Estimation of measurement uncertainties - an alternative to the ISO Guide

Estimation of measurement uncertainties – an alternative to the ISO *Guide*

M. Grabe

Abstract. The procedure for uncertainty assessment set out in the *Guide to the Expression of Uncertainty in Measurement*, published by the International Organization for Standardization (ISO), is compared with an alternative system. The ISO *Guide* links the overall uncertainty to a so-called k_P factor, implicitly expressing the degree of confidence that the measured result should cover the true value of the physical quantity in question. Here it is argued that this standard procedure, in which the magnitude of a given k_P factor is tied to the degree of confidence, presents the experimenter with difficulties. These arise from the need to associate a probability density with so-called *unknown systematic errors*. Consequently, the recommended procedures may not be taken up by experimenters who carry out laboratory measurements on a daily basis.
A simplification of the ISO procedure suggested by the *Guide* involves the assignment of the values 1 or 2 (or even 3) to k_P. The alternative approach presented here argues that this simplification does not reflect the prevailing physical situation, and that the degree of confidence obtained lacks physical objectivity. Stationary measurement processes, strictly separating random and unknown systematic errors, are considered. Random errors are assumed to be normally distributed, and no probability distribution density is attributed to the unknown systematic errors. As random and systematic errors are kept separate, it is possible to express the influence of random errors by generalized confidence intervals (from Student's t-distribution) and the influence of systematic errors by worst-case estimates. No confidence statement is associated with the overall uncertainties, and the intrinsic problem of the *Guide* is circumvented. The formalisms are robust, transparent, and locate measured quantities with reasonable reliability.

1. Introduction

When Karl Friedrich Gauss conceived his contributions to error calculus, random errors dominated the field. Though he recognized so-called unknown systematic errors, he was in a position to ignore them. Today, metrology is different. Unknown systematic errors may no longer be neglected and, two decades or so ago, the questions most passionately discussed by the metrological community involved their classification and treatment.

Such time-constant errors, whose magnitude and sign are unknown, are expected to lie within specific boundaries. Should a systematic error be treated as a constant bias or as a random variable? According to the *Guide* [1], unknown systematic errors should be processed on essentially the same basis as random errors. As a consequence, the *Guide* is in a position to maintain at least parts of the classical structures of error calculus as laid down by Gauss.

The idea of attributing a postulated probability density to unknown systematic errors goes back to a proposal of Wagner [2]. Such a density was expected to exhaustively exploit the information implied within the measured data. But Wagner also pointed out certain practical difficulties stemming from such a probability density.

In the present paper, we suggest that unknown systematic errors could be treated just as they are, as *constant quantities*. This divergent view, probably explicitly proposed for the first time during a seminar held at the Physikalisch-Technische Bundesanstalt (Braunschweig, Germany) in 1978 [3], was supported by the edifying drawings of C. Eisenhart [4], and by occasional industrial practices [2].

Clearly, if we do not provide unknown systematic errors with a probability statement, their influence would have to be assessed by absolute values and triangle inequalities in the sense of *worst-case* estimates. These suggestions are discussed in the following sections. As we shall see, if unknown systematic errors are formally kept constant, then the way in which random and systematic errors are combined necessarily changes from geometric to arithmetic.

Another basic point relates to the propagation of random errors and here to the numbers of repeated measurements. For simplicity, let us look at just two

M. Grabe: Am Hasselteich 5, 38104 Braunschweig, Germany.

variables, x and y, being normally distributed. The idea is to postulate the existence of a *Fisher-Romanowsky* density [5], for the five empirical quantities $\bar{x}$, $\bar{y}$, s_x^2, s_y^2, s_{xy}, denoting the arithmetic means, the variances and the covariance. Clearly, the prerequisite is the strict requirement to have equal numbers of repeated measurements in x and y! As shown below, this postulate puts the propagation of random errors on sound ground, whether or not x and y are dependent.

On the other hand, to admit only equal numbers of repeated measurements is strictly rejected by experimentalists, as they do not want to lose "useful information" – where the latter refers to the *excessive number* of repeated measurements in either x or y. So the choice is as follows: either we back the propagation of random errors on a sound formalism, i.e. accept the postulate of a Fisher-Romanowsky density, thereby processing equal numbers of repeated measurements in x and y; or we run into tremendous difficulties when attempting to propagate random errors, so that, as a consequence, we have to live with doubtful uncertainty statements.

Furthermore, the decision to admit only equal numbers of repeated measurements and the additional precaution of relying exclusively on *empirical* variances and *empirical* covariances leads directly to *Hotelling's distribution density* [5], a well-established tool of statistics which strangely enough has been ignored in metrology. Hotelling's distribution density elucidates and solves many seemingly intricate problems of error calculus [6-8].

The authors of the *Guide* also note the premise of equal numbers of repeated measurements, but only in relation to the definition of empirical covariances between random errors and not as a means of introducing the postulate of a Fisher-Romanowsky density.

2. Basic relationships

An officially recommended procedure for incorporating unknown systematic errors in uncertainty assignments was not available until Recommendation 1 of the International Committee of Weights and Measures in 1981 [1]. Wagner's proposal [2], based on the *principle of equal ignorance* (Bayes and Laplace), further elaborated by Wöger [9], by invoking the *principle of maximum entropy,* assigned a rectangular distribution density to such errors. These considerations were accepted by the *Guide* [1], thus formally preserving parts of the Gaussian error calculus.

The following alternative solution is proposed in this paper. According to the drawings by Eisenhart [4], the processing of random errors is strictly separated from that of biases [10]. In other words, the alternative model of data evaluation is based on the assumption of a *stationary measurement process*, and consequently a natural, intrinsic separation between random and systematic errors can be expected. If this assumption did not apply, then the experimenter would have to make use of the cruder methods of robust statistics.

But as long as unknown systematic errors are constant in time, the measurement process is within *statistical control*, as Eisenhart has explained, so that the experimental set-up does not exhibit a *drift*. This is the best situation an experimenter can hope for. If systematic errors drifted, the measured data would constitute a time series.

A systematic error f, unknown in magnitude and sign, is expected to lie within an interval whose boundaries are symmetric to zero. All we know is that f is a constant and

$$-f_s \le f \le f_s \qquad (f_s \ge 0). \tag{1}$$

If such an error interval is known to be asymmetrical to zero, i.e. $f_{s_1} \le f \le f_{s_2}$, the experimenter should use this knowledge and subtract a correction $(f_{s_1} + f_{s_2})/2$ from f_{s_1}, f_{s_2} *and* the data set x_i, $(i = 1, \ldots, n)$.

Let $i = 1, \ldots, n$; then the basic equation of error analysis reads

$$x_i = x_0 + \varepsilon_i + f \qquad (\text{where } f \text{ is a constant}). \tag{2}$$

Here, x_i is the measured value, x_0 the true value, ε_i the random error and f the unknown systematic error.

Clearly, random errors enter the data while measurements are repeated. In contrast to this, the unknown systematic error f is caused a priori by the particular features of the experimental set-up, i.e. f is activated before the measurement process begins and before any of the x_i are recorded.

Unknown systematic errors are to be attributed to imperfect adjustments, to biases of the detectors, to boundary conditions, to environmental influences, etc. In no case is the experimenter in a position to control and eliminate any such temporally constant perturbation by simple arithmetic subtractions. Even Gauss was confronted with this situation. At his time, however, random errors dominated and so he ignored unknown systematic errors. Today the latter may no longer be neglected and all the experimenter can hope for is that they remain constant while n repeated measurements are carried out. Should this apply, we deduce from (2), considering the most basic estimator, the arithmetic mean $\bar{x}$; an arbitrary number of measurements does not eliminate or reduce the bias f, because

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = x_0 + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i + f.$$

Repeated measurements are apt to reduce the influence of random errors but they do not alter the bias f. Moreover, from a purely practical point of view, too many repeated measurements might be so time-consuming that the experimental set-up might start to drift. What we should keep in mind is that f is

constant during the time it takes to make n repeated measurements; the value of n may have any reasonable value but it must not be too high.

At this point, the *Guide* and the alternative error model diverge. In view of the expectation $E\{\bar{X}\}$, where $\bar{X}$ designates a random variable formally associable with the arithmetic mean $\bar{x}$, the *Guide* arrives at

$$E\{\bar{X}\} = x_0,$$

i.e. the expectation $E\{\bar{X}\}$ equals the *true value* x_0, because the symmetrical rectangular distribution density, introduced for f, erases the bias. However, if f is considered an unknown constant bias, the expectation $E\{\bar{X}\}$ yields something quite different:

$$E\{\bar{X}\} = \mu,$$

where μ is the "centre of gravity" of the distribution density of the random errors. From (2) we find

$$\mu = x_0 + f. \tag{3}$$

In the latter interpretation, the arithmetic mean $\bar{x}$ is biased[1] with respect to the true value x_0 so that the repeated measurements x_i scatter with respect to the population parameter μ. However, as the actual value of f remains unknown, both views are justified and ultimately lead to different error models.

To proceed, the combination of (2) and (3) relates the scattering of the ε_i to the expectation μ according to

$$\varepsilon_i = x_i - \mu \qquad (i = 1, \ldots, n). \tag{4}$$

For practical reasons, we include the two identities

$$x_i = x_0 + (x_i - \mu) + f$$

and

$$\bar{x} = x_0 + (\bar{x} - \mu) + f.$$

Finally, from the arithmetic mean $\bar{x}$ and (2), we find that the difference

$$x_i - \bar{x} = \varepsilon_i - \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \tag{5}$$

does not include f, i.e. the empirical variance

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

is insensitive to an unknown systematic error.

The error equations (1) to (5) constitute the basis of the idea to design uncertainties beyond the *Guide*. As these equations do not include a probability density for f, it is suggested that we should have recourse to appropriately designed worst-case estimates [10].

1. Strictly speaking, only if $f \neq 0$.

3. Equal numbers of repeated measurements

We intend to separate strictly random errors from unknown systematic errors. The influence of the first is expressed by means of generalized confidence intervals according to W. S. Gosset (who published Student's t-distribution) and the influence of the latter by means of worst-case estimates.

We now construct these confidence intervals by postulating the existence of a Fisher-Romanowsky density. In the case of two normally distributed variables, x and y, which are to be combined within a given function $\phi(x, y)$, there is a statistically based relationship between the five empirical quantities $\bar{x}$, $\bar{y}$, s_x^2, s_y^2, s_{xy}, the arithmetic means

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{6}$$

and the variances and the covariance

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2,$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2,$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})\,(y_i - \bar{y}). \tag{7}$$

The Fisher-Romanowsky density [5], $p(\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy})$ factorizes according to

$$p(\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}) = p_1(\bar{x}, \bar{y}) \times p_2(s_x^2, s_y^2, s_{xy}),$$

where $p_1(\bar{x}, \bar{y})$ is the two-dimensional normal probability density of the means $\bar{x}$ and $\bar{y}$ while

$$p_2(s_x^2, s_y^2, s_{xy}) \propto [s_x^2 s_y^2 - s_{xy}^2]^{(n-4)/2}\, \mathrm{e}^{\frac{n-1}{2|\sigma|}\, q}$$

with

$$|\boldsymbol{\sigma}| = \sigma_x^2\sigma_y^2 - \sigma_{xy}^2$$

and

$$q = \sigma_y^2 s_x^2 - 2\sigma_{xy}s_{xy} + \sigma_x^2 s_y^2$$

designates the density of the second-order empirical moments. Here, we have

$$E\{S_x^2\} = \sigma_x^2, \quad E\{S_y^2\} = \sigma_y^2, \quad E\{S_{xy}\} = \sigma_{xy},$$

the capital letters denoting random variables S_x^2, S_y^2, S_{xy} formally associable with the empirical quantities s_x^2, s_y^2, s_{xy}. The Fisher-Romanowsky density clearly presupposes equal numbers of repeated measurements for x and y.

Surprisingly, the second factor $p_2(s_x^2, s_y^2, s_{xy})$ *does not factorize, even if the random variables* X *and* Y *are independent,* which means that the three empirical quantities s_x^2, s_y^2 and s_{xy} are generally dependent – should the postulate of that density be accepted. Of course, the empirical covariance s_{xy} is a mathematically induced auxiliary quantity; nevertheless, the use of this quantity alone will avoid Welch's [6] so-called numbers of effective degrees of freedom – an awkward concept to handle and deficient in cases where X and Y are dependent (Welch's generalization of Student's problem relies on the χ^2 distribution density, which implies independent data). In the converse case, with independent data, Welch's approach may be acceptable, but it might also be considered inconsistent. On the one hand, Welch's argument is basically statistical, based on a χ^2 *distribution density*. On the other hand, when treating measured pairs (x, y), another basic requirement is to have samples of the kind (x_l, y_l), $(l = 1, 2, \ldots, n)$; for x_l, $(l = 1, 2, \ldots, n_x)$ and y_l, $(l = 1, 2, \ldots, n_y)$ with $n_x \neq n_y$ a joint distribution density would necessarily remain pathological. Consequently, a *statistical* evaluation of such data would be dubious: Welch's formalism stands firmly on statistical grounds while the present author argues that the treatment of *unequal numbers* of repeated measurements violates basic statistical ideas.

Ignoring systematic errors for the moment, we consider n repeated measurements $x_l\,(l = 1, \ldots, n)$ and n repeated measurements $y_l\,(l = 1, \ldots, n)$. Furthermore let X_l, Y_l and $\Phi(X_l, Y_l)$ designate random variables[2] formally associable with x_l, y_l, $\phi(x_l, y_l)$. How can we construct a confidence interval for $\mu_\phi = E\{\Phi(X_l, Y_l)\}$? The usual linearization leads to

$$s_\phi^2 = \frac{1}{n-1}\sum_{l=1}^{n}[\phi(x_l, y_l) - \phi(\bar{x}, \bar{y})]^2$$
$$= \left(\frac{\partial\phi}{\partial\bar{x}}\right)s_x^2 + 2\left(\frac{\partial\phi}{\partial\bar{x}}\right)\left(\frac{\partial\phi}{\partial\bar{y}}\right)s_{xy} + \left(\frac{\partial\phi}{\partial\bar{y}}\right)^2 s_y^2, \quad (8)$$

where $\partial\phi/\partial\bar{x}$ is used as an abbreviation for $\partial\phi/\partial\bar{x}|_{\bar{x},\,\bar{y}}$, etc. We consider the derivatives as constants, but this can only be a tentative approximation. Even if the function $\phi(x, y)$ behaves well enough and even if the relative uncertainties are, say, 10^{-6} or less, such an approximation calls for closer inspection. Only if the neighbourhood of the expansion point turns out to be linear would the cumbersome inclusion of higher-order terms, for which at best highly intricate approaches exist, be superfluous.

According to [11], the values $\phi(x_l, y_l)$ may be regarded as realizations of a normally distributed random variable $\Phi(X_l, Y_l)$, whether or not X and Y are dependent. Let S_ϕ denote the random variable corresponding to s_ϕ, then

$$\chi^2(n-1) = \frac{(n-1)\,S_\phi^2}{\sigma_\phi^2}$$

is definable with $\sigma_\phi^2 = E\{S_\phi^2\}$. Finally, as $\mu_\phi = E\{\Phi(X, Y)\} = E\{\Phi(\bar{X}, \bar{Y})\}$, we arrive at

$$T(n-1) = \frac{\Phi(\bar{X}, \bar{Y}) - \mu_\phi}{S_\phi/\sqrt{n}},$$

i.e. at Student's random variable T. Clearly,

$$\Phi(\bar{X}, \bar{Y}) \pm \frac{t_P(n-1)}{\sqrt{n}}\,S_\phi \quad (9)$$

is a confidence interval covering μ_ϕ with probability $P\{t_P(n-1)\}$, where t_P denotes the Student factor. This basic, simple mechanism offers a "solution" to an otherwise insoluble problem if different numbers of repeated measurements were admitted in $\bar{x}$ and $\bar{y}$. Different numbers of repeated measurements indicate, in our view, a non-statistical, ill-posed problem. Remarkably, (9) is valid whether or not X and Y are dependent. In contrast, the approximation given in [6] breaks down in the case of dependence.

On the other hand, the approach outlined here seems to contradict common practice. Experimenters tend to take advantage of all their measurements $x_l\,(l = 1, \ldots, n_x)$ and $y_l\,(l = 1, \ldots, n_y)$, where in general $n_x \neq n_y$. Thus, the processing of *unequal* numbers of repeated measurements turns out to be a pyrrhic victory: experimenters would have to resort to numbers of effective degrees of freedom to design confidence intervals.

To give further support to the view that only equal numbers of repeated measurements should be admitted, consider the following:

- If the repeated measurements x_l and y_l are simultaneously recorded in the same experiment, we have $n_x = n_y = n$, so that normally distributed pairs (x_l, y_l), $(l = 1, 2, \ldots, n)$ are automatically connected within a Fisher-Romanowsky density $p(\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy})$ – where, for the uncertainty analysis, it remains irrelevant whether or not x and y are dependent.
- If, on the other hand, the repeated measurements x_l and y_l belong to two different experiments, carried out at different locations and at different times, we have to expect, as experimenters are used to working independently, $n_x \neq n_y$. Pairs (x_l, y_l) no longer exist and, a priori, there is no density $p(\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy})$. Nevertheless, such a density may be introduced via a *postulate* provided we are willing to ignore the excessive number of repeated measurements in either x or y, so that *artificial* pairs (x_l, y_l), $(l = 1, 2, \ldots, n)$ can be created. In this situation, it goes without saying

2. Though capital and small Greek letters, e.g. Φ and ϕ, are not always easily distinguished, we keep this notation where possible.

that X and Y are expected to be independent so that the empirical covariance s_{xy} merely expresses the mathematically induced dependency between the three quantities s_x^2, s_y^2, s_{xy} maintained by the density $p_2\left(s_x^2, s_y^2, s_{xy}\right)$. Nevertheless, s_{xy} should not be ignored.

The claim to admit only equal numbers of repeated measurements is here referred to as the *compound density postulate.*

Given a postulated Fisher-Romanowsky density, the formalism leading to (9) applies to both independent and dependent variables X, Y, ... entering error propagation. Consequently, it is no longer necessary to make a distinction between dependence and independence.

However, on closer inspection, a new problem becomes apparent. If x and y are independent, we should not have just one empirical covariance but $n!$ covariances:

$$\begin{aligned}&(x_1-\bar{x})\ (x_2-\bar{x})\ \ldots\ (x_n-\bar{x}) \quad \text{fixed order}\\ &(y_{l_1}-\bar{y})\ (y_{l_2}-\bar{y})\ \ldots\ (y_{l_n}-\bar{y})\\ &\qquad\qquad\qquad n!\ \text{permutations.}\end{aligned}$$

Which covariance is the correct one? As long as the variables X and Y are independent, *each possible* combination is necessarily valid. By their very nature, confidence intervals are never of fixed lengths but are dependent on the particular sample under consideration.

Incidentally, we should bear in mind that s_{xy} is limited by

$$-s_x s_y \le s_{xy} \le s_x s_y. \tag{10}$$

But before we make use of this inequality, we should take a cursory look at Hotelling's distribution density [5]. Let us consider m arithmetic means, each involving n repeated measurements. The probability of finding a realization of Hotelling's T within a given interval t to $t+\mathrm{d}t$ is given by

$$p_T(t;\, m,\, n)\,\mathrm{d}t = \frac{2\Gamma(n/2)}{(n-1)^{m/2}\,\Gamma((n-m)/2)\,\Gamma(m/2)} \times \frac{t^{m-1}}{[1+t^2/(n-1)]^{n/2}}\,\mathrm{d}t, \tag{11}$$

where $\Gamma(\bullet)$ denote values of the Γ-function. Hotelling's density is the multidimensional analogue of Student's density, therefore the letter T is retained, writing $T(n-1)$ for one mean and $T(m,\, n)$ for m means, each encompassing n repeated measurements. For $m=1$, Hotelling's density turns over into Student's density (apart from a factor of 2).

For $m=2$, Hotelling's variable is given by

$$t^2(2,\, n) = \frac{n}{|\boldsymbol{S}|}\left[s_{yy}(\bar{x}-\mu_x)^2 - 2\,s_{xy}(\bar{x}-\mu_x)\times(\bar{y}-\mu_y) + s_{xx}(\bar{y}-\mu_y)^2\right], \tag{12}$$

where $|\boldsymbol{S}|$ designates the determinant of the empirical variance-covariance matrix of the $s_{xx}\equiv s_x^2$, $s_{yy}\equiv s_y^2$ and s_{xy} values, and

$$\boldsymbol{S} = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{yx} & s_{yy}\end{pmatrix}.$$

Given a fixed t, the variables $\bar{x}$ and $\bar{y}$ define an ellipse. When the x_l, y_l are interchanged, the variances s_{xx}, s_{yy} remain fixed while the covariance s_{xy} varies, causing the ellipse to rotate within a geometrically *fixed* rectangle. This is what happens if the empirical covariance is submitted to the aforementioned $n!$ permutations.

By means of Hotelling's density, it is ultimately possible to define realistic confidence ellipsoids regarding *physical* applications. As we see, Hotelling's ellipsoid is *not based on unknown expectations* as $\sigma_x^2 = E\{S_x^2\}$, $\sigma_y^2 = E\{S_y^2\}$, $\sigma_{xy} = E\{S_{xy}\}$, and this implies a considerable improvement over those confidence ellipsoids customarily used in error calculus. As is known, the latter are taken from the exponent of the multidimensional normal probability density and therefore they *rely on the unknowns* σ_x^2, σ_y^2, σ_{xy}.

Though (9) is quite appropriate, it may sometimes be useful to go beyond the structure of a confidence interval. Using (10), we may assess

$$2\left(\frac{\partial\phi}{\partial\bar{x}}\right)\left(\frac{\partial\phi}{\partial\bar{y}}\right)s_{xy} \le 2\left|\frac{\partial\phi}{\partial\bar{x}}\right|\left|\frac{\partial\phi}{\partial\bar{y}}\right|s_x s_y,$$

so that

$$\frac{t_P(n-1)}{\sqrt{n}}\,S_\phi \le \frac{t_P(n-1)}{\sqrt{n}} \times \left[\left|\frac{\partial\phi}{\partial\bar{x}}\right|s_x + \left|\frac{\partial\phi}{\partial\bar{y}}\right|s_y\right], \tag{13}$$

which is clearly a simpler expression. On the other hand, as $\boldsymbol{S}$ is now singular, (13) no longer expresses statistics but is simply an easy estimate, which may be useful in the context of *overall uncertainties* developed in Section 4.

4. A building kit for uncertainty assessments

The treatment of unknown systematic errors as constants entails the following: the formulae to assign measurement uncertainties need to be rigorously redesigned, the analysis of variance will fail, and in least-squares fitting, the Gauss-Markoff theorem will break down. At present, we present the new formalisms merely as options, as outlined below.

4.1 Error propagation

Given a function $\phi(x, y)$ and measurements

$$x_l = x_0 + (x_l - \mu_x) + f_x,$$
$$y_l = y_0 + (y_l - \mu_y) + f_y; \quad (l = 1, \ldots, n),$$

what is the uncertainty $u_{\bar{\phi}}$ of $\phi(\bar{x}, \bar{y})$? Linearization of $\phi(\bar{x}, \bar{y})$ with respect to the true values x_0, y_0 yields

$$\phi(x_l, y_l) = \phi(x_0, y_0) + \left[\left(\frac{\partial\phi}{\partial\bar{x}}\right)(x_l - \mu_x) + \left(\frac{\partial\phi}{\partial\bar{y}}\right)(y_l - \mu_y)\right] + \left[\left(\frac{\partial\phi}{\partial\bar{x}}\right) f_x + \left(\frac{\partial\phi}{\partial\bar{y}}\right) f_y\right]. \quad (14)$$

Again, the derivatives $\partial\phi/\partial\bar{x}$, $\partial\phi/\partial\bar{y}$ abbreviate $\partial\phi/\partial x|_{\bar{x}, \bar{y}}$, $\partial\phi/\partial y|_{\bar{x}, \bar{y}}$, here approximating $\partial\phi/\partial x|_{x_0, y_0}$, $\partial\phi/\partial y|_{x_0, y_0}$. Furthermore, the derivatives shall be considered as constants. Obviously, the expectation

$$E\{\Phi(X, Y)\} = \mu_\phi = \phi(x_0, y_0) + \left[\left(\frac{\partial\phi}{\partial\bar{x}}\right) f_x + \left(\frac{\partial\phi}{\partial\bar{y}}\right) f_y\right] \quad (15)$$

is biased with respect to the true value $\phi(x_0, y_0)$. Setting

$$\phi(x_l, y_l) = \mu_\phi + \left[\left(\frac{\partial\phi}{\partial\bar{x}}\right)(x_l - \mu_x) + \left(\frac{\partial\phi}{\partial\bar{y}}\right)(y_l - \mu_y)\right],$$

provides an opportunity to strictly separate the propagation of random errors from that of unknown systematic errors. The overall uncertainty $u_{\bar{\phi}}$ of $\phi(\bar{x}, \bar{y})$ is given by [8]:

$$u_{\bar{\phi}} = \frac{t_P(n-1)}{\sqrt{n}} \times \sqrt{\left(\frac{\partial\phi}{\partial\bar{x}}\right)^2 s_x^2 + 2\left(\frac{\partial\phi}{\partial\bar{x}}\right)\left(\frac{\partial\phi}{\partial\bar{y}}\right) s_{xy} + \left(\frac{\partial\phi}{\partial\bar{y}}\right)^2 s_y^2} + \left|\frac{\partial\phi}{\partial\bar{x}}\right| f_{s,x} + \left|\frac{\partial\phi}{\partial\bar{y}}\right| f_{s,y}. \quad (16)$$

Clearly, this expression is based on an *arithmetic combination* of random and systematic errors, as we simply added to the statistically based term $t_P(n-1)S_\phi/\sqrt{n}$ the worst-case estimate $|\partial\phi/\partial\bar{x}|\, f_{s,x} + |\partial\phi/\partial\bar{y}|\, f_{s,y}$ of the propagated systematic error $(\partial\phi/\partial\bar{x})\, f_x + (\partial\phi/\partial\bar{y}) f_y$. No probability statement should be associated with the overall uncertainty so defined.

The uncertainty (16) is to be compared with the Gaussian *geometric combination* of random and systematic errors

$$u_{\bar{\phi}} = k_P \sqrt{\left[\left(\frac{\partial\phi}{\partial\bar{x}}\right)^2 \left(\frac{s_x^2}{n_x} + \frac{f_{s,x}^2}{3}\right) + \left(\frac{\partial\phi}{\partial\bar{y}}\right)^2 \left(\frac{s_y^2}{n_y} + \frac{f_{s,y}^2}{3}\right)\right]}, \quad (17)$$

recommended by the *Guide*. The factor k_P expands or compresses the overall uncertainty expressing different *degrees of confidence*. Unfortunately, the formalisms backing the selection of k_P appear rather intricate, so that it is doubtful whether they will really be taken up to evaluate daily laboratory measurements.

To simplify the situation, the *Guide* addresses $k_P = 1$ as the $1\,\sigma$ *standard uncertainty* and $k_P = 2$ as the *expanded uncertainty*, where detailed analyses of the associated degrees of confidence remain open. Though it may appear superfluous to prove by means of computer simulations that $1\,\sigma$ *standard uncertainties* generally turn out to be too small with regard to the basic aim of reliably localizing the true values of the physical quantities to be measured, we shall nevertheless do so, remarking that the set of fundamental physical constants is based on $1\,\sigma$ *standard uncertainties*. On the other hand, as we shall see, expanded uncertainties corresponding to $k_P = 2$ appear, at least in general, to be fairly reliable.

To demonstrate the properties of the *Guide* and alternative uncertainties by means of computer simulations we "measure" the area of a rectangle by simulating data (Figure 1). Let the true value of the horizontal side be $x_0 = 1$ cm and the true value of the vertical side be $y_0 = 2$ cm.

The scattering of the random errors of the simulated x and y data has been so adjusted that their variances

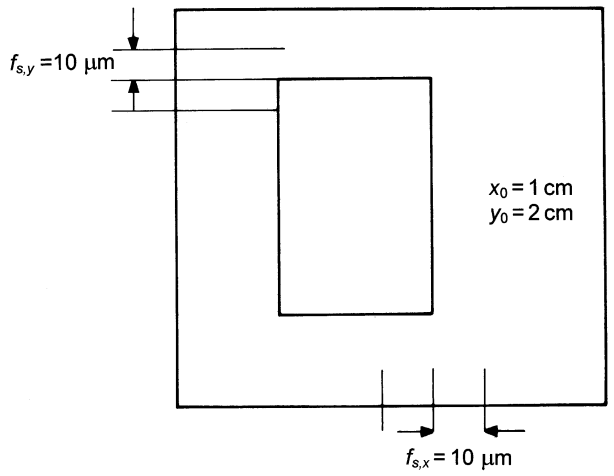

Figure 1. Simulated measurement of a rectangle of area $x_0 y_0$; $f_{s,x}$ and $f_{s,y}$ designate the boundaries of the unknown systematic errors f_x and f_y.

σ_x^2, σ_y^2 are equal to the *Guide* variances $\sigma_{f_x}^2 = f_{s,x}^2/3$ and $\sigma_{f_y}^2 = f_{s,y}^2/3$ – the latter indicating the variances of f_x, f_y in the sense of attributed rectangular distribution densities.

Let the boundaries of the unknown systematic errors be $f_{s,x} = 10\ \mu\text{m}$ and $f_{s,y} = 10\ \mu\text{m}$; *these boundaries* are to be inserted into the uncertainty formulae. But on to the data we superpose *any values* f_x, f_y lying *within* the assigned boundaries $\pm f_{s,x}$ and $\pm f_{s,y}$. Then we check whether the "measured" result $\bar{x}\,\bar{y} \pm u_{\bar{x}\,\bar{y}}$ covers the a priori known true value $x_0\,y_0$. We perform $n = 10$ repeated "measurements". For the *Guide* uncertainties we choose $k_P = 1$ and $k_P = 2$. For the alternative uncertainties we propose $t_{68\%}(9) = 1.04$ and $t_{95\%}(9) = 2.26$.

The uncertainty of the "measured" result $\bar{x}\,\bar{y} \pm u_{\bar{x}\,\bar{y}}$ is either given by

$$u_{\bar{x}\,\bar{y}} = k_P \sqrt{\bar{y}^2 \left[\frac{s_x^2}{n_x} + \frac{f_{s,x}^2}{3}\right] + \bar{x}^2 \left[\frac{s_y^2}{n_y} + \frac{f_{s,y}^2}{3}\right]} \quad \textit{(Guide)}$$

or, alternatively, by

$$u_{\bar{x}\,\bar{y}} = \frac{t_P\,(n-1)}{\sqrt{n}} \sqrt{\bar{y}^2\,s_x^2 + 2\,\bar{x}\,\bar{y}\,s_{xy} + \bar{x}\,s_y^2} + \bar{x}\,f_{s,x} + \bar{y}\,f_{s,y} \qquad \text{(as proposed here).}$$

Tables 1 and 2 show the results.

It is reasonable to assume that not every experimenter will be content with the *Guide* uncertainty, as it strongly depends on the factor k_P. As stressed above, the formalisms establishing a relationship between k_P values and corresponding degrees of confidence are indeed difficult to deal with. On the other hand, if treated inappropriately, physical objectivity may be affected.

A disadvantage of the approach proposed here is that (16) can occasionally be expected to express overestimations. Nevertheless the uncertainty expressed is perhaps the smallest uncertainty that covers the true value with reasonable reliability. This suggestion is closely linked to the premise that the given experiment fits the underlying error model, namely

- f is a constant, corresponding to stationary statistics, and
- random errors are normally distributed.

I say *reasonable reliability* because the worst-case property of the triangle inequality used in (16) may overestimate the effects of systematic errors and because a Student factor t_P referring to, say, a 95 % level[3] exhausts the range of scattering of random errors of experimentally implemented stationary random processes. After all, (16) should express, if not mathematical security, at least reasonable reliability, provided (to reiterate this prerequisite) the prevailing experimental situation fits the error model.

This section closes with the insertion of (13) into (16). The uncertainty thus reduced would be

$$\begin{aligned} u_{\bar{\phi}} &\le \left|\frac{\partial\phi}{\partial\bar{x}}\right| \left(\frac{t_P\,(n-1)}{\sqrt{n}}\,s_x + f_{s,x}\right) + \left|\frac{\partial\phi}{\partial\bar{y}}\right| \left(\frac{t_P\,(n-1)}{\sqrt{n}}\,s_y + f_{s,y}\right) \\ &= \left|\frac{\partial\phi}{\partial\bar{x}}\right| u_{\bar{x}} + \left|\frac{\partial\phi}{\partial\bar{y}}\right| u_{\bar{y}}, \end{aligned} \tag{18}$$

implying a singular empirical variance-covariance matrix $\boldsymbol{S}$.

Table 1. Comparison of the successes and failures of the *Guide* and the proposed alternative for $k_P = 1$ and $t_{68\%}$, respectively.

per 1000 simulations		$k_P = 1$		$t_{68\%}(9) = 1.04$	
		Guide		Alternative model	
f_x/μm	f_y/μm	Successes	Failures	Successes	Failures
Compensation					
10	–10	815	185	1000	0
5	–5	986	14	1000	0
Amplification					
10	10	0	1000	842	158
5	5	363	637	1000	0
Symmetry					
10	0	59	941	1000	0
0	10	841	186	1000	0
5	0	796	204	1000	0
0	5	978	22	1000	0

Table 2. Comparison of the successes and failures of the *Guide* and the proposed alternative for $k_P = 2$ and $t_{95\%}$, respectively.

per 1000 simulations		$k_P = 2$		$t_{95\%}(9) = 2.26$	
		Guide		Alternative model	
f_x/μm	f_y/μm	Successes	Failures	Successes	Failures
Compensation					
10	–10	1000	0	1000	0
5	–5	1000	0	1000	0
Amplification					
10	10	232	768	974	26
5	5	999	1	1000	0
Symmetry					
10	0	966	34	1000	0
0	10	1000	0	1000	0
5	0	1000	0	1000	0
0	5	1000	0	1000	0

3. In (16), the assumed 95 % level refers to the effects of *random errors alone*, as no probability in a statistical sense should be attributed to that overall uncertainty.

4.2 Least-squares adjustment

Let us consider a least-squares adjustment with input data x_i:

$$a_{i1}\beta_1 + a_{i2}\beta_2 + \ldots + a_{ir}\beta_r \approx x_i, \quad (i = 1, \ldots, m) \tag{19}$$

where we assume that the theoretical variances of the input data are equal. Then, according to common interpretation, the minimized sum of squared residuals should yield an empirical estimate of that unknown variance. But as the right-hand side reads

$$x_i = x_{0,i} + \varepsilon_i + f_i,$$

in general, there is no chance of finding such an estimate.

On the other hand, the i-th measurement could be repeated n times, yielding n different random errors $\varepsilon_{i,l}$ $(l = 1, \ldots, n)$, while f_i remains constant. Then, instead of single observations x_i, arithmetic means

$$\bar{x}_i = x_{0,i} + (\bar{x}_i - \mu_i) + f_i \qquad (i = 1, \ldots, m)$$

can be introduced for the right-hand side of (19). Now, the empirical variances of the $\bar{x}_i$ values, whether or not they imply a common theoretical variance, are calculable a priori and it is no longer necessary to refer to the minimized sum of squared residuals.

To be more general, the linear system under consideration,

$$a_{i1}\beta_1 + a_{i2}\beta_2 + \ldots + a_{ir}\beta_r \approx \bar{x}_i, \tag{20}$$

should be weighted by means of a matrix, chosen according to the *Gauss-Markoff theorem*. The weight-matrix is $\boldsymbol{G}^{-1/2}$, where $\boldsymbol{G}$ is the *theoretical* (i.e. unknown) variance-covariance matrix of the input data.

However, as we know, the Gauss-Markoff theorem breaks down in cases where the means $\bar{x}_i$ are biased. In the alternative error model there are biases, thus there is no weighting procedure according to the Gauss-Markoff model. In contrast, the *Guide* maintains the Gaussian type of error handling. A weighting matrix, formally corresponding to the Gauss-Markoff theorem, should therefore exist. With such a weighting procedure, the minimized sum of squared residuals is expected to yield a reasonable estimate for the degrees of freedom of the weighted linear system. This is rarely the case, however; see for example [12]. To artificially induce a "consistent" set of input data, so-called *Birge factors* have been introduced. As a result, the unsatisfactory behaviour of the minimized sum of squared residuals has produced a considerable level of frustration within the metrological community [13].

If we suspend the Gauss-Markoff theorem because we know about the influence of unknown systematic errors, then we need to ask whether it is nevertheless possible to improve the adjustment.

As has been practically shown in [14], instead of choosing the weighting matrix according to Gauss-Markoff, it would be possible to multiply the linear system by a matrix of the form

$$\boldsymbol{G} = \mathrm{diag}\,\{g_1, g_2, \ldots, g_m\}. \tag{21}$$

Varying the elements g_i *by trial and error* appreciably reduces the uncertainties $u_{\bar{\beta}_k}$ of the estimators $\bar{\beta}_k$. Additionally, the procedure reveals any discrepancy of a given input datum.

A potential argument against the empirical variation of the weights to reduce the uncertainties – that this procedure would violate the objectivity of the evaluation – can be refuted as follows. Let $\bar{\beta}_k \pm u_{\bar{\beta}_k}$ $(k = 1, \ldots, r)$ denote the adjusted least-squares parameters. Then, as long as each interval covers the true value $\beta_{0,k}$, the least-squares adjustment is undeniably correct. The alternative formalism of uncertainty assignment fulfills this premise: independently of whether or not weights are used at all and, in the affirmative, which particular set of weights is used, the uncertainty intervals *always* cover the true values, the intervals being broader or smaller depending on the features of the weights [14].

Finally, the couplings between the estimators $\bar{\beta}_k$ $(k = 1, \ldots, r)$ may be formalized. Those stemming from random errors lead to *confidence ellipsoids* as defined by Hotelling's density; those from unknown systematic errors give rise to spatial objects that the present author has called *security polytopes*. The combination of the two objects leads to spatial uncertainty regions somewhat similar to *convex potatoes* [7].

Many numerical examples of least-squares adjustments are given in [15].

5. Key comparisons

In key comparisons, carried out for the mutual recognition of national measurement standards, the compatibility of different arithmetic means, aiming at *one and the same* physical quantity, is scrutinized. A very simple procedure would be to carry out pairwise comparisons of two of these means at a time. Consider the means $\bar{x}$ and $\bar{y}$ and let $\bar{\phi} = \bar{x} - \bar{y}$. Then, any difference $|\bar{x} - \bar{y}|$ satisfying the relation

$$|\bar{\phi}| \leq \frac{t_P\,(n-1)}{\sqrt{n}} \sqrt{s_x^2 - 2\,s_{x,y} + s_y^2} + [f_{s,x} + f_{s,y}] \tag{22}$$

would signal the compatibility of $\bar{x}$ and $\bar{y}$. With reference to (10) and (13), (22) yields

$$|\bar{\phi}| < \left[\frac{t_P\,(n-1)}{\sqrt{n}}\, s_x + f_{s,x}\right] + \left[\frac{t_P\,(n-1)}{\sqrt{n}}\, s_y + f_{s,y}\right] = u_{\bar{x}} + u_{\bar{y}},$$

where $u_{\bar{x}}$ and $u_{\bar{y}}$ designate the overall uncertainties of the two means $\bar{x}$ and $\bar{y}$. Though this appears a rather tough estimation, it nevertheless underscores that in the case of compatibility the two intervals $\bar{x} \pm u_{\bar{x}}$ and $\bar{y} \pm u_{\bar{y}}$ should overlap to define a "non-empty intersection". As the same has to apply to any two means, a non-empty intersection should exist, established by all the means involved.

Within the model of probability-free treatment of unknown systematic errors it is clear that the classical tool of *analysis of variance* would fail: unknown systematic errors entering the analysis as constants would obscure the result.

6. Appropriate application of the triangle inequality

If we assume that unknown systematic errors are constants, we can comment on the appropriate application of the triangle inequality and the *uniqueness* of error propagation. Overall uncertainties should be independent of the specific path taken. To verify this independence, consider the expression[4]

$$\Gamma[y, \phi(x, y)]. \tag{23}$$

As the quantity y appears in two different places, we have to expect covariances. For better understanding, let us distinguish between "pseudo covariances" and "true covariances". Then, in (23), the correlation introduced by the twofold appearance of the quantity y would be a pseudo covariance as dependencies of such a kind are easily removed by the sorting of partial derivatives [16]. If correlations are of physical origin, however, they generally turn out to be mathematically immovable, thus establishing true covariances.

We expand $\phi(x, y)$:

$$\begin{aligned}\phi(\bar{x}, \bar{y}) - \phi(x_0, y_0) = \\ \left[\frac{\partial\phi}{\partial\bar{x}}(\bar{x} - \mu_x) + \frac{\partial\phi}{\partial\bar{y}}(\bar{y} - \mu_y)\right] + \\ \left[\frac{\partial\phi}{\partial\bar{x}} f_x + \frac{\partial\phi}{\partial\bar{y}} f_y\right].\end{aligned} \tag{24}$$

To abbreviate, we set

$$\mu_\phi = \phi(x_0, y_0) + f_\phi, \tag{25}$$

where

$$f_\phi = \frac{\partial\phi}{\partial\bar{x}} f_x + \frac{\partial\phi}{\partial\bar{y}} f_y. \tag{26}$$

4. Here Γ designates *any* function.

The parameter μ_ϕ designates the expectation $E\{\Phi(\bar{X}, \bar{Y})\}$ and f_ϕ specifies the propagated systematic error. We have

$$\begin{aligned}\phi(\bar{x}, \bar{y}) - \mu_\phi = \\ \left[\frac{\partial\phi}{\partial\bar{x}}(\bar{x} - \mu_x) + \frac{\partial\phi}{\partial\bar{y}}(\bar{y} - \mu_y)\right].\end{aligned} \tag{27}$$

An estimation for f_ϕ would be

$$\begin{aligned}-f_{s,\phi} \le f_\phi \le f_{s,\phi}, \\ f_{s,\phi} = \left|\frac{\partial\phi}{\partial\bar{x}}\right| f_{s,x} + \left|\frac{\partial\phi}{\partial\bar{y}}\right| f_{s,y}.\end{aligned} \tag{28}$$

Now, expanding $\Gamma[y, \phi(x, y)]$ yields

$$\begin{aligned}\Gamma(\bar{y}, \bar{\phi}) = \Gamma(y_0, \phi_0) + \\ \left[\frac{\partial\Gamma}{\partial\bar{y}}(\bar{y} - \mu_x) + \frac{\partial\Gamma}{\partial\bar{\phi}}(\bar{\phi} - \mu_\phi)\right] + \\ \left[\frac{\partial\Gamma}{\partial\bar{y}} f_y + \frac{\partial\Gamma}{\partial\bar{\phi}} f_\phi\right].\end{aligned} \tag{29}$$

Inserting f_ϕ and $(\bar{\phi} - \mu_\phi)$ from (26) and (27), we find

$$\begin{aligned}\Gamma(\bar{y}, \bar{\phi}) = \Gamma(y_0, \phi_0) + \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{x}}(\bar{x} - \mu_x) + \\ \left[\frac{\partial\Gamma}{\partial\bar{y}} + \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{y}}\right](\bar{y} - \mu_y) + \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{x}} f_x + \\ \left[\frac{\partial\Gamma}{\partial\bar{y}} + \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{y}}\right] f_y.\end{aligned} \tag{30}$$

As (30) shows, there are no more covariances. Still, we should take a final look at the propagated systematic error

$$f_\Gamma = \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{x}} f_x + \left[\frac{\partial\Gamma}{\partial\bar{y}} + \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{y}}\right] f_y. \tag{31}$$

A reasonable estimation would be

$$\begin{aligned}-f_{s,\Gamma} \le f_\Gamma \le f_{s,\Gamma}, \\ f_{s,\Gamma} = \left|\frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{x}}\right| f_{s,x} + \left|\frac{\partial\Gamma}{\partial\bar{y}} + \frac{\partial\Gamma}{\partial\bar{\phi}}\frac{\partial\phi}{\partial\bar{y}}\right| f_{s,y}.\end{aligned} \tag{32}$$

From (29),

$$f_\Gamma = \frac{\partial\Gamma}{\partial\bar{y}} f_y + \frac{\partial\Gamma}{\partial\bar{\phi}} f_\phi,$$

we may have concluded

$$f^*_{s,\Gamma} = \left|\frac{\partial\Gamma}{\partial\bar{y}}\right| f_{s,y} + \left|\frac{\partial\Gamma}{\partial\bar{\phi}}\right| f_{s,\phi}$$

and inserted (28)

$$f^*_{s,\Gamma} = \left|\frac{\partial \Gamma}{\partial \bar{y}}\right| f_{s,y} + \left|\frac{\partial \Gamma}{\partial \bar{\phi}}\right| \left[\left|\frac{\partial \phi}{\partial \bar{x}}\right| f_{s,x} + \left|\frac{\partial \phi}{\partial \bar{y}}\right| f_{s,y}\right].$$

However, this would be an inconsistent application of the triangle inequality. A "hidden dependence" would have been overlooked.

The worst-case approach has often been criticized for provoking uncertainties that are too large. But is this reproach really justified? Metrology, in its unique objective, is hardly comparable with any other field of statistics, as metrology wants to know *where* to locate the true value of a physical quantity. If this basic question remains unanswered, the investigation has been futile.

Consequently,

- in the case of just one variable, an unknown systematic error f should be estimated by the boundaries $\pm f_s$ of its limiting interval,
- in the case of several variables, let us consider a simple example: the calculation of a grand mean $\bar{\beta} = \sum_{i=1}^{3} w_i \bar{x}_i$ from three arithmetic means $\bar{x}_i$ from three different laboratories. The least-squares construction of $\bar{\beta}$ implies the weights $g_i = 1/u_{x_i}$, yielding $w_i = g_i^2 / \sum_{i=1}^{3} g_i^2$.

With regard to the question of whether or not the estimate of the propagated systematic errors is too large, we consider two situations. For

(a) $u_{x_1} = u_{x_2} = u_{x_3}$ we find

$$f_{s,\beta} = \frac{1}{3}(f_{s,1} + f_{s,2} + f_{s,3}),$$

while for

(b) $u_{x_1} = a, \quad u_{x_2} = 2a, \quad u_{x_3} = 3a$ we obtain

$$f_{s,\beta} = \frac{1}{49}(36 f_{s,1} + 9 f_{s,2} + 4 f_{s,3}).$$

Both estimates seem to be reasonable.

7. Conclusions

The postulate of a Fisher-Romanowsky density for the treatment of random errors and the suggestion that no probability density function should be assigned to unknown systematic errors lead to one-dimensional uncertainty intervals and multidimensional uncertainty spaces. These cover the true value(s) of the physical quantities aimed at with reasonable reliability – given that the experimental situations fit the presumed error model. The formalisms work like a building kit: they are transparent, and they are easy to control and check. *The associated error budgets reliably admit traceabilities of units and measures*, though occasional overestimations cannot be excluded.

By comparison, the formalisms of the *Guide* suffer from the absence of a comfortable algorithm that would easily put us in a position to connect any given k_P factor with its associated degree of confidence so that the overall uncertainty reliably covers the true value of the physical quantity in question. The vagueness stemming from the simplified choices $k_P = 1$ or $k_P = 2$ could impede the interpretation of *Guide* uncertainties, thereby even calling into question physical objectivity.

The alternative formalism proposed here is not an attempt to improve the intricate concept of connecting k_P factors with degrees of confidence. Rather, it circumvents the problem by strictly separating random and systematic errors, and by arithmetically combining well-defined Student's confidence intervals with safe worst-case estimates. As a result, reliable measurement uncertainties can be established.

References

1. *Guide to the Expression of Uncertainty in Measurement*, Geneva, International Organization for Standardization, 1993. The *Guide* is based on Recommendation INC-1 (1980) of the "Working Group on the Statement of Uncertainties" of the BIPM and on Recommendation 1 (CIPM-1981) of the CIPM. It is a joint publication of the BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML.
2. Wagner S., *PTB-Mitteilungen*, 1969, **79**, 343-347.
3. *Seminar über die Angabe der Meβunsicherheit*, Physikalisch-Technische Bundesanstalt, Braunschweig, 20-21 February 1978.
4. Eisenhart C., In *Precision Measurement and Calibration* (Edited by H. H. Ku), NBS Spec. Publ. 300, Vol. 1, Washington, D.C., US Government Printing Office, 1969.
5. Cramer H., *Mathematical Methods of Statistics,* Princeton, Princeton University Press, 1954; Fisz M., *Probability Theory and Mathematical Statistics*, New York, Wiley & Sons, 1967.
6. Welch B. L., *Biometrica*, 1974, **34**, 28-35.
7. Grabe M., *Phys. Lett.*, 1992, **A165**, 124-132; Erratum, 1995, **A205**, 425.
8. Grabe M., *Towards a New Standard for the Assignment of Measurement Uncertainties*, National Conference of Standard Laboratories, Chicago, 31 July to 4 August 1994.
9. Wöger W., *IEEE*, 1987, **IM-36**, 655-658.
10. Grabe M., *Metrologia,* 1986/87, **23**, 213-219.
11. Graybill F. A., *An Introduction to Linear Statistical Models*, New York, McGraw-Hill, 1961.
12. Cohen E. R., Taylor B. R., *The 1986 Adjustment of the Fundamental Physical Constants*, CODATA Bulletin No. 63, 1986.
13. Bender P. L. et al., Should Least Squares Adjustment of the Fundamental Constants be Abolished? In *Precision Measurement and Calibration*, NBS Spec. Publ. 343, Washington, D.C., US Government Printing Office, 1971.
14. Grabe M., *Phys. Lett.,* 1996, **A213**, 125-137.
15. Grabe M., *Estimation of Measurement Uncertainties in Science and Technology*, 300 p., unpublished.
16. Clifford A. A., *Multivariate Error Analysis*, London, Applied Science Publishers Ltd, 1973.

Received on 27 October 1999 and in revised form on 26 September 2000.