

# A comparison of statistical methods for the correction of the systematic effects and for uncertainty assessment

F Pavese

INRIM, 10135 Torino, Italy, E-mail: [f.pavese@inrim.it](mailto:f.pavese@inrim.it)

Since 1993, the expression of uncertainty in the field of metrology has been based on the ISO *Guide to the expression of uncertainty in measurement* (GUM). According to the GUM (3.2.4), “it is assumed that the results of a measurement have been corrected for all recognised significant systematic effects”. The problem of performing a correction is not trivial at all in many cases, and also raises other commonly asked questions that are not addressed clearly by the GUM: should we make corrections when the uncertainty in the correction is large? The GUM indicates that only “significant” effects should be corrected, but what does significant mean in practice? Are there sensible statistical methods for handling uncorrected results? The GUM only considers symmetric distributions; how should we calculate a correction and uncertainty for effects characterised by asymmetric distributions?

This paper will aim at considering the issues related to obtaining reliable and scientifically sound corrections, reviewing statistical methods proposed for either obtaining on the most firm bases the value of a correction and the uncertainty to be associated to it, or to be used should the correction not be applied for insufficient confidence on its value.

*Modelling the effects of influence quantities quantitatively (QCM).* Let us assume that we can model the effects of influence quantities by a linear expression  $\Delta Y_{\text{QCM}} = \sum_i (m_i \cdot X_i)$ ,  $i = 1, \dots, I$ , where  $\Delta Y$  is the overall effect of all  $I$  influence quantities  $X_i$  on the measured value  $Y$ , and  $m_i$  are generally called the sensitivity coefficients: when a model for the relation  $Y = f(X_1, \dots, X_I)$  is available, they can be expressed analytically; when it is not, their values are obtained experimentally with the variational method especially popular in engineering. Should a correction be applied for this effect, it has the form  $DY_{\text{QCM}} \approx -\Delta Y$ , where  $DY$  indicates an *estimate* of  $\Delta Y$ , preventing from using the equal symbol. The  $X_i$  are random variables, so are both  $\Delta Y$  and  $DY$ .

Additionally to this fact, the correction is affected by an uncertainty. In general,  $u^2(DY) \approx u^2(\Delta Y) = \sum_i [u^2(m_i)X_i^2 + u^2(X_i)m_i^2]$ , where the symbol  $\approx$  is used since the estimation also involves the uncertainty.

*Modelling the effects of the influence quantities qualitatively (NQCM).* In the total absence of detailed information about the effect of each and all the influence factors, one can only resort to a non-quantitative correction-method (NQCM).

Extreme values (large and small) of the possible variability of the influence quantities are preferentially removed, and the experimenter is making a judgment either about the overall result of all possible variations or about the variation of each individual influence quantity. As a result, it is assumed that total effect lies in the interval  $|Y_M - Y_m| < \Delta Y_{\text{NQCM}} = m_{\text{ave}} \cdot X_{\text{tot}}$ , where  $Y_M$  and  $Y_m$  are the bounds of the variation interval of  $Y$ , max and min respectively, and depending on the circumstances, one may be able to estimate separately the total effect of the influence parameters  $X_{\text{tot}}$  and the average sensitivity coefficient  $m_{\text{ave}}$  of the influence parameters over  $Y$ , or can only estimate the product  $m_{\text{ave}} \cdot X_{\text{tot}}$ . The corresponding correction equation is  $DY_{\text{NQCM}} = -m_{\text{ave}} \cdot X_{\text{tot}}$ .

If one assumes that the interval  $|Y_M - Y_m|$  is characterised by a rectangular (uniform) probability distribution centred on  $Y$ , then *the value of the correction to be applied is zero*, so  $E(^cY) = E(Y)$ , and the uncertainty is given by  $u(\Delta Y_{\text{NQCM}}) = (m_{\text{ave}} \cdot X_{\text{tot}})/\sqrt{3}$ .

*Combining QCM-NQCM method.* If the knowledge on the effect of all the influence quantities is partial, there may be useful data for some of them. It is therefore possible to combine the QCM method for the known effects, leaving the application of the NQCM for the remaining individually unknown effects.

*One-sided NQCM method.* The problem of assessing uncertainty in the presence of effects of known sign and unknown magnitude frequently occurs. This is a special case where the resulting overall effect of the influence quantities may or may not be symmetric, but is known to not be centred about zero. Assuming that all effects are such that they influence  $Y$  with the same sign, individually or overall, the possible variability of  $Y$  is bounded according to the following expression, showing the case where the effect is only negative  $0 < (Y_M - Y_m) < \Delta Y_{\text{PQCM}} = m_{\text{ave}} \cdot X_{\text{tot}}$  (PQCM means partially quantitative correction method).

If one assumes that the interval  $(Y_M - Y_m)$  is characterised by a rectangular (uniform) probability distribution centred on  $Y$ , then the value of the correction to be applied is  $DY_{\text{PQCM}} = -(m_{\text{ave}} \cdot X_{\text{tot}})/2$ , and its uncertainty  $u(DY_{\text{PQCM}}) = (m_{\text{ave}} \cdot X_{\text{tot}})/2\sqrt{3}$ , which yields a non-zero correction and an uncertainty half that of the straight NQCM approach. In addition, because of the one-sided effects of all relevant impurities,  $(Y_M - Y_m)_+$  could not exceed zero (the reverse in the case of all-positive effects), the uncertainty should not extend outside the interval  $(0, u_-)$ , bringing to an asymmetric interval.

On the contrary, the experimenter *may choose to not perform the correction*, but the uncertainty has to be increased by taking into account the variability of the influence quantities.

*Gauß & Legendre approach.* The basic approach, called *RSSu*, is said to take as reference the independent work of Gauß and Legendre on this subject matter. The derivation is quite complex and omitted here. It brings to the following expression of uncertainty  $u$ , based on the assumption that, when  $y'$  is an estimate of  $Y$ , the best estimate of a quantity  $Y$  is determined by minimising the expression  $E(Y - y')^2 = \text{Var}(Y) + [E(Y) - y']^2$  and omitting to express also the observation uncertainty  $u^2(q) = DY^2 + u^2(DY)$ , where  $q$  is the best estimate of the quantity  $Q$ . Therefore, when no correction is applied, its estimated value is said to be still required to compute the resulting increase of uncertainty.

On the other hand, the above approach brings to the paradox that, in the expression of the expanded uncertainty  $U = ku$  (with  $k \approx 2$ ), also the estimated correction  $DY$  is multiplied by  $k$ . This occurrence is not considered correct by other authors, who preferred to directly express the expanded uncertainty.

*RSSU approach.* This approach differs from the previous approach because considers more correct to express directly the expanded uncertainty  $U$ , thus considers the value of the correction  $DY$  not included in the standard uncertainty but in the expanded uncertainty, whence having to be divided by the coverage factor  $k$  in the expression of  $u$  in the previous approach.

*SUMU approach.* This approach first combines in quadrature the uncertainty of the statistical components (the observation and the correction) to obtain the standard uncertainty  $u$ , then computes the corresponding expanded uncertainty  $U = ku$ . Finally, the value of the correction  $DY$  is added to it: in this way, the correction is treated as weighting  $1/k$  in  $u$ . *SUMU* also constrains the uncertainty interval, which is asymmetric, to not extend to negative values. *SUMU<sub>MAX</sub> approach* is similar to the *SUMU* approach, except that transforms the interval into a symmetric one by using *SUMU<sub>+</sub>* value.

The merits of the above methods, developed in both fields of metrology and testing, will be compared and some conclusions attempted.