10th IMEKO TC7 International Symposium
June 30–July 2, 2004, Saint-Petersburg, Russia

HANDLING AND TREATMENT OF MEASUREMENT DATA FROM DIFFERENT GROUPS OR SOURCES

Anna G Chunovkina[1], Patrizia Ciarlini[2], Maurice G. Cox[3] and Franco Pavese[4]

[1]The D.I.Mendeleyev Institute for Metrology (VNIIM), St.Petersburg, Russia
[2]IAC-CNR, Rome, Italy
[3]National Physical Laboratory, Teddington, Middlesex
[4]Istituto di Metrologia "G.Colonnetti", Torino, Italy

Abstract – This paper addresses the problem of evaluation of data acquired from different groups or sources. It concentrates on a grouping of measurement tasks into (1) measurand value estimation, (2) quantification of the quality of the measurement method, (3) quantification of the comparison of measurement quality. The emphasis is on appropriate use of statistical methods for parameter estimation, experiment design and testing hypothesis for measurement data evaluation.

Keywords: interlaboratory experiment, data processing algorithm, repeatability and reproducibility of measurement results

1. INTRODUCTION

In metrology it is often necessary to treat measurement data received from different sources. According to the objective of the measurement, problems can be divided into three main groups:

1. The estimation of the measurand value or the value of some parameter on the basis of measurement results received from a number of sources. Examples are reference material certification based on the results of an interlaboratory experiment and the adjustment of the fundamental constants.
2. The assessment of the quality of the measurement method used. Quality here relates to the repeatability and reproducibility [1] of the measurement procedure.
3. The assessment of the quality of measurement results obtained in different laboratories. An example is establishing the degree of equivalence of national measurement standards [2] or professional testing.

2. OBJECTIVE

The aim of the investigation is to provide a valid choice of the statistical approach and algorithm for the joint treatment of groups of measurement data. This choice should be made according to the measurement task as well as the available *a priori* information concerning the data submitted and the measurement conditions provided. The reason for writing this paper is to compare interlaboratory data processing for typical tasks of the above problems:

- A. Certification of a reference material through an interlaboratory experiment
- B. Assessment of the reproducibility and repeatability of the measurement method
- C. Evaluation of key comparison data according to MRA when a single stable standard is used for the comparison.

3. DATA EVALUATIOM PROCEDURE

Data evaluation procedure comprises the following stages:

- *Measurement model building.* In this case the model [3] is a result of the aggregation of the measurement models in every group (local models). The local model comprises the model of the value of the measurand itself and models for random and systematic effects. Additional assumptions about the data are typically required, e.g., relating to the covariance matrix associated with the results from different groups.
- *Measurement experiment design.* The requirements for the each group and the number of repeated observations in that group are formulated in order to provide as far as possible a stipulated accuracy in the final results of data processing. How to choose the participating laboratories in an interlaboratory experiment when the measurement procedure is under investigation is discussed below. The motivation is to provide a methodology for extending the results received for a particular group of laboratories to all laboratories that use this measurement method. If this stage is omitted (i.e., the data has already been submitted), the measurement design is reduced to the choice of the data processing algorithm.
- *Estimation of the measurand value and the associated uncertainty.* The data model dictates the choice of the estimation algorithm to be used It differs for problems A, B and C. The estimation of the measurand value should be accompanied by

evaluation of the uncertainty associated with the estimate. Analytical methods and computational tools such as Monte Carlo simulation can be used for this purpose.

- *Check of the consistency of the data with the measurement model*. The measurement model is always based on certain assumptions, which can be difficult to check before the measurement experiment. In cases where the measurand value is to be estimated, it is necessary to confirm the consistency of the data to the model. The omission of this check could result in invalid conclusions about the measurement uncertainty as a consequence of the use of an inadequate model.

4. STAGE I. MEASUREMENT MODEL BUILDING

The basic statistical model for the groups of measurement results from *N* participating laboratories can be described by the model given in ISO 5725:

$$x_i = a + m_i + \varepsilon_i \,, \qquad (1)$$

where x_i, a, m_i and ε_i are, respectively, the measurement result (the beat estimate of the measurand obtained from repeated observations in the i^{th} laboratory), the measurand value, the systematic effect or bias, and the random error in x_i for the i^{th} of the participating laboratories. The measurand value is unknown and to be estimated, except for the case B where the estimation of repeatability and reproducibility variances are the task.

It is important to outline that there is a measurand *"a"* in all tasks considered, but its meaning is different in the three tasks. Consequently, the results obtained in participating laboratories are treated with different objectives. In task A we intend to provide the best estimate of the unknown "*a*". One of the possible ways to solve this problem is to replace the model (1) by the simplified model:

$$\begin{aligned} &x_i = a + w_i, \\ &Ew_i = 0, \quad Ew_i w_j = 0, \quad i \neq j. \\ &Ew_i^2 = u^2(x_i) \end{aligned} \qquad (2)$$

The results from different laboratories are assumed to be uncorrelated. The validity of this model can be established by means of a proper measurement design and should be checked. For task A it is important to stress that systematic biases are treated here as random quantities. Moreover, we should validate the model by designing the measurement experiment in a way that will achieve the randomization of systematic biases. Such a design relates to the choice of participating laboratories, the different measurement methods used for making measurements, and so on.

In task B the reproducibility S_R and the repeatability S_r of the measurement method under investigation are assessed. In this case we operate with some agreed value *a* as an accepted reference value. The basic model is following:

$$\begin{aligned} &x_{ij} = a + m_i + \varepsilon_{ij}, \ i = 1,...N, \ j = 1,...n_i \\ &S_R^2 = S_r^2 + S_L^2, \end{aligned} \qquad (3)$$

where S_R^2 is the reproducibility variance, S_r^2 the repeatability variance and S_L^2 the between-laboratory variance. To apply this model the repeatability variances are assumed to be the same in each laboratory using this method.

In task C the degree of equivalence of results from different laboratories is to be established. For two laboratories no estimate of the measurand "*a*" is needed. Here, the degree of equivalence is expressed by two quantities: the difference between the results and uncertainty associated with this difference. This difference is in fact the estimate of the systematic deviation between the results from these laboratories and the associated uncertainty is a measure of the accuracy of this estimate. For two laboratories the model is the following:

$$\begin{aligned} &x_i - x_j = d_{ij} + \varepsilon_{ij}, \\ &d_{ij} = m_i - m_j, \quad \varepsilon_{ij} = \varepsilon_i - \varepsilon_j, \end{aligned} \qquad (4)$$

since no estimate for the value of the measurand *a* is needed. Actually, the degree of equivalence between pairs of results from a number of the laboratories participating can be expressed by the corresponding matrix: $\{d_{ij}, u(d_{ij})\}_{i,j=1}^{N}$. However, the MRA requires also the degree of equivalence of results from participating laboratories. It is expressed by the deviation of the result from the key comparison reference value (KCRV) and the uncertainty associated with this deviation. The interpretation of the KCRV remains the matter of discussion.

It may be interpreted as a best estimate of the measurand *"a"*. In this case we actually return to problem A and could operate with the model (2). However, in the given case the assumption that the correlation coefficients between the measurement data items are equal to zero seems to be an unjustified simplification. Hence, the problem of quantifying the covariance matrix arises. Often we do not have the required additional information to quantify this matrix. So other interpretations of the KCRV and subsequently another models of the data are of interest.

Below we discuss the mixture of probability distributions as a model for the total population of data obtained in participating laboratories. The dispersion of the data for the i^{th} laboratory can be described by a distribution function $F_i(x)$, with expectation $a + m_i$. Each laboratory provides an estimate of the expectation value, viz., an estimate of the sum $a + m_i$. It is important to stress that frequency distributions of measurement results from the participating laboratories may have different expectations and these expectations are consequently shifted from the measurand value a. The model below expresses this fact. The mixture distribution can be used to describe the distribution of the pooled data from the Laboratories:

$$F(x) = (1/N)\sum F_j(x) \quad (5)$$

Letting $\bar{m} = (1/N)\sum m_i$, the expectation $EF(x)$ and variance of $F(x)$ are, respectively,

$m = a + \bar{m}$,

and

$\sigma^2 = (1/N)\sum \sigma_i^2 + (1/N)\sum (m_i - \bar{m})^2$.

The KCRV can be interpreted as the expectation of the distribution mixture $EF(x)$:

$x_{ref} = a + \bar{m}$.

In the approach here the degree of equivalence

$d_i = EX_i - EX = a + m_i - a - \bar{m} = m_i - \bar{m}$,

for the i^{th} laboratory, can be interpreted as a difference between the "laboratory reference value" and the KCRV, which characterizes the systematic deviation of the results of that laboratory from the KCRV.

5. STAGE II. MEASUREMENT EXPERIMENT DESIGN

Measurement experiment design aims at:

- optimizing the parameters of a measurement experiment: the number of laboratories participating, the number of repeated observations in each laboratory to reduce the random component
- providing conditions under which the proposed data model can be regarded as valid.

To discuss the first item we need an *a priori* estimate of the measurand value and a value for the associated uncertainty. We need to understand the dependence of the uncertainty on the measurement experiment parameters: the number of observations, the number of participating laboratories, the uncertainties associated with the input data, the data processing algorithm, etc. We return to this matter at the next stage of data processing after obtaining an estimate of the measurand value and a value for the associated uncertainty. For the second item it seems very important because it is directly connected with the correct interpretation of the results obtained in the interlaboratory experiment.

As to task A, for the laboratories that use different measurement procedures, different equipment should be involved in the interlaboratory experiment to provide the condition $Ew_i = 0$ in (2). From the mathematical point of view the idea is very simple: randomise the systematic biases in the results from different laboratories

As to task B, participating laboratories should have significant experience in applying the method and the method itself should be thoroughly investigated. By doing so it will be reasonable to expect that the repeatability variance will be the same.

As to task C, when establishing the degree of equivalence of national standards, it is unreasonable to regard the choice of laboratories as a random selection from an ideal general population. The degree of equivalence established is a characteristic of a particular group of laboratories: the KCRV is a consequence of the information provided by this particular group, i.e., it is determined from CIPM KC data only [4]. At present, the choice of participating laboratories cannot be restricted only by laboratories that realise the SI unit independently. Consequently, there can be correlations between data from different laboratories. Hence at the measurement design stage we should provide the following.

- Careful analysis of the uncertainty budgets provided by all participating laboratories, with the purpose of revealing those uncertainty components corresponding to effects that are correlated. However, there is likely to be a degree of arbitrariness in the quantitative expression of the extent of the correlation by means of a covariance matrix.
- A uniform measurement procedure and the algorithm for data processing and uncertainty evaluation to be used by all participating laboratories. This aspect is especially important when the weighted mean is used as the KRCV. In this case the weights attached to the participating laboratory measurements when forming the KCRV are proportional to the inverse squares of the standard uncertainties associated with these measurements [2]. If the uncertainty budget is not agreed by all participating laboratories, the introduction of weighed mean may not be justified. Some incorrectness in the uncertainty associated with the KCRV can result, and consequently in the uncertainty associated with the degrees of equivalence. This aspect should be taken into account when using the deviation from the KCRV and the associated uncertainty to support calibration and measurement capabilities of participating laboratories.

6. STAGE III. ESTIMATION OF THE MEASURAND VALUE AND THE ASSOCIATED UNCERTAINTY

The best estimate of a measurand value in task A for the simplified model (2) is given by:

$$\hat{a} = \frac{\sum \varpi_i x_i}{\sum \varpi_i}, \ \varpi_i = \frac{1}{u^2(x_i)}, \quad (6)$$

$$u^2(\hat{a}) = \frac{1}{\sum \varpi_i}, \quad (7)$$

where $u(x_i)$ is the standard uncertainty associated with the measurement provided by the i^{th} laboratory. The equation (7) is a tool for reducing the uncertainty by enlarging the number of laboratories. But one should have in mind that these laboratories must be "independent" laboratories and a number of such laboratories is restricted. So doing we want to estimate a common mean value from all the data that are therefore supposed to be drawn from different populations having equal means.

Otherwise a data model accounting for correlation effects should be applied.

Formulae for repeatability and reproducibility variances (task B) are available [1].

The estimates of reproducibility and repeatability are used widely in practice for the determination of reproducibility and repeatability limits, for checking stability, for assessment of laboratories, and so on. So these estimates should generally be accepted as having the required accuracy. The choice of the number of laboratories and of the number of repeated observations is based on the following expression [1]:

$$P\left[-A < \frac{s-\sigma}{\sigma} < A\right] = 0{,}95,$$

where for estimates of repeatability and reproducibility variances the limits A are equal to A_r and A_R, respectively:

$$A_r = 1{,}96\sqrt{\frac{1}{2N(n-1)}},$$

$$A_R = 1{,}96\sqrt{\frac{N\{1+n(\gamma^2-1)\}+(n-1)(N-1)}{2\gamma^4 n^2 (N-1)N}},$$

where

$$\gamma^2 = \sigma_R^2 / \sigma_r^2,$$

the ratio of the reproducibility and repeatability variances.

For values A_r and A_R given in advance, the appropriate numbers of laboratories and repeated observations can be calculated from the above expressions. Note that here the equality of the number of repeated observations as well as an *a priori* value of the ratio γ is assumed.

Concerning the evaluation of the degree of equivalence from N participating laboratories we consider only the mixture model. In the approach here the unbiased estimate of the KCRV is the simple mean $\hat{x}_{ref} = (1/N)\sum x_i$ and $u(\hat{x}_{ref})$ is given by

$$u^2(\hat{x}_{ref}) = u^2((1/N)\sum x_i) = (1/N^2)\sum u_i^2,$$

where u_i denotes the reproducibility standard uncertainty s_i for laboratory i. Accordingly, the degree of equivalence d_i with associated uncertainty $u(d_i)$ are given by

$$d_i = x_i - \bar{x}, \ \ u^2(d_i) = (1-2/N)u_i^2 + (1/N^2)\sum u_j^2.$$

In this case the associated uncertainty relates only to the random dispersions, with standard uncertainty u_i, of the laboratory's data. Note that the mutual dependencies between the data due to similar systematic biases are revealed automatically in the form of a mixture distribution.

7. CHECK OF CONSISTENCY OF DATA AND MEASUREMENT MODEL

To check the consistency of the results obtained and the data model applied, a chi-squared test is used for task A. First, one should calculate the observed chi-squared value $\chi^2_{obs} = \sum_{i=1}^{N} \frac{(x_i - \hat{a})^2}{u_i^2}$ Then, if $P(\chi^2_{obs} < \chi^2_{N-1}) < 0{,}05$, where P denotes probability and $\chi^2_{N-1;p}$ is the fractile of the chi-squared distribution with $N-1$ degrees of freedom corresponding to probability p, the consistency check is regarded as failing. For the purposes of the test, the data x_i is regarded as a sample from a Gaussian pdf.

For task B the assumption about the equality of repeatability variances of the data obtained in all participating laboratories seems to be very strong. To check the validity of this assumption Cochran's one-sided test for variance heterogeneity based on the statistic $S^2_{\max} \Big/ \sum_{i=1}^{N} S_i^2$, where S_i^2 – repeatability variance for laboratory i, $S^2_{\max}$ – the larges S_i^2 can be used. For this test the same number of repeated observations in each laboratory is assumed.

As to task C, the mixture distribution model suggested requires the least *a priori* information. No consistency check is required.

CONCLUSIONS

This paper has overviewed methodologies that are regarded as important in the context of handling and treating measurement data from different groups or sources. It has concentrated on grouping measurement tasks into (1) measurand value estimation, (2) quantification of the quality of the measurement method, and (3) quantification of the comparison of measurement quality. The stages

- Measurement model building
- Measurement experiment design
- Estimation of the measurand value and the associated uncertainty
- Check of the consistency of the data with the measurement model

in the data evaluation procedure were considered. The importance of the first stage, measurement model building, for the choice of the data processing and uncertainty evaluation algorithms, is stressed. The content of the data evaluation procedure stages were made concrete concerning the above three measurement tasks.

REFERENCES

[1] ISO. ISO 5725, Accuracy (trueness and precision) of measurement methods and results, International Organization for Standardization, Geneva, Switzerland, 1994.

[2] M. G. Cox, The evaluation of key comparison data, *Metrologia*, **39**, 589–595, 2002.

[3] A. G. Chunovkina, M. G. Cox, A model-based approach to key comparison data evaluation, XVII IMEKO World Congress, Metrology in the 3rd Millennium, Dubrovnik, Croatia, 22–27 June, 2003.

[4] BIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes. Technical report, Bureau International des Poids et Mesures, Sèvres, France, 1999.

[5] Patrizia Ciarlini, Maurice Cox, Franco Pavese, Giuseppe Regoliosi, The use of a mixture of probability distributions in temperature interlaboratory comparisons**,** *Metrologia*, in press.

[6] Franco Pavese, Patrizia Ciarlini, Classes of inter-comparisons and the case of temperature standards, PTB-Bericht, Series IT – Information Technology, Berlin, 2003.

[7] Patrizia Ciarlini, Franco Pavese, Some basic questions in the treatment of intercomparison data, Izmeritelnaya Technika, **N4**, 68-71, 2003.

Author(s):

Dr. Anna G Chunovkina, The D.I.Mendeleyev Institute for Metrology (VNIIM), Moskovsky pr., 19, 190005 St.Petersburg, Russia,
Tel. +7 812 113 01 10, Fax +7 812 113 01 14,
A.G.Chunovkina@vniim.ru

Dr. Patrizia Ciarlini, IAC-CNR, Viale del Policlinico 137, 00161 Rome, Italy,
Tel. +39 06 88470 243, Fax +39 06 4404306,
ciarlini@iac.rm.cnr.it

Prof. Maurice G. Cox, National Physical Laboratory, Teddington, Middlesex TW11 0LW, UK,
Tel. +44 20 8943 6096, Fax +44, 20 8977 7091,
maurice.cox@npl.co.uk

Franco Pavese, CNR - Istituto di Metrologia "G.Colonnetti", strada delle Cacce 73, 10135 Torino, Italy, Tel. +39 011 3977 341, Fax +39 011 3977 347, f.pavese@imgc.cnr.it