

## **Draft B Report:**

*EURAMET Project no. 1079*

**“Volumetric and gravimetric calibration of a 20 L proving tank”**

**September, 2011**

*Giorgio Cignolo and Andrea Malengo  
INRIM, Italy*

### *Abstract*

*The Volume Subgroup of the EURAMET TC “FLOW” carried out EURAMET Project 1079, a comparison of the results obtained by 15 European National Measurement Institutes that agreed to calibrate with water a 20 litre-capacity stainless-steel proving tank of the type commonly used to extensively check, both in laboratory and in the field, the correct measurement and delivery of given amounts of water as well as (especially for custody transfer purposes) of liquids other than water. At nine laboratories, among the participants, the most accurate method of calibration, implemented by weighing the contents of water at known temperature and density, was coupled to the more common method of calibration by water transfers from standard vessels of higher accuracy. Indeed, the Project was essentially aimed at comparing results and accuracy of the latter method to those of the former, which is considered as a primary method, already tested in several international comparisons.*

*During the Project, launched in 2008 and closed in 2011, the proving tank underwent a large decrease in its capacity soon after the middle point of its circulation among participants; the possible (so far unknown) causes of this instability are discussed. The volume change of the proving tank made it difficult to draw generally valid conclusions; therefore, it should be advisable to repeat the comparison, or part of it, in the next years using an intrinsically more stable transfer standard.*

## Contents

	Page
1. Introduction.....	3
2. Participants.....	3
3. Transfer Standard and test conditions.....	5
4. <i>Volumetric</i> Calibration Procedure.....	7
4.1. Definitions.....	7
4.2. Meniscus setting.....	9
4.3. Calculations for <i>volumetric</i> calibration.....	10
5. <i>Gravimetric</i> Calibration Procedure.....	11
5.1. Procedure and computations.....	11
5.2. TS cleaning and water preparation.....	12
5.3. Determination of water density.....	12
6. Results.....	13
7. Possible causes of the TS instability.....	16
8. Analysis of results.....	19
8.1. General.....	19
8.2. <i>Gravimetric</i> results.....	19
8.2.1. Analysis of the <i>gravimetric</i> results of Group 1.....	20
8.2.2. Analysis of the <i>gravimetric</i> results of Group 2.....	22
8.3. <i>Volumetric</i> results.....	24
8.3.1. Comparing the <i>volumetric</i> results of double-task participants to their own <i>gravimetric</i> data.....	24
8.3.2. Analysis of the <i>volumetric</i> results of Group 1.....	25
8.3.3. Analysis of the <i>volumetric</i> results of Group 2.....	26
8.4. Youden plot.....	28
9. Degrees of equivalence between participants.....	30
10. Analysis of the main components of uncertainty budgets.....	32
10.1. Type A, Type B overall components and their composition.....	32
10.2. Details of individual Type B uncertainty components.....	35
11. Remarks and conclusions.....	38
12. References.....	40
Appendix ( <i>Reporting your results.xls</i> spreadsheet).....	41

## 1. Introduction

During the EUROMET/EURAMET TC “Flow” meeting, held in Istanbul in March 2007, it was agreed to start a comparison of liquid volume measurements devoted to the method of vessel calibration alternative to the “gravimetric” one. It consists in determining the volume of a vessel by filling it up with water (not necessarily *pure* water) delivered by one or more calibrated volume standards, namely by other proving tanks or pipettes (not excluding in principle other possible devices, such as pistons).

The “gravimetric” method (weighing of pure water contents at known temperature) had been used and tested already in several comparisons, such as CCM.FF-K4 and its continuations within three RMOs (EUROMET.M.FF-K4b in Europe), as well as in two comparisons of a same 5 L capacity glass flask, namely the EUROMET Project 88/51 (1988) and the more recent EUROMET Project 691.

However, it was suggested that all participant laboratories able to work with both methods (that will be conventionally called *volumetric* and *gravimetric* in this document) would do so in this Project. The goal was to obtain the following advantages:

- 1) Availability of a double series of correlated data;
- 2) Determination of a Comparison Reference Value (CRV) obtained by the *gravimetric* method that can be regarded as a reference for the *volumetric* measurements, as the traceability chain of the *volumetric* method is longer (in the most favourable situations it includes one more step at least);
- 3) Assessment of intra-laboratory compatibility of the participants able to work with both methods (*double-task participants*);
- 4) Assessment of the differences between values and uncertainties of the two methods at the same double-task laboratories.

The recognition of 16 prospective participants showed that: a) 6 participants could perform *gravimetric* calibration only; b) the remaining 10 participants were able to use both methods (see Table 1).

Seven participant Countries had already filed their CMCs for the calibration of this type of standard and range of volume into the BIPM database (see Table 2).

The volume chosen for the Transfer Standard (TS) was 20 litres (see chapter 3).

## 2. Participants

Until November 2008 the following Laboratories (hereafter listed in Country alphabetical order) asked to join the Project, later registered at EURAMET as Project 1079:

Table 1- Participants

	<b>Country</b>	<b>EURAMET Acronym, Laboratory full name, location</b>	<b>Responsible Person</b>	<b>E-mail address</b>	<b>Method</b>
1	BELGIUM (BG)	SMD (FPS Economy, Scientific Metrology SMD, Brussels)	William Vincke	William.Vincke@economie.fgov.be	<i>Volumetric and Gravimetric</i>
2	BOSNIA & HERZEGOVINA (BA)	IMBiH (Institute of Metrology of Bosnia & Herzegovina, Sarajevo)	Zijad Džemić	zijad.dzemic@met.gov.ba	<i>Volumetric and Gravimetric</i>
3	BULGARIA (BG)	BIM (GD NCM, BIM, Sofia)	Mariana Miteva	m.miteva@bim.government.bg	<i>Gravimetric</i>
4	DENMARK (DK)	FORCE (FORCE Technology, Brøndby)	Lene Savstrup Kristensen	lsk@force.dk	<i>Volumetric and Gravimetric</i>
5	GREECE (GR) (Co-pilot)	EIM (Hellenic Institute of Metrology, Thessaloniki)	Zoe Metaxiotou	zoe@eim.gr	<i>Volumetric and Gravimetric</i>
6	HUNGARY (HU)	MKEH (Hungarian Trade Licensing Office, Budapest)	Csaba Czibulka	czibulkacs@mkeh.hu	<i>Gravimetric</i>
7	ITALY (IT) (Pilot)	INRIM (Istituto Nazionale di Ricerca Metrologica, Turin)	Giorgio Cignolo Andrea Malengo	g.cignolo@inrim.it a.malengo@inrim.it	<i>Gravimetric</i>
8	LITHUANIA (LT)	VMT (Vilnius Metrology Centre, Vilnius)	Tadas Juodelis	head_metrologist@vmc.lt	<i>Gravimetric</i>
9	MACEDONIA (MK) (Assisting lab)	BoM (Bureau of Metrology, Skopje)	Anastazija Sarevska	anastazija.sarevska@bom.gov.mk	<i>Volumetric and Gravimetric</i>
10	MONTENEGRO (ME)	MBM (Bureau of Metrology, Podgorica)	Mirjana Mihailovic	mira.m@t-com.me	<i>Gravimetric</i>
11	NORWAY (NO)	JV (Justervesenet, Kjeller)	Gunn Kristin Svendsen	gks@justervesenet.no	<i>Volumetric and Gravimetric</i>
12	PORTUGAL (PT) (Co-pilot)	IPQ (Instituto Português da Qualidade, Lisbon)	Elsa Batista	ebatista@mail.ipq.pt	<i>Volumetric and Gravimetric</i>
13	SERBIA (RS)	DMDM (Directorate of Measures and Precious Metals, Belgrade)	Branislav Tanasic	tanasic@dmdm.rs	<i>Volumetric and Gravimetric</i>
14	SLOVENIA (SI)	MIRS (Metrology Institute of the Republic of Slovenia, Celje)	Matjaž Korošec	M.Korosec@gov.si	<i>Volumetric and Gravimetric</i>
15	SPAIN (ES)	CEM (Centro Español de Metrología, Tres Cantos, Madrid)	Nieves Medina	mnmedina@cem.mityc.es	<i>Volumetric and Gravimetric</i>
16	TURKEY (TK)	UME (TUBITAK-Ulusal Metroloji Enstitüsü, Gebze Kocaeli)	Umit Akcadag	umit.akcadag@ume.tubitak.gov.tr	<i>Gravimetric</i>

The pilot was INRIM, Italy; Greece (EIM) and Portugal (IPQ) volunteered as co-pilots. R. Macedonia (BoM) assisted all of them to ease circulation of the Transfer Standard (TS) in former

Yugoslav Countries. A “daisy” type circulation was chosen, with two “petals”, divided by a recalibration of the TS at INRIM about the middle point of the circulation.

A Protocol [ref. 1] was drafted by INRIM during spring 2008 and in due time approved by co-pilots and participants, together with a standard spreadsheet, meant to report results in as similar as possible formats (see in Appendix the *Reporting\_your\_results.xls* worksheets). The technical contents of the Protocol is amply recalled in the present Report.

The Project measurements were started in November 2008. Each participant was allowed to use two weeks for measurements, and one week more to ensure transportation and delivery of the TS to the next participant, at her/his own care and expenses.

TS transportation had to be made by express air freight, wherever possible. Ground transportation was allowed just in case of short distances. An ATA Carnet for countries where this is accepted was provided by the pilot laboratory INRIM; it always travelled along with the TS, without causing bureaucratic troubles. Unfortunately, one prospective participant Country, Bosnia & Herzegovina, at the time did not accept a free temporary import/export of goods under the arrangement of ATA Carnets.

This fact, with the relevant supplement of costs, in a first time caused Bosnia & Herzegovina’s participation to be shifted to the end of circulation; at last, as the technical problems caused by the TS instability (see chapter 7) showed up, the cost/benefit ratio for that Country increased considerably. Therefore, in spite of everyone’s good will, the TS was not sent to Bosnia & Herzegovina for their participation and measurements.

In the end, the Laboratories and Countries that received and measured the TS were 15, and the measurement “sessions” were 17, as the pilot acted three times. Therefore the total submitted results (volume estimates with an associated uncertainty) are 17 *gravimetric* plus 9 *volumetric*, total 26 from 15 Countries.

### 3. Transfer Standard and test conditions

The proving tank used as Transfer Standard is the property of INRIM. It is made as follows (see Figures 1 and 2):

- Hand-held type, with a graduated neck and no draining valve, nor included thermometer
- stainless steel construction
- 20 litres nominal volume at 15 °C
- double windows (glass plates) in the neck (front and back)
- scale extending from 19,9 L to 20,1 L, with a length of about 59,2 mm
- 20 graduation marks engraved on both glass plates (corresponding to increments of 0,05%, or 10 mL each)
- Maximum Permissible Error (MPE): 0,1%
- approximate mass: 9,7 kg (\*)
- diameter: about 26,5 cm
- height: about 59 cm
- inner diameter of the neck: 64,5 mm
- coefficient of cubical thermal expansion of the TS:  $48 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$
- 

**(\*) Note:** the empty TS was weighed at INRIM before the start of the Project after several weeks of rest in totally dry conditions (this was called State 0). The conventional mass [ref. 2], average of 5

weighings, was 9677,86 g, with an estimated expanded ( $k = 2$ ) uncertainty  $U = 0,08$  g. However, the mass of the supposedly dry TS is not stable to such levels, as moisture or even droplets may hide inside it, or in crevices behind glasses, nameplate and other mechanical parts. For the same reason, and owing to the difficulty of determining either average density, or total volume of the tank body, it is difficult to compute exactly the “true” mass. However, the above datum was used to monitor: 1) possible deposition or removal of dirt or other foreign material inside the TS; 2) by difference of weights, the amount of water wetting the inside walls after pouring and draining the TS according to the prescribed procedure, a factor which is responsible for most of the non-repeatability of ordinary proving tanks.

**By agreement, the volume of the TS that had to be determined in this comparison was the TS volume “to contain” at + 20 °C (not at the + 15 °C temperature marked on the nameplate).**

The uncertainty of the above quoted thermal expansion coefficient is just assumed, as no experimental determination was carried out on the specific stainless steel used, nor on this individual proving tank; besides, any not-annealed welded construction may undergo unpredictable thermal distortions caused by residual stresses (see chapter 7). A relative standard uncertainty of 5% (namely  $2,4 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ ) was suggested by the pilot as a conservative estimate.

In order to reduce the influence of this uncertainty, participants were asked to carry out measurements with a water temperature as close as possible to 20 °C, with ambient air temperatures preferably between 18 °C and 23 °C. However, in order to avoid a large uncertainty on the actual temperature of the TS walls, (which by approximation is usually assumed to be the same of water), air temperature had to be kept within  $\pm 1,5$  °C of the water temperature. Relative humidity of ambient air plays a limited role, however it was recommended to avoid extreme values.



Fig. 1 – The Transfer Standard





Fig. 2 – Front window and front scale

The TS was shipped in a “flight case”, whose dimensions are 74 cm x 43 cm, height 40 cm. The TS lay inside it in horizontal position, together with an aluminium support plate of trapezoidal shape (whose use is described in clause 4.2) and three threaded buttons, to be used as adjustable feet of the support plate in order to get an easy levelling of the TS. The total weight of the loaded flight case was 30 kg.

The participants were asked to clean the TS upon arrival according to their usual procedures, assuming that the TS may be still contaminated by unknown impurities (for example, salts introduced by tap water possibly used by previous participants). Of course none of the cleaning products or actions taken by participants had to cause any adverse or permanent effect on the TS and on its volume.

#### 4. Volumetric Calibration Procedure

##### 4.1. Definitions

The participants were asked to determine the volume of water (at 20 °C) that is needed to pass from State 1 to State 2 (in the *collection*, or *filling mode*, or “*to contain*”), as they are described below:

**State 1, Empty:** the clean TS is previously filled with tap or pure water up to (approximately) the 20 L graduation mark, then emptied. Emptying must be made *slowly*, as necessary during use of proving tanks when the delivered water is introduced into a narrow-neck receiving proving tank of

similar design through a funnel, in order to avoid any sprays. When preparing State 1, the participant does not really need to discharge water into one of her/his volume standards. However s/he is required to do so all the same, in order to simulate a typical use of proving tanks and to get comparable emptying times, and a comparable amount of residual water wetting the inside walls of the TS as well.

*Note: The experience of INRIM with this TS, and with proving tanks of the same make and denomination is as follows:*

- *the difference between the two volume values obtained with the described slow emptying (plus 30 s dripping time) or with the fastest emptying (about 15 s if the tank is turned abruptly and kept upside down, then dripped for 30 s) is impressive: about +9 mL*
- *The slow emptying method usually lasts between 168 and 195 seconds plus dripping time; **the participants were asked to discard and repeat the filling procedure when the emptying phase happened to be shorter than 160 s or longer than 200 s.***

At the end of the main flow, the TS had to be hand-held upside down and still for 30 seconds in order to let it drip, before being turned again into its normal position. During those 30 seconds, the tank had to be kept slightly inclined (by about 10°, in any case by not more than 15°) with respect to verticality. This detail is at variance with [ref. 3] as well as with [ref. 4], whose clause 3.4.1 (essentially referring to volume standards *for liquids other than water* not fitted with a drain valve), states:

#### *3.4.1. Standard flasks and standard test measures*

*A standard flask or a standard test measure, after being filled to the appropriate scale mark, shall be emptied by pouring so that the liquid flows out at only one point on the rim. After the flow breaks into drops, the standard flask or test measure should be drained by **holding it vertically for 30 seconds**, then quickly returning it to the upright position.*

The pilot estimated it necessary to slightly incline this TS during dripping time because the bottom of the tank may be slightly convex inside; owing to this fact, several droplets might simply collect at the centre of the bottom plate, and would not be able to drain along the TS walls. A slight inclination can prevent this phenomenon. The TS had not to be otherwise moved or rotated during the 30 s dripping time.

After dripping, State 1 was achieved. If necessary, it could be maintained in time, by putting a lid on the neck to prevent evaporation.

**State 2, Full:** the clean TS is filled with clean or pure water, exactly up to the 20 L graduation mark; air bubbles inside it have been previously eliminated, or reduced as far as possible. During the filling in the *volumetric* mode of calibration very little can be done to reduce bubbles, as (differently from the *gravimetric* method) it is not advisable to fit a funnel with a plastic hose, to lead quietly the water poured by the participant's standard(s) directly down to the TS bottom. The wetting of the hose, in addition to that of the funnel, would cause an appreciable error, unless both of them are pre-wetted in a "normal" way. A gentle percussion by hand on the TS body had to be made for a while, until no residual air bubbles were observed to rise in the neck.

In the process to achieve State 2 in the *volumetric* calibration mode, the participants could either use a single 20 L volume standard of theirs, or one or more standards of lower denomination, as many



times as necessary to transfer approximately 20 L of water into the TS. After this, the air-water meniscus had to be further treated as described in the next clause.

#### 4.2. Meniscus setting

At the end of such water transfers from the participants' own standard(s) the air-water meniscus generally did not coincide exactly with the 20 L graduation mark of the TS. Then, the participants were asked:

- either to adjust the level exactly to the 20 L mark, by adding or extracting a small amount of water, measured as they would usually do (for instance with a calibrated pipette, or syringe, or by weighing).
- or to calculate the actual volume at the 20 L mark (and 20 °C) on the basis of their scale readings and of the geometrical characteristics of the neck and scale, as they have been determined once for all by the pilot Laboratory [ref. 1].

The former procedure is better, because:

- the most accurate setting and reading of the meniscus level is that made at the 20 L graduation mark;
- the result does not depend upon the features and accuracy of the scales

However, four laboratories out of nine preferred the latter choice (see clause 10.2).

The definition of the proper level was that given in clause 7 of [ref. 4] for *volumetric* glassware used with transparent liquids:

*“.....The meniscus shall be set so that the plane of the upper edge of the graduation line is horizontally tangential to the lowest point of the meniscus, the line of sight being in the same plane.....”.*

The sketch in Figure 3 is similar to the one given in the quoted ISO Standard, which actually refers to cylindrical necks. The sketch below should be seen instead as the representation of one of the two flat glass plates in the windows of the TS.

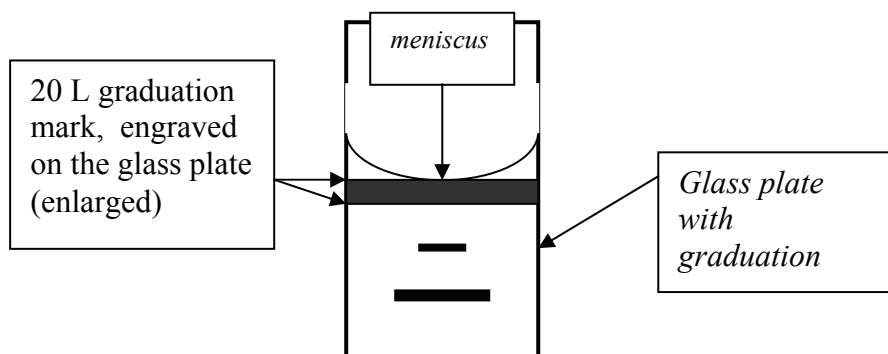


Fig. 3 – Meniscus setting according to ISO 4787

As a rule, the proper setting of the meniscus on the 20 L mark engraved on the *front* window has to be observed by looking at it from the *back* window, and vice-versa: both of them must appear correct as seen from the *opposite side*, in order to avoid the optical disturbance caused by the curvature and often variable thickness of the meniscus. The operator's line of sight *through the liquid* should start from *below* the graduation mark, then it should be raised slowly and carefully up to the upper edge of the opposite 20 L graduation mark, where the horizontal plane tangential to the lowest point of the meniscus is to be adjusted.

The reading, or position of meniscus with reference to the 20 L marks must be the same for both windows and scales. This may not be ensured by simply putting the TS on a truly horizontal plane, owing to a slight imperfection of window positioning in this proving tank, as well as to variable meniscus shapes.

Therefore, the pilot prepared (and shipped along together with the TS, as already mentioned) a tilting aluminium plate, of trapezoidal shape, supported by three threaded feet. The TS could be put on the plate, which could be finely and slightly tilted once for all by each participant by turning the threaded feet/buttons appropriately, until the reading on both windows was the same. Not only a front-rear alignment was necessary, also a lateral (left-right) adjustment had to be made. Of course, after adjustment of the plate, the TS had always to be positioned on the plate with a same orientation.

### 4.3. Calculations for *Volumetric* calibration

The TS volume had to be measured 10 times. The volume of the TS at 20 °C might be computed by calculating sequentially the following volumes:

- volume of the *participant's standard (PS)* at its reference temperature, or of each standard, if more than one are used (source: calibration certificates)
- calculation of the volume of it/them at the temperature measured in its/their water contents, just before pouring water into the TS
- (*pouring into the TS*)
- (*adjustment of water level just to the 20 L graduation mark of the TS by introduction/removal of small measured quantities of water, or reading of the scale indication*)
- (*measurement of the new water temperature in the TS*)
- calculation of the volume exhibited by water, after delivery into the TS, on the basis of the newly measured water temperature
- calculation of the TS volume at 20 °C, based upon: 1) the difference between the new water temperature and 20 °C; 2) on  $\gamma$ , the coefficient of cubical thermal expansion of the TS, assumed to be  $48 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$

However, most people use just a single formula, that is a very good and sufficient approximation of the above process. The Protocol recommended to use that formula, that is:

$$V_{20} = V_0 \left[ 1 - \alpha (t_0 - t_p) + \beta (t_w - t_p) + \gamma (20 - t_w) \right] + \delta = V_1 + \delta \quad (1)$$

where:

$V_{20}$  is the volume of the TS (Transfer Standard) at 20 °C

$V_0$  is the volume of the PS (Participant's Standard) at its reference temperature  $t_0$

$t_0$  is the reference temperature of the PS

- $t_p$  is the temperature of water in the filled PS before pouring
- $t_w$  is the temperature of water in the TS after its filling
- $\alpha$  is the coefficient of cubical thermal expansion of the PS
- $\beta$  is the coefficient of thermal expansion of water at the average test temperature, namely  $0,5 (t_p + t_w)$
- $\gamma$  is the coefficient of cubical thermal expansion of the TS
- $V_1$  is the computed volume after application of the three thermal corrections
- $\delta$  is a correction corresponding either: 1) to the amount of water added (or removed) to adjust the meniscus level exactly to the 20 L graduation mark, or: 2) plus or minus (see below) the absolute value of the difference between the actual reading of the scale and 20 000 mL (with the difference being corrected according to the scale calibration data given by the pilot [ref. 1]).

**The  $\delta$  correction must be input in Eq. (1) with positive sign if the water level is lower than the 20 L mark, and with negative sign if it is higher.**

**Note:** there are several equivalent formulations of Eq. (1). At the time, a single and recognized reference format, to be adopted and quoted, was not found.

If the participant did not use as PS a single 20 L standard, the individual water volumes  $V_i$  contributed by more than one delivery from her/his own smaller standard(s) with different initial temperatures  $t_0$  had to be computed separately, then added together. On the contrary, water temperature in the TS,  $t_w$ , in this case, had to be measured just once (at the end of filling of the TS, and as soon as possible after meniscus setting or reading of the scale); this final value had to be adopted in all individual equations (1), applied to each delivery from the smaller PS (or PSs).

It was recommended to try and determine the *average* temperature inside each volume standard, as there can be appreciable vertical gradients; the stirring of water with a plastic rod just before making level adjustment in the PS, and after TS level reading was recommended. Temperature measurements had to be carried out rather timely, in order to limit temperature drifts.

## 5. Gravimetric Calibration Procedure

### 5.1. Procedure and computations

The *gravimetric* calibration of the TS had to be carried out by weighing its contents of pure water at known temperature.

States 1 and 2 were defined just the same as in the *volumetric* calibration, with the same preparation and handling of the TS.

The participants were required to use their own Procedures, to perform 10 measurements.

The ISO 4787 [ref. 4] and EURAMET [ref. 5] standard formula to be used for volume calculation is:

$$V_{20} = (I_L - I_E) \cdot \left( \frac{1}{\rho_w - \rho_A} \right) \cdot \left( 1 - \frac{\rho_A}{\rho_B} \right) \cdot (1 - \gamma(t_w - 20)) \quad (2)$$

where:

$V_{20}$  is the volume of the TS at 20 °C

$I_L$  is the balance reading (or result of the substitution – double substitution - or other - method of weighing) of the tank filled with water, in grams (State 2)

$I_E$  is the balance reading (or result of the substitution – double substitution - or other - method of weighing) of the empty (wetted) tank, in grams (State 1)

$\rho_A$  is the density of air, in grams per millilitre

$\rho_B = 8,0 \text{ g/mL}$  is the reference density of weights adjusted in compliance with OIML D 28 [ref. 2] document

$\rho_w$  is the density of water at the temperature  $t_w$ , in grams per millilitre

$\gamma$  is the coefficient of cubical thermal expansion of the TS, assumed to be  $48 \times 10^{-6} \text{ }^\circ\text{C}^{-1}$

$t_w$  is the temperature of the water used in testing, in degrees Celsius

(the units kilogram and cubic metre can also be used, with the obvious changes).

## 5.2. TS cleaning and water preparation

The following participants mentioned what kind of cleaning, if any, they used upon receiving the TS:

- INRIM used a glassware cleaning agent with warm water, then rinsed many times the TS with deionized water
- FORCE did not clean the TS, having found it not necessary
- JV just rinsed the TS three times with water, having found it very clean
- MKEH used pure alcohol, then rinsed the TS with deionized water.
- SMD rinsed with deionized water
- IPQ used a cleaning agent, abundant tap water, then rinsed with distilled water and dried with dry air.

Just 8 participants gave the required details about the water used by them to implement the *gravimetric* method. They used:

- INRIM, JV, MKEH and SMD: deionized and air-saturated (aerated) water
- DMDM: distilled and disaerated water
- EIM: *pure* water (quoted as *fresh and clean*, presumably aerated) together with an electronic density meter
- UME: single-distilled water (presumably aerated)
- IPQ: tap water, together with an electronic density meter

## 5.3. Determination of water density

The density meters used by EIM and IPQ are among the *Anton Paar* frequency-based electronic instruments. As many as 3 equations for “pure” water density at different temperatures have been mentioned and adopted by those, among the participants, who gave such details:

- Tanaka et al. [ref. 6] equation: INRIM, DMDM, SMD
- Bettin and Spieweck [ref. 7] equation: FORCE, UME and MKEH
- IAPWS equation, as given in [ref. 8]: JV

## 6. Results

As already noted, the claimed best uncertainties (at 95% confidence level) *in routine calibration work* of seven among the participants have been filed for years in the BIPM database as a part of the Calibration and Measurement Capabilities, or CMCs of those Institutes: the values, spread in a very wide range, are listed in the following Table 2:

Table 2 – Best expanded uncertainties claimed by some participants, as per BIPM Database

Institute	Country/Service	Range	Claimed best uncertainty for 20 L capacity (%)	Claimed best uncertainty for 20 L capacity (mL)
INRIM	IT14	1 to 20 L	0,01%	2
FORCE	DK24	6 to 26 L	0,007%	1,4
JV	NO12	5 to 500 L	0,006%	1,2
UME	TR10	1 to 200 L	0,02%	4
MKEH	HU6	10 to 500 L	0,02%	4
VMT	LT5	5 to 500 L	0,08%	16
IPQ	PT7	20 to 50 L	0,04% (*)	8 (*)

(\*) *A new figure (0,02% or 4 mL) is currently undergoing the international acceptance process*

Among the results initially submitted by participants, those of several laboratories appeared immediately, or at a later stage of data processing, at variance with the expected Comparison Reference Value(s). These participants were invited to check their own submissions (of course without disclosing any information about values) and most of them actually made amendments and changed their data.

The whole set of confirmed and revised results was presented in Draft A of this Report (April, 2011), a set (according to the CIPM MRA-D-05 document) not to be changed any more. The data are as shown in the following Table 3; they are plotted, together with uncertainty bars referring to the expanded uncertainties  $U_{\text{grav}}$  or  $U_{\text{vol}}$  in the plot of Figure 4. **The accepted figures are those given by each participant in two rows of the (final) front sheet of the *Reporting\_your\_results.xls* standard spreadsheet (see Appendix), under the captions:**

- **Average Volumes (*volumetric* and/or *gravimetric*)**
- **Expanded Uncertainty**

As requested in the Protocol, not more than 2 decimal digits are retained in both entries.



Table 3 - Results submitted by the participating laboratories.

No.	Institute	Country	Date of Measur. (approx.)	GRAVIMETRIC		VOLUMETRIC	
				Volume $V_{\text{grav}}$ mL	Expanded uncertainty $U_{\text{grav}}$ mL	Volume $V_{\text{vol}}$ mL	Expanded uncertainty $U_{\text{vol}}$ mL
<b>1<sub>1</sub></b>	<b>INRIM<sub>1</sub></b>	<b>IT</b>	<b>Oct. 2008</b>	<b>20001,59</b>	<b>1,57</b>		
2	MIRS	SI	Dec. 2008	20000,32	0,94	20000,62	2,76
3	DMDM	RS	Dec. 2008	20000,94	3,02	20001,02	5,93
4	BoM	MK	Jan. 2009	19994,26	2,38	19994,53	4,01
5	MBM	ME	Feb. 2009	19999,60	2,93		
6	EIM	GR	Mar. 2009	19998,64	1,14	19998,05	1,88
7	BIM	BG	Apr. 2009	19999,19	2,17		
<b>1<sub>2</sub></b>	<b>INRIM<sub>2</sub></b>	<b>IT</b>	<b>Aug. 2009</b>	<b>20003,83</b>	<b>1,69</b>		
8	FORCE	DK	Sept. 2009	19998,97	2,08	19998,6	4,89
9	JV	NO	Sept. 2009	19996,8	3,82	19995,7	4,45
10	UME	TR	Nov. 2009	19989,77	1,14		
11	MKEH	HU	Nov. 2009	19981,60	4,02		
12	SMD	BE	Dec. 2009	19983,6	1,1	19973,9	2,2
13	VMT	LT	Jan. 2010	19983,81	1,65		
14	CEM	ES	Feb. 2010	19984,49	2,63	19988,57	3,63
15	IPQ	PT	Mar. 2010	19982,33	2,10	19979,34	5,05
<b>1<sub>3</sub></b>	<b>INRIM<sub>3</sub></b>	<b>IT</b>	<b>Apr. 2010</b>	<b>19985,62</b>	<b>1,66</b>		

*Note: FORCE sent also by mail two printed and signed Calibration Certificates of their own, in which the same above four values are rounded up to just one decimal digit. SMD sent two Calibration Reports (not Certificates), from which the pilot drew the above figures.*

The three measurements at the pilot's laboratory INRIM are evidenced in bold, in order to underline the boundaries of the two petals of TS circulation: beside those of the pilot, results from six and eight laboratories are included in the two petals, respectively.

Details about Type A and B uncertainties, degrees of freedom and coverage factors are given in chapter 10.

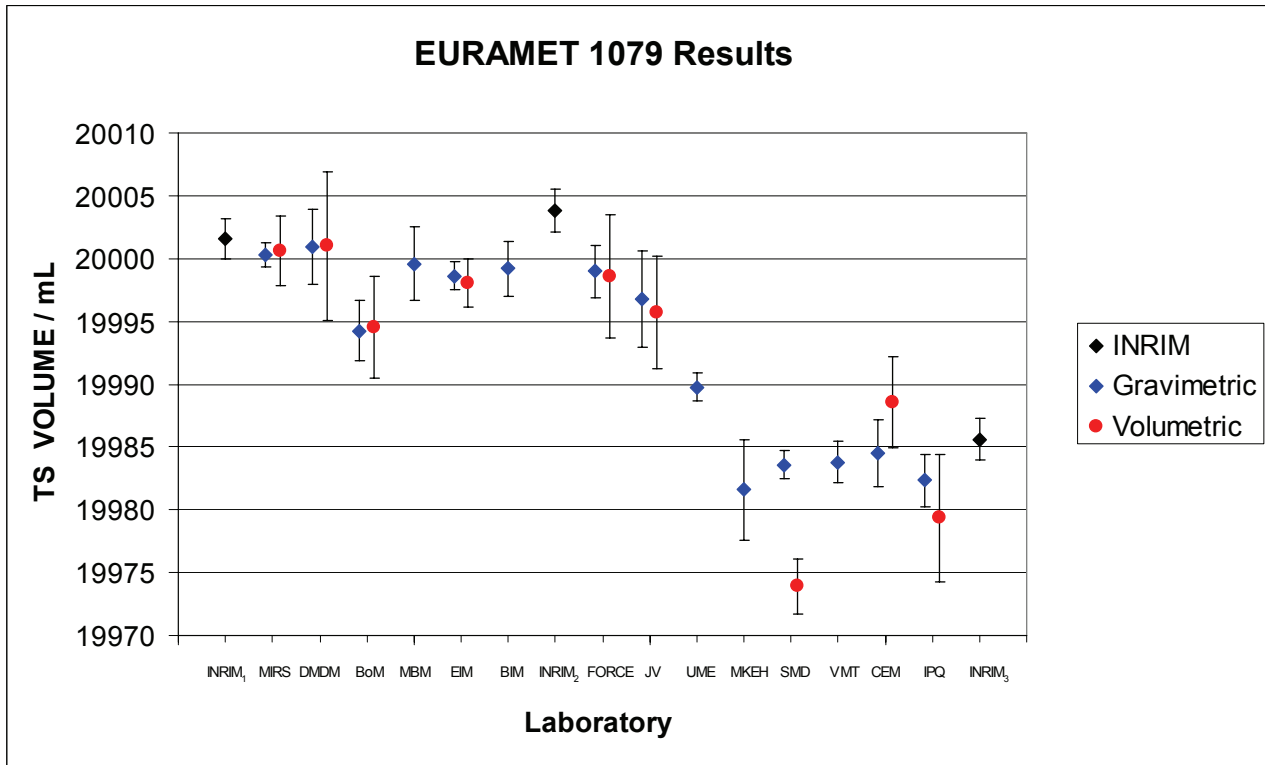


Fig. 4 – Plot of results with expanded uncertainty bars

The comparison between the three pilot's measurements, as well as the plot of all results, shows that the TS underwent an extraordinary decrease in volume, a little less than a couple of divisions (namely about 17 mL), which is one order of magnitude larger than any reasonable expectations. In fact, the stability of proving tanks is normally considerably better than usual calibration accuracies, a sample of which has been shown in Table 2.

Looking at the plot in Figure 4, it can be hypothetically admitted that the TS was rather steady, or at least moderately unsteady until measurements by JV were made, then it decayed (it is not known whether slowly or abruptly) to a lower, but rather steady again volume. Under this hypothesis, the steady final value can be estimated by averaging data from INRIM<sub>3</sub> backwards, seemingly down to MKEH which, by exhibiting one of the lowest absolute values, provides probably the best sign or proof of the termination of the suggested TS drift.

Unfortunately, this approach leaves aside UME, whose result happens to be just in the middle of the two extreme volume values shown by the TS, and not compatible with either. No cause could be really ascertained for the change of volume (see chapter 7), as none of the participants, including those in the middle of the circulation, are supposed to have used any less than usual handling or cleaning processes.

In addition, it should be noted that the first and second pilot's results, although overlapping, are not mutually compatible. Indeed, if they were mutually independent, their normalized compatibility index  $E_n$  would be an acceptable  $E_n = 0,97$ . However, the largest part of Type B uncertainty contributions is certainly shared by both measurements. Under the worst assumption, that the whole of the  $u_B$  contributions is common, the covariance term is  $0,53 \text{ (mL)}^2$ , therefore the difference between the two values (simply indicated as INRIM<sub>1</sub> and INRIM<sub>2</sub>) is:

$$(\text{INRIM}_1) - (\text{INRIM}_2) = -2,24 \text{ mL} \quad (3)$$

and:

$$U(\text{INRIM}_1 - \text{INRIM}_2) = 2 \cdot \sqrt{u^2(\text{INRIM}_1) + u^2(\text{INRIM}_2) - 2 \cdot \text{cov}(\text{INRIM}_1, \text{INRIM}_2)} = 1,04 \text{ mL} \quad (4)$$

Therefore, the maximum value of their ratio, namely the worst normalized compatibility index  $E_n$  is:

$$(E_n)_{\max} = 2,15 \quad (5)$$

Such figures suggest that the TS stability was questionable during the first measurement trial already. In addition, Fig. 4 shows that the third pilot's measurement is obviously not compatible with the two previous ones.

## 7. Possible causes of the TS instability

The TS was constructed in Italy in 2003. It was chosen as TS for EURAMET project 1079, because, differently from CCM.FF-KC4 and its following regional comparisons, the aim was to check compatibility of results in a very ordinary calibration task, performed on an artefact of traditional type and average quality. Indeed, the TS had been already used by the pilot as a travelling standard in national Italian audits, showing a good stability.

In order to try and explain the very large volume decrease of the TS, several hypotheses were put forward. One of the first investigations focused on the repeatability of the amount of water left in the wetted tank.

FORCE, upon pilot's request, contributed 6 pictures of the interior of the TS; the most significant among them are shown in Figures 5 and 6. Fig. 5 shows the rather rough surface finish inside the neck, that could also be noted by any participant by visual inspection. The pilot can state that this surface did not undergo any significant change since construction, therefore it cannot have caused any repeatability problem.

Fig. 6 shows part of the inside cavity of the TS and (centre-bottom) of the volume correction body (displacer, or adjustment plunger). FORCE commented by e-mail that "*the inside edges (were) irregular and the inside walls and the displacer was rough and not smooth.....*". This comparatively poor surface quality, if continuously deteriorating, in principle might have increased the residual amount of water after dripping, thus decreasing the TS volume. However, the quantity of residual water was measured (by weighing) at least at four different times and found in average lower than 10 mL, with very limited deviations (see Note 2 below). Such figures definitely rule out the changes in quantity of retained water as a possible cause for the large volume change shown by the TS.



Fig. 5 - Inner surface of the TS neck



Fig. 6 - Insight into the TS cavity with bottom plate and displacer

*Note 1* – The TS was weighed again at INRIM in its totally dry condition after the end of circulation. The conventional mass (according to [ref. 2]) was 9677,63 g ( $U = 0,08$  g). Although not compatible, this figure is not far from the 9677,86 g value obtained in 2008 (see chapter 3.). A loss of mass in the order of a few tenths of a gram has been easily admitted.

*Note 2* – The following data for residual water are available, and surely many more, if deemed necessary, can be derived by all (or any of) participants, as differences between the TS weight in wetted and dry state:

INRIM, start (Oct. 2008): residual water was 8,98 g, std. deviation of 10 measurements:  $s = 0,24$  g

BoM (Jan. 2009): residual water 10,33 mL, st. deviation of 10 measurements:  $s = 0,56$  mL

INRIM, middle term (Aug. 2009): residual water 8,92 g, std. deviation of 10 measurements:  $s = 0,27$  g

IPQ (March 2010): residual water 8,23 g, standard deviation of 10 measurements:  $s = 0,21$  g

INRIM, final (April 2010): residual water 8,30 g, standard deviation of 10 measurements:  $s = 0,19$  g

Another possible cause for the observed change might be an accidental bump received by the TS in its bottom circular plate. Indeed, no bump can be observed on the side and top walls of the TS, that to date are completely exempt from visible damages, as they were at the Project start. The bottom plate, instead, was (and still is) somewhat irregular in planarity, as it is extensively welded to the side walls, as well as to a strengthening beam and to the flange which the displacer is bolted to. Therefore, a possible bump received during circulation may have produced a change of shape, which can pass undetected, being just a change of already existing curvatures.

However, this explanation is not very much considered, because of the care with which the TS has surely been treated by all participants. A damage cannot possibly have occurred during transportation of the closed case, owing to the very safe packaging. A possible, although very unlikely event, is an extremely rough handling at a Customs office, since they retain the right to open and check any case.

This hypothesis is further weakened by the fact that a bump causes a sudden change, whereas the TS seems to have changed quite rapidly, but not abruptly: indeed, the result of UME lies just in the middle of the transition.

A third hypothesis suggests exposure to extreme temperatures, either in the cargo bay of an aircraft, or during a cleaning procedure using very hot water. It should be considered that welding causes large temperature gradients in the metal bodies. As soon as the melted metal cord solidifies, the mechanical stresses stemming from differential temperatures and quick cooling are retained as if they were “frozen” inside the welded structure. This is the reason why welded structures to be further precision machined are always annealed, namely heated for many hours at temperatures above 100 °C (depending upon the material). This treatment allows the irreversible relieving of mechanical stresses, and gives durable stability to the structure.

Like probably all provers of the same type, the TS was not annealed after welding. This may have two consequences:

- 1) as already suggested in the Protocol, the actual thermal expansion coefficient of the *internal volume* of this artefact, differently from that of a massive or stress-relieved body, can differ from that of bulk metal, because at slightly variable ambient temperatures a different elastic



reaction of the material against its own residual stresses might alter not only the overall dimensions of the tank, but the shape itself of its cavity;

- 2) a severe heating (or – less probably – cooling) can break the equilibrium condition between the residual stresses and the material strength in the elastic range, thus triggering larger and *permanent* changes in shape of the concerned welded components (this is exactly what is purposely sought in the annealing thermal treatment). It is thought that this phenomenon might have affected at least the base plate of the TS, owing to its poor planarity. However, no evidence of this process is available; much less it is possible to tell whether it was slow or abrupt. It should also be considered that INRIM alone declared to have cleaned the TS with warm water (see 5.2), and the volume change did not definitely occur on, or soon after such occasions.

## 8. Analysis of results

### 8.1. General

Following the previous considerations, two groups of results have been identified. The former (Group 1) consists of the 10 measurements (or couples of measurements) between INRIM<sub>1</sub> and JV (included), whereas in the latter group (Group 2) the 6 measurements (or couples) between MKEH and INRIM<sub>3</sub> are considered.

The *gravimetric* volume result (not accompanied by a *volumetric* one) of UME lies just in between the two groups; in principle, it can be treated in three different ways:

- 1) if added to Group 1, it is clearly (Fig. 4) not compatible with any of the individual results;
- 2) if added to Group 2, it is not compatible with any of the individual results, but the *volumetric* result of CEM;
- 3) owing to these facts, on the hypothesis that the UME result was obtained just in the middle of the TS drift or change, the option suggested by the pilot and accepted by UME is not to include it in either group.

**The above defined two groups (*both without UME*) were analyzed separately, as if they came from measurements of two different TSs. In order to determine the degree of equivalence between laboratories belonging to different groups, they were linked to each other through the pilot's measurements (see chapter 9).**

### 8.2. Gravimetric results

Since the *gravimetric* method is, in principle, more accurate than the *volumetric* method and, in addition, the *gravimetric* results are available from all participants, the CRV, Comparison Reference Value ( $x_{CRV}$ ) of each group is determined by the *gravimetric* results alone, later to be compared to the *volumetric* results and to their weighted mean (or other), as a kind of lower-ranking *volumetric* reference value.

### 8.2.1. Analysis of the *gravimetric* results of Group 1

#### a) Results and determination of CRV(1)

For the calculation of the reference value a single datum from the pilot laboratory is adopted, namely the arithmetic mean of the first and second pilot's measurement results  $x_{\text{INRIM}}^G$  (where the G superscript refers to *gravimetric* measurement)

$$x_{\text{INRIM}}^G = \frac{\text{INRIM}_1 + \text{INRIM}_2}{2} = 20002,71 \text{ mL} \quad (6)$$

with its associated standard uncertainty:

$$u(x_{\text{INRIM}}^G) = \sqrt{\frac{u^2(\text{INRIM}_1) + u^2(\text{INRIM}_2)}{4} + \frac{\text{cov}(\text{INRIM}_1, \text{INRIM}_2)}{2}} = 0,78 \text{ mL} \quad (7)$$

As shown in chapter 6 (Eq. 5), the first and second pilot's measurements are not mutually compatible. This fact suggests a possible drift of the TS in the course of the first petal circulation already, therefore it was decided to introduce into the uncertainty budgets of each participant to this group a supplementary uncertainty contribution  $u_{\text{drift}}$ , so that:

$$u(x_i^G) = \sqrt{u^2(x_i) + u_{\text{drift}}^2} \quad (8)$$

where  $u_{\text{drift}}$  is computed as the standard deviation of a rectangular distribution:

$$u_{\text{drift}} = \left| \frac{\text{INRIM}_1 - \text{INRIM}_2}{2\sqrt{3}} \right| = 0,65 \text{ mL} \quad (9)$$

In Table 4 the results  $x_i^G$  and the so-increased standard uncertainties  $u(x_i^G)$  used to determine the reference value  $x_{\text{1CRV}}$  for Group 1 are shown.

Table 4 – Results of Group 1 (*gravimetric* method) with  $u_{\text{drift}}$  increased uncertainty

Participant Laboratory	Country	Volume ( <i>gravim. det.</i> ) $x_i^G$ / mL	Standard Uncertainty $u(x_i^G)$ / mL
INRIM	IT	20002,71	1,015
MIRS	SI	20000,32	0,802
DMDM	RS	20000,94	1,644
BoM	MK	19994,26	1,356
MBM	ME	19999,60	1,603
EIM	GR	19998,64	0,865
BIM	BG	19999,19	1,265
FORCE	DK	19998,97	1,226
JV	NO	19996,8	2,018

The reference value is first calculated by using the weighted mean [ref. 9]; the consistency of the result is checked using the  $\chi^2$  test, which fails if:

$$Pr\{\chi^2(\nu) > \chi_{obs}^2\} < 0,05 \quad (10)$$

where  $\nu$  are the degrees of freedom. According to the data in Table 4 the weighted mean  $x_{wm}$  using all nine laboratories ( $\nu = 8$ ) is:

$$x_{wm} = 19999,46 \text{ mL} \quad (11)$$

$$u(x_{wm}) = 0,39 \text{ mL} \quad (12)$$

The observed  $\chi^2$  ( $\chi_{obs}^2 = 29,8$ ), compared to the expected one ( $\chi^2 = 15,5$ ), shows that the consistency test fails. However, if the datum by BoM is removed, the consistency test with the 8 remaining laboratories ( $\nu = 7$ ) passes ( $\chi_{obs}^2 = 13,7$  and  $\chi^2 = 14,1$ ). Therefore, the newly calculated weighted mean value is taken as the Comparison Reference Value for Group 1, CRV(1) or with mathematical symbol  $x_{1CRV}$ :

$$x_{1CRV} = 19999,92 \text{ mL} \quad (13)$$

$$u(x_{1CRV}) = 0,40 \text{ mL} \quad (14)$$

CRV(1) and its expanded uncertainty band is represented in the plot of Fig. 7.

### b) Degrees of equivalence with respect to the CRV(1) of Group 1

The degree of equivalence between the laboratory  $i$ -th and the reference value CRV(1) is determined on the basis of [ref. 9]:

$$\Delta x_i^G = x_i^G - x_{1CRV} \quad (15)$$

$$U(\Delta x_i^G) = 2\sqrt{u^2(x_i^G) - u^2(x_{1CRV})} \quad (16)$$

Since  $x_{1CRV}$  and the result from BoM are not correlated to each other,

$$U(\Delta x_{BoM}^G) = 2\sqrt{u^2(x_{BoM}^G) + u^2(x_{1CRV})} \quad (17)$$

The results are presented in the following Table 5:

Table 5 - Group 1, degrees of equivalence (*gravimetric* method)

	INRIM	MIRS	DMDM	MBM	EIM	BIM	FORCE	JV	BoM
$\Delta x_i^G / \text{mL}$	2,79	0,40	1,02	-0,32	-1,28	-0,73	-0,95	-3,12	-5,66
$U(\Delta x_i^G) / \text{mL}$	1,86	1,39	3,19	3,10	1,53	2,40	2,33	3,95	2,83

The results of INRIM as well as those of BoM are not consistent with the CRV(1) (given by eq. 13), since the (absolute) difference between their volume data and CRV(1) is larger than the expanded uncertainty of the said difference. They are coloured in yellow.

In August, 2011 BoM mentioned to the pilot a possible cause for their discrepancy. Having performed some tests by calibrating *gravimetrically* a volume standard of theirs, they found that its volume, as determined using the water supplied by their own distillation apparatus, was different from that obtained by using a sample of certainly pure water of external origin. In fact, the purity of the in-house distilled water had not been checked (by measuring its conductivity, or otherwise) at the time of BoM measurements.

## 8.2.2. Analysis of the *gravimetric* results of Group 2

### a) Results and determination of CRV(2)

In this group only one result from the pilot laboratory is available (INRIM<sub>3</sub>), so that no check of the possible final instability of the TS could be made. Therefore, no addition was made to the original uncertainty budgets submitted by participants. In the following table, the standard uncertainties were computed by the pilot as expanded uncertainties divided by the relevant coverage factors  $k$ :

Table 6 – Results of Group 2 (*gravimetric* method)

Participant Laboratory	Country	Volume ( <i>gravim. det.</i> ) $x_i^G /$ mL	Standard Uncertainty $u(x_i^G) /$ mL
MKEH	HU	19981,60	2,010
SMD	BE	19983,6	0,54 (*)
VMT	LT	19983,81	0,825
CEM	ES	19984,49	1,315
IPQ	PT	19982,33	1,050
INRIM <sub>3</sub>	IT	19985,62	0,830

(\*) SMD uses a coverage factor  $k = 2,1$  for its *gravimetric* measurement alone. All other coverage factors in this Report were equal to 2.

According to the data in Table 6 the weighted mean using the 6 laboratories ( $\nu = 5$ ) is determined. Since  $\chi_{\text{obs}}^2 = 8,3$  is less than the expected value  $\chi^2 = 11,1$ , the result is consistent, and the calculated value  $x_{2\text{CRV}}$  is taken as CRV(2), or Comparison Reference Value for Group 2 participants:

$$x_{2\text{CRV}} = 19983,86 \text{ mL} \quad (18)$$

$$u(x_{2\text{CRV}}) = 0,35 \text{ mL} \quad (19)$$

### b) Degrees of equivalence with respect to the CRV(2) of Group 2

The degree of equivalence between the laboratory  $i$ -th and the reference value  $x_{2\text{CRV}}$  is determined on the basis of:

$$\Delta x_i^G = x_i^G - x_{2\text{CRV}} \quad (20)$$

$$U(\Delta x_i^G) = 2\sqrt{u^2(x_i^G) - u^2(x_{2CRV})} \quad (21)$$

The results are presented in the following Table 7:

Table 7 - Group 2, degrees of equivalence (*gravimetric method*)

	MKEH	SMD	VMT	CEM	IPQ	INRIM <sub>3</sub>
$\Delta x_i^G$ /mL	-2,26	-0,26	-0,05	0,63	-1,53	1,76
$U(\Delta x_i^G)$ /mL	3,96	0,82	1,49	2,53	1,98	1,50

The result INRIM<sub>3</sub> is not consistent with CRV(2).

Both Comparison Reference Values and their expanded uncertainty bands are represented in the following plot (Fig. 7).

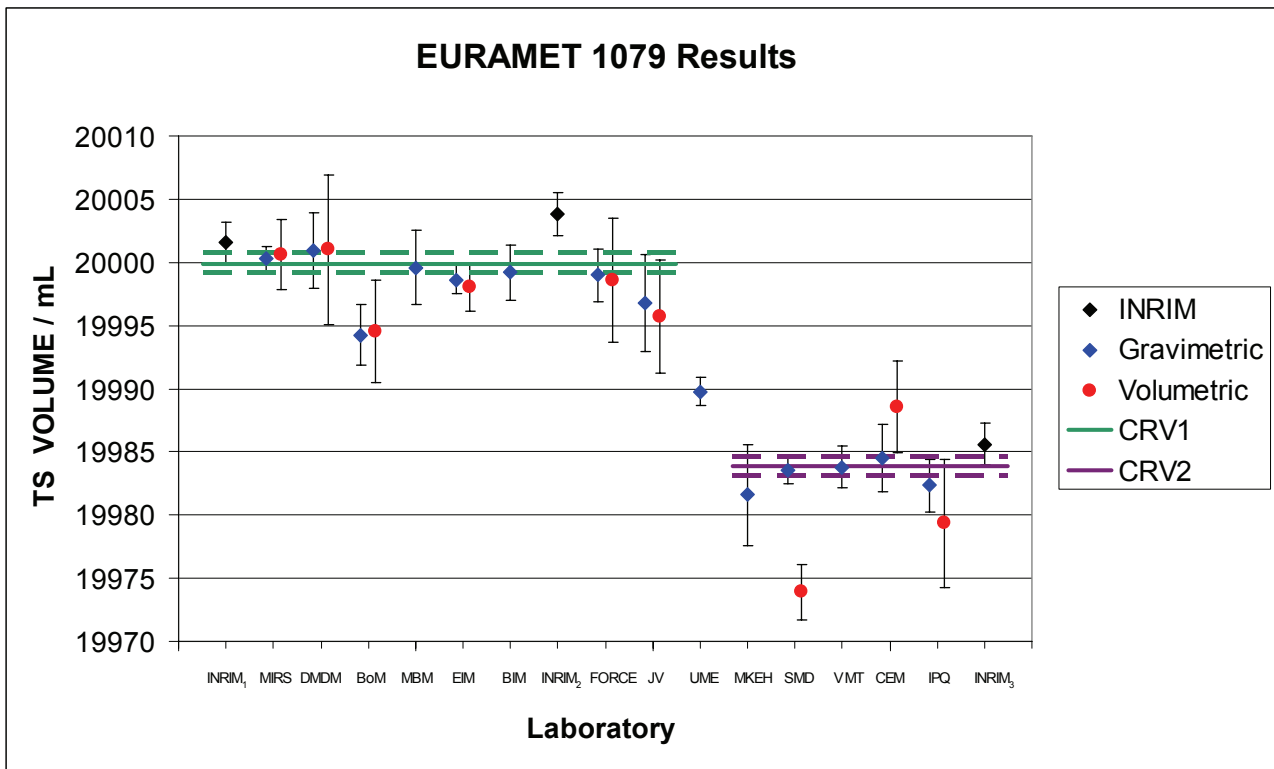


Fig. 7 – Plot of results with lines corresponding to the two Comparison Reference Values (based on *gravimetric* measurements alone) with their own expanded uncertainty bands



### 8.3. Volumetric results

#### 8.3.1. Comparing the volumetric results of double-task participants to their own gravimetric data

Nine laboratories, the so-called *double-task participants*, were able to perform measurements with the *volumetric* method in addition to the *gravimetric* one.

In order to compare the two methods, the volume data and associated uncertainties obtained by each laboratory using both methods are shown in Table 8. As expected, the uncertainties associated with the *volumetric* method are always larger than the *gravimetric* ones. The percent increase ranges from just + 16 % up to + 194 % (+ 87 % in average), as shown in the last but two column. More details are given in chapter 10.

Table 8 - Double-task participants, results of measurements with both methods

	Laboratory & Country		Volume (gravim.)	Std. Unc.	Volume (volum.)	Std. Unc.	Ratio betw. uncertainties	Volume difference between methods	Expanded uncertainty of difference
			$x_i^G$ / mL	$u(x_i^G)$ / mL	$x_i^V$ / mL	$u(x_i^V)$ / mL	$u(x_i^V)/u(x_i^G)$	$(x_i^G - x_i^V)$ / mL	$U(x_i^G - x_i^V)$ / mL
Group no. 1	MIRS	SI	20000,32	0,47	20000,62	1,38	2,94	-0,30	2,92
	DMDM	RS	20000,94	1,51	20001,02	2,97	1,96	-0,08	6,65
	BoM	MK	19994,26	1,19	19994,53	2,01	1,68	-0,27	4,66
	EIM	GR	19998,64	0,57	19998,05	0,94	1,65	0,59	2,20
	FORCE	DK	19998,97	1,04	19998,6	2,45	2,35	0,37	5,31
	JV	NO	19996,8	1,91	19995,7	2,23	1,16	1,10	5,86
Group no. 2	SMD	BE	19983,6	0,54	19973,9	1,10	2,04	9,70	2,45
	CEM	ES	19984,49	1,32	19988,57	1,82	1,38	-4,08	4,48
	IPQ	PT	19982,33	1,05	19979,34	2,53	2,40	2,99	5,47

The last column, worked out by the pilot, should be considered just as indicative: indeed, the *gravimetric* and *volumetric* measurement chains at a same laboratory most probably exhibit some degree of correlation (as confirmed by the Youden plot in clause 8.4), for instance due to the use of a same set of auxiliary instruments, or the traceability to a same national realization of the scale of masses, of temperatures, pressures and so on.

In principle, each participant alone is in such a position as to be able to evaluate her/his own covariance terms, but this evaluation was not requested by the Protocol; therefore, the values in the last column are computed by the pilot as if the two sets of measurements were mutually independent. This means that such values are actually the upper limit of the uncertainty of the difference between the two methods at each laboratory (since any common uncertainty contribution will decrease the uncertainty of the difference, thus increasing the chances of discrepancies).

The couple of results from SMD shows a volume difference (last but one column) larger than the uncertainty of the difference itself (last column); see discussion at the end of clause 8.3.3 (a).

### 8.3.2. Analysis of the *volumetric* results of Group 1

#### a) Results and degrees of equivalence

Let us first recall that the data analysis presented in this Report is based upon the following assumptions:

- the *gravimetric* measurements, besides being the only ones common to all participants, are more reliable, therefore apt to be used as references for the *volumetric* ones;
- according to the only repeated measurements that are available, those of the pilot, the TS underwent a slight, but true drift during the former part of the circulation already;
- as well as a dramatic volume change during the latter part of circulation;
- there are no repeated measurements in the latter part of circulation, therefore a possible drift cannot be evaluated.

Under those hypotheses, the supplementary uncertainty contribution  $u_{\text{drift}}$ , given by (Eq. 9) must be introduced into the *volumetric* uncertainty budgets of the Group 1 participants, and just in them alone.

In Table 9 the *volumetric* measurement results of Group 1,  $x_i^V$ , and the (so increased) standard uncertainties  $u(x_i^V)$  are shown:

Table 9 – Results of Group 1 (*volumetric* method) with increased uncertainty

Participant Laboratory	Country	$x_i^V$ / mL	$u(x_i^V)$ / mL
MIRS	SI	20000,62	1,525
DMDM	RS	20001,02	3,035
BoM	MK	19994,53	2,108
EIM	GR	19998,05	1,143
FORCE	DK	19998,6	2,530
JV	NO	19995,70	2,318

The degree of equivalence between the laboratory  $i$ -th and the *gravimetrically determined* Comparison Reference Value  $x_{1\text{CRV}}$  (Eq. 13), alias CRV(1), is computed on the basis of the two following quantities:

$$\Delta x_i^V = x_i^V - x_{1\text{CRV}} \quad (22)$$

As the  $x_{1\text{CRV}}$  and the  $x_i^V$  values are assumed to be mutually uncorrelated, the uncertainty of the difference is:

$$U(\Delta x_i^V) = 2\sqrt{u^2(x_i^V) + u^2(x_{1\text{CRV}})} \quad (23)$$

The results are shown in Table 10:

Table 10 - Group 1, degrees of equivalence (*volumetric* method) with reference to CRV(1)

	MIRS	DMDM	BoM	EIM	FORCE	JV
$\Delta x^V_i / \text{mL}$	0,70	1,10	-5,39	-1,87	-1,32	-4,22
$U(\Delta x^V_i) / \text{mL}$	3,16	6,12	4,29	2,42	5,12	4,71

The volume difference in the first line is larger than the corresponding expanded uncertainty for BoM.

**b) Group 1: Comparison between the *gravimetrically* determined CRV(1) and the weighted mean of *volumetric* measurements**

In order to compare the *volumetric* method to the *gravimetric* method, a reference value based on the *volumetric* results is determined, too. **This is intended just as an auxiliary datum, the only accepted CRV for Group 1 being the quantity CRV(1), namely:**

$$x_{1\text{CRV}} = 19999,92 \text{ mL} \quad (13 \text{ and } 24)$$

The *volumetric* weighted mean is:

$$x^V_{1\text{wm}} = 19998,20 \text{ mL} \quad (25)$$

$$u(x^V_{1\text{wm}}) = 0,73 \text{ mL} \quad (26)$$

Taking into consideration that the observed  $\chi^2$  ( $\chi_{\text{obs}}^2 = 7,6$ ) is lower than the expected one ( $\chi^2 = 11,1$ ) the consistency test is passed.

When comparing the two results obtained with the two methods, the difference between them is:

$$\Delta x_1 = x^V_{1\text{wm}} - x^G_{1\text{CRV}} = -1,72 \text{ mL} \quad (27)$$

and the corresponding uncertainty is:

$$U(\Delta x_1) = 2\sqrt{u^2(x^V_{1\text{wm}}) + u^2(x^G_{1\text{CRV}})} = 1,67 \text{ mL} \quad (28)$$

These figures show that the *volumetric* datum is lower than the *gravimetric* one, and that the two results are (marginally) not compatible with each other.

**8.3.3. Analysis of the *volumetric* results of Group 2**

**a) Results and degrees of equivalence**

In the second loop only three laboratories performed calibration with the *volumetric* method; their results are shown in Table 11 as submitted (namely without the additional  $u_{\text{drift}}$  contribution to uncertainty):

Table 11 – Recall of results of Group 2 (*volumetric* method)

Participant Laboratory no.	Country	$x^V_{i/}$ / mL	$u(x^V_{i/})$ / mL
SMD	BE	19973,9	1,100
CEM	ES	19988,57	1,815
IPQ	PT	19979,34	2,525

Following the same procedure as for Group 1, the degrees of equivalence between each laboratory and the Comparison Reference Value CRV(2), namely  $x_{2CRV}$  (Eq. 18), have been determined, as shown in Table 12:

Table 12 - Group 2, degrees of equivalence (*volumetric* method)

	SMD	CEM	IPQ
$\Delta x^V_{i/}$ / mL	-9,96	4,71	-4,52
$U(\Delta x^V_{i/})$ / mL	2,31	3,70	5,10

The *volumetric* datum from SMD has been long recognized by the laboratory itself as largely discrepant from their own *gravimetric* result; however, in spite of every effort, the cause of this discrepancy could not be found nor corrected before the issue of Draft A (April, 2011).

Soon after, SMD sent a mail of acceptance of Draft A in which (having known and examined the results by some other participants), they also say that the problem with their *volumetric* measurements was probably caused by the use of a very “uncomfortable” hand-operated pipette to measure the volume of water in excess of 20 L (the quantity  $\delta$  in Eq. 1 of the Protocol). SMD believes that the adoption of some other device, maybe an automatic one, would improve their measurements.

Looking at the plot in Fig. 4, it can be easily observed that the *gravimetric* result of SMD is compatible with all but one of the other *gravimetric* results in Group 2, as well as with the CRV(2). Therefore, the outlier is surely their *volumetric* result, which is no longer considered in the following clause.

The difference between the *volumetric* result of CEM and CRV(2), too, is larger than the relevant uncertainty.

**b) Group 2: Comparison between the *gravimetrically* determined CRV(2) and the *volumetric* results of CEM and IPQ**

The *volumetric* results of the (just two) remaining participants (CEM and IPQ) are not mutually compatible, as shown in Fig. 4, making it almost pointless to compute any sort of *volumetric* reference value. However, for sake of completeness, it is possible to carry out the following computations. The weighted mean of the two results is:

$$x^V_{2wm} = 19985,43 \text{ mL} \quad (29)$$

$$u(x^V_{2wm}) = 1,47 \text{ mL} \quad (30)$$

but the  $\chi^2$  test fails ( $\chi_{\text{obs}}^2=8,8$  , expected  $\chi^2=3,8$ ).

The mean value of the two results is:

$$x_{2\text{mean}}^{\text{V}} = \frac{x_{\text{CEM}}^{\text{V}} + x_{\text{IPQ}}^{\text{V}}}{2} = 19983,96 \text{ mL} \quad (31)$$

with an associated uncertainty:

$$u(x_{2\text{mean}}^{\text{V}}) = \sqrt{\frac{u^2(x_{\text{CEM}}^{\text{V}}) + u^2(x_{\text{IPQ}}^{\text{V}})}{4}} = 1,55 \text{ mL} \quad (32)$$

The comparison of these data to CRV(2) yields:

$$\Delta x_2 = x_{2\text{mean}} - x_{2\text{CRV}} = 0,09 \text{ mL} \quad (33)$$

$$U(\Delta x_2) = 2\sqrt{u^2(x_{2\text{mean}}^{\text{V}}) + u^2(x_{2\text{CRV}})} = 3,19 \text{ mL} \quad (34)$$

Therefore, in spite of the above highlighted inconsistencies, the arithmetic mean of the two retained *volumetric* results  $x_{2\text{mean}}$  is in good agreement with the relevant CRV, namely CRV(2).

#### 8.4. Youden plot

Youden plots have been widely adopted to show (in a very effective graphical way) results with their bias, uncertainties and correlations wherever two sets of data have been obtained, more or less simultaneously, on a same measurand quantity.

In the present case, plotting the sets of 9 volumes obtained by double-task participants with the *volumetric* method vs. those obtained *gravimetrically* would lead to two separate clusters of experimental points, with too few laboratories in each cluster, owing to the change of TS volume.

Therefore, it seemed better to plot the Degrees of Equivalence (DoEs): with CRV(1) for Group 1 participants and with CRV(2) for Group 2. Tables 5 and 7 provide entries for the  $x$ -axis, whereas Tables 10 and 12 are used for  $y$ -axis.

In Fig. 8, the laboratories closer to the origin are obviously those with the best agreement, both *gravimetrically* and *volumetrically*, with the CRV of their own group. The horizontal and vertical separations from the two coordinate axes show deviations of the *gravimetric* or *volumetric* method respectively. However, the feature that Youden plots very often are able to highlight is the correlation between the two sets of data, as well as the presence of biases. Indeed, in this case, as in very many other applications, the trend of the cluster of experimental points to align along the bisecting line of the first and third quadrant shows two facts:

- the presence of an appreciable intra-laboratory correlation between the two calibration methods in at least 6 out of 9 laboratories;

- the presence, in the same 6 laboratories, of a similar bias in the two methods (namely comparable overestimation or underestimation).

Of course the position of SMD and CEM, whose points are close to the  $y$ -axis, but comparatively far away from the  $x$ -axis, shows a low intra-laboratory correlation, with the implementation of the *volumetric* method appearing not as good as that of the *gravimetric* method.

Looking at the whole of these 9 double-task participants, it can be seen that the optimum configuration of the cluster, with experimental points evenly (and closely) distributed inside a circle centred in the origin has not been achieved. Such a desirable configuration would mean that all deviations, in both methods, are comparable, as well as randomly distributed, therefore there is no appreciable, definite and possibly permanent (one would say *systematic*) deviation between laboratories.

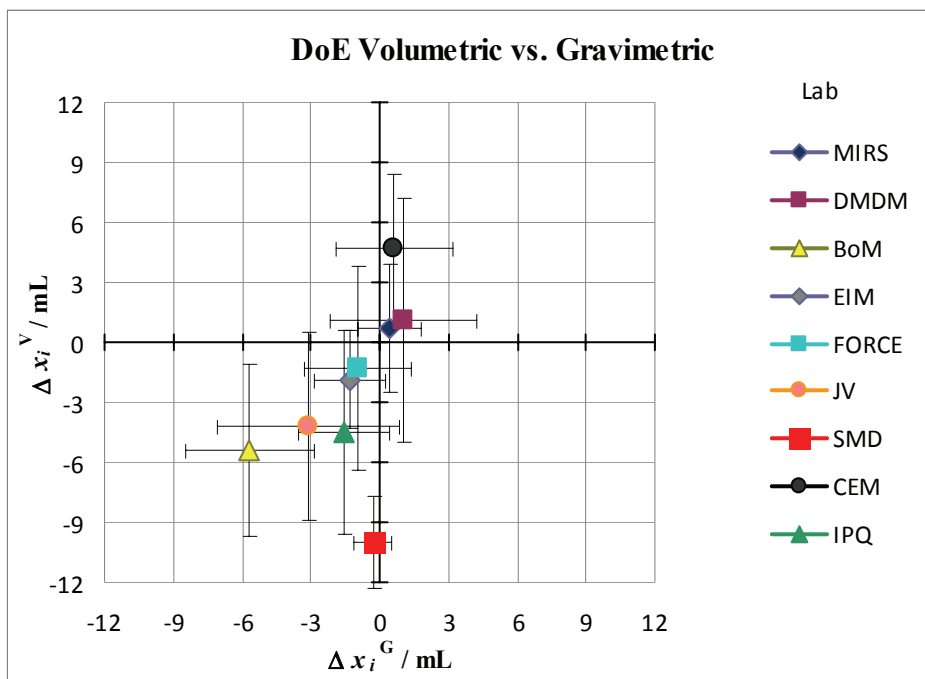


Figure 8 – Youden plot of the Degrees of Equivalence for double-task participants. The differences ( $\Delta x_i^G$  and  $\Delta x_i^V$ ) between each laboratory's measurements and the relevant CRV (CRV1 or CRV2) are plotted, with their expanded uncertainty bars, on the  $x$  and  $y$  axes respectively.



## 9. Degrees of equivalence between participants

The degree of equivalence between any  $i$  and  $j$  laboratories belonging to a same Group is evaluated on the basis of the two following data:

$$\Delta x_{ij} = x_i - x_j \quad (35)$$

$$U(\Delta x_{ij}) = 2\sqrt{u^2(x_i) + u^2(x_j)} \quad (36)$$

For the laboratories belonging to Group 1 the supplementary uncertainty  $u_{\text{drift}}$  (see Eq. 9) is taken into account and combined with the original participants' uncertainty contributions.

When the pilot laboratory is involved, the selected result is that of the corresponding group,  $x_{1\text{INRIM}}$  (given by eq. 6) for Group 1 and  $x_{2\text{INRIM}}$  (namely the result of the final measurement INRIM<sub>3</sub>) for Group 2, respectively.

Laboratories belonging to different Groups are linked to each other by taking into account the assumed change of the TS volume,  $\Delta\text{group}$ , as it can be estimated by means of the *gravimetric* measurements carried out by the pilot:

$$\Delta\text{group} = x_{1\text{INRIM}}^{\text{G}} - x_{2\text{INRIM}}^{\text{G}} = 17,09 \text{ mL} \quad (37)$$

$$u(\Delta\text{group}) = \sqrt{u^2(x_{1\text{INRIM}}^{\text{G}}) + u^2(x_{2\text{INRIM}}^{\text{G}}) - 2 * \text{cov}(x_{1\text{INRIM}}^{\text{G}}, x_{2\text{INRIM}}^{\text{G}})} = 0,81 \text{ mL} \quad (38)$$

**Note:** the value in Eq. (37) is slightly higher than the difference between CRV(1) and CRV(2) given by Eq. (13) and (18), namely 16,06 mL, whose standard uncertainty (if we neglect the very weak correlation contributed by the pilot's multiple measurements) is 0,53 mL.

As an example, the degree of equivalence between the  $i$ -th laboratory of Group 1 and the laboratory  $j$ -th of Group 2 is given by:

$$\Delta x_{i2j} = x_{i1} - x_{2j} - \Delta\text{group} \quad (39)$$

$$U(\Delta x_{i2j}) = 2\sqrt{u^2(x_i) + u^2(x_j) + u^2(\Delta\text{group})} \quad (40)$$

The same procedure is used to determine the degrees of equivalence with both methods, *gravimetric* and *volumetric*. The results are shown in Tables 13 and 14. The entries where the difference between two laboratories is higher than the relevant uncertainty are colored in yellow.

With the method chosen in the present analysis no degree of equivalence could be determined for UME.

Table 13 – Degrees of equivalence between laboratories for *gravimetric* volume measurements (all data are given in mL; the figures in column and line headings refer to the Laboratory own number)

i \ j		INRIM	MIRS	DMDM	BoM	MBM	EIM	BIM	FORCE	JV	MKEH	SMD	VMT	CEM	IPQ
INRIM	$\Delta x_{ij}$		2,39	1,77	8,45	3,11	4,07	3,52	3,74	5,91	4,02	2,02	1,81	1,13	3,29
	$U(\Delta x_{ij})$		2,59	3,86	3,39	3,79	2,67	3,24	3,18	4,52	4,35	1,98	2,34	3,11	2,68
MIRS	$\Delta x_{ij}$	-2,39		-0,62	6,06	0,72	1,68	1,13	1,35	3,52	1,63	-0,37	-0,58	-1,26	0,90
	$U(\Delta x_{ij})$	2,59		3,66	3,15	3,58	2,36	3,00	2,93	4,34	4,62	2,53	2,82	3,48	3,10
DMDM	$\Delta x_{ij}$	-1,77	0,62		6,68	1,34	2,30	1,75	1,97	4,14	2,25	0,25	0,04	-0,64	1,52
	$U(\Delta x_{ij})$	3,86	3,66		4,26	4,59	3,71	4,15	4,10	5,21	5,44	3,82	4,02	4,51	4,23
BoM	$\Delta x_{ij}$	-8,45	-6,06	-6,68		-5,34	-4,38	-4,93	-4,71	-2,54	-4,43	-6,43	-6,64	-7,32	-5,16
	$U(\Delta x_{ij})$	3,39	3,15	4,26		4,20	3,22	3,71	3,66	4,86	5,11	3,34	3,57	4,11	3,80
MBM	$\Delta x_{ij}$	-3,11	-0,72	-1,34	5,34		0,96	0,41	0,63	2,80	0,91	-1,09	-1,30	-1,98	0,18
	$U(\Delta x_{ij})$	3,79	3,58	4,59	4,20		3,64	4,08	4,04	5,15	5,39	3,75	3,95	4,45	4,16
EIM	$\Delta x_{ij}$	-4,07	-1,68	-2,30	4,38	-0,96		-0,55	-0,33	1,84	-0,05	-2,05	-2,26	-2,94	-0,78
	$U(\Delta x_{ij})$	2,67	2,36	3,71	3,22	3,64		3,06	3,00	4,39	4,67	2,61	2,89	3,54	3,17
BIM	$\Delta x_{ij}$	-3,52	-1,13	-1,75	4,93	-0,41	0,55		0,22	2,39	0,50	-1,50	-1,71	-2,39	-0,23
	$U(\Delta x_{ij})$	3,24	3,00	4,15	3,71	4,08	3,06		3,52	4,76	4,75	2,75	3,02	3,65	3,29
FORCE	$\Delta x_{ij}$	-3,74	-1,35	-1,97	4,71	-0,63	0,33	-0,22		2,17	0,28	-1,72	-1,93	-2,61	-0,45
	$U(\Delta x_{ij})$	3,18	2,93	4,10	3,66	4,04	3,00	3,52		4,72	4,98	2,68	2,96	3,60	3,23
JV	$\Delta x_{ij}$	-5,91	-3,52	-4,14	2,54	-2,80	-1,84	-2,39	-2,17		-1,89	-3,89	-4,10	-4,78	-2,62
	$U(\Delta x_{ij})$	4,52	4,34	5,21	4,86	5,15	4,39	4,76	4,72		5,92	4,48	4,65	5,08	4,83
MKEH	$\Delta x_{ij}$	-4,02	-1,63	-2,25	4,43	-0,91	0,05	-0,50	-0,28	1,89		-2,00	-2,21	-2,89	-0,73
	$U(\Delta x_{ij})$	4,35	4,62	5,44	5,11	5,39	4,67	5,02	4,98	5,92		4,16	4,35	4,80	4,54
SMD	$\Delta x_{ij}$	-2,02	0,37	-0,25	6,43	1,09	2,05	1,50	1,72	3,89	2,00		-0,21	-0,89	1,27
	$U(\Delta x_{ij})$	1,98	2,53	3,82	3,34	3,75	2,61	3,19	3,13	4,48	4,16		1,97	2,84	2,36
VMT	$\Delta x_{ij}$	-1,81	0,58	-0,04	6,64	1,30	2,26	1,71	1,93	4,10	2,21	0,21		-0,68	1,48
	$U(\Delta x_{ij})$	2,34	2,82	4,02	3,57	3,95	2,89	3,43	3,37	4,65	4,35	1,97		3,10	2,67
CEM	$\Delta x_{ij}$	-1,13	1,26	0,64	7,32	1,98	2,94	2,39	2,61	4,78	2,89	0,89	0,68		2,16
	$U(\Delta x_{ij})$	3,11	3,48	4,51	4,11	4,45	3,54	3,99	3,95	5,08	4,80	2,84	3,10		3,37
IPQ	$\Delta x_{ij}$	-3,29	-0,90	-1,52	5,16	-0,18	0,78	0,23	0,45	2,62	0,73	-1,27	-1,48	-2,16	
	$U(\Delta x_{ij})$	2,68	3,10	4,23	3,80	4,16	3,17	3,67	3,61	4,83	4,54	2,36	2,67	3,37	

Table 14 – Degrees of equivalence between laboratories for *volumetric* volume measurements (all figures are given in mL)

i \ j		MIRS	DMDM	BoM	EIM	FORCE	JV	SMD	CEM	IPQ
MIRS	$\Delta x_{ij}$		-0,40	6,09	2,57	2,02	4,92	9,63	-5,04	4,19
	$U(\Delta x_{ij})$		6,79	5,20	3,81	5,91	5,55	4,10	5,01	6,12
DMDM	$\Delta x_{ij}$	0,40		6,49	2,97	2,42	5,32	10,03	-4,64	4,59
	$U(\Delta x_{ij})$	6,79		7,39	6,49	7,90	7,64	6,66	7,26	8,06
BoM	$\Delta x_{ij}$	-6,09	-6,49		-3,52	-4,07	-1,17	3,54	-11,13	-1,90
	$U(\Delta x_{ij})$	5,20	7,39		4,80	6,59	6,27	5,02	5,80	6,78
EIM	$\Delta x_{ij}$	-2,57	-2,97	3,52		-0,55	2,35	7,06	-7,61	1,62
	$U(\Delta x_{ij})$	3,81	6,49	4,80		5,55	5,17	3,56	4,59	5,78
FORCE	$\Delta x_{ij}$	-2,02	-2,42	4,07	0,55		2,90	7,61	-7,06	2,17
	$U(\Delta x_{ij})$	5,91	7,90	6,59	5,55		6,86	5,75	6,44	7,33
JV	$\Delta x_{ij}$	-4,92	-5,32	1,17	-2,35	-2,90		4,71	-9,96	-0,73
	$U(\Delta x_{ij})$	5,55	7,64	6,27	5,17	6,86		5,38	6,11	7,05
SMD	$\Delta x_{ij}$	-9,63	-10,03	-3,54	-7,06	-7,61	-4,71		-14,67	-5,44
	$U(\Delta x_{ij})$	4,10	6,66	5,02	3,56	5,75	5,38		4,24	5,51
CEM	$\Delta x_{ij}$	5,04	4,64	11,13	7,61	7,06	9,96	14,67		9,23
	$U(\Delta x_{ij})$	5,01	7,26	5,80	4,59	6,44	6,11	4,24		6,22
IPQ	$\Delta x_{ij}$	-4,19	-4,59	1,90	-1,62	-2,17	0,73	5,44	-9,23	
	$U(\Delta x_{ij})$	6,12	8,06	6,78	5,78	7,33	7,05	5,51	6,22	

## 10. Analysis of the main components of uncertainty budgets

### 10.1. Type A, Type B overall components and their composition

Table 15 recalls the expanded uncertainty data already given in Table 3, with additional details about Type A and B components, degrees of freedom and coverage factors. A few notes are required:

- In the front page of the standard spreadsheet sent out and approved before the launch of the Project in order to submit results, it is directly suggested that the Experimental Standard Deviation (ESD) of the 10 repeat measurements constitutes the Type A component of uncertainty. It appears, on the contrary, that the Procedures of many participants adopt the Standard Deviation of the Mean (SDoM) of the 10 measurements. The entries of the participants that adopted SDoM instead of ESD, as their Type A component, are evidenced in bold;
- This discrepancy reflects actual and significant differences not only in the present comparison, but also in the process leading to various national calibration certificates, the ESD being in the present case as much as 3,16 times larger than SDoM. The discussions that took place during the preparation of the recently issued EURAMET Guide devoted to the *gravimetric* method [ref. 5, clause 5.3.9], and of the similar Guide expected to be released soon about the *volumetric* method for volume calibrations, suggested to prefer the adoption of SDoM in calibration certificates;

- JV and CEM entered, as Type A component in their front page: SDoM for their *gravimetric* measurement and ESD for *volumetric* measurement; the reason for this double choice was not explained;
- Of course the GUM definition of Type A components (those determined by statistical calculations) in principle may include contributions other than ESD or SDoM. However none among participants claimed different, or extra, Type A contributions.

Table 15 –Type A and B uncertainty contributions and their composition

	Laboratory	INRIM	MIRS	DMDM	BoM	MBM	EIM	BIM	FORCE	JV	UME	MKEH	SMD	VMT	CEM	IPQ
Grav.	Type A / mL	0,39	0,16	0,38	<b>0,23</b>	<b>0,26</b>	<b>0,30</b>	<b>0,23</b>	<b>0,62</b>	<b>0,25</b>	0,03	1,74		0,41	<b>0,18</b>	<b>0,34</b>
	Type B / mL	0,75	0,44	1,46	1,17	1,44	0,49	1,06	0,83	1,89	0,57	1,01		0,71	1,30	0,99
	Std. unc. / mL	0,85	0,47	1,51	1,19	1,46	0,57	1,08	1,04	1,91	0,57	2,01	0,54	0,82	1,31	1,05
	v not less than:	50		2139	354	100	60	280	$\infty$	249	60	146	34	140	60	723
	<i>k</i>	2	2	2	2	2	2	2	2	2	2	2	2,1	2	2	2
	Exp. Unc. /mL	1,69	0,94	3,02	2,38	2,93	1,14	2,17	2,08	3,82	1,14	4,02	1,1	1,65	2,63	2,10
Vol.	Type A / mL	-	0,79	0,28	<b>0,12</b>	-	<b>0,63</b>	-	<b>0,29</b>	0,24	-	-		-	0,23	<b>0,48</b>
	Type B / mL	-	1,13	2,95	2,00	-	0,70	-	2,43	2,21	-	-		-	1,80	2,48
	Std. unc / mL	-	1,38	2,96	2,00	-	0,94	-	2,45	2,22	-	-	1,1	-	1,81	2,52
	v not less than:	-		18546	77	-	40	-	$\infty$	35	-	-	66	-	40	96
	<i>k</i>	-	2	2	2	-	2	-	2	2	-	-	2	-	2	2
	Exp. Unc. /mL	-	2,76	5,93	4,01	-	1,88	-	4,89	4,45	-	-	2,2	-	3,63	5,05

Further notes:

- The uncertainty data of INRIM are those of the mid term measurement (INRIM<sub>2</sub> elsewhere in the text), that are those with the largest uncertainty;
- type B uncertainty of the *volumetric* calibration alone of MIRS was found to be the arithmetic sum of its components, given in the detailed budget (see Table 18). In spite of this deviation from the GUM method of composition, the declared 1,13 mL overall figure is retained;
- all other entries are correctly derived as square root composition of their components, or differ from the exact figure by not more than a few hundredths of a millilitre (from 0,01 to 0,04 mL), owing to rounding approximations;
- SMD made calculations by means of a dedicated computer program of their own. The program and the already mentioned two calibration reports did not provide to the pilot the list of the 2 x 10 separate volume data, nor the value, expressed in volume units, of each component of the uncertainty budget. Therefore, some cells in Table 15 are left blank.

Figures 9 and 10 show graphically the Type A and Type B components of uncertainty and their composition (standard uncertainty) for the *gravimetric* and *volumetric* method respectively.

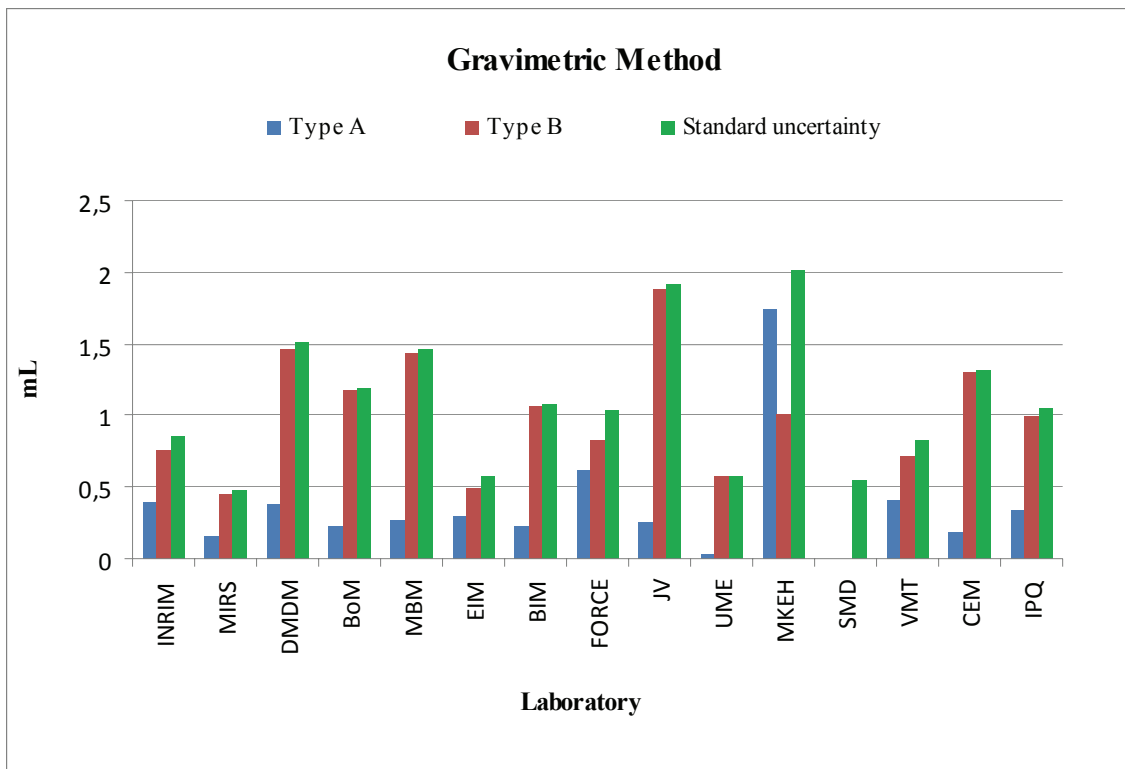


Figure 9 – Gravimetric method: Type A and Type B uncertainties and their composition

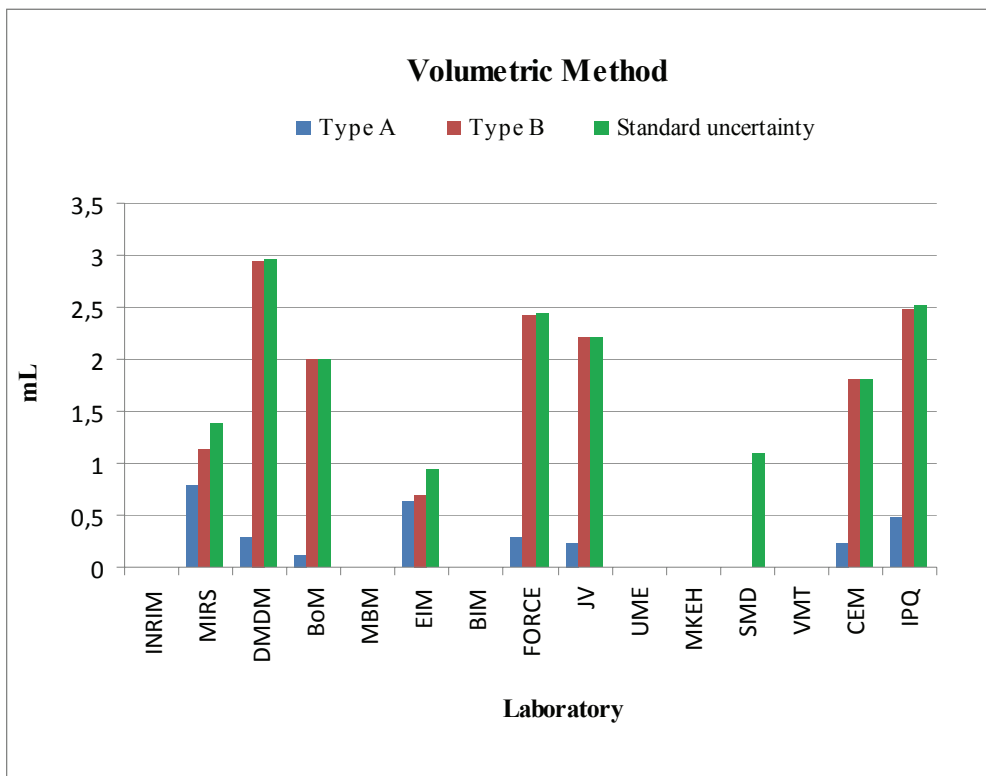


Figure 10 - Volumetric method : Type A and B uncertainties and their composition

The following table shows some statistics about standard uncertainties and their components.

Table 16 - Statistics of Type A (not increased by  $u_{\text{drift}}$ , as declared) and Type B components, together with standard uncertainties

	Minimum value mL	Maximum value mL	Arithm. mean mL	Standard deviation mL
<i>Gravimetric</i> Type A	0,03 (0,16)	1,74 (0,62)	0,39 (0,31)	0,41 (0,13)
<i>Volumetric</i> Type A	0,12	0,79	0,38	0,23
<i>Gravimetric</i> Type B	0,44	1,89	1,07	0,49
<i>Volumetric</i> Type B	0,7	2,95	1,96	0,74
<i>Gravimetric</i> Standard Uncertainty	0,47	2,01	1,09	0,47
<i>Volumetric</i> Standard Uncertainty	0,94	2,96	1,93	0,69

The first line alone shows two extreme values (both min and max) that produce an exceptionally large scatter; the figures between parentheses are the values obtained if such two extreme entries are not considered. The last column indicates the magnitude of scatter around arithmetic means.

Roughly speaking, it can be seen that with the *volumetric* method the Type A components are normally quite good, whereas the Type B components (and, as a consequence, the standard uncertainties) are in average about 80% higher than with the *gravimetric* method. The scatter of standard uncertainties with the *volumetric* method, too, is about 50% larger.

In all cases, but one (a *gravimetric* one using ESD as Type A contribution) the Type A components (blue bars in Figures 9 and 10) are smaller, and mostly below 50%, than the total Type B contributions. This shows a generally good short-term and intra-laboratory repeatability of measurements. As a consequence, in more than one-half of cases the brown (Type B components) and the green (combined or standard uncertainty) bars are almost equally long.

## 10.2. Details of individual Type B uncertainty components

The Protocol was accompanied by the already mentioned spreadsheet, the second and third sheets of which were meant to declare all components of the participants' uncertainty budget (see Appendix); the Protocol allowed participants to make additions and changes both in the definition and in the grouping of individual components. However, it was required that all uncertainties associated with the influence quantities had to be multiplied by the relevant sensitivity coefficients, to be finally given in volume units.

Most participants followed the pattern suggested by the pilot through the standard spreadsheet. For the *gravimetric* method, the proposed and listed uncertainty components were related to: correct filling (meniscus level adjustment and reading), TS weighing (filled up and empty), air density, water density, TS thermal expansion. For the *volumetric* method the main contributions were: certified uncertainty of the participant's standard tank (PS), its volume at working temperature, water thermal expansion, reduction of the TS volume at 20 °C, level adjustment or reading on the TS scale, with (scale) correction and uncertainty. Tables 17 and 18 show the entries.



Table 17 - Gravimetric method: uncertainty components (all figures are given in mL).

Laboratory:	INRIM	MIRS	DMDM	BoM	MBM	EIM	BIM	FORCE	JV	UME	MKEH	SMD	VMT	CEM	IPQ
Type A	0,390	0,160	0,385	0,225	0,260	0,300	0,231	0,620	0,251	0,034	1,740		0,413	0,185	0,338
<b>Type B:</b>															
Filling, level reading	0,707	0,202	1,448	0,776	1,440	0,300	0,400	0,480	1,700	-	1,001	-	0,180	0,709	0,964
Mass (weighing)	0,071	0,012	0,094	0,508	0,071	0,381	0,981	0,480	0,834	0,125	0,029	0,470	0,451	0,537	0,237
Air density	0,016	0,020	0,047	0,012	0,015	0,009	0,007	0,058	0,014	0,010	0,087	0,060	0,090	0,003	0,005
Water dens.	0,116	0,393	0,110	0,102	0,094	0,023	0,029	0,444	0,049	0,464	0,099	0,260	0,500	0,950	0,026
Thermal exp.	0,206	0,003	0,045	0,043	0,008	0,001	0,030	0,175	0,043	0,315	0,019	0,065	0,070	0,023	0,038
Other	-	-	-	0,700	0,026	-	-	-	-	-	-	0,016	-	0,077	-

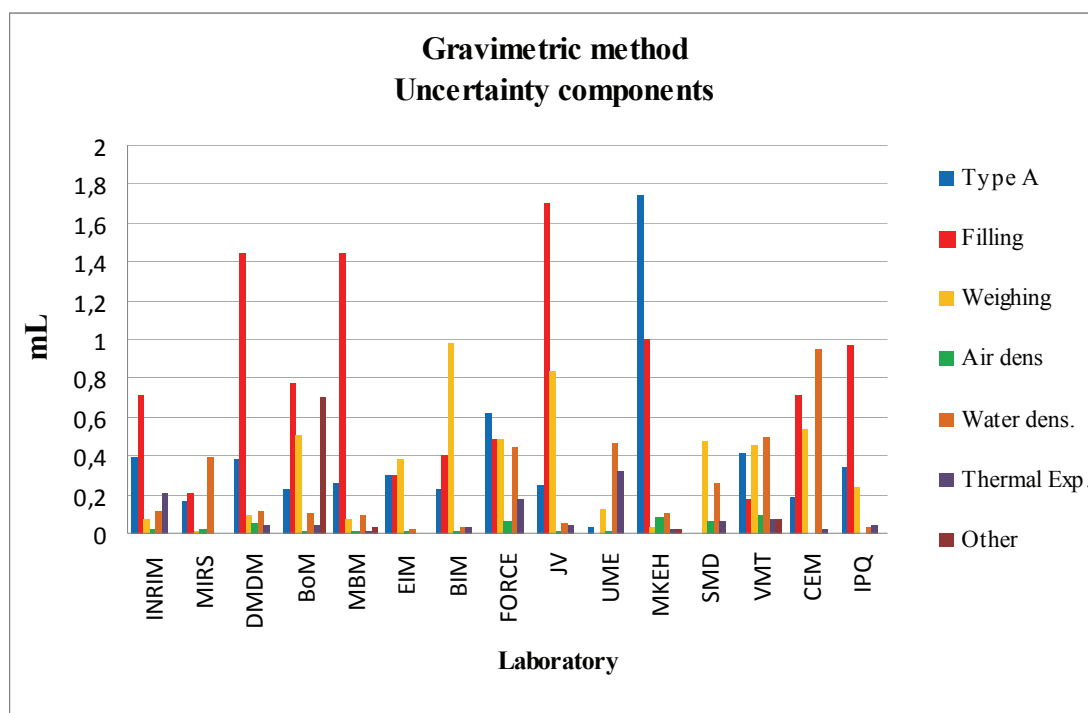


Figure 11 - Gravimetric method: histogram of the uncertainty components

It can be seen in Fig. 11 that the most important contribution among Type B is (in about 50% of cases) the one due to the correct filling, meniscus setting and level reading (red bars). Two laboratories did not consider this source (separately); in the remaining cases, it ranges from 0,18 mL up to 1,7 mL. The mean value of the 13 declared figures is 0,66 mL, with a standard deviation of 0,46 mL. If this average contribution is interpreted uniquely in terms of *uncertainty in meniscus level setting and reading*, in the given proving tank it translates into a very reasonable standard uncertainty of 0,20 mm.

The uncertainties in weighings (State 1 and State 2 weighings are put together) and in actual water density (especially as regards its temperature) rank second (yellow) and third (brown), respectively.

Table 18 - *Volumetric* method: uncertainty components (all figures are given in mL).

Laboratory:	MIRS	DMDM	BoM	EIM	FORCE	JV	SMD	CEM	IPQ
<b>Type A</b>	0,787	0,277	0,125	0,634	0,290	0,239		0,230	0,483
<b>Type B</b>									
Certified volume of Participant's Standard (PS)	0,600	0,334	0,671	0,583	1,312	0,608	0,770	1,000	0,728
PS volume at its working temperature	0,049	0,269		0,050	0,350			0,277	2,160
Water expansion between PS and TS	0,126	0,233		0,071				0,123	0,223
Conversion to 20 °C of measured TS volume	0,040	0,301		0,206	0,350			0,096	0,033
Scale (adjustment to 20 L graduation mark)	-	-	2,501	0,300	-	1,695	0,510	1,430	-
Scale (reading along the scale graduation)	0,210	2,896	-	-	2,009	-	-	-	0,943

**Note:** as already mentioned, the Type B components of MIRS were arithmetically added to yield the Type B contribution already shown in Table 15.

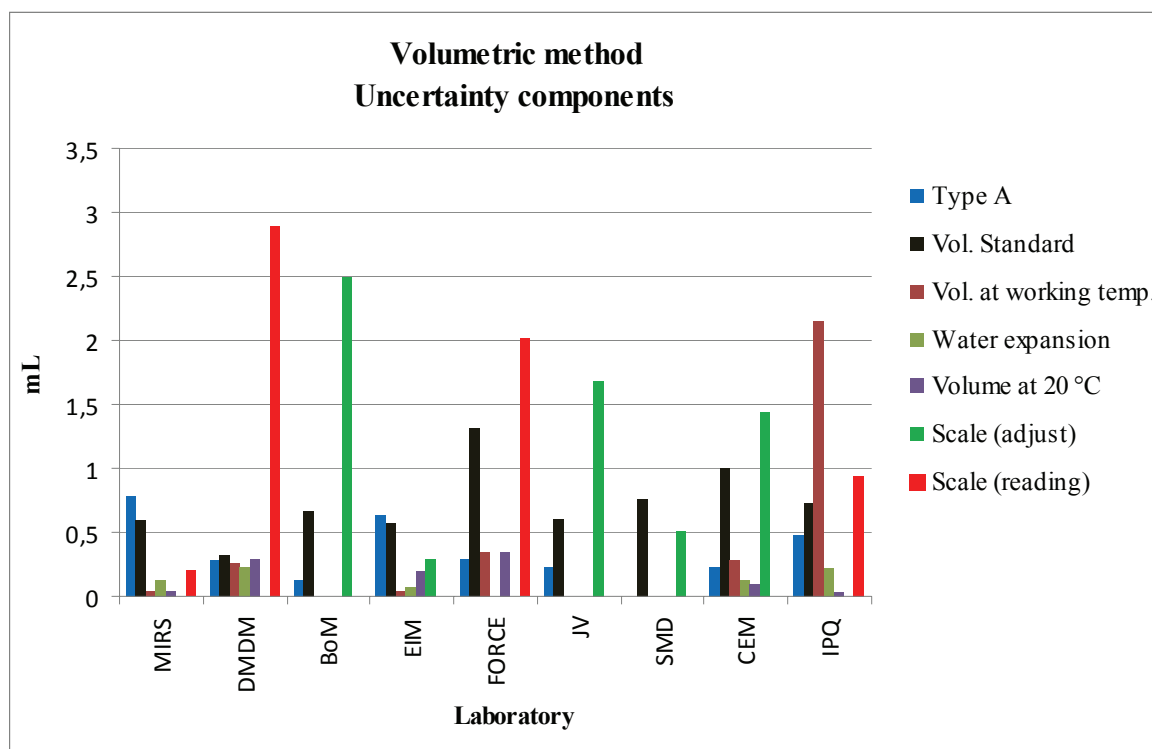


Figure 12 – *Volumetric* method: histogram of the uncertainty components

As for the *gravimetric* method, the most important contribution is due to the filling, to be intended:

- either as the uncertainty with which the meniscus has been brought (and read) to correspond to the 20 L nominal graduation mark, by adding or extracting measured amounts of water (method called *scale adjustment*, in Table 18 and Figure 12 – green bars);
- or as the uncertainty of the reading of meniscus level along the scale, where it happens to be, including interpolation uncertainties and the scale correction with the relevant uncertainty given in the Protocol (method called *scale reading* in the table and figure – red bars).

Five participants chose to adjust the level, with a mean uncertainty of 1,29 mL and a large scatter (between 0,3 and 2,5 mL). The remaining four participants read the meniscus position along the graduated scale, obtaining a somewhat larger (as expected) mean uncertainty of 1,51 mL, with a wider scatter (0,2 to 2,9 mL).

The second-ranking contribution (black bars) is the certified uncertainty of the Participants' standards at their nominal temperature. Here again the spread is wide (0,33 to 1,31 mL), with a mean value of 0,73 mL.

Differently from the *gravimetric* method, the three thermal contributions altogether play a minor role.

## 11. Remarks and conclusions

- The Project, originally requested in order to check results and accuracy of the *volumetric* method of calibration at nine laboratories, thanks to a massive implementation of the *gravimetric* method (all participants) acquired the further significance of a wide comparison based on the latter method, too;
- The Project was carried out by as many as 15 countries, several of which have acquired a new expertise in international comparisons of measurements, as they had never been involved before in any (or similar) comparisons of volume calibrations;
- All the presented results, but two, are within the Maximum Permissible Error (MPE) for this type of standard (0,1%, which allows a maximum deviation of 20 mL from nominal value); however, this legal metrology requirement of the instrument has no influence on the data processing and evaluation of results of the comparison;
- The outcome of the Project was adversely affected by the large volume change shown by the TS around the middle of its circulation;
- The causes of this change have been investigated and discussed, but not ascertained; it is thought that only an annealed, perfectly shaped and finished metal-made prover might be safer as regards stability. Or, of course, a piece of glassware;
- Having considered the dramatic lack of stability shown by the TS, the DoE results outlined in Tables 13 and 14 should be considered as a reasonable and likely, but not totally sure and definitive picture of mutual compatibilities. The inconsistencies, in particular, call for further, more reliable checks to be conducted using a Transfer Standard whose stability should be beyond any doubts;
- As a consequence of the volume change of the TS, two different Comparison Reference Values (CRVs), both based on *gravimetric* measurements alone, were determined, the former for the first 9 participants (Group 1), the latter for the last 5 participants (Group 2). An intermediate laboratory could not possibly be included in either group, as they made their measurements just during the TS volume decrease, so that they obtained a result not compatible with either group;
- The difference between the two CRVs (before and after) is 16,06 mL, the latter CRV being the lower one;

- It was decided not to declare any “official” CRV from the nine *volumetric* results, because a) their uncertainty was found almost the double (in average) of the *gravimetric* one, and b) the smaller number of participants, being splitted in two groups, did not ensure a wide enough basis for statistics, at least in the latter group;
- The (weighted) means of the *volumetric* results were however computed for both groups, and they have been compared to the corresponding *gravimetric* CRV, finding just a marginal discrepancy and a difference of 1,72 mL (the *volumetric* datum being lower than the *gravimetric* one) for the former group of participants. A good agreement was found in the latter group, which besides yielded just two (out of three) acceptable results;
- Another problem experienced in this Project was that the measurements carried out *gravimetrically* by the pilot INRIM showed discrepancies (values in excess) with both CRVs;
- This fact was of course unexpected, as INRIM, with the same expert experimenter and equipment, had been found fully compatible in three previous European *gravimetric* comparisons, and just marginally discrepant in CCM.FF-K4, a worldwide comparison where all uncertainties were in the order of just a few tenths of a millilitre;
- In addition to the pilot, just one more laboratory got *gravimetric* data discrepant with the relevant CRV;
- Three laboratories got *volumetric* results discrepant with the relevant CRV;
- The mutual (intra-laboratory) compatibility between *gravimetric* and *volumetric* results of each double-task participant was assured in all but one laboratory;
- The degrees of equivalence between the results of all laboratories, but the one in between the two groups, were computed by admitting a volume decrease of 17,09 mL (as determined on the basis of the pilot’s measurements), and correcting the results accordingly;
- The degree of mutual equivalence was satisfactory between most couples of participants, however the number of discrepancies is not negligible;
- The uncertainties with the *volumetric* method have been found about 80% higher than with the *gravimetric* method;
- Type B uncertainties have proved much larger than Type A uncertainties with both methods;
- The main uncertainty component in both methods is one of Type B, namely the uncertainty either in viewing and setting properly the meniscus level at the 20 L nominal graduation mark (*gravimetric* method), or (with the *volumetric* method) in reading the position of meniscus along the scale, or in adjusting it to the nominal graduation mark of 20000 mL;
- The second largest uncertainty contribution is that caused by weighing, or that of the participant’s certified volume standard, in the two methods respectively;
- The declared expanded (95% confidence level) uncertainties range between approximately 1 and 4 mL with the *gravimetric* method, and between 2 and 6 mL with the *volumetric* method;

- Last, the most significant quantities in uncertainty components and uncertainty budgets show very high ratios (very often 1:10 or more) between minimum and maximum participants' evaluations, showing a real need to further improve contacts to bring closer to each other both procedures and performances of all participants.

## 12. References

- [1] Cignolo, G., Technical Protocol for EURAMET Project no, 1079, "*Volumetric and gravimetric calibration of a 20 L proving tank*", November, 2008 (sent to all participants)
- [2] OIML D 28 "*Conventional value of the result of weighing in air (Revision of R 33)*", 2004
- [3] OIML R 120 "*Standard capacity measures for testing measuring systems for liquids other than water*", 2010
- [4] ISO 4787 Int. Standard "*Laboratory glassware – Volumetric glassware – Methods for use and testing of capacity*", 1984
- [5] EURAMET Calibration Guide cg-19/v.2.0 "*Guidelines on the determination of uncertainty in gravimetric volume calibration*", March 2011
- [6] Tanaka, M. et al. "*Recommended table for the density of water between 0 °C and 40 °C based on recent experimental reports*", Metrologia, 2001, 38, 301-309
- [7] Bettin, H. and Spieweck, F. "*Die Dichte des Wassers als Funktion der Temperatur nach Einführung der Internationalen Temperaturskala von 1990*", PTB Mitteilungen, 100, 3/1990, 195-196
- [8] OIML R 49-2, Annex F "*Water meters intended for the metering of cold potable water and hot water*" Part 2: Test methods, 2006
- [9] Cox, M,G, "*Evaluation of key comparison data*", Metrologia, 2002, 39, no, 6, 589-595 (on line)





1	B	C	D	E	F	G	H	I	J	K	L	
2	<b>PARTICIPANT:.....</b>											
3	<b>UNCERTAINTY BUDGET OF GRAVIMETRIC MEASUREMENTS</b>											
4	The list below is given just as a guidance or as a non-exhaustive "check list". Participants are free to leave blank or rename, or group together, or add entries, according to their own judgement or practice, on the condition that they do not change the overall format. Of course, the least is changed, the largest becomes the (normally very small) chance to make significant comparisons of figures and structures of the uncertainty budgets submitted by a number of participants.											
5	Individual contributions in the third column can be given in relative format, (percent, or ppm or other), as well as in absolute units (mm, mL, °C etc.). However, the components to the uncertainty of volume given in one of the last two columns <b>MUST BE CONVERTED AND GIVEN IN MILLILITRES</b> . For sake of simplicity it is not required to write down the often complex sensitivity coefficients, although they must obviously be known to the participant.											
6	NOTE: the resolution of the balance should not be included among Type B components: with 10 repeated measurements, its effect shows up as a part of the experimental standard deviation. The same applies obviously to the repeatability of the balance. Please note that some other components may exhibit the same feature, in part or totally.											
7												
8												
9												
10												
11	<b>Source Group</b>	<b>Individual source of uncertainty</b>	<b>estim.std. uncertainty</b>	<b>unit</b>	<b>sensitivity coefficient (OPTIONAL)</b>	<b>its unit (OPTIONAL)</b>	<b>PREFERRED: volume uncertainty components from individual</b>	<b>ALTERNATIVELY: volume uncertainty components from each of the 8 source groups</b>	<b>PARTICIPANT'S NOTES</b>			
12	<b>Type A component, u(A):</b>							<b>(mL)</b>	<b>(mL)</b>			
13	1	<b>REPEATABILITY</b>	Experimental standard deviation	See the relevant entry in the "VOLUMES" Sheet								
14	<b>Type B components:</b>							<b>(mL)</b>	<b>(mL)</b>			
15	2	<b>CORRECT FILLING</b>	Meniscus adjustment and level reading Residual droplets in the neck, possible bubbles									
16	3	<b>WEIGHING, STATE 1 (EMPTY)</b>	Calibration of mass standards used Density of mass standards Weighing scheme (?) Calibration of balance Linearity of balance Load eccentricity and magnetic effects									
17	4	<b>WEIGHING, STATE 2 (FILLED UP)</b>	Calibration of mass standards used Density of mass standards Weighing scheme (?) Calibration of balance Linearity of balance Load eccentricity and magnetic effects									
18	5	<b>AIR DENSITY (BUOYANCY)</b>	Atmospheric pressure Atmospheric temperature Relative humidity									
19	6	<b>WATER DENSITY</b>	Purity of water (uncertainty of its density at 20.000 °C) Water temperature measuring system Correct sampling of the average water temperature (vertical gradients and drifts)									
20	7	<b>THERMAL EXPANSION OF THE STANDARD</b>	Temperature of the walls of the TS (unknown difference from water temperature, under the influence of air temperature) Thermal expansion coefficient of the TS	2.4*10E-6	1/°C							
21	8	<b>OTHERS</b>										
22	<b>COMPOSITION OF ALL TYPE B COMPONENTS, u(B):</b>							<b>mL</b>	<b>0</b>	<b>0</b>		

Figure 24 – Excel Sheet 2 for the submission of the uncertainty budget in *gravimetric* measurements  
(spreadsheet *Reporting\_your\_results.xls*)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

A B C D E F G H I

PARTICIPANT:.....

CHECK LIST OF UNCERTAINTY SOURCES IN VOLUMETRIC MEASUREMENT					
The lists is given just as a guidance or as a non-exhaustive "check list". The participants are asked to develop it into their own uncertainty budget, written in the form and format they prefer, on the condition that it be understandable and complete. Of course the best format of the uncertainty budget is the one given in the GRAVIMETRIC UNCERTAINTY sheet, that participants can copy and rewrite, if they like.					
The five groups of sources of uncertainty refer to the five terms in the right-end term of Equation 1 (groups 5a and 5b are alternative)					
Source Group	Individual source of uncertainty	Would you like to enter here your estimates?		PARTICIPANT'S NOTES	
		Std. Uncertainty	Unit		
<b>Type A component, <math>u(A)</math>:</b>					
REPEATABILITY	Experimental standard deviation				
<b>Type B components:</b>					
1 VOLUME OF PARTICIPANT'S OWN STANDARD TANK(S) AT ITS/THEIR REFERENCE TEMPERATURE	Calibration of the participant's standard(s) (PS)				
	Possible drift of it/them, if calibration is not recent				
	Residual droplets in the neck, possible bubbles				
	Evaporation and micro sprays during pouring into the TS				
2 VOLUME OF PARTICIPANT'S OWN STANDARD TANK(S) AT ITS/THEIR WORKING TEMPERATURE	Water temperature measuring system in the PS				
	Correct sampling of water temperature in the PS (vertical gradients and drifts)				
	Temperature gradient between water and walls of the PS				
	Thermal expansion coefficient of the PS ( $\alpha$ )				
3 THERMAL EXPANSION OF WATER BETWEEN PARTICIPANT'S STANDARD(S) AND TRANSFER STANDARD	Water temperature measuring system in the PS				
	Correct sampling of water temperature in the PS (vertical gradients and drifts)				
	Water temperature measuring system in the TS				
	Correct sampling of water temperature in the TS (vertical gradients and drifts)				
4 COMPUTATION OF THE VOLUME OF THE TRANSFER STANDARD AT 20 °C	Thermal expansion coefficient of water ( $\beta$ )				
	Water temperature measuring system in the TS				
	Correct sampling of water temperature in the TS (vertical gradients and drifts)				
	Temperature gradient between water and walls in the TS				
5a CORRECTION "delta" IN Eq. 1, WHEN THE MENISCUS IN TS IS ADJUSTED EXACTLY TO 20 L	Thermal expansion coefficient of the TS ( $\gamma$ )				
	Adjustment and visual reading of meniscus exactly at the 20 °C graduation mark of the TS				
5b CORRECTION "delta", Eq. 1 WHEN THE TS SCALE IS USED	Uncertainty of the small water volume used for the said adjustment				
	Visual reading of the TS scale, uncertainty of the visual interpolation included				
		Uncertainty of the correction of the scale (see Protocol § 6)			
THE COMPOSITION OF ALL TYPE B COMPONENTS, $u(B)$ MUST BE ENTERED IN THE "VOLUMES" SHEET, WHERE IT IS COMBINED WITH THE EXPERIMENTAL STANDARD DEVIATION, $u(A)$ , TO YIELD THE STANDARD UNCERTAINTY $u$					

Figure 35 – Excel Sheet 3 for the submission of the uncertainty budget in volumetric measurements (spreadsheet *Reporting\_your\_results.xls*)