is necessary to represent the result of a measurement in relation to conventional true values in order to gain international confidence in and acceptance of

- measurement and test results,
- manufacturers' specifications,
- the statement of values in standards, (norms as well as laws and regulations).

being a prerequisite for

- trade free from technical barriers,
- international agreements and treaties,
- subcontracting and delivery of parts by industries,
- efficient use of raw materials,
- improvement in health care, environment and safety,
- technical and scientific developments,
- quality assurance, accreditation and certification of products and systems".

The EA is the successor organisation of the EAL (European co-operation for accreditation of laboratories) which was preceded by the WECC. It is the umbrella organisation for the national European calibration services, which was also joined by a number of non-European services, and guarantees mutual metrological recognition. Information about the EA and the documents published by it can be obtained in the Internet under http//:www.european-accreditation.com. An overview of the tasks and publications of the Deutscher Kalibrierdienst (DKD—German Calibration Service) which is a member of the EA and makes German translations of almost all EA documents available which can be accessed under http//:www.dkd.ptb.de.

The representation of measurement results in relation to values recognised as conventionally true requires a hierarchy of traceability as schematically shown in Fig. 1. It should be kept in mind that

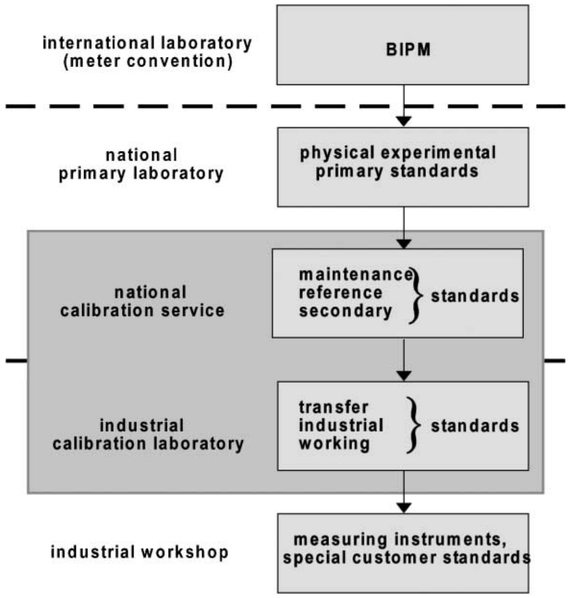

Fig. 1. Schematic representation of the traceability chain (traceability hierarchy) with the definition, realisation/transfer and application levels.

according to the VIM "conventional true values" are understood to be "values recognised as correct", i.e. values which are consistent with the generally accepted scientific knowledge and state no contradiction to these.

The measurement uncertainty as defined in an internationally recognised form, is the measure by which the confidence in the result of the comparisons is expressed at each level of the traceability chain. Confidence in the traceability is based on a realistic statement of the measurement uncertainty. With the publication of a harmonized GUM, the uncertainty definition generally accepted today was adopted; it roughly reads ((VIM 3.9),(GUM 2.2.3),(DIN 1319-1 3.6)):

> *Uncertainty of measurement is a parameter, associated with the result of measurement, that (based on the available knowledge about the measurement) characterises the dispersion of the values that could reasonably be attributed to the measurand.*

3. The measurement

As mentioned in the introduction, it is the aim and objective of a measurement from the operational point

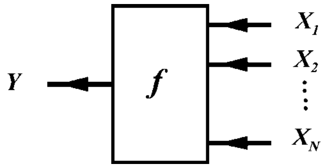

Fig. 2. Schematic representation of the relation between the measurand Y and the quantities relevant to the measurement X_1, $X_2,\ldots, X_N$.

of view to attribute a value to the measurand Y, and this by comparing the measurand with a known quantity or with a quantity formed by other quantities $X_1, X_2, \ldots,$ X_N which are easy to determine. For this, a model of evaluation is needed which expresses this assignment of a value to the measurand in mathematical terms:

$$Y = f(X_1, X_2, \ldots, X_N) \tag{3.1}$$

For illustration of the physical relations only, in most cases a graphical representation as schematically shown in Fig. 2 is used.

Taking the calibration of a mercury-in-glass thermometer at 20 °C as an example the measuring task is to determine the error or systematic deviation of the indication. According to the VIM, it is the difference between the temperature value the thermometer indicates and the temperature value the thermometer should indicate under the conditions given. The measurement principle underlying this calibration consists in the determination of the temperature indication of the object to be calibrated in a medium whose temperature has a known, uniform value. The method of measurement is a comparison of two temperatures. The measurement procedure uses the comparison of the indications of two mercury-in-glass thermometers (object to be calibrated and standard) in a stirred water-bath as schematically shown in Fig. 3.

A way frequently used to establish the model of evaluation starts from the cause-effect relations realised in the measurement procedure. The temperature of the water-bath is so adjusted that the desired indication results on the object to be calibrated. The correct temperature is then inferred from the indication on the standard thermometer. Fig. 4 graphically shows the relations. The water-bath acts as the cause SRC which produces the indication both in the measuring branch (index X) and in the reference branch (index S). The statement includes also the relevant

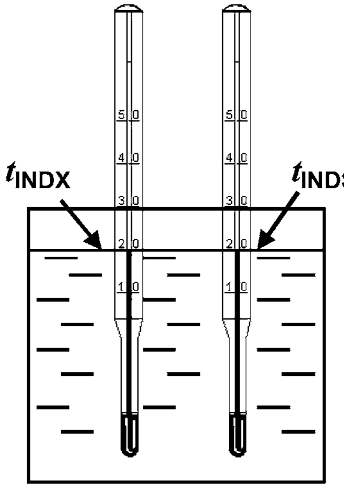

Fig. 3. Schematic representation of the measurement set-up used for the calibration of a mercury-in-glass thermometer.

quantities effective in the various points (temperatures and temperature deviations).

From the graph, the equations

$$\Delta t_X = t_{IND\,X} - t_X - \delta t_{IND\,X},$$
$$t_X = t_S - \delta t_{BATH},$$
$$t_S = t_{IND\,S} - \delta t_{IND\,S} - \Delta t_{IND\,S} - t_{SD} \tag{3.2}$$

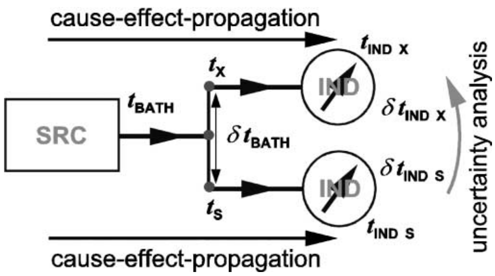

Fig. 4. Cause-effect propagation in the calibration of a mercury-in-glass thermometer (index X) by comparison with a standard thermometer (index S): $t_{IND\ X}$: temperature indicated by sample, $\delta t_{IND\ X}$: deviation due to the finite resolution of the indication of the sample, t_X: temperature in the bath at the location of the sample, δt_{BATH}: difference between the temperature at the location of the sample and the temperature at the location of the working standard, $t_{IND\ S}$: temperature indicated by the working standard, $\delta t_{IND\ S}$: deviation due to the finite resolution of the indication of the working standard, $\Delta t_{IND\ S}$: error of indication of the working standard at the time of its calibration, δt_{SD}: deviation of the error of indication of the working standard from the value at the time of its calibration (drift), t_S: temperature in the bath at the location of the working standard.

can be read immediately that form the model of evaluation. After insertion of the second and third equations into the first equation, the model assumes the form of Eq. (3.1).

4. The naive view—uncertainty intervals

The naive approach to the measurement uncertainty regards the whole range of compatible values as an *uncertainty interval* without weighting them:

$$I_X = [x_-; x_+] \tag{4.1}$$

(x_+ and x_- are the upper and lower limits of the uncertainty interval). Fig. 5 illustrates this approach.

In this approach, the arithmetic mean of the two limits

$$\bar{x} = \tfrac{1}{2}(x_+ + x_-) \tag{4.2}$$

is considered to be the measurement value, and the half-width of the uncertainty interval

$$\Delta a = \tfrac{1}{2}(x_+ - x_-) \tag{4.3}$$

is regarded as the measure of the uncertainty.

This approach to measurement uncertainty is useful for measurements in which only one influence quantity or two occur. As the uncertainty in the value of the measurand is due to the uncertainty in the values of the influence quantities, the limits to the result quantity are to be calculated from the limits to the input quantities using Eq. (3.1). The formulae for the arithmetic operations of addition, subtraction, multiplication and division first appear to be rather simple:

$$\begin{aligned}
I_{X_1} + I_{X_2} &= [(x_1 + x_2)_-; (x_1 + x_2)_+] \\
I_{X_1} - I_{X_2} &= [(x_1 - x_2)_-; (x_1 - x_2)_+] \\
I_{X_1} \times I_{X_2} &= [(x_1 \times x_2)_-; (x_1 \times x_2)_+] \\
\frac{I_{X_1}}{I_{X_2}} &= \left[\left(\frac{x_1}{x_2}\right)_-; \left(\frac{x_1}{x_2}\right)_+\right]
\end{aligned} \tag{4.4}$$

If one enters into the details of the calculations, however, equations are obtained which contain confusing exchanges of values and require a lot of case analyses:

$$\begin{aligned}
(x_1 + x_2)_- &= x_{1-} + x_{2-} \\
(x_1 + x_2)_+ &= x_{1+} + x_{2+} \\
(x_1 - x_2)_- &= x_{1-} - x_{2+} \\
(x_1 - x_2)_+ &= x_{1+} - x_{2-} \\
(x_1 \times x_2)_- &= \min(x_{1-}x_{2-}, x_{1+}x_{2-}, x_{1-}x_{2+}, x_{1+}x_{2+}) \\
(x_1 \times x_2)_+ &= \max(x_{1-}x_{2-}, x_{1+}x_{2-}, x_{1-}x_{2+}, x_{1+}x_{2+}) \\
\left(\frac{x_1}{x_2}\right)_- &= \min\left(\frac{x_{1-}}{x_{2-}}, \frac{x_{1+}}{x_{2-}}, \frac{x_{1-}}{x_{2+}}, \frac{x_{1+}}{x_{2+}}\right) \\
\left(\frac{x_1}{x_2}\right)_+ &= \max\left(\frac{x_{1-}}{x_{2-}}, \frac{x_{1+}}{x_{2-}}, \frac{x_{1-}}{x_{2+}}, \frac{x_{1+}}{x_{2+}}\right)
\end{aligned}$$

Even if only a small number of influence quantities are involved, the calculations get rather complex and, e.g. for trigonometric functions they cannot be carried out without additional decision rules and distinction of cases. The naive approach to the measurement uncertainty can therefore wisely be used only for certain coarse classifications. It is not suitable as a basis for an versatile measure of quality.

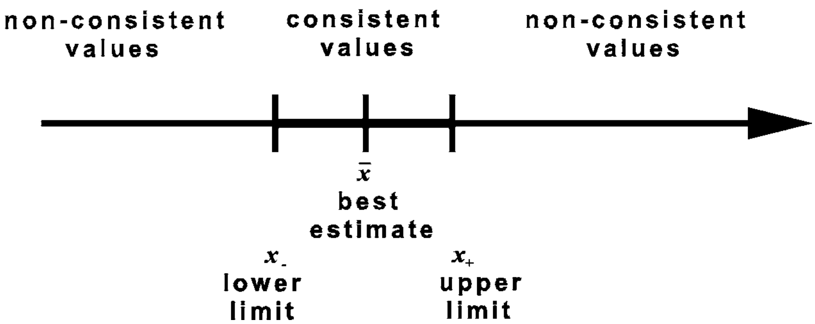

Fig. 5. Characteristics of an uncertainty interval.

In one point, the naive approach reveals, however, the basic nature of incomplete knowledge: the evidence of conformity of a value with a specification. When the uncertainty interval is taken into account as it should, the position of the measured attribute with respect to a specified limit cannot in any case be unambiguously inferred from a measurement result. Fig. 6 schematically shows the situation for various measurement values at an uncertainty interval of equal width. If the whole uncertainty interval is situated on the left of the boundary value, the value of the attribute will certainly also lie on this side. If, the other way round, the whole uncertainty interval lies on the right side, the value of the attribute will certainly also lie there.

But if the uncertainty interval is so situated that it encompasses the boundary value, a well-defined statement cannot be made. There is rather a risk that the measurement value and the criterion lie on different sides of the boundary value. The knowledge of the measurement result in this case does not allow a safe conclusion to be drawn; the measurement value lies in the so-called indifference or *uncertainty range*.

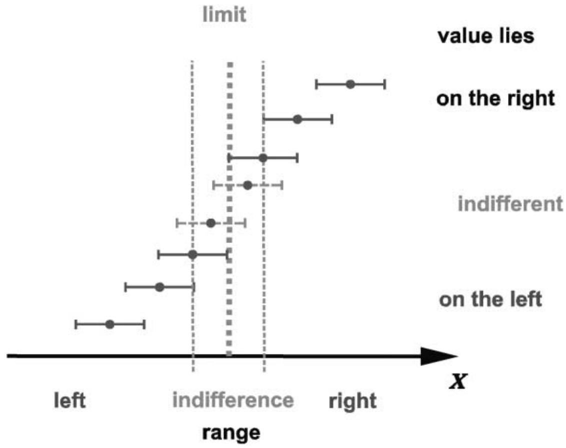

Fig. 6. Schematic representation of the statements which can be made on the basis of the knowledge of the measurement value x and the uncertainty interval as regards the position of a attribute with reference to a limiting value.

Fig. 7 shows the effects these considerations have on conformity statements for which compliance with a specification is to be checked. Clear conformance or non-conformance will be given only if the measurement

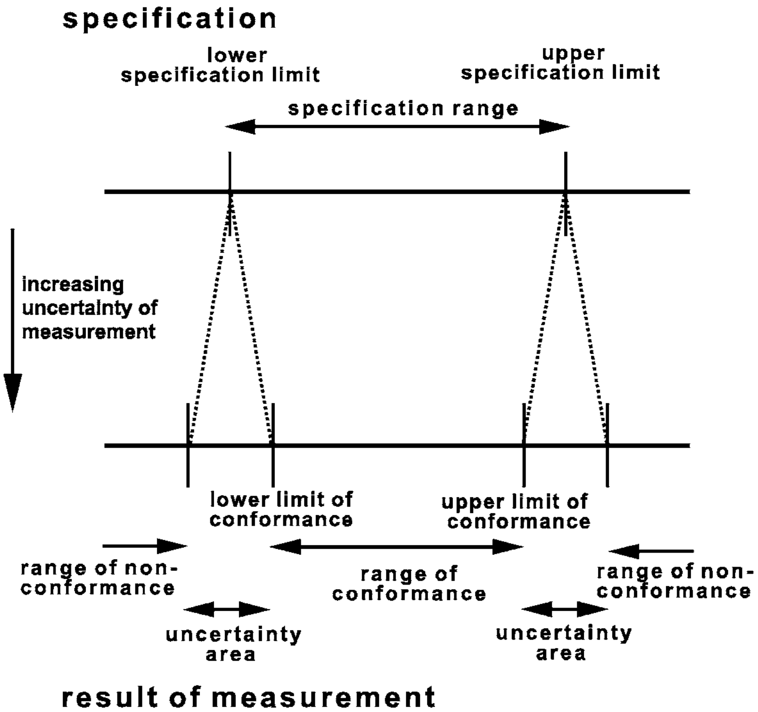

Fig. 7. Ranges of conformance and non-conformance for a measured criterion with a specification and ranges of the uncertainty.

value lies in the corresponding regions. Values in the uncertainty range, however, must be either assessed more specifically or determined more precisely. On this basis, the standard DIN EN ISO 14253–1 develops rules assisting manufacturers and buyers in proving conformance or non-conformance, respectively [12]. They require that for reasons of unambiguous statements the partner wanting to prove conformance or non-conformance should take into account the uncertainty interval. The consequences of these rules were discussed in more detail within the scope of the accompanying program of MICROTECH 1996 [13].

The limits of the naive approach appear where precise measurements with many influence quantities are concerned. In these cases Eq. (4.4) furnish for the measurand an uncertainty interval whose half-width usually equals the sum of the half-widths of the uncertainty intervals of the individual influence quantities. The increase in precision, which is reflected by the increased number of influence quantities, does not result in a substantial reduction of the uncertainty interval. The refined analysis does not lead to a substantial curtailment of the uncertainty range. This conclusion is in contradiction to the general view that a more detailed analysis of a measurement should also furnish a result with a smaller uncertainty.

5. The GUM view: standard uncertainty of measurement

Quite another approach to the problem of measurement uncertainty is taken in the GUM. The statement of the unweighted uncertainty interval is too coarse. It does not take into account that not every value compatible with the conditions of measurement has the same chance of realisation. It is frequently known from specific experience or fundamental considerations that values close to the centre of the uncertainty interval are more likely than values close to the boundaries. There are also cases in which this is the other way around. The GUM is based on a description of the more or less complete knowledge of the possible values of relevant quantities by distributions or more exactly probability distributions. From this exact conclusions are drawn using the knowledge of the physical relations realised or to be realised in the measurement and which are mostly quite exactly known. It thus follows the Bayesian approach to non-complete knowledge and the inference that can be made from this. The logical basis of this approach has been presented, i.e. by Spencer [14], Cox [15] and Tribus [16]. It has its origins in the investigations of the Swiss mathematician Jakob Bernoulli (1654–1705) that were the starting point of the development of theory of probabilities already 50 years before the famous paper of referent Thomas Bayes (1702–1761). Bernoulli wrote the book ‘‘Ars conjectandi’’ (The art of conjecture) which was published after his death in 1713. Bayes again dealt in greater detail with this problem in his ‘‘Essay towards Solving a Problem in the Doctrine of Chance’’. The appraisal of more or less complete knowledge by probability distributions is the object of the Bayesian statistics (see [16]).

Starting from the principle of maximum entropy which is well-known in the theory of probabilities and in statistics [18], Weise and Wöger [17] showed that the more or less complete knowledge of a quantity X can always be described by a probability distribution of the values compatible with the knowledge. The best estimate derived from the distribution is the expectation of the distribution, and the measurement uncertainty associated with it is its standard deviation. It is obtained as the positive square root from the variance:

$$x = E[X] \tag{5.1}$$

$$u(x) = \sqrt{\mathrm{Var}[X]} \tag{5.2}$$

The measurement uncertainty thus determined is called *standard measurement uncertainty* because on the one hand it is the standard deviation of the distribution and on the other hand it is such a fundamental description of the mean width of the distribution that it appears in many formulae.

The distributions which are needed to represent the knowledge of the input data in the evaluation are relatively simple. Mostly one of the following five cases is given:

1. It is known only that the value of the quantity X lies somewhere between a lower limit a_- and an upper limit a_+. As this knowledge does not give preference to a value between the limits, a rectangular distribution of the half-width

$$\Delta a = a_+ - a_- \tag{5.3a}$$

with the expectation

$$E[X] = \tfrac{1}{2}(a_+ + a_-) \tag{5.3b}$$

and the variance

$$\mathrm{Var}[X] = \tfrac{1}{3}(\Delta a)^2. \tag{5.3c}$$

is obtained.

2. It is known that the quantity X is the sum or the difference of two quantities X_1 and X_2

$$X = X_1 \pm X_2 \tag{5.4}$$

for which it is also known that their values are rectangular distributed with equal half-width

$$\Delta a_1 = \Delta a_2 = \Delta a', \tag{5.5}$$

and that there is no correlation in the knowledge about their variability. The convolution of the two rectangular distributions yields a triangular distribution of half-width (sum of the half-widths of the two rectangular distributions) (Fig. 8)

$$\Delta a = \Delta a_1 + \Delta a_2 = 2\Delta a' \tag{5.6}$$

with the expectation

$$E[X] = E[X_1] \pm E[X_2] \tag{5.7a}$$

and the variance

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \mathrm{Var}[X_1] = \frac{(\Delta a)^2}{6} \tag{5.7b}$$

3. It is known that the quantity X is the sum or the difference of two quantities X_1 and X_2

$$X = X_1 \pm X_2 \tag{5.8}$$

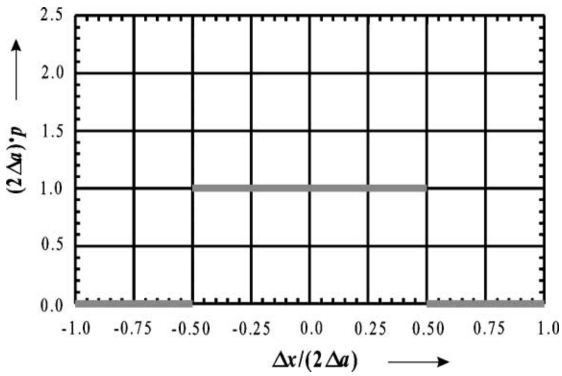

Fig. 8. Rectangular distribution of half-width Δa for the deviations from the expectation value of the quantity X.

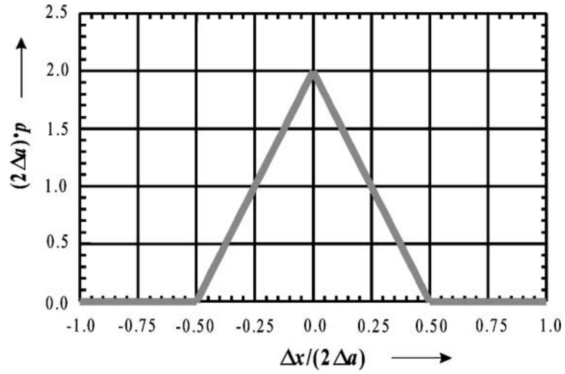

Fig. 9. Triangular distribution of the half-width Δa for the deviations from the expectation value of the quantity X.

and it is also known that their values are rectangular distributed but with different half-widths, and that there is no correlation in the knowledge about their variability. In this case, the convolution of the two rectangular distributions yields a trapezoidal distribution with half-width

$$\Delta a = \Delta a_1 + \Delta a_2 \tag{5.9}$$

equal to the sum of the half-widths of the two rectangular distributions and a form parameter

$$\beta = \frac{|\Delta a_1 - \Delta a_2|}{\Delta a_1 + \Delta a_2} \tag{5.10}$$

which equals the ratio of the difference between the half-widths of the two rectangular distributions and the sum of the half-widths. It also has the expectation value (Fig. 9):

$$E[X] = E[X_1] \pm E[X_2] \tag{5.11a}$$

but the variance:

$$\mathrm{Var}[X] = \mathrm{Var}[X_1] + \mathrm{Var}[X_1] = \frac{(\Delta a)^2}{6}(1 + \beta^2). \tag{5.11b}$$

The rectangular and the triangular distribution are special cases of trapezoidal distributions with the form parameter β=1 and β=0, respectively (Fig. 10).

4. It is known that the quantity X dependent on the sine of the phase Φ:

$$X = \Delta a \cdot \sin(\Phi), \tag{5.12}$$

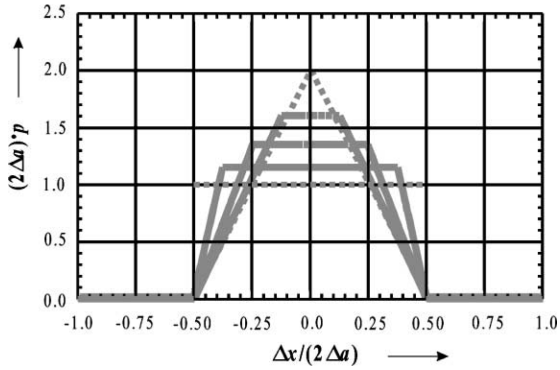

Fig. 10. Trapezoidal distributions of half-width Δa for the deviations from the expectation value of the quantity X for different form parameters β.

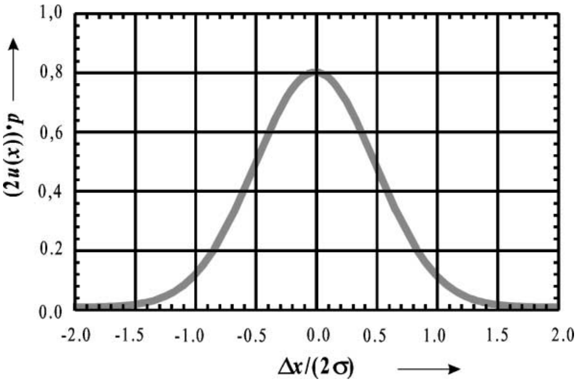

Fig. 12. Bell-shaped normal distribution of the standard deviation σ for the deviations from the expectation value of the quantity X.

whose value is not known at all. It thus is known only that the values are rectangular distributed in the interval $-\pi \ldots +\pi$. As a result, a U-shaped distribution (Fig. 11) of half-width Δa with the expectation:

$$E[X] = 0 \tag{5.13a}$$

and the variance:

$$\mathrm{Var}[X] = \tfrac{1}{2}(\Delta a^2) \tag{5.13b}$$

is obtained.

5. For the quantity X the best estimate μ and the standard deviation σ attributed to it are known. A bell-type normal distribution with the expectation:

$$E[X] = \mu \tag{5.14a}$$

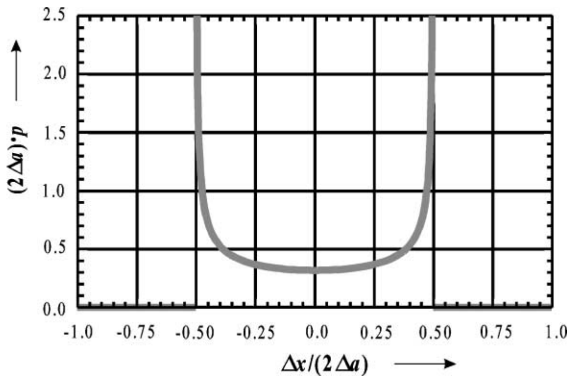

Fig. 11. U-shaped distribution of half-width Δa for the deviations from the expectation value of the quantity X.

and the variance:

$$\mathrm{Var}[X] = \sigma^2 \tag{5.14b}$$

results.

The latter case is encountered especially in the evaluation of repeated observations of a quantity—such as the indication of a measuring instrument—by statistical methods. In this case the best estimate is identified with the arithmetic mean and the associated standard measurement uncertainty with the experimental standard deviation of a single observation (type A method of evaluation).

6. The method of the GUM

This section gives an overview only how a straight forward method of evaluating measurement uncertainties is derived from the concept of the GUM and stated in the last section. It largely follows the description [19] and is given here for reasons of completness only.

The data relevant to the measurement and the measurement value to be assigned to the measurand are linked by the model of evaluation Eq. (3.1). According to the GUM approach, the expectation values of the distributions

$$x_1 = E[X_1], x_2 = E[X_2], \ldots, x_N = E[X_N] \tag{6.1}$$

are the best estimates of the input quantities for the evaluation from which the searched value y of the result quantity of the evaluation is obtained by insertion into the model of evaluation:

$$y = f(x_1, x_2, \ldots, x_N). \tag{6.2}$$

The standard uncertainties of measurement to be attributed to the input values are obtained as square roots from the variances of the distributions:

$$u(x_1) = \sqrt{\mathrm{Var}[X_1]}, u(x_2) = \sqrt{\mathrm{Var}[X_2]}, \ldots,$$
$$u(x_N) = \sqrt{\mathrm{Var}[X_N]} \tag{6.3}$$

The standard measurement uncertainty which is to be attributed to the measurement result is obtained in two steps: first the uncertainty contributions

$$u_1(y) = c_1 u(x_1), u_2(y) = c_2 u(x_2), \ldots,$$
$$u_N(y) = c_N u(x_N) \tag{6.4}$$

stemming from the individual influence quantities, are to be calculated with the aid of the sensitivity coefficient

$$c_1 = \left.\frac{\partial f}{\partial X_1}\right|_{\underline{X}=\underline{x}}, \quad c_2 = \left.\frac{\partial f}{\partial X_2}\right|_{\underline{X}=\underline{x}}, \ldots, c_N = \left.\frac{\partial f}{\partial X_N}\right|_{\underline{X}=\underline{x}} \tag{6.5}$$

to be derived from the model of evaluation. Subsequently, the searched standard measurement uncertainty is obtained as square root from the quadratic sum of all uncertainty contributions

$$u(y) = \sqrt{\sum_{i,j=1}^{N} u_i(y) r(x_i, x_j) u_j(y)} \tag{6.6}$$

in which the correlation coefficients of the input quantities occur. They are defined as the ratio:

$$r(x_i, x_k) = \frac{Cov[X_i, X_k]}{\sqrt{\mathrm{Var}[X_i]} \times \sqrt{\mathrm{Var}[X_i]}} \tag{6.7}$$

with $\mathrm{Cov}[X_i, X_k]$ being the covariance of the input quantities X_i and X_k determined from the distributions. The correlation coefficients are the measure in which potential dependencies of the knowledge of the input quantities for the evaluation are expressed. Their value lies in the range:

$$-1 \le r(x_i, x_k) \le 1 \tag{6.8}$$

In the most frequent case that the knowledge of the input quantities can be considered non-correlated, the correlation coefficients have the values:

$$r(x_i, x_j) = \begin{cases} 0, \text{ when } & i \neq j \\ 1, \text{ when } & i = j \end{cases} \tag{6.9}$$

Eq. (6.6) then changes into the well-known equation:

$$u(y) = \sqrt{u_1^2(y) + u_2^2(y) + \ldots + u_N^2(y)} \tag{6.10}$$

which is used in most cases, i.e. the standard measurement uncertainty attributed to the measurement result is obtained as square root from the sum of the squares of the uncertainty contributions.

The values and standard uncertainties of measurement for the input quantities are determined according to the knowledge available by two different methods. Evaluation method B directly uses procedures of the theory of probabilities. It is applied when a statistical evaluation cannot be made but profound metrological knowledge allowing the distribution of the values of an input quantity to be estimated is available. Knowledge falling into this category are:

- data from previous measurements,
- experience or general knowledge of the behaviour and the characteristics of measuring instruments or materials,
- manufacturer's specifications for measuring instruments and standards,
- statements in calibration certificates and other certificates,
- statements of values in manuals
- etc.

An appropriate realisation and use of the information available naturally will be possible only if sufficient experience and general knowledge of the technical area in question is available. The expert estimate is a skill which is acquired in the metrological practice. A well-founded, realistic estimate of the measuring situation can furnish values for the standard measurement uncertainty which are not inferior to those obtained by type A evaluation. This will be true especially if type A evaluation can be carried out only for a very small number of observations or on principle cannot be used. The latter is the case, e.g. for (digitally) indicating measuring instruments. Here, a residual ignorance of the conventional true value due to the finite scale interval will persist even if repeated observations do not reveal fluctuations.

For the application a distinction is made between the two following cases:

1. If only an individual value for the quantity X_i is known such as, e.g. an individual measurement value, a value resulting from a former measurement, a reference value from literature or a correction, this value is to be used as best estimate x_i. If the standard measurement uncertainty $u(x_i)$ is also given, this is to be used. Otherwise, it is to be determined from unambiguous knowledge of potential deviations from the value x_i. If data of this kind are not available, a suitable value for the standard measurement uncertainty must be empirically estimated.
2. If for the quantity X_i a distribution of the values can be stated from theoretical or empirical knowledge, the expectation value derived from it is to be used as best estimate x_i and the square root of the variance as the standard uncertainty $u(x_i)$ assigned to it. If for the potential values only upper and lower limits a_+ and a_- can be estimated (e.g. manufacturer's specification for a measuring instrument, a range of variability or vagueness of temperature, a rounding or truncation error due to data pre-processing), this knowledge corresponds to the rectangular distribution stated in Section 5.

The type A method of evaluation uses statistical procedures to obtain the best estimate and the standard measurement uncertainty. It will be used if one or several input quantities are observed in the measurement several times under unchanged conditions and different values are determined. In this case, the distribution of the values and the uncertainty about the value of the quantity are obvious. If a number of independent observations are carried out under unchanged conditions for only one of the input quantities and if the measurement procedure has a sufficient resolution, the values observed generally show a scatter. This is due to the fact that the conditions of the measurement as regards the resolution are not kept sufficiently constant. If Q is the input quantity repeatedly observed and if n statistically independent observations ($n>1$) were carried out, the best estimate is the arithmetic mean of the observed values q_j ($j = 1, 2, \ldots, n$)

$$\overline{q} = \frac{1}{n}\sum_{j=1}^{n} q_j \tag{6.11}$$

whereas the standard measurement uncertainty is determined by one of the procedures that are discussed below.

Strictly speaking, a frequency distribution is seen in the observations. From this the properties of the process leading to the scatter must be derived. This in turn is an estimate of the parameters of the underlying probability distribution. These estimates are to be distinguished from the estimates on which the procedure of the GUM is based. The additional estimates to be made for repeated observations are necessary to draw conclusions from the values observed for the background process. When this step is made and the knowledge of the values observed is converted via statistical evaluation into a best estimate and the standard measurement uncertainty associated to it, the procedure of the GUM is used again.

1. The variance of the distribution underlying the observations is estimated from the empirical variance $s^2(q)$ of the observed values q_j. It is given by

$$s^2(q) = \frac{1}{n-1}\sum_{j=1}^{n}(q_j - \overline{q})^2 \tag{6.12}$$

   Its (positive) square root is the empirical standard deviation of the individual observation. It characterises the width of the distribution of the observations. The estimate for the variance needed in the uncertainty analysis is the empirical variance of the mean value. The standard measurement uncertainty is its (positive) square root. It is referred to as empirical standard deviation of the mean value:

$$u(\overline{q}) = \frac{s(q)}{\sqrt{n}} \tag{6.13}$$

   If the number n of the repeated observations is small ($n<10$), the statistical reliability of the value of the empirical standard deviation of the mean value must be taken into account. If the number n of the observations cannot be increased, other statistical methods such as the method of the effective degrees of freedom must be included in the determination of the measurement uncertainty [7].
2. If for a measurement which is carried out repeatedly under the same conditions and in a statistically controlled way a combined variance

s_p^2 of the individual observation is available, this characterises the statistical influences of the respective measurement procedure. Its value will therefore generally better describe the required variance than a value obtained from a small number of observations. In this case, one will continue to assign to the quantity Q the arithmetic mean of the smaller number n of observations as the best estimate but define the standard measurement uncertainty by the equation:

$$u(\bar{q}) = \frac{s_p}{\sqrt{n}} \tag{6.14}$$

Here s'_p is the combined standard deviation from series of measurements carried out previously under the same conditions and n' the total number of combined observations.

7. The industrial view: expanded measurement uncertainty

As briefly discussed in Section 4, it is a common task of business to compare a measurement value with boundary values given in specifications or by regulations. When decisions on the conformity are made, it is to be taken into account whether the value is reliable or only just lies within the boundaries. According to the GUM approach, the standard measurement uncertainty is the universal parameter for characterising the quality of a measurement or of the measurement value but it is not suitable for proving the conformity. This proof requires not only a quality characteristic but also a range encompassing a large fraction of the values which are compatible with the conditions of measurement and can be regarded as a value of the measurand. A range of this kind is necessary when health or safety aspects are involved. For this reason, in industry and trade, the *expanded measurement uncertainty* which is derived from the standard measurement uncertainty is used. It is defined as the product

$$U = k_P \cdot u(y) \tag{7.1}$$

with the *coverage factor* k_P. The coverage factor is chosen such that the uncertainty interval

$$I_Y = [y - U; y + U] \tag{7.2}$$

covers the desired high fraction of values. The covered fraction is called *coverage probability P.* As with the naive approach, a range of values is available which can be used for comparisons. The determination of the expanded measurement uncertainty is, however, carried out via the determination of the standard measurement uncertainty and thus relies on a procedure of the theory of probabilities. On the one hand, this allows the aspects of frequency statistics and evaluation probability to be consistently reconciled with one another, and on the other hand, any refinement of the considerations leads to the measurement uncertainty being reduced as the influences are split into non-correlated components due to the statistical superposition.

In their calibrations, the calibration services forming the EA—among them the DKD—use the coverage factor for the coverage probability $P = 0.95$. Though when the GUM method is used the distribution of the values of the measurand Y is not determined, the coverage factor can be determined in more precise measurements on the basis of the central limiting value theorem. Generally, in the uncertainty analysis for these measurand several essential influence quantities with uncertainty contributions of the same order of magnitude occur. In these cases the distribution resulting for the measured is in good approximation a bell-shaped normal distribution (Fig. 12). For this the standard coverage factor $k_{0.95}=2$ is obtained. An expansion of the procedure to many cases in which the stated condition does not apply—if, e.g. only one or two dominant uncertainty contributions are made out—is shown by the examples of EA-4/02, [9].

The complete measurement result consists of the determined measurement result y and the expanded measurement uncertainty calculated according to Eq. (7.1). In some cases, it may be reasonable to use the relative standard measurement uncertainty

$$W = \frac{U}{|y|} \tag{7.3}$$

related to the measurement value. In practice one will find relative uncertainty statements related to quite different things: measurement values, nominal values, indicated values etc. To give an unambiguous statement the value or quantity of reference to which the uncertainty statement relates should always made clear. This avoids also an incorrectness often found in practice: the use of the adjective 'absolute' to characterise the normal uncertainty statement not related to any value. Since the uncertainty evaluation is based on subjective judgements this is nonsense per se.

In the following, by the example of the thermometer calibration from Section 3, the three forms are given in which the complete measurement result can be stated (measurement result and expanded measurement uncertainty with the same unit, measurement result and expanded measurement uncertainty with different units or measurement result and relative expanded measurement uncertainty). Together with the complete result the conditions are to be stated under which the measurement result has been obtained.

At the indication of 20 °C, the error of indication of the calibrated mercury-in-glass thermometer is

$$(-0.14 \pm 0.1)\,^{\circ}\mathrm{C},\ -0.14\,^{\circ}\mathrm{C} \pm 100\,\mathrm{mK},$$
$$\text{or } -0.14\,^{\circ}\mathrm{C}(1 \pm 0.7)$$

The measurement uncertainty stated is the expanded uncertainty (relative expanded uncertainty for the third form) which is obtained from the standard uncertainty by multiplication by the coverage factor k=2. For a normal case it corresponds to a coverage probability of 95%. The standard measurement uncertainty has been determined in accordance with the Guideline EA-4/02.

Two points must be observed when the values are stated, [7]: On the one hand, the measurement uncertainty is based on a probability statement. Therefore, two significant digits should be stated at most. Probabilities can generally be stated only with an accuracy of 3–4%. Statistical statements such as, e.g. the probability for the occurrence of one among several events can be given with an accuracy of 1% only in most favourable circumstances. To find out, e.g. with the accuracy stated whether a given dice behaves ideally, i.e. its sides with the different numbers of spots face upwards with the same probability of 1/6, the dice would have to be thrown more than 10,000 times.

If the uncertainty calculations yield more than two significant digits, mathematical rounding has to be made to the suitable number of digits. If the change for the rounding exceeds 5%, however, rounding up of the value should be used. In the sense of the uncertainty analysis, the number of the stated digits of the measurement result is to be limited to the least significant digits of the standard measurement uncertainty.

8. Uncertainty analysis—how?

The GUM method described in Section 6 leads to a procedure which is based on the evaluation of the knowledge of the influence quantities of a measurement and the applications of Eq. (6.2) and Eq. (6.6). The result is a basic structure for the uncertainty analysis which is given in the standard DIN 1319–3 (Section 4.2, [2]) and reads more or less as follows:

1. setting-up of an model of evaluation relating the measurand of interest—the output quantity of the evaluation—to the other quantities involved in the measurement—the input quantities of the evaluation;
2. preparation of the given measurement values and other available data from the particular knowledge of the measurement and its conditions;
3. calculation of the measurement result and the measurement uncertainty to be attributed to it from the prepared data by means of the model of evaluation;
4. statement of the complete measurement result for the measurand, consisting of the measurement value and the measurement uncertainty attributed.

The discussions about the collections of examples of the EA [8,9], have shown that these points are to be specified by a more detailed subdivision.

1. setting-up of an model of evaluation:
   - title characterising the measurement;
   - brief description of the measurement procedure with reference to the measurement principle and the sequence of the measurement;
   - model of evaluation as a mathematical relation with the aid of which the value of the measurand is determined from the values of the other quantities involved in the measurement. It can consist of several equations describing sub-models. This form of representation is favourable for ample models as it offers increased clearness.

A model can consist of a calculation algorithm or instructions for action in the form of a descriptive text. What is essential is that the model of evaluation describes the relation between the output quantity and the input quantities for the evaluation with which the values of the output quantity can be unambiguously determined from the values of the input quantities;
- ○ list of the symbols and quantities used in the model of evaluation with a brief definition or explanation of their meaning.

2. preparation of the given measurement values and other data required for evaluation:
   representation of the knowledge of the input quantity of the evaluation such as
   - ○ statement of sources (calibration certificate, manufacturer's certificate and the like) from which the values have been taken;
   - ○ estimates of the ranges of variation of the potential deviations whose values are not exactly known;
   - ○ table of observed values;
   - ○ statement of standard uncertainties of measurement derived. Here, the list form has proved to be suitable as it can be adjusted to different requirements and metrological situations. For each quantity occurring in the model of evaluation a separate section should be provided. For the type A evaluations, not only the list of values observed but also the characteristics of the statistical evaluation;
   - ○ arithmetic mean;
   - ○ experimental standard deviation of individual observation; and
   - ○ experimental standard deviation of arithmetic mean should be stated.
3. calculation of the measurement result and of the standard measurement uncertainty to be attributed to it, from the prepared data:
   This part which is to be considered the uncertainty budget in the narrow sense is advantageously realised in the tabular form. It is recommended to extend the table given in EA-4/02 (DKD-3) to the columns
   - ○ quantity,
   - ○ value,
   - ○ associated standard measurement uncertainty,
   - ○ effective degrees of freedom,
   - ○ form of distribution,
   - ○ sensitivity coefficient,
   - ○ uncertainty contribution.
4. statement of the complete measurement result:
   statement of the conditions of measurement essential to the definition of the measurand and of the measurement result, together with the associated expanded measurement uncertainty and the coverage factor chosen for its determination, e.g. in the form
   measurement result $\pm$ expanded measurement uncertainty (k = coverage factor).
   The expanded measurement uncertainty is calculated here as the half-width of an interval of values covering the (great) fraction of 95% of the values which, under the conditions of the measurement, can reasonably be attributed to the measurand.

By their nature, these instructions apply only to a detailed uncertainty analysis which certainly need not be made for each routine measurement. They rather apply to the general description of the measurement procedure and should be contained in the quality manual or an equivalent record transparently documenting the measurement procedure. The calculation steps to be made under (3) and (4) are a necessary result of the GUM method. They can be easily translated into a computer program. This also applies to the part of point (2) dealing with the determination of the value, and of the standard uncertainty to be attributed to it, of the individual input quantities for the evaluation. As soon as the distribution of the values and of their parameters is determined, the calculations are established. The translation into professional support by the computer is shown in Fig. 13 by the example of the calibration of a mercury-in-glass thermometer from Section 3, [19] (Fig. 13).

The translation into a computer program provides human resources for other tasks, i.e. for those parts of the uncertainty analysis for which suitable support by the computer is not yet known: the *setting-up of the model of evaluation* and the *preparation of the data*, in particular the evaluation of the knowledge. The evaluation of the conditions of measurement is closely linked with the respective measurement procedure. As it generally plays an essential part in the preparation of the data, it can be made only on the site by the

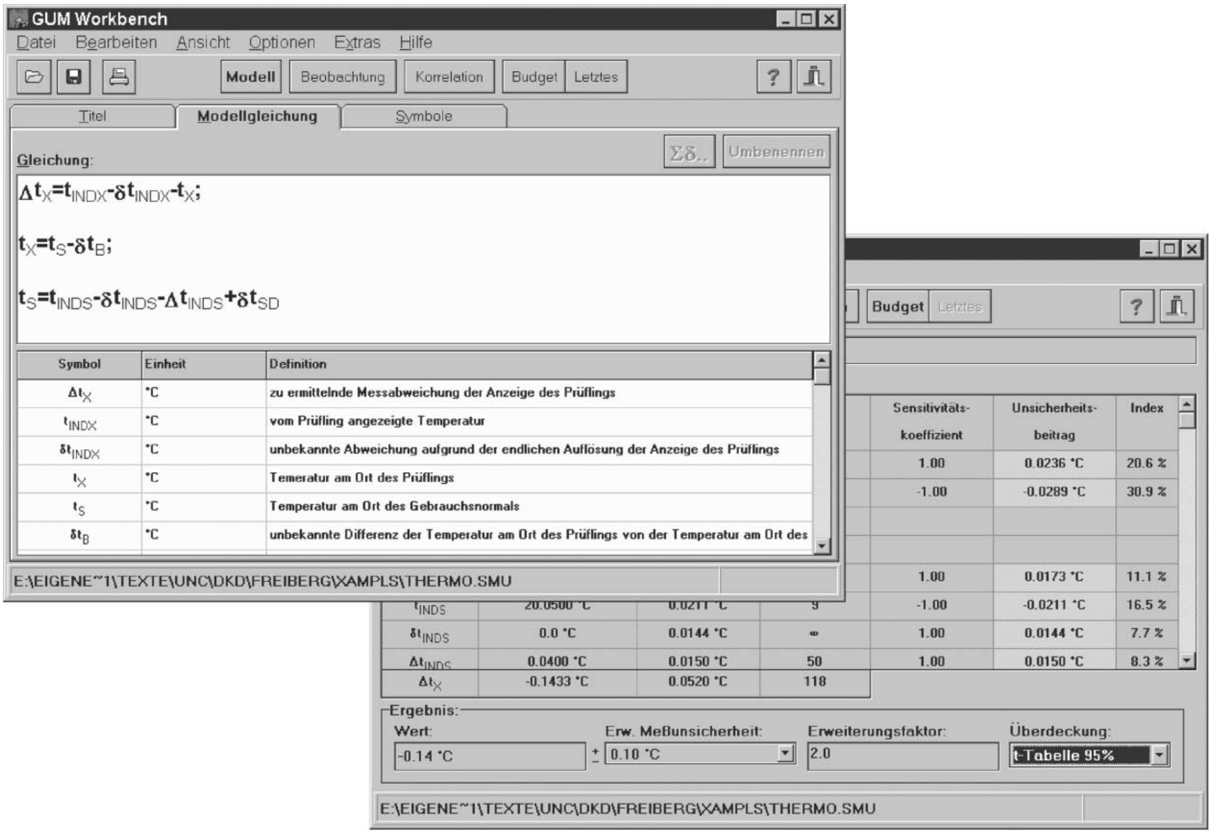

Fig. 13. Screen views "Model equation" and "Measurement uncertainty budget" of the GUM workbench program package for the calibration of a mercury-in-glass thermometer, example discussed in Section 3.

expert with his metrological experience. Apart from the well-known checklists, general support is scarcely possible here. This is not so for the setting-up of the model of evaluation. Each measurement is based on a specific realisation and thus an individual model of evaluation but there are typical partial tasks in almost all measurement procedures. In most cases they differ only by the quantities involved and are to be implemented in the individual models in a similar way. This may be demonstrated by the example of the calibration of a mercury-in-glass thermometer from Section 3. It is a special example of a class of problems: the calibration of an indicating instrument by use of another indicating instrument that has be calibrated before and serves as working standard in the actual calibration. Calibration examples of this type realising the same cause-effect structure may easily be found in many other fields of metrology such as dimensional, electrical or optical metrology.

Acknowledgements

I wish to thank my PTB colleagues who never tired of answering my questions about their measurement procedures. I wish to thank also the members of the PTB's Measurement Uncertainty Steering Group who helped me by discussions of a great number of examples to find a more transparent approach to certain problems of the measurement uncertainty. I owe particular thanks to the European colleagues of the EA Task Force for Measurement Uncertainty and W. Wöger (PTB) for many stimulating discussions which have contributed to my understanding of the GUM.

And last not least I owe special thanks to Mrs. H. Mewes from PTB's translation office for her support to say what I want to say in English.

References

[1] International Vocabulary of Basic and General Terms in Metrology. 2nd Edition, 1993, International Organisation for Standardisation, Geneva, Switzerland.
[2] DIN 1319 Grundlagen der Messtechnik, Teil 1: Grundbegriffe (1995), Teil 2: Begriffe für die Anwendung von Messgeräten (in Arbeit), Teil 3: Auswertung von Messungen einer einzelnen Messgröße, Messunsicherheit (1996), Teil 4: Auswertung von Messungen, Messunsicherheit (1999). Deutsches Institut für Normung—Beuth-Verlag.
[3] Guide to the Expression of Uncertainty in Measurement, 1st Edition, 1993,: corrected and reprinted edtion, 1995, International Organisation for Standardisation, Geneva, Switzerland.
[4] W. Kessel, European and International standards for statements of uncertainty, Eng. Sci. Educ. J. 7 (1998) 201.
[5] W. Kessel, Messunsicherheit und Messwert nach der neuen ISO/BIPM-Leitlinie, tm-Technisches Messen 62 (1995) 306.
[6] W. Kessel, Der ISO/BIPM-Leitfaden zur Ermittlung der Messunsicherheit, tm-Technisches Messen 68 (2001), Seite 5–13.
[7] EA-4/02 Expression of the Uncertainty of Measurement in Calibration, requirements document, 1st Edition, 1996, European co-operation for Accreditation, Utrecht, The Netherlands.
[8] EA-4/02-S1 Supplement 1 to EA-4/02 (Examples), guidance document, 1st Edition, 1996, European co-operation for Accreditation, Utrecht, The Netherlands.
[9] EA-4/02-S2 Supplement 2 to EA-4/02 (additional Examples),guidance document, 1st Edition, 1999, European co-operation for Accreditation, Utrecht, The Netherlands.
[10] W. Kessel, Messunsicherheit, ein wichtiges Element der Qualitätssicherung, PTB Mitteilungen 108 (1998), Seite 377–382.
[11] R. Kaarls, Measurement Uncertainty in Calibration, Measurement and Testing in proceedings of the Eurolab Workshop and Seminar on Measurement Uncertainty in Testing, Barcelona, 1992.
[12] ISO EN DIN 14 253 Geometrical Product Specification (GPS)—Inspection by Measurement of workpieces and measuring equipment, Part 1: Decision rules for proving conformance or non-conformance with specifications, 1998, International Organisation for Standardisation, Geneva, Switzerland.
[13] Tischler, K. Entscheidungsregeln zur Festlegung der Übereinstimmung mit geometrischen Produktspezifikationen (GPS), MICROTECH 1996, Rahmenprogramm im Auditorium, Frankfurt Oktober 1998 VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, Düsseldorf 1997.
[14] G. Spencer, Probability and Scientific Inference, Longmans, Ginn & Co, London, 1957.
[15] R. T. Cox, The Algebra of Probable Inference, The Johns Hopkins University Press, Baltimore, 1961.
[16] M. Tribus, Rational Descriptions, Decisions and Designs, Pergamon Press, New York, 1969.
[17] K. Weise, W. Wöger, A Bayesian theory of measurement uncertainty, Meas. Sci. Technol. 4 (1993) 1–11.
[18] E. T. Jaynes, Where do we stand on maximum entropy? in: E.T. Jaynes, R.D. Rosenkrantz (Eds.), Proceedings of the Papers on Probability, Statistics and Statistical Physics, Reidel Publishing Company, Dordrecht, 1983, pp. 210–314.
[19] At the 139th PTB Seminar Typical problems of the dissemination of HF measurands (May 1998), the professional GUM workbench program package was presented which assists the user in the performance of an uncertainty analysis. In the determination of the complete measurement result it follows the recommendations of the Guideline EA-4/02 (DKD-3) and additionally furnishes a printed report which meets the requirements for documentation in quality assurance. More detailed information as well as demo versions can be obtained under http://www.metrodata.de.