

**Traceability for computationally – intensive metrology:
Benchmarking reference pairs for Measurement
Uncertainty Evaluation**

P M Harris

MAY 2015

Traceability for computationally-intensive metrology: Benchmarking reference pairs for Measurement Uncertainty Evaluation

P M Harris
Mathematics and Modelling Group

May 2015

ABSTRACT

Ten benchmarking reference pairs for a particular instance of the computational aim of Measurement Uncertainty Evaluation are described, and instructions for how the benchmarking reference pairs can be used to test a candidate software implementation of the computational aim are given. The report is a deliverable of the EMRP joint research project NEW06 “Traceability for computationally-intensive metrology” (TraCIM).

NPL Report MS 17

© NPL Management Limited, 2015

ISSN 1754–2960

National Physical Laboratory,
Hampton Road, Teddington, Middlesex, United Kingdom TW11 0LW

Extracts from this report may be reproduced provided the source is acknowledged and the extract is not taken out of context

We gratefully acknowledge the financial support of the UK Department for Business,
Innovation and Skills (National Measurement Office)

Approved on behalf of NPLML by Prof Alistair Forbes,
Science Area Leader for the Mathematics & Modelling Group

Contents

1	Introduction	1
1.1	Disclaimer	1
2	Computational aim	2
2.1	Input and output parameters	2
2.2	Mathematical models	3
2.3	Remarks	4
3	Reference pairs	5
4	Performance metrics	5
5	Numerical uncertainty	6
6	Benchmarking reference pairs	7
6.1	MUE_test01	7
6.2	MUE_test02	7
6.3	MUE_test03	7
6.4	MUE_test04	7
6.5	MUE_test05	7
6.6	MUE_test06	8
6.7	MUE_test07	8
6.8	MUE_test08	8
6.9	MUE_test09	8
6.10	MUE_test10	8
6.11	Data and results file formats	9
7	Procedure	9
	Acknowledgements	10
	References	10

1 Introduction

The report describes ten benchmarking reference pairs for a particular instance of the computational aim of Measurement Uncertainty Evaluation (MUE) [1, 2]. It also describes how the benchmarking reference pairs can be used to test a candidate software implementation of the computational aim.

The report is organised as follows. The computational aim addressed in this work is recorded in section 2, and the form of the reference pairs is described in section 3. Performance metrics for comparing reference and test results are defined in section 4, and considerations of the numerical uncertainty associated with the reference pairs are given in section 5. Details of the particular reference pairs that are provided are given in section 6 together with information about the formats of the files used to store the reference pairs. Finally, a procedure for using the benchmarking reference pairs to assess the numerical performance of test software for the computational aim is presented in section 7.

1.1 Disclaimer

The reference pairs are made available from the TraCIM Computational Aims Database and Website. NPL does not offer any validation or certification against the results obtained by using the reference pairs. Use of the website and the reference pairs is subject to the terms laid out in the following Disclaimer.¹

1. Although care has been taken in preparing the information supplied on the website, to the fullest extent permitted at law, NPL acting on behalf of EURAMET is providing this website and its contents “as is” and makes no (and expressly disclaims all) representations or warranties of any kind, express or implied, with respect to the website and the information, content, materials or products included therein, including, without limitation, warranties of merchantability and fitness for a particular purpose. In addition, NPL does not represent or warrant that the information accessible via the website is accurate, complete or current.
2. To the fullest extent permitted at law, neither NPL or any of its affiliates, directors, employees or other representatives will be liable for damages arising out of or in connection with the use of this website or the information, content, materials or products included on this site. This is a comprehensive limitation of liability that applies to all damages, loss of data, income or profit, loss or damage to property and claims of third parties. For avoidance of doubt, NPL does not limit its liability for death or personal injury to the extent that it arises as a result of the negligence of NPL, its affiliates, directors, employees, or other representatives.

In addition:

¹See also www.tracim-cadb.npl.co.uk/conditions.php.

- (a) NPL does not and cannot guarantee the accuracy, reliability, or completeness; nor can NPL be held responsible for any errors or omissions; nor can NPL be held liable for any infringement of the rights of any third party by information posted to the website.
- (b) Your use of the information held on this site is at your sole risk.
- (c) NPL shall not be liable for any direct, indirect, incidental, special, consequential or exemplary damages, including but not limited to damages for loss of profits, goodwill, use, data or other intangible losses, resulting from:
 - i. the use of or inability to use the information held on this website;
 - ii. the cost of procurement of goods and services resulting from any data, information, or services obtained through or from the website;
 - iii. unauthorised access or alternation to your contributions;
 - iv. statements or conduct of any third party on the website;
 - v. any other matter relating to the information held on the website.

2 Computational aim

2.1 Input and output parameters

The input parameters for the problem of Measurement Uncertainty Evaluation (MUE) addressed here are as follows:

1. Measurement model

$$Y = c_1 X_1 + \dots + c_N X_N = \mathbf{c}^\top \mathbf{X},$$

defined by sensitivity coefficients $\mathbf{c} = (c_1, \dots, c_N)^\top$, relating an output quantity (or measurand) Y to N input quantities $\mathbf{X} = (X_1, \dots, X_N)^\top$ assumed to be independent;

2. Probability density functions $g_i(\xi_i)$, $i = 1, \dots, N$, for the input quantities X_i , each of which can take one of the following possible forms:

- Gaussian (or normal) distribution $N(\mu, \sigma^2)$ with expectation μ and standard deviation σ ;
- Rectangular (or uniform) distribution $R(a, b)$ on the interval $[a, b]$;
- Scaled and shifted t -distribution $t_\nu(\mu, s^2)$ with shift parameter μ , scale parameter s and degrees of freedom ν ;

3. Coverage probability p satisfying $0 < p < 1$;

4. Number n_{dig} of significant decimal digits regarded as meaningful in the standard deviation of Y .

The first three of the above parameters define the problem of MUE, and the last parameter is used to specify the requirements on the numerical correctness of the solution to the defined problem, which is usually to obtain a solution to a degree of approximation that is modest. In practice, n_{dig} will be chosen to be one or two (or, in exceptional circumstances, three).

The output parameters of the problem of MUE are an estimate y of Y , the associated standard uncertainty $u(y)$ and a coverage interval $I = [\underline{I}, \bar{I}]$ for coverage probability p , each to be evaluated to a degree of approximation corresponding to n_{dig} . The coverage interval is selected to be the probabilistically symmetric interval, i.e., the interval for which

$$\Pr(Y < \underline{I}) = \Pr(Y > \bar{I}) = (1 - p)/2.$$

Since the measurement model is a linear function of the input quantities $\mathbf{X}$ and the probability distributions are symmetric, it follows that the probability distribution for the output quantity is also symmetric, and consequently the probabilistically symmetric coverage interval is also the shortest coverage interval.

2.2 Mathematical models

The computational task is undertaken according to two mathematical models as follows:

Model 1 implements the ‘GUM uncertainty framework’ [3].

For each input quantity X_i , $i = 1, \dots, N$, set x_i , $u(x_i)$ and ν_i as follows:

- For the Gaussian distribution $N(\mu, \sigma^2)$, $x_i = \mu$, $u(x_i) = \sigma$ and $\nu_i = \infty$;
- For the rectangular distribution $R(a, b)$, $x_i = (a+b)/2$, $u(x_i) = (b-a)/(2\sqrt{3})$ and $\nu_i = \infty$;
- For the t -distribution $t_\nu(\mu, s^2)$, $x_i = \mu$, $u(x_i) = s$ and $\nu_i = \nu$.

Evaluate

$$\begin{aligned} y_1 &= c_1 x_1 + \dots + c_N x_N, \\ u^2(y_1) &= c_1^2 u^2(x_1) + \dots + c_N^2 u^2(x_N), \\ \underline{I}_1 &= y_1 - k_p u(y_1), \\ \bar{I}_1 &= y_1 + k_p u(y_1), \end{aligned}$$

where k_p is the 100 p -percentile of the (unshifted and unscaled) t -distribution $t_{\nu_{\text{eff}}}$ with ν_{eff} degrees of freedom, i.e.,²

$$\Pr(-k_p \leq T \leq k_p) = p, \quad T \sim t_{\nu_{\text{eff}}},$$

and

$$\nu_{\text{eff}} = \frac{u^4(y_1)}{\sum_{i=1}^N \frac{c_i^4 u^4(x_i)}{\nu_i}}.$$

Model 2 implements the ‘propagation of distributions’ [4].

Evaluate

$$\begin{aligned} y_2 &= \mathbb{E}(Y), \\ u^2(y_2) &= \mathbb{V}(Y), \end{aligned}$$

the expectation and variance, respectively, of (the probability distribution $g(\eta)$ for) Y implied by the measurement model and the probability distributions for the input quantities, i.e.,

$$\begin{aligned} y_2 &= \int_{-\infty}^{\infty} \eta g(\eta) \, \mathrm{d}\eta, \\ u^2(y_2) &= \int_{-\infty}^{\infty} (\eta - y_2)^2 g(\eta) \, \mathrm{d}\eta, \end{aligned}$$

where

$$g(\eta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta[\eta - \mathbf{c}^\top \boldsymbol{\xi}] g_1(\xi_1) \cdots g_N(\xi_N) \, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N$$

and $\delta(\cdot)$ is the Dirac delta function. Furthermore, evaluate $\underline{I}_2$ and $\bar{I}_2$ to satisfy

$$\Pr(Y < \underline{I}_2) = \Pr(Y > \bar{I}_2) = (1 - p)/2,$$

also in terms of the probability distribution for Y , i.e.,

$$\int_{-\infty}^{\underline{I}_2} g(\eta) \, \mathrm{d}\eta = \int_{\bar{I}_2}^{\infty} g(\eta) \, \mathrm{d}\eta = (1 - p)/2.$$

2.3 Remarks

Although the problem of MUE defined above concerns a restricted class of problems specified by a linear (or linearized) measurement model and particular probability distributions

²The symbol ‘ $\sim$ ’ is used as shorthand for ‘is distributed as’. Therefore, ‘ $T \sim t_{\nu_{\text{eff}}}$ ’ means that (the random variable) T is distributed as a t -distribution with ν_{eff} degrees of freedom.

for the input quantities, that problem is often assumed to apply and is the basis for the widely-applied GUM uncertainty framework [3]. Consequently, the restricted class constitutes an important set of problems. Model 1 can be used as the basis for testing the claim that software implements the ‘GUM uncertainty framework’ for that restricted class of problems. Similarly, model 2 can be used to test the claim that software implements the ‘propagation of distributions’ for that class. Models 1 and 2 together can be used to validate the ‘GUM uncertainty framework’ as an approach to MUE for linear measurement models and the particular set of input probability distributions considered.

3 Reference pairs

Reference data consists of the information

$$\{N, \mathbf{c}, \mathbf{D}, p, n_{\text{dig}}\}$$

where N is the number of input quantities, $\mathbf{c}$ is a vector of dimension $N \times 1$ that contains the sensitivity coefficients c_i , $i = 1, \dots, N$, defining the measurement model, $\mathbf{D}$ is a matrix of dimension $N \times 4$ that encodes the probability distributions for the input quantities $\mathbf{X}$, p is the coverage probability, and n_{dig} is the number of significant decimal digits regarded as meaningful in the standard deviation of Y . The matrix $\mathbf{D}$ encodes information about $\mathbf{X}$ in the following way:

- $X_i \sim N(\mu, \sigma^2) \Leftrightarrow D_{i1} = 1, D_{i2} = \mu, D_{i3} = \sigma$;³
- $X_i \sim R(a, b) \Leftrightarrow D_{i1} = 2, D_{i2} = a, D_{i3} = b$;
- $X_i \sim t_\nu(\mu, s^2) \Leftrightarrow D_{i1} = 3, D_{i2} = \mu, D_{i3} = s, D_{i4} = \nu$.

Corresponding reference results consist of the information

$$\{y_k, u(y_k), I_k \equiv [\underline{I}_k, \bar{I}_k]\}, \quad k = 1, 2.$$

4 Performance metrics

Let n_{dig} denote the number of significant decimal digits regarded as meaningful in the reference value $u(y_k)$ of the standard uncertainty defined by model k . A corresponding numerical tolerance δ_k for the value $u(y_k)$ is given as follows [4, clause 7.9.2]:

³The element D_{i4} can contain a value, e.g., zero or ‘NaN’, that is used to ‘encode’ a degrees of freedom that is infinite. It can generally be assumed that it is not necessary to access the element during the generation of a reference pair or in the application of test software for the reference data. A similar comment applies to the case of the rectangular distribution.

1. Express $u(y_k)$ in the form $r \times 10^\ell$, where r is an n_{dig} decimal digit integer and ℓ is an integer;
2. Set

$$\delta_k = \frac{1}{2} \times 10^\ell.$$

Suppose $\{y_k^t, u(y_k^t), I_k^t \equiv [\underline{I}_k^t, \bar{I}_k^t]\}$ are the results returned by a test software implementation for model k . These test results are compared with corresponding reference results $\{y_k, u(y_k), I_k \equiv [\underline{I}_k, \bar{I}_k]\}$ by evaluating the following metrics:

$$\begin{aligned} D[y_k^t] &= |y_k^t - y_k|, \\ D[u(y_k^t)] &= |u(y_k^t) - u(y_k)|, \\ D[\underline{I}_k^t] &= |\underline{I}_k^t - \underline{I}_k|, \\ D[\bar{I}_k^t] &= |\bar{I}_k^t - \bar{I}_k|, \end{aligned}$$

and the test results are considered to meet the requirements on the numerical correctness of the solution to the defined problem if

$$\Delta_k \equiv \max\{D[y_k^t], D[u(y_k^t)], D[\underline{I}_k^t], D[\bar{I}_k^t]\} \leq \delta_k,$$

i.e., each test result agrees with the corresponding reference result to the number n_{dig} of significant decimal digits regarded as meaningful in the (reference value of) the standard uncertainty.

5 Numerical uncertainty

As part of the generation of the reference pairs an assessment is made of their quality in order to ensure that the values of the metrics defined in section 4 will provide reliable information about the performance of test software implementations of models 1 and 2. Consequently, the reference pairs are not provided with any quantitative information about their numerical uncertainty.

The assessment involves determining the number n of significant decimal digits of the reference results that are unchanged by small perturbations in the reference data. Provided that $n > n_{\text{dig}}$, the numerical uncertainty of the reference pair is considered adequate to decide whether a test software implementation of model 1 or model 2 returns results that agree with the reference results to n_{dig} significant decimal digits. Because the generation of reference pairs is based on an implementation of the *forward* calculation of reference results from reference data, there is no way to assess comprehensively the numerical correctness of the reference software and the results generated by the software, which would include demonstrating that the software is free of mistakes and errors. However, confidence in the correctness is provided by (a) considering reference data for which the solution is known

analytically (e.g., additive models defined by $c_i = 1$, $i = 1, \dots, N$, with input probability distributions that are all Gaussian or all the same rectangular distribution), and (b) comparing with the results from other (independently developed) implementations of the forward calculation (as in a ‘measurement comparison’).

6 Benchmarking reference pairs

Benchmarking reference pairs are provided with properties defined by the nominal values given in sections 6.1–6.10. The reference data are specified by values (for the sensitivity coefficients c and the parameters of the probability distributions for the input quantities) that are small perturbations of those nominal values. Specifications of the formats of the ASCII data and results files for each benchmarking reference pair are given in section 6.11.

6.1 MUE_test01

$N = 2$, $c = (1, 1)^\top$, $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 1)$, $p = 0.95$, $n_{\text{dig}} = 2$.

6.2 MUE_test02

$N = 2$, $c = (1, 1)^\top$, $X_1 \sim R(-1, 1)$, $X_2 \sim R(-2, 2)$, $p = 0.95$, $n_{\text{dig}} = 2$.

6.3 MUE_test03

$N = 4$, $c = (1, 1, 1, 1)^\top$, $X_1 \sim R(-1, 1)$, $X_2 \sim R(-1, 1)$, $X_3 \sim R(-1, 1)$, $X_4 \sim R(-1, 1)$, $p = 0.95$, $n_{\text{dig}} = 2$.

6.4 MUE_test04

$N = 3$, $c = (1, 1, 1)^\top$, $X_1 \sim N(0, 1)$, $X_2 \sim R(-2, 2)$, $X_3 \sim t_{10}(0, 1)$, $p = 0.95$, $n_{\text{dig}} = 2$.

6.5 MUE_test05

$N = 3$, $c = (1, 1, 1)^\top$, $X_1 \sim N(0, 1)$, $X_2 \sim R(-2, 2)$, $X_3 \sim t_5(0, 5)$, $p = 0.95$, $n_{\text{dig}} = 1$.

6.6 MUE_test06

$N = 3$, $\mathbf{c} = (1, 1, 1)^\top$, $X_1 \sim \text{N}(0, 1)$, $X_2 \sim \text{R}(-2, 2)$, $X_3 \sim t_{10}(0, 1)$, $p = 0.90$,
 $n_{\text{dig}} = 2$.

6.7 MUE_test07

$N = 3$, $\mathbf{c} = (1, 1, 1)^\top$, $X_1 \sim \text{N}(0, 1)$, $X_2 \sim \text{R}(-2, 2)$, $X_3 \sim t_{10}(0, 1)$, $p = 0.99$,
 $n_{\text{dig}} = 2$.

6.8 MUE_test08

$N = 3$, $\mathbf{c} = (1, 1, 1)^\top$, $X_1 \sim \text{N}(0, 1)$, $X_2 \sim \text{R}(-2, 2)$, $X_3 \sim t_{10}(0, 1)$, $p = 0.95$,
 $n_{\text{dig}} = 2$.

6.9 MUE_test09

$N = 3$, $\mathbf{c} = (1, 1, 1)^\top$, $X_1 \sim \text{N}(0, 1)$, $X_2 \sim \text{R}(-2, 2)$, $X_3 \sim t_{10}(0, 1)$, $p = 0.95$,
 $n_{\text{dig}} = 2$.

6.10 MUE_test10

$N = 3$, $\mathbf{c} = (1, 1, 1)^\top$, $X_1 \sim \text{N}(0, 1)$, $X_2 \sim \text{R}(-2, 2)$, $X_3 \sim t_{10}(0, 1)$, $p = 0.95$,
 $n_{\text{dig}} = 2$.

6.11 Data and results file formats

The file ‘fname_data.txt’ containing reference data has the following format:

$$\begin{array}{cccc} N & & & \\ c_1 & & & \\ \vdots & & & \\ c_N & & & \\ D_{11} & D_{12} & D_{13} & D_{14} \\ \vdots & \vdots & \vdots & \vdots \\ D_{N1} & D_{N2} & D_{N3} & D_{N4} \\ p & & & \\ n_{\text{dig}} & & & \end{array}$$

The file ‘fname_res1.txt’ containing reference results for model 1 has the following format:

$$\begin{array}{cc} y_1 & \\ u(y_1) & \\ \underline{I}_1 & \bar{I}_1 \end{array}$$

The file ‘fname_res2.txt’ containing reference results for model 2 has the following format:

$$\begin{array}{cc} y_2 & \\ u(y_2) & \\ \underline{I}_2 & \bar{I}_2 \end{array}$$

7 Procedure

A procedure for using the benchmarking reference pairs described in section 6 to assess the numerical performance of test software for MUE is given below. It can be applied with $k = 1$ or $k = 2$ or $k = 1, 2$ according to the claim made about the test software (that it implements model 1 or model 2 or both models).

1. Read reference data $\{N, \mathbf{c}, \mathbf{D}, p, n_{\text{dig}}\}$ from a file ‘fname_data.txt’ (see section 6.11);
2. Read corresponding reference results $\{y_k, u(y_k), I_k \equiv [\underline{I}_k, \bar{I}_k]\}$ from the file ‘fname_resk.txt’ (see section 6.11);
3. Apply test software to obtain test results $\{y_k^t, u(y_k^t), I_k^t \equiv [\underline{I}_k^t, \bar{I}_k^t]\}$;
4. Evaluate the tolerance δ_k in terms of the reference value $u(y_k)$ and n_{dig} (see section 4);

5. Evaluate $\Delta_k = \max\{D[y_k^t], D[u(y_k^t)], D[\underline{I}_k^t], D[\bar{I}_k^t]\}$ (see section 4);
6. Report the values Δ_k and δ_k , and the interpretation of those values: the test results for model k are considered to meet the requirements on the numerical correctness of the solution to the defined problem if $\Delta_k \leq \delta_k$.

Acknowledgements

The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

References

- [1] Barzdajn, B. 2013 Computational Aim ‘en/L/0/000021’ Propagation of uncertainty for explicit, univariate measurement models: estimate, associated standard uncertainty and coverage factor *TraCIM Computational Aims Database*
- [2] Barzdajn, B. 2013 Computational Aim ‘en/L/0/000018’ Propagation of distributions for general, univariate measurement models: estimate, associated standard uncertainty and quantiles *TraCIM Computational Aims Database*
- [3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 *Guide to the expression of uncertainty in measurement (GUM:1995 with minor corrections)* Bureau International des Poids et Mesures, JCGM 100
- [4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 *Evaluation of measurement data — Supplement 1 to the ‘Guide to the expression of uncertainty in measurement’ — Propagation of distributions using a Monte Carlo method* Bureau International des Poids et Mesures, JCGM 101