Why should correction values be better known than the measurand true value?

Franco Pavese

Torino, Italy

E-mail: frpavese@gmail.com

Abstract. Since the beginning of the history of modern measurement science, the experimenters faced the problem of dealing with systematic effects, as distinct from, and opposed to, random effects. Two main schools of thinking stemmed from the empirical and theoretical exploration of the problem, one dictating that the two species should be kept and reported separately, the other indicating ways to combine the two species into a single numerical value for the total uncertainty (often indicated as 'error'). The second way of thinking was adopted by the GUM, and, generally, adopts the method of assuming that their expected value is null by requiring, for all systematic effects taken into account in the model, that corresponding 'corrections' are applied to the measured values before the uncertainty analysis is performed. On the other hand, about the value of the measurand intended to be the object of measurement, classical statistics calls it 'true value', admitting that a value should exist objectively (e.g. the value of a fundamental constant), and that any experimental operation aims at obtaining an ideally exact measure of it. However, due to the uncertainty affecting every measurement process, this goal can be attained only approximately, in the sense that nobody can ever know exactly how much any measured value differs from the true value. The paper discusses the credibility of the numerical value attributed to an estimated correction, compared with the credibility of the estimate of the location of the true value, concluding that the true value of a correction should be considered as imprecisely evaluable as the true value of any 'input quantity', and of the measurand itself. From this conclusion, one should derive that the distinction between 'input quantities' and 'corrections' is not justified and not useful.

1. Introduction

Since the beginning of the history of modern measurement science, the experimenters faced the problem of dealing with systematic effects, as distinct from, and opposed to, random effects.

Two main schools of thinking stemmed from the empirical and theoretical exploration of the problem, one dictating that the two species should be kept and reported separately, the other indicating ways to combine the two species into a single numerical value for the total uncertainty (often indicated as 'error' in past years). The second way of thinking was adopted by the GUM, and, generally, adopts the method of randomising the systematic effects (i.e. assuming that their expected value is null), by requiring, for all systematic effects taken into account in the model, that corresponding 'corrections' are applied to the measured values before the uncertainty analysis is performed.

On the other hand, about the *value* of the measurand (the quantity intended to be the object of measurement), classical statistics calls it 'true value', admitting that the value should exist objectively (e.g. the value of a fundamental constant), and that any experimental operation aims at obtaining an ideally exact measure of it—GUM omits 'true' as an unnecessary specification. However, due to the

uncertainty affecting every measurement, this goal can be attained only approximately, in the sense that a range of values is measured and nobody can ever know exactly how much any measured value differs from the true value. A frequent way to try to overcome the issue is to use a "confidence interval" or a "degree of belief", assumed to allow one to estimate the probability for the true value to fall within the aimed interval.

The paper discusses the credibility of the numerical value attributed to an estimated correction, as compared with the credibility of the estimate of the location of the true value of a quantity.

2. Input quantities and corrections

The *Guide to the expression of uncertainty in measurement* (GUM) [1] introduced since 1995 methods to be used in the field of metrology for evaluating the measurement uncertainty. According to this Guide, "*it is assumed that the results of a measurement have been corrected for all recognised significant systematic effects and that every effort has been made to identify such effects*" (3.2.4), before the issue of evaluating their uncertainty is tackled.

In addition, in the GUM it is also required that a model is used, expressing the functional dependence of the measurand on the identified input quantities X_i: (GUM, 4.1.1) "*In most cases, a measurand Y is not measured directly, but is determined from N other quantities* $X_1, X_2, ..., X_N$ *through a functional relationship f :* $Y = f(X_1, X_2, ..., X_N)$"; (GUM, 4.1.2) "*The* ***input quantities*** $X_1, X_2, ..., X_N$ *upon which the output quantity Y depends may themselves be viewed as measurands and may themselves depend on other quantities, including corrections and correction factors for systematic effects, thereby leading to a complicated functional relationship f that may never be written down explicitly. Further, f may be determined experimentally … or exists only as an algorithm that must be evaluated numerically. The function f as it appears in this Guide is to be interpreted in this broader context, in particular as that function which contains every quantity,* ***including all corrections and correction factors*** *that can contribute a significant component of uncertainty to the measurement result*" (boldface added here).

The function may remain implicit $f(Y, X_i) = 0$, where the effect of each X_i on Y is obtained through a Type A method of variation of each influence parameter for the determination of the "sensitivity coefficients", or based on Type B information; or, the function can be an explicit mathematical expression, the values of whose parameters are determined experimentally, or via Type B information.

The wished effects of *correction* for systematic effects are, as expressed in the GUM, that: "*it is assumed that, after correction, the expectation or expected value of the error arising from a systematic effect is zero*" (3.2.3); this outcome *"… is based on the concept that there is no inherent difference between an uncertainty component arising from a random effect and one from a correction for a systematic effect*." (E.1.1), because "*… this Guide treats uncertainty components arising from random effects and from corrections for systematic effects in exactly the same way in the evaluation of the uncertainty of the result of a measurement*." (E.3).

However, in the GUM, after having admitted in clause (4.1.2) that corrections are included in the model, they are then treated differently from the other 'input quantities' when clause (3.2.4) requires that the results of a measurement must *preliminarily* be "corrected". This induces an ambiguity in the GUM modelling procedure. To show it, let us take a simple explicit model $Y = X_1 \cdot X_2$, where Y is measured indirectly by measuring X_1 and to X_2, with corrections included according to clause (4.1.2):

$$Y = (X_1 + C_1) \cdot (X_2 \cdot C_2) \,, \tag{1}$$

where X_1 and X_2 are two measured 'input quantities', and C_1 and C_2 are 'corrections',[1] respectively to X_1 and to X_2. The corrections can also be measured quantities, or their values and uncertainties can be obtained from Type B evaluations.

The expression (1) is the GUM model according to clause (4.1.2). However, according to clause

[1] C_1 is an offset correction, typically small in value ($E(C_1) \approx 0$ relative to $E(X_1)$. C_2 is a scale-factor correction, typically with $E(C_2) \approx 1$

(3.2.4), model (1) is only the *initial* model. In fact, after the required corrections are applied, model (1) should be rewritten as

$$Y \approx (X_1^* + C_1^*) \cdot (X_2^* \cdot (1 + C_2^*))^{2} \quad (2)$$

where all quantities on the right of (2) are different from those in (1): $X_1^* = X_1 + \mathrm{E}(C_1)$; $X_2^* = X_2 \cdot \mathrm{E}(C_2)$; $C_1^* = C_1 - \mathrm{E}(C_1)$; $C_2^* = C_2 - \mathrm{E}(C_2)$, where, in general, $\mathrm{E}(C_1) \neq 0$ and $\mathrm{E}(C_2) \neq 0$.[3] According to GUM, the new quantities C_1^* and C_2^* have zero expectation: $\mathrm{E}(C_1^*) = 0$ and $\mathrm{E}(C_2^*) = 0$. Expression (2) is the actual GUM model subject to the uncertainty analysis.

3. Can the value of a correction be known exactly?

Correction for systematic effects may actually be affected by limitations, as also recognised in limited cases in the GUM: "*It is now widely recognized that, when all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured*." (0.2); "*The result of a measurement after correction for recognized systematic effects is still only an estimate of the [true] value of the measurand because of the uncertainty arising from random effects and from imperfect correction of the result for systematic effects*." (3.3.1); "*A corrected measurement result is not the value of the measurand—that is, it is in error—because of imperfect measurement of the realized quantity due to random variations of the observations (random effects), inadequate determination of the corrections for systematic effects, and incomplete knowledge of certain physical phenomena (also systematic effects). Neither the value of the realized quantity nor the value of the measurand can ever be known exactly; all that can be known is their estimated values*." (D.4).

As a consequence, a problem arises about the way to comply with the GUM requirements, and, more in general, to comply with the need of performing corrections for known 'biases', or for 'biases' estimated at various levels of confidence—as occurring also in the testing field. In fact, the above caveats should concern not only the uncertainty, but also the value of the correction, since a model is likely to be imperfect for several reasons, most *not* of random nature:

(i) Non-uniqueness of the measurand (epistemic uncertainty, intrinsic);
(ii) Incomplete list of influence quantities (epistemic uncertainty);
(iii) Imperfect knowledge of the input quantities (re model functional form and the values of its parameter).

In addition, in tackling the problem, one is often confronted with issues arising from practical difficulties, especially when:

(a) Applying a significant recognised correction with large uncertainty;
(b) Handling uncertainty for uncorrected results.

Concerning issue (a), two conditions have to be noted: to be "recognised" (be known) and, to be "significant", also embedded in the GUM: "*In some cases, … although the uncertainty has been evaluated, it may be ignored if its contribution to the combined standard uncertainty of the measurement result is insignificant. If the value of the correction itself is insignificant relative to the combined standard uncertainty, it too may be ignored*" (3.4.4). However, e.g., CITAC [2] recommends instead correcting also for non-significant effects. As to issue (b), the preference for this choice is widespread.

Actually, the corrections are fundamental good practice in science, because they are to be considered as an integral part of the process of sufficiently understanding and modelling an experiment. They are as essential as the practice of replicating the measurements.

In practice, there are several conditions under which the correction of the measured value is not desirable, and this is not explicitly deprecated in the GUM but allowed in special cases. In the GUM

[2] Notice that $\mathrm{E}(1 + C_2^*) \approx 1$, so that the second term is approximately $(X_2^* \cdot C_2)$.

[3] Obviously here we assume that a correction is not a fixed value as in the VIM [3], see Footnote 5.

the difficulties in applying a correction are recognised: "*The error arising from imperfect compensation of a systematic effect cannot be exactly known*" (3.2.3. Note), but the specified way-out is "*The uncertainty of a correction for a known systematic effect may in some cases be obtained by a Type A evaluation while in other cases by a Type B evaluation, as may the uncertainty characterizing a random effect*" and, then, the "*Guide treats uncertainty components arising from random effects and from corrections for systematic effects in exactly the same way in the evaluation of the uncertainty of the result of a measurement*." (E.3).

GUM's recipe is basically Gauß' one: should one recognise a systematic effect, a correction must be applied to measured data before the uncertainty analysis is undertaken. In fact, fixed effects are assumed in the GUM to be fully taken into account by the corrections (zero-expectation after correction), so that only random effects apply to the corrected values.

4. The true value of a correction

The very distinction between 'input quantities' relevant to a measurand and corrections to these quantities is basically artificial, though customary: the fact that a 'correction' may be ignored, or that its value can be zero [4], has possibly promoted this distinction.

However, in principle, 'corrections', when performed, are simply *compensations* for deviations of measured from reference states of specified quantities due to perturbations caused by *other* influence quantities or by off-set conditions. These deviations are customarily called 'bias': in VIM bias is defined "*estimate of a systematic measurement error*" (2.18). Actually, "systematic" should rather mean that bias is relative to a standard condition that is not met (bias value different from zero), or to an unexpected deviation from a standard condition, so that the measurement results are systematically 'off centre' in the variations of their values.

These other quantities are also measured, or the deviations from the reference state are estimated from Type B evaluations, exactly as for the previous 'input quantities': in the vast majority of cases the values of the corrections are non-zero, as are the values of the input quantities. In all respects, as recognised by the GUM, they are part of the input quantities. By definition, once the influence quantities have been identified, no other quantity outside the set of 'input quantities' plus 'corrections' can influence the value of the measurand.

In actuality, due to the fact that *no value of a quantity can be exempt from an uncertainty* about the actual location of the true value, irrespective to the process (either Type A or Type B) bringing to its evaluation, there is not in principle a difference between 'input quantities' X_i and 'biases' B_n—the 'corrections' merely being $C_n = -B_n$:

$$X_i = \mathrm{E}(X_i) + X_i^*, \qquad B_n = \mathrm{E}(B_n) + B_n^*, \tag{3}$$

where, as seen in the expressions (1) and (2) of the initial example, the star quantities have zero mean (the normally non-zero mean value having been 'extracted' from the not-starred quantities) and take into account the variability of the results.

Furthermore, for additive biases, $X_i = X_i^{\ominus} + B_n$—the symbol of 'standard state' $\ominus$ is borrowed here from physical chemistry to indicate the reference condition for which $\mathrm{E}(B_n) =: 0$. Thus $X_i^{\ominus} = X_i - B_n = X_i + C_n$.[4]

The measurement equation for a measurand Y, written in the general way, is:

$$Y = \mu_X + \Delta_X + \varepsilon_\mathrm{r} + \varepsilon_\mathrm{s} \tag{4}$$

where $X = f(X_i)$ with X_i the 'input quantities', μ_X is the mean, Δ_X (sometimes indicated with b or B) is the deviation from the mean, ε_s is the zero-mean random error associated to the systematic effects, and ε_r is the zero-mean random error associated to X (often only $\varepsilon = \varepsilon_\mathrm{r} + \varepsilon_\mathrm{s}$ is indicated). Comparing with

[4] The only functional characteristics of a 'correction' C_n is that it applies to an 'input quantity' X_i, or to another correction.

(3), one can write the following equivalence expressions: $\mu_X = \mathrm{E}(X)$, $\varepsilon_r = X^*$, $\Delta_X = \sum_n B_n$, $\varepsilon_s = \mathrm{conv}(B_n^*)$—the convolution of the B_n^*. In addition, $X^e = X - \Delta_X$.

The existence of an uncertainty *always* associated to a correction,[5] should prevent to omit any correction from the initial model, unless, according to the GUM, one wants to label it as non-significant, so neither including its contribution to uncertainty. Thus, as a random variable, any summary statistics of the true value of a correction is only providing a 'best estimate'. *The (true) value of a correction remains as unknowable as that of any 'input variable', and that of the measurand itself*. However, the bias terms should remain explicit also in the model for the *corrected* measurement results [5, 6] because they carry in an uncertainty component.

5. Possible ways out and conclusions

In [4] the main reasons why the most difficult issue, trusting a correction value, cannot be solved easily, or universally, are summarised.

When the decision is to apply a correction, an edge-cutting solution is to consider, and treat, all 'corrections' as they actually are: a type of 'input quantities', to be included in the model—e.g. model (1)—and treated as any other input quantity. From this viewpoint, one does not need anymore to make a distinction between X_i and C_n, and the uncertainty analysis should concern all relevant quantities in a single step.

Therefore, the model (1) should be explicitly modified as

$$Y = f(W_1, W_2, \ldots, W_m, \ldots, W_M),\ W_m = X_1, \ldots, X_I, C_1, \ldots, C_N, \text{ with } i = 1\ldots I,\ n = 1\ldots N,\ M = I + N \quad (5)$$

where the distinction between the X_i and the C_n is left here only for clarity.

Basically, the practical difference between what is considered a (primary) 'input quantity', X_i, and a correction, C_n, is the fact that the measurement results concerning the former always involve Type A uncertainty components, while the evaluation of a correction may involve only Type B components.

Let us make two examples, one for an 'indirect' measurement, another for a 'direct' measurement.

Gas thermometry. The basic model for the indirect measurement of the intensive quantity T is: $p \cdot V = n \cdot R \cdot T$, with p, pressure; V, volume; n, amount of substance; T, temperature; R, gas constant. In the constant-volume type, V and n are assumed constant during the measurement of p, for the purpose of obtaining measurement results for $T_b = f(p_b)$, where T_b and p_b are the thermodynamic values of the gas *in the thermometer bulb*. In the actual measurements, many "corrections" are to be applied, some due to the gas thermodynamics, some to the technical apparatus, some to the measuring system. Part of them (i) are obtained from measurement of subsidiary quantities, different from p_b, so affected (also) by at least a Type A component of uncertainty; another part (ii) are obtained from calibration certificates, handbook, etc, i.e. by means of Type B components of uncertainty. Among (i), for example, all the corrections for temperature-dependent parameters require the measurement of T in the relevant locations (room, apparatus, equipment, …), using calibrated thermometers affected by calibration and measurement uncertainty components. Among (ii), one can list, e.g.: (a) as to gas, impurities (chemical and isotopic), non ideality, changes in n (adsorption on walls, diffusion outside the bulb, leaks), amount of gas outside the bulb, difference between p_{meas} and p_b; (b) as to the experimental apparatus, changes in V_b due to thermal expansion or finite stiffness of the bulb material, temperature distribution in the dead-volume and non-uniformities; (c) as to the measuring apparatus, pressure gauge corrections and calibration concerning p_{meas}, calibration of instruments measuring electrical quantities (V, I, R), laboratory standards, and vacuum gauges—also mechanical standards for volume V_b measured dimensionally.

Mass measurement. A model for the direct measurement of the extensive quantity m is $m = \sum_i m_{ci}$, with the mass m placed on one balance pan and calibrated masses m_{ci} put on the other pan, and where i

[5] Corrections are variables in the GUM, though the 'bias' intended to be corrected is often considered instead a fixed value—the most notable case is VIM [3], where bias is defined in (2.18)—see above—where the systematic error is a "component of measurement error that in replicate measurements remains constant or varies in a predictable manner" (2.17).

of the latter masses are necessary to obtain the exact balance ($\Delta m \equiv 0$). In the actual measurements, many "corrections" are to be applied, some due to m itself, other to the m_{ci}, other to the technique used. Apart from the possible calibration corrections of each m_{ci} for deviations from the nominal values, all other corrections requires the measurement (or knowledge) of the value of subsidiary quantities. A list of the main, not in a particular order is: for buoyancy, involving air thermodynamic state (p, T and density d) and volume of m and each m_{ci}; for mass magnetism, involving permanent magnetisation and magnetic susceptibility; for mass and balance cleanliness, involving also adsorption; for linearity, sensitivity, drift, hysteresis, eccentricity of the balance.

Incidentally, as shown in these typical examples, in most cases the actual measurement procedure requires a multidimensional capability—in particular temperature and pressure being often involved even in 'direct' measurement, carrying a Type A component—and, traceability of the measurements always requires involving calibration certificates, carrying a Type B component.

Customary, corrections—measured, computed or stipulated—are applied separately. However, an overall model could obviously be constructed by merging all in a single analytical expression whenever an explicit model is feasible. Sometimes the resulting model may look exceedingly complicated. This fact is mitigated today by the use of automatic computing. However, some simple corrections not placing critical problems might remain applied separately.

In many cases, this overall-merging procedure is more straightforward, not needing to separate a first step of application of the corrections of the measurement results from a subsequent distinct step of taking into account its uncertainty. In some cases, it may also overcome the dilemma introduced in [4] and the consequent need for methods of uncertainty-increase evaluation in the absence of a correction. It would also avoid the confusion often observed about the correct handling of 'corrections' [4].

In conclusion, Y should be computed by using the overall model (5), including the explicit functional form of all corrections. This is possible in many cases.

When the model remains instead in an implicit form, or when one likes to avoid embarking in exceedingly complicated analytical computations of uncertainty propagation, a suggestion for a simple way can be to apply the Monte Carlo method. In this case, in addition to the variability distributions of the X_i, one would include, for each 'correction' term, C_i, its probability distribution; E(C_i) is the contingent value attributed to 'bias'—zero, only when justified—i.e. the one that would derive from the GUM clause (3.2.4). In this way, E(Y) directly embeds also the effect of the current 'corrections'.

References

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, *Guide to the expression of uncertainty in measurement* (GUM), and its Supplement 1: *Propagation of distributions using a Monte Carlo method*, 2nd edn, Bureau International des Poids et Mesures (BIPM), Sèvres, 1995. Free downloading from http://www.bipm.org/en/publications/guides/gum .
[2] CITAC (2000) *Eurochem Guide 2000*, end edn. http://www.eurachem.org
[3] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, *International Vocabulary of Basic and General Terms in Metrology*, BIPM, Sèvres, 2008. Free downloading from http://www.bipm.org/en/publications/guides/vim .
[4] F. Pavese, On the difference of meaning of 'zero correction': zero value versus no correction, and of the associated uncertainties, in *Advanced Mathematical and Computational Tools in Metrology IX* (F. Pavese *et al*., Eds), vol. 9, World Scientific, Singapore, pp. 297–309, 2012.
[5] F. Pavese, An introduction to data modeling principles in metrology and testing, in *Advances in data modeling for measurements in metrology and testing*, Birkhauser-Springer, Boston, 2009, 1–30.
[6] F. Pavese, *Mathematical and statistical tools in metrological measurements*, in Physical Methods, Instruments and Measurements, Eds. in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK, [http://www.eolss.net], pp 68, in publication.