XVI IMEKO World Congress
Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future
Vienna, AUSTRIA, 2000, September 25-28

APPLICATION OF THE GUM TO MEASUREMENT SITUATIONS IN METROLOGY

S. D'Emilio and F. Galliana
Accreditation of Laboratories Department - IEN Galileo Ferraris, Torino, Italy

Abstract: Following the Guide to the Expression of Uncertainty in Measurement (GUM) and other recent documents on uncertainty, we analyse typical measurement situations met in a set of Measurement Systems: measurement modelling and comparison, calibration and traceability of measuring devices, dissemination model and interlaboratory comparison, test of measurement compatibility, evaluation of dissemi-nation processes, improved evaluation methods of interlaboratory comparisons.
Keywords: uncertainty, comparison, calibration, compatibility, dissemination process.

1 INTRODUCTION

The GUM and other recent documents on uncertainty [1, 2] have been widely discussed and applied in the analysis of measurement situations of increasing interest and met:

- in a set of *k Measurement Systems* $\{MS_i,\ i = 1, 2, ..., k\}$, each MS_i comprising a national metrology institute NMI_i, a network of calibration laboratories, a network of testing laboratories and a large number of organisations for which the traceability of measuring devices has to be assured;
- in a set of *k Regional Metrology Organisations* $\{RMO_i,\ i = 1, 2, ..., k\}$, each *RMO* comprising several national metrology institutes NMIs.

Nevertheless, several issues have to be settled in connection with:

- the key comparisons organised by the International Committee of Weights and Measures (CIPM) or by *RMO*s in the framework of the Mutual Recognition of national measurement standards and of calibration and measurement certificates issued by NMIs, i.e. the Arrangement (MRA) [3] drawn up by the CIPM in view of the 21st General Conference of Weights and Measures, held in October 1999;
- the Multilateral Agreement (MLA) defined in the framework of the European cooperation for Accreditation (EA), concerning, in particular, the equivalence of the calibration certificates and of the test reports issued by laboratories accredited by National Accreditation Bodies members of EA.

Scope of this paper is to present main results previously obtained by the authors [4 ÷ 10] and, possi-bly, to supply contributions on some issues related to the MRA and the MLA.

2 MEASUREMENT MODELLING AND COMPARISON

For the product X_1X_2 of two independent random variables X_1 and X_2 it is:

$$E\{X_1X_2\} = E\{X_1\}E\{X_2\} \qquad s^2(X_1X_2)=E^2\{X_2\}s^2(X_1)+E^2\{X_1\}s^2(X_2)+s^2(X_1)s^2(X_2) \quad (1)$$

being $E\{X_i\}$ and $s^2(X_i)$ their mean and variance, respectively. If $E\{X_1\} \neq 0$ and $E\{X_2\} \neq 0$, from (1) we derive the relative variance of X_1X_2: $\quad s_r^2(X_1X_2) = s_r^2(X_1) + s_r^2(X_2) + s_r^2(X_1)\, s_r^2(X_2) \quad (2)$

If $s_r^2(X_1)$ or $s_r^2(X_2)$ is <<1, from (2) it follows: $s_r^2(X_1X_2) \cong s_r^2(X_1) + s_r^2(X_2)$.

A quantity involved in a *measurement process MP* (i.e. the measurand) may be modelled by a ran-dom variable *M* having mean $E\{M\}$, variance $s^2(M) = E\{(M - E\{M\})^2\}$, standard deviation $s(M)$, and if $E\{M\} \neq 0$, relative standard deviation $s_r(M) = s(M)/E\{M\}$. Estimates *m* of $E\{M\}$ and *u* of $s(M)$ obtained through *MP* identify the measurement result [*m*, *u*] that *MP* associates to *M*, being *m* and *u* its measurement value and standard uncertainty respectively. In general, a measurement requires:

- ***Modelling the measurement process*** *MP*. The model is fundamental in the design and analysis of every measurement situation and comprises a *model of the measured system*, of which the measurand is a particular parameter, a *model of the measuring system*, and a *model of their interaction* (the measurement method). We express the output quantity (measurand) *Y* in terms of the input quantities $(X_1,X_2,..,X_N)$: $Y = f(X_1,X_2,..,X_N)$ and then we consider the relations between:

- the measurement values *y* of *Y* and $(x_1,x_2,...,x_N)$ of $(X_1,X_2,..,X_N)$: $y = f(x_1,x_2,...,x_N)$ (3)
- the combined standard uncertainty $u_c(y)$ of *y*, the standard uncertainty $u(x_i)$ of x_i and the sensitivity coefficients c_i of *y* versus x_i:

$$u_c^2(y)=\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i)+2\sum_{i=1}^{N-1}\sum_{j=1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u(x_i,x_j)=\sum_{i=1}^{N}c_i^2u^2(x_i)+2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}c_i\,c_j\,u(x_i)u(x_j)r(x_i,x_j) \quad (4)$$

- the effective number of degrees of freedom $n_{eff}(y)$ of $u_c(y)$ and the numbers of degrees of freedom $n(x_i)$ of $u(x_i)$ (formula of Welch-Satterthwaite): $\frac{u_c^4(y)}{n_{eff}(y)}=\sum_{i=1}^{N}\frac{c_i^4u^4(x_i)}{n(x_i)}$ (5)

- ***Evaluation of*** $(X_1, X_2, ..., X_n)$: $(x_1, x_2, ..., x_n)$, $u(x_i)$, $n(x_i)$.
- ***Evaluation of Y***: *y*, $u_c(y)$, $n_{eff}(y)$, expanded uncertainty (for a confidence level *CL*) $U(y) = k_{CL}\, u_c(y)$.

The comparison of two measurement results $[m_1, u_1]$ and $[m_2, u_2]$ of the same measurand obtained through two measurement processes MP_1 and MP_2 may be based on the standardised normal statistic:

$$D^* = [D - E\{D\}]/s(D) \quad (6)$$

being $D = (M_1 - M_2)$ the difference between the two variables M_1 and M_2 modelling the measurand in MP_1 and MP_2, $E\{D\}$ the mean of *D* and $s(D)$ the standard deviation of *D* given by:

$$s^2(D) = s^2(M_1) + s^2(M_2) - 2s(M_1)\, s(M_2)\, R_{12} \quad (7)$$

being R_{12} the correlation coefficient R_{12} between M_1 and M_2. It is: $-1 \le R_{12} \le 1$.

D is estimated by $(m_1 - m_2)$, while $s(D)$ is estimated substituting in (7) $s(M_1)$, $s(M_2)$ and $E\{M\}$ with u_1, u_2 and $(m_1+m_2)/2$, yielding the standard uncertainty $u(D)$ associated to *D*:

$$u^2(D) = u_1^2 + u_2^2 - 2\, u_1\, u_2\, R_{12} \quad (8)$$

From (8) relations are derived for two results: *uncorrelated*, e.g. coming from two independent *MS*s or obtained through independent paths in a *MS* whose top reference standard has negligible uncertainty ($R_{12} \cong 0$ and $u(D) \cong (u_1^2 + u_2^2)^{1/2}$); *with complete positive correlation*, e.g. when a very high uncertainty component is imported in both MP_1 and MP_2 ($R_{12} \cong 1$ and $u(D) \cong |u_1 - u_2|$); *with complete negative correlation* ($R_{12} \cong -1$ and $u(D) \cong u_1 + u_2$).

3 CALIBRATION AND TRACEABILITY OF MEASURING DEVICES

A measuring device MD may be represented as a functional block, which either:

a) supplies an output reading (or quantity) dependent on the quantity applied to its input, or

b) supplies an output quantity dependent on the input reading (nominal value or setting).

The type *a)* devices include indicating and recording instruments, transducers and converters, while the type *b)* devices comprise material measures, generators and calibrators.

The *calibration of* MD is the set of operations which establishes the relation, under specified conditions, between the reading *R* and the measurement result *M* of the quantity applied to or supplied by MD. The results of the calibration are presented in a *calibration diagram* (Fig. 1) or a *calibration table* (Fig. 2). The mid line of the calibration diagram (calibration curve) is expressed by a parameter, i.e. the *calibration factor* $F = M/R$, that summarises the characteristics of MD. If MD is calibrated by comparison with a reference measuring device MD_r, it is: $F = F_r\, R_r / R$, being $F_r = M/R_r$ the calibration factor of MD_r and R_r its reading when the same quantity is applied to MD and MD_r.

The *calibration uncertainty* of MD is given by the uncertainty $u(F)$ of *F*, which is obtained by combining all the uncertainty components imported in the dissemination process from the national standards to the used calibration procedure. MD should be calibrated using reference measuring devices with uncertainties lower than its *instrumental uncertainty*, equal to a limit value given: for a type *a)* MD, by the uncertainty it introduces in measuring a quantity having a negligible intrinsic uncertainty; for a type *b)* MD, by the intrinsic uncertainty of the supplied quantity, i.e. the uncertainty given for this quantity by a MD having a negligible instrumental uncertainty.

In the utilisation of a measuring device MD the result $[m, u]$ associated by MD to the measurand *M* is derived from the equation $M = F{\cdot}R$ and two additional definitions are useful: *uncertainty limit*, i.e. the upper limit of the instrumental uncertainty of MD operating under specified conditions; *utilization uncertainty*, i.e. the uncertainty of a measurement produced by MD.

If the measurement of a quantity *Y* is derived from the measurements of *N* quantities X_1, X_2, ..., X_N through a relation $Y = f(X_1, X_2, ..., X_N)$, it is $Y = f(F_1 R_1, F_2 R_2, ..., F_N R_N)$, being F_i the calibration factor and R_i the reading of the measuring device used for measuring or generating X_i.

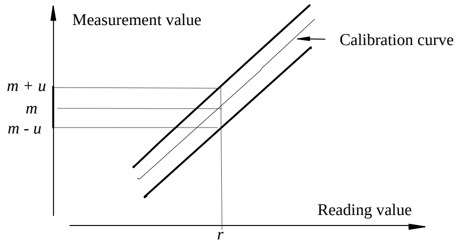

Figure 1. Calibration diagram of a measuring device.

Type a) device

Calibration point		Applied quantity *M*		Output reading	
Measurement function	Measurement range	Value *m* [*m*]	Uncertainty *u*(*m*) [*m*]	Value *r* [*r*]	Uncertainty *u*(*r*) [*r*]

Type b) device

Calibration point		Input reading		Supplied quantity *M*	
Measurement function	Measurement range	Value *r* [*r*]	Uncertainty *u*(*r*) [*r*]	Value *m* [*m*]	Uncertainty *u*(*m*) [*m*]

Figure 2. Calibration tables: [*m*] is the measurement unit of the relevant quantity, [*r*] is the reading unit.

The measuring devices of laboratories operating within a *MS* are traceable to national or international standards, the traceability paths are defined and represented in *traceability diagrams*, and for each echelon of the traceability diagram the *traceability level* can be evaluated. It represents the minimum uncertainty that the laboratory can reach and declare in its certificates and is obtained by combining the uncertainties of the MDs used in the measurement procedure with the uncertainties due to the procedure itself, i.e. all the uncertainties that are independent of the measurand.

The software used to process readings and recorded data is part of the measurement chain. Thus its uncertainty components are evaluated and combined with the other uncertainty components.

4 DISSEMINATION MODEL AND INTERLABORATORY COMPARISONS

Let's model the dissemination processes in a set of Measurement Systems $\{MS_i,\ i = 1, 2, ..., k\}$, belonging to the same or to different geographical regions (Fig. 3, where for simplicity it is $k = 2$):

- S_{i0}, $i = 1, 2, \ldots, k$, is the reference standard in MS_i, made available by a NMI_i. The main parameter of S_{i0} is the calibration factor F_{i0}. In general the F_{i0} are not independent. For example the reference standards for volt and ohm depend on the uncertainty of the physical constants $2e/h$ and h/e^2 met in the reproduction of volt and ohm using the Josephson effect and the quantum Hall effect respectively. It is $F_{i0} = F_C F_i$, being F_C and F_i related to the physical constant and the reproduction of the unit in MS_i respectively.

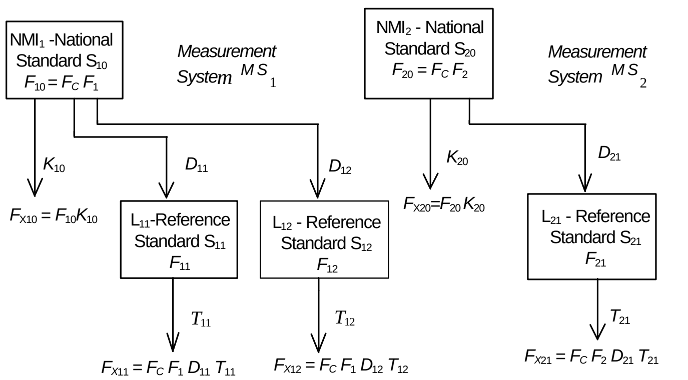

Figure 3. Dissemination in a set of Measurement Systems.

- S_{ij}, $j = 0, 1, 2, ..., n_i$, is the reference standard of the laboratory L_{ij}, modelled by $F_{ij} = F_{i0} D_{ij} = F_C F_i D_{ij}$, being D_{ij} a factor modelling the dissemination process from NMI_i to L_{ij} (L_{i0} is identified with NMI_i).

In order to test and quantify the agreement of measurement performed in different laboratories we perform a comparison, where a measuring standard S_X is circulated between $N = n_1 + n_2 + ... + n_k$ laboratories L_{ij}, belonging to $\{MS_i,\ i = 1, 2, ..., k\}$ being $j = 1, 2, ..., n_i$. The calibration factor F_{Xij} of S_X measured by L_{ij} is given by:

$$F_{Xij} = F_C F_i D_{ij} T_{ij} = F_C F_i K_{ij} \qquad (9)$$

where T_{ij} is a factor modelling the calibration performed by L_{ij} and $K_{ij} = D_{ij} T_{ij}$ accounts for both the dissemination from S_{i0} to S_{ij} and the comparison of S_X with S_{ij}. As we expect the various F_C, F_i, D_{ij}, T_{ij} and K_{ij} to be independent, mean and variance of F_{Xij} are given by:

$$E\{F_{Xij}\} = E\{F_C\}\, E\{F_i\}\, E\{D_{ij}\}\, E\{T_{ij}\} = E\{F_C\}\, E\{F_i\}\, E\{K_{ij}\} \qquad (10a)$$

$$s_r^2(F_{Xij}) \cong s_r^2(F_C) + s_r^2(F_i) + s_r^2(D_{ij}) + s_r^2(T_{ij}) = s_r^2(F_C) + s_r^2(F_i) + s_r^2(K_{ij}) \qquad (10b)$$

A reference result F_{Xi} within each MS_i, to be used as reference in the evaluation of each of the results F_{Xij} obtained by the n_i laboratories L_{ij} belonging to MS_i, is given by their weighted mean:

$$F_{Xi} = \frac{1}{w_i}\sum_{j=0}^{n_i} w_{ij} F_{Xij} = \frac{F_C F_i}{w_i}\sum_{j=0}^{n_i} w_{ij} K_{ij} \qquad w_i = \sum_{j=0}^{n_i} w_{ij} \qquad s_r^2(F_{Xi}) \cong s_r^2(F_{i0}) + \frac{1}{w_i^2}\sum_{j=0}^{n_i} w_{ij}^2 s_r^2(K_{ij}) \qquad (11)$$

A suitable choice of the weights w_{ij} which minimise the relative variance $s_r^2(F_{Xi})$ of F_{Xi} is:

$$w_{ij} \cong s_r^{-2}(K_{ij}) \quad \Rightarrow \quad s_r^2(F_{Xi}) = s_r^2(F_{i0}) + w_i^{-1} \qquad (12)$$

From (11) and (12) we derive relations valid within MS_i and expressing:

- the equivalence of F_{Xij} with F_{Xi} (deviation $(F_{Xij} - F_{Xi})$ and associated variance $s^2(F_{Xij} - F_{Xi}) \cong F_{Xi}^2[s_r^2(K_{ij}) - w_i^{-1}]$) or the equivalence between pair of results F_{Xij} and F_{Xil} (difference $(F_{Xij} - F_{Xil})$ and associated variance $s^2(F_{Xij} - F_{Xil}) \cong F_{Xi}^2[s_r^2(K_{ij}) + s_r^2(K_{il})]$);

- the relations valid in specific cases (comparison involving only two laboratories L_{ij} and L_{il}, uncertainty $s_r(F_{i0})$ of the top reference standard negligible, uncertainties $s_r(K_{ij})$ negligible).

A reference result F_X for the whole comparison, to be used as reference in the evaluation of each of the reference results F_{Xi} obtained by the k systems MS_i, is given by their weighted mean:

$$F_X = \frac{1}{p}\sum_{i=1}^{k} p_i F_{Xi} = \frac{F_C}{p}\sum_{i=1}^{k}\frac{p_i F_i}{w_i}\sum_{j=0}^{n_i} w_{ij}K_{ij} = \frac{F_C}{p}\sum_{i=1}^{k} p_i K_i \qquad p = \sum_{i=1}^{k} p_i \qquad s_r^2(F_X) \cong s_r^2(F_C) + \frac{1}{p^2}\sum_{j=0}^{n_i} p_i^2 s_r^2(K_i) \tag{13}$$

A suitable choice of the weights p_i which minimise the variance of F_X is:

$$p_i \cong s_r^{-2}(K_i) \qquad \Rightarrow \qquad s_r^2(F_X) = s_r^2(F_C) + p^{-1} \tag{14}$$

From (13) and (14) we derive relations valid in a set of systems $\{MS_i, i = 1, 2, ..., k\}$ and expressing:

- the equivalence of F_{Xi} with F_X (deviation $(F_{Xi} - F_X)$ and associated variance $s^2(F_{Xi} - F_X) \cong \cong F_X^2[s_r^2(K_i) - p^{-1}]$) or the equivalence between pair of results F_{Xi} and F_{Xl} (difference $(F_{Xi} - F_{Xl})$ and associated variance $s^2(F_{Xi} - F_{Xl}) \cong F_X^2[s_r^2(K_i) + s_r^2(K_l)]$);
- the relations valid for comparison involving only two NMIs, two MSs, two laboratories or k NMI_i, $i = 1, 2, .., k$ (in this case it is: $s^2(F_{Xi0} - F_X) \cong F_X\,[s_r^2(F_{Xi0}) - s_r^2(F_X)]$).

5 MEASUREMENT COMPATIBILITY

Related to uncertainty is the concept of *measurement compatibility*, which can be applied to the analysis of measurement comparisons [5, 6]. First we consider a criterion for compatibility based on hypotheses testing used in statistical inference: two results $[m_1, u_1]$ and $[m_2, u_2]$ are compatible if the null hypothesis H_0: $E\{M_1\} = E\{M_2\}$ versus the alternative hypothesis H_1: $E\{M_1\} \neq E\{M_2\}$ is accepted at a significance level a chosen for testing H_0. H_1 is two-sided because we shall want to reject H_0 if m_1 is significantly less or greater than m_2. H_0 is tested using $D = (M_1 - M_2)$ and - under the assumption that D is normal and its standard deviation $s(D)$ is known - D^* defined by (6). H_0 is accepted if:

$$\Pr\{|d^*| \leq q_{1-(a/2)} \,||\, H_0\} = 1 - a \tag{15}$$

The quantiles $q_{1-(a/2)}$ and $q_{a/2}=-q_{1-(a/2)}$ of D^* define the acceptance region $q_{a/2} \leq d^* \leq q_{1-(a/2)}$, having an area $1-a$, and the two critical regions $d^*<q_{a/2}$ and $d^*>q_{1-(a/2)}$ for which we reject H_0. If $a = 0.05$ it is $-q_{a/2} = q_{1-(a/2)} = 1.96$, and if $|d^*| \leq 1.96$, H_0 cannot be rejected. In terms of $[m_1, u_1]$, $[m_2, u_2]$ and the compatibility index

$$I_C = (m_1 - m_2)/u(D) \tag{16}$$

H_0 cannot be rejected and $[m_1, u_1]$ is compatible with $[m_2, u_2]$ if:

$$|I_C| = |m_1 - m_2|/u(D) \leq q_{1-(a/2)} \tag{17}$$

This compatibility criterion can be used for evaluating a comparison involving only: NMI_i and L_{ij}, L_{ij} and L_{il}, NMI_i and NMI_l, MS_i and MS_l, L_{ij} and L_{lm}. However, the compatibility criterion defined by (17) is a weak test for equivalence: if H_0 is not rejected, the results are not in disagreement, while the confidence reaches the 95% level only in the most favourable circumstance of $|m_1 - m_2| = 0$ [11].

6 EVALUATION OF DISSEMINATION PROCESSES

The equivalence of the dissemination processes within k MS_i, may be checked by comparing again the results obtained by $N = n_1 + n_2 + ... + n_k$ laboratories L_{ij} belonging to $\{MS_i, i = 1, 2, ..., k\}$ in the calibration of a standard S_X. The N results are analysed using the random variables F_{Xij} modelling the calibration factor of S_X (i.e. the measurand) in the measurement processes MP_{ij} used by L_{ij}:

$$F_{Xij} = F_C\, F_i\, K_{ij} = m_i + e'_{ij} = m + a_i + e_{ij} \tag{18}$$

being m_i the mean of the variables F_{Xij} related to the n_i laboratories L_{ij} belonging to MS_i and assuming e_{ij} to be independent normal random variables $N(0, s_i^2)$, i.e. with zero mean and variance s_i^2. It is:

$$F_{Xij} \equiv N(m + a_i,\ s_i^2) \qquad \sum_{i=1}^{n_i} a_i = 0 \tag{19}$$

The null hypothesis H_0 that the k systems MS_i have equal mean, i.e. that the dissemination processes within the k MS_i are equivalent, is expressed by

$$H_0\text{: } a_1 = a_2 = ... = a_k = 0 \tag{20}$$

The alternative hypothesis is that at least two of the MS_i are unequal is equivalent to the alternative hypothesis H_1: $a_i \neq 0$ for some i. To test H_0, we use the analysis of variance ANOVA for the one-way classification by comparing two estimates of the variance: one based on the variation *between* the samples (i.e. the MS_i) means, and one based on the variations *within* the samples. First we introduce the mean F_{Xi} of the F_{Xij} related to the n_i laboratories belonging to MS_i and the grand mean F_X

$$F_{Xi} = \frac{1}{n_i}\sum_{j=1}^{n_i} F_{Xij} \qquad F_X = \sum_{i=1}^{k}\frac{n_i F_{Xi}}{N} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{F_{Xij}}{N} \tag{21}$$

The total sum of squares of deviations from F_X is

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(F_{Xij} - F_X)^2 \tag{22}$$

SST has (*N* -1) degree of freedom. Then we define the sum of squares related to H_0, identified also as "between-sample sum of squares", as it is a weighted measure of the spread of the sample means:

$$SSH = \sum_{i=1}^{k} n_i (F_{Xi} - F_X)^2 \qquad (23)$$

with (*k* -1) degrees of freedom. Under H_0 it is: $E\{SSH \| H_0: a_1 = a_2 = ... = a_k = 0\} = k - 1$ (24)
A first estimate of the variance is provided by the ratio of *SSH* and the between-sample degrees of freedom (*k* - 1). Finally, we define the "within-sample sum of squares", because it is a weighted measure of the spread of the observations from within each of the *k* samples:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (F_{Xij} - F_{Xi})^2 \qquad (25)$$

with (*N* - *k*) degrees of freedom. It is expected: $E\{SSE\} = N - k$ (26)
A second estimate of the variance is provided by the ratio of *SSE* and the within-sample degrees of freedom (*N*- *k*). It is: $SST = SSH + SSE$ (27)
with expected value, under the null hypothesis H_0: $E\{SST \| H_0: a_1 = a_2 = .. = a_k = 0\} = N - 1$ (28)
To test H_0 we consider the statistic:

$$\boldsymbol{F} = \frac{SSH/(k-1)}{SSE/(N-k)} \qquad (29)$$

which has an $\boldsymbol{F}$-distribution with (*k* -1), (*N* - *k*) degrees of freedom. The observed value $\boldsymbol{F_o}$ of $\boldsymbol{F}$ is derived from (29) substituting the measurement value f_{Xij} for F_{Xij}, the grand mean f_X of the *N* values f_{Xij} for F_X, the mean f_{Xi} of the n_i values f_{Xij} for F_{Xi}. Thus H_0 is accepted at the a level of significance if $\boldsymbol{F_o}$ is so that

$$\boldsymbol{F}_{1-a/2}(k-1, N-k) = \frac{1}{\boldsymbol{F}_{a/2}(N-k, k-1)} < \boldsymbol{F}_o < \boldsymbol{F}_{a/2}(k-1, N-k) \qquad (30)$$

where $\boldsymbol{F}_{a/2}(k\text{-}1, N\text{-}k)$ is the (*a*/2)-quantile of an $\boldsymbol{F}$-distribution with (*k*-1), (*N* - *k*) degrees of freedom.

7 IMPROVED METHODS FOR INTERLABORATORY COMPARISONS

In par. 4 we have discussed a method for deriving the reference result of an interlaboratory comparison, the degree of equivalence of each result with the reference result and the degree of equivalence between pair of results. Interesting proposals are presented in documents issued by international and regional metrology organisations (see, for example, [11]). However, the discussion is still open on agreed methods to be adopted in different measurement situations and for discrepant data.

When in a comparison not compatible results appear (i.e. the hypothesis H_0 defined by (15) is rejected), the statistical procedures and the evaluation of the reference results described in par. 4 are not valid, because the hypothesis that the results belong the same population does not hold. Similarly, if for a comparison the hypothesis H_0 defined by (20) is rejected, the dissemination processes within the systems MS_i cannot be considered equivalent. In both cases, the very first step is to revise the measurement model and to re-evaluate the input and the output quantities. This revision should reduce cases and degrees of non compatibility and make applicable the standard statistical procedures. However, the new results could remain to some extent non compatible.

Specifically, the use of medians has been proposed as robust approach against the effect of outliers [12]. Following Cabiati [13], before the use of standard statistical procedures an additional uncertainty component $u_{Ci} = |m_i - m_a|$ is summed in quadrature to the uncertainty u_i stated by each laboratory, being m_a the arithmetic mean of the values m_i obtained by each laboratory; the new reference result is sufficiently robust with respect either to accurate or rough discrepant result. A more efficient method has been proposed by Wood and Douglas [14]: from a pair of results it is possible to evaluate, for any confidence level *CL*, a Quantified Demonstrated Equivalence (QDE), defined as the half-width of confidence interval d_{CL} within which the two results are expected to agree with confidence *CL*.

A simple method for dealing with comparison with discrepant data is presented in Tab 1 *a)*, where six cases, each comprising 11 independent results [m_i, u_i], are considered: in the first four cases the values m_i are uniformly distributed between 0 and 10 and the uncertainties u_i are equal to 0.3, 1, 3 and 10 respectively; in the fifth case it is m_{11} = 15 (instead of 10) and the u_i are equal to 3; in the sixth case it is again m_{11} = 15, but it is u_{11} = 15 (instead of 3). In the last 3 columns we have evaluated for each case, using the statistical procedures described in par. 4: the weighted value *m* and standard uncertainty *u* of the reference result; the compatibility index defined by (16) of each m_i with *m*. With the exception of the third and fourth case the compatibility criterion (17) is not satisfied for a = 0.05.

The cases and degrees of non compatibility can be reduced if we combine the information coming from [m_i, u_i] with the one obtained *a posteriori* when the results are put together, which takes into account unrecognised factors affecting the results. In Tab. 1 *b)* we consider u_i' instead of u_i, being u_i' the quadratic sum of u_i and the experimental standard deviation $u(m_i)$ of m_i. In the last 3 columns of Tab. 1 *b)* we have derived for each case: the weighted mean *m'* and the associated standard deviation

u', using relations (11) and (12); the compatibility index defined by (16) of each m_i with m'. With the exception of the last point of the fifth case, the compatibility criterion (17) is satisfied for $\alpha = 0.05$.

Let's observe that in the first case the uncertainties u_i are negligible and the results [m_i, u_i] are highly non compatible. Thus the value m' and the uncertainty u' of the reference result are near to the arithmetic mean m_a and to the standard deviation associated to m_a and equal to $u(m_i)/\sqrt{n}$. In the forth case the uncertainties u_i are predominant and the results [m_i, u_i] are highly compatible. Thus the reference result [m', u'] is near to the originally guessed reference result [m, u].

Table 1. Evaluation of the reference result of a comparison using: *a)* standard statistical procedure; *b)* combining a priori and a posteriori uncertainty components.

a)

m_i	u_i	m	u	I_c
0 1 … 9 10	0.3 0.3 … 0.3 0.3	5.00	0.09	-17.5 -14.0 … 14.0 17.5
0 1 … 9 10	1 1 … 1 1	5.00	0.30	-5.24 -4.20 … 4.20 5.24
0 1 … 9 10	3 3 … 3 3	5.00	0.90	-1.75 -1.40 …1.40 1.75
0 1 … 9 10	10 10 … 10 10	5.00	3.02	-0.52 -0.42 …0.42 0.52
0 1 … 9 15	3 3 … 3 3	5.45	0.90	-1.91 -1.56 … 1.24 3.34
0 1 … 9 15	3 3 … 3 15	4.54	0,95	-1.60 -1.24 … 1.57 0.70

b)

m_i	u_i	m_a	$u(m_i)$	u'_i	m'	u'	I_c'
0 1 … 9 10	0.3 0.3 … 0.3 0.3	5.00	3.32	3.33 3.33 ... 3.33 3.33	5.00	1.00	-1.64 -1.31 … 1.31 1.64
0 1 … 9 10	1 1 … 1 1	5.00	3.32	3.46 3.46 … 3.46 3.46	5.00	1.04	-1.57 -1.25 … 1.25 1.57
0 1 … 9 10	3 3 … 3 3	5.00	3.32	4.47 4.47 … 4.47 4.47	5.00	1.35	1.17 -0.94 … 0.94 1.17
0 1 … 9 10	10 10 … 10 10	5.00	3.32	10.5 10.5 … 10.5 10.5	5.00	3.18	-0.48 -0.38 … 0.38 0.48
0 1 … 9 15	3 3 … 3 3	5.45	4.27	5.22 5.22 … 5.22 5.22	5.45	1.57	-0.99 -0.79 … 0.79 1.98
0 1 … 9 15	3 3 … 3 15	5.45	4.27	5.22 5.22 ... 5.22 15.6	4.62	1.64	-0.99 -0.79 … 0.79 0.64

REFERENCES

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML: Guide to the expression of uncertainty in measurement, First edition, 1993, ISBN 92-67-10188-9.
[2] EA-4/02: Expression of the uncertainty of measurement in calibration. Edition 1, April 1997.
[3] BIPM: *Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes*, 14 October 1999.
[4] E. Arri et al.: *On the application of the Guide to Expression of Uncertainty in Measurement to measuring instruments.* Measurement, Vol. 16, pp. 51-57, 1995.
[5] S. D’Emilio et al.: *Compatibility criterion based on hypothesis testing*. Proc. XIV IMEKO World Congress, Vol. V, pp. 155-160, Tampere (Finland), June 1997.
[6] S. D’Emilio, F. Galliana, E. Arri, G. Iuculano: *Compatibility of measurement results within measurement systems using hypotheses testing.* Proc. 8th International Metrology Congress, pp. 589-594, Besançon (France), October 1997.
[7] S. D'Emilio (editor): Proc. IMEKO TC 8 Workshop "Evaluation and Check of Traceability: Basic Aspects and Experimental Check", Torino, September 1998 (to appear).
[8] S. D'Emilio et al.: *Traceability in the dissemination of SI units: modelling, quantitative evaluation and experimental check*. Proc. IMEKO TC 8 Workshop, Torino, September 1998 (to appear).
[9] G. Crotti, S. D’Emilio, G. La Paglia, C. Cherbaucich, G. Rizzi: *Common rules and experience in the accreditation of high voltage calibration laboratories*. Proc. ERA Technology Conference Measurements and Calibration in High Voltage Testing, pp. 4.5.1-4.5.12, London, October 1998.
[10] V. Tuninsky, S. D’Emilio, F. Cabiati: *Concepts and rules in the Guide to the expression of uncertainty: additional proposals*. Proc. 9th International Metrology Congress, pp. 361-364, Bordeaux, October 1999.
[11] NORAMET Document no. 8*: Mutual Recognition of Calibration Services of National Metrology Institutes*. Edition 26 May 1998.
[12] J. W. Muller: *Possible advantages of a robust evaluation of comparisons.* Rapport BIPM-95/2.
[13] F. Cabiati: *Two possible approaches to the evaluation of collective measurement results.* Proc. IMEKO TC 8 Workshop, Torino, September 1998 (to appear).
[14] B.M. Wood, R.J. Douglas: *Confidence-interval interpretation of a measurement pair for quantifying a comparison*. Metrologia, 1998, 35, pp. 187-196.

AUTHORS: S. D'EMILIO, F. GALLIANA, IEN G. Ferraris, Strada delle Cacce, 91 – 10135 Torino, Italy
Phone Int ++39 011 3919 546, Fax Int ++39 011 34 63 84, E-mail: demilio@me.ien.it