Transverse Disciplines in Metrology

Transverse Disciplines in Metrology

Proceedings of the 13th International Metrology Congress, 2007 – Lille, France

French College of Metrology

ISTE WILEY

First published in Great Britain and the United States in 2009 by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

International Metrology Conference (13th : 2007 : French College of Metrology)
Transverse disciplines in metrology : proceedings of the 13th International Metrology Congress, 2007, Lille, France / French College of Metrology.
p. cm.
Includes bibliographical references.
ISBN 978-1-84821-048-6
1. Measurement--Congresses. I. Collège Français de métrologie.
T50.I585 2007
620'.0044--dc22

2008035181

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-048-6

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne

Table of Contents

Preface

The *International Metrology Congress* is organized every two years by the French College of Metrology in collaboration with the Laboratoire National de Métrologie et d'Essais (LNE).

In 2007 the congress was organized with the scientific support of Nederlands Meetinstituut (NMI), the national metrology institute of the Netherlands.

The aims of the congress are:

– the evolution of metrology, and its applications in industry, environment and safety, health and medicine, economy and quality, and new information and communication technologies;
– the improvement of measurement procedures to guarantee the quality of products and processes;
– the development of metrology linked to innovating technologies.

The themes of the congress (quality and reliability of measurement, measurement uncertainties, calibration, verification, accreditation, sensory metrology, regulations and legal metrology) are developed either in a general way or applied to a specific economic sector or to a specific scientific field.

An *exhibition of the latest technical improvements* is also located at the congress: manufacturers of measurement equipments, providers of services and the official organizations of metrology are present.

Technical visits to major firms are possible. Each visit is oriented towards metrological activities connected to product quality or to services provided by the firms.

In this document you could find a selection of conferences presented during the event in June 2007 in Lille.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Chemistry – Statistics/Gas Analysis

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Certification of a reference material for herbicides in water by way of inter laboratory comparison

B. Lalère[a], V. Le Diouron[a], M. Désenfant[a], K. El Mrabet[a,b] V. Pichon[b], J. Vial[b], G. Hervoüet[a]

[a] Laboratoire National de Métrologie et d'Essais (LNE), 1 rue Gaston Boissier, 75724 Paris cedex 15, France

[b] Laboratoire Environnement et Chimie Analytique (LECA) UMR-CNRS 7121, Ecole Supérieure de Physique et Chimie Industrielles, 10 rue Vauquelin, 75231 Paris Cedex 05, France

ABSTRACT: The certified reference materials (CRMs) are among the most appropriate tools for the traceability and validation of analytical methods. Although their number seems to be high 20,000, they only cover a small part of the analytical needs for environmental monitoring. Until now, none were available for analyzing pesticides in water.
After an intralaboratory feasibility study and an evaluation of the behavior of this material during interlaboratory testing, the LNE has produced and certified a CRM for the triazine and phenylurea analysis in water samples.

Introduction

In accordance with the European outline directive on water (DCE), the department of ecology published memorandum service instruction DCE 2006/16 relating to the constitution and implementation of a monitoring program for the different water categories. Because of this, quality of rivers and ground water is regularly monitored. Pesticide content, a micropollutant considered as a priority by the European Union in particular is verified.

In order to achieve this monitoring, several measures are taken daily by different laboratories. It is important to be able to compare results. This objective can only be reached by ensuring traceability of these analyses with the help of different tools such as certified reference material (CRM), the link to national standards through uninterrupted chains of comparison and a participation in interlaboratory testing. CRMs have the advantage of also being able to evaluate accuracy during validation of analytical methods implemented. They can be of two types: those destined to calibrate measurement systems and "matrices" making it possible to consider all steps from the preparation of the sample.

Until now, there were none for the analysis of pesticides in water. That is why LNE in cooperation with LECA decided to develop one. This project started four years ago with a feasibility study of such a material. The behavior of products developed was then evaluated during an interlaboratory comparison.

A group of CRMs was developed in March 2006 and certified in June 2006 for the analysis of triazines and phenylureas in drinking water.

Previous studies: feasibility and behavior of RM

The following components which belong to two pesticide families were selected:

- triazines: deisopropylatrazine (DIA), deethylatrazine (DEA), simazine, atrazine, terbutylazine and terbutryne.
- phenylureas: chlortoluron, diuron, isoproturon and linuron.

Completed in 2002, this choice, except for terbutylazine and DIA, is based on the frequency of detection of these triazines and phenylureas in water [1]. In addition, they are indexed in the list of 50 pesticides sold in larger quantities than 500 tons per year in Europe. Terbutylazine, following the ban on the use of atrazine, is part of the mix used for its replacement in corn cultures and thus was added. Despite its usage ban, atrazine can still be detected in water; because of this, two of its metabolites, DEA and DIA were included in this study.

Feasibility study

In order to achieve the most complete kit possible, two types of reference materials (RM) were retained:

- sealed vials: targeted pesticides are stored:
  - in a solution of acetonitrile,
  - dried after evaporation of the solvent used for their preparation.
- solid phase extraction (SPE) cartridges called cartridges in this document: a water sample spiked with each compound is percolated on the support in order to retain them with other molecules in the sample. Two polymeric supports were evaluated: a divinylbenzene copolymer functionalized with N-vinylpyrrolidone (Oasis HLB, Waters) groups and a polystyrene divinylbenzene copolymer (ENVI Chrom P, Supelco). Analysis laboratories will then have to carry out the elution of compounds.

These materials were prepared with two levels of concentration, targeted in relation to regulation (0.15 µg/l for drinking water and 0.50 µg/l in surface water)

and they are stored under different temperature conditions: Tambiant, T=0.5°C and T=-18°C for a year. Every month, the evolution of compound concentration was studied by the analysis of vial and cartridge contents.

The detailed presentation of results for each compound, each type of material and each storage condition was the subject of publication [2]. Between the beginning and ending of the study, the evolution was quantified in percentage of the quantity found in relation to the one initially introduced (rate of recovery). To summarize, compounds are classified into three families of behaviors:

- non-usable family (Figure 1): either the compounds were totally degraded or their recovery level after nine months of study are lower than 10%;
- family with trend (Figure 2): compound concentration evolves over time;
- family without trend (Figure 3): pesticide concentration does not evolve significantly over time.

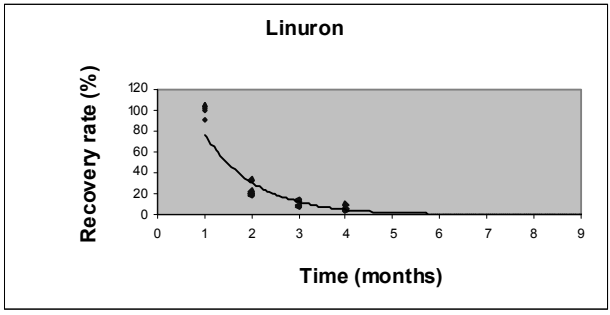


Figure 1. *"Non-usable family" example: Linuron stored in vial after dry evaporation and conservation at ambient temperature*

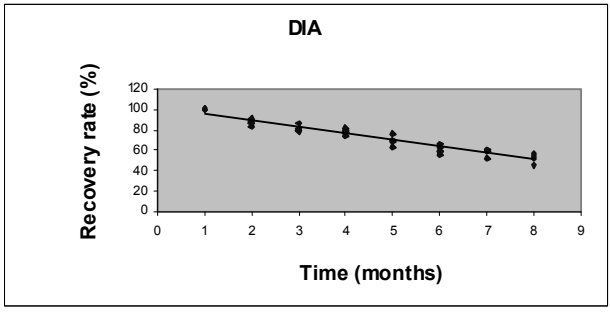


Figure 2. *"Family with trend" example DIA stored in vial after dry evaporation and conservation at ambient temperature*

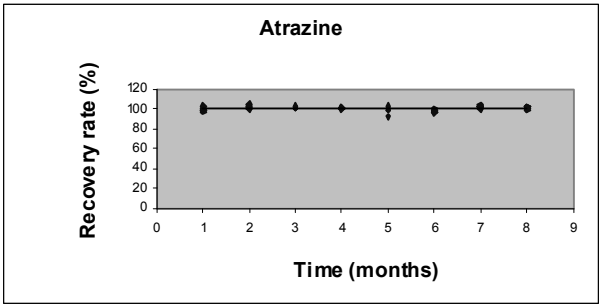


Figure 3. *"Family without trend" example Atrazine stored in vial after dry evaporation and conservation at T=-18°C*

The results have shown that compounds are stable when then are in a solution of acetonitrile when temperature becomes ambient, whereas it is necessary to store them at -20°C when they are dried in a vial. When they are fixed on cartridges, storage temperature must be lower than 0.5°C. Observed behaviors are identical regardless of the level of concentration.

The retained RM was then a kit with:

- vials containing pesticides in acetonitrile;
- cartridges (Oasis HLB, Waters) treated by percolation of drinking water containing selected pesticides.

As a precaution, they will be stored at -20°C.

RM behavior during an interlaboratory comparison

In order to test the conditions of RM usage (sending conditions, reception and analysis with different methods), an interlaboratory test involving about 15 participants was organized [3].

The variation factors of both components of the material tested vary between 14 and 30% depending on compounds, which is very satisfactory for a circuit involving laboratories using different methods compared to other campaigns [4,5].

In view of the results (feasibility and interlaboratory test), a certification campaign was then completed.

Material production and certification

This reference material, in the form of a kit is made up of:

- two vials containing approximately 1.2 ml of a herbicides solution (atrazine, deethylatrazine, deisopropylatrazine, simazine, terbutryne, terbutylazine, chlortoluron, linuron, diuron, and isoproturon) in acetonitrile (concentration for each pesticide ≈ 0.1 mg/l);
- three cartridges on which 250 ml of a drinking water spiked with a solution of herbicides in acetonitrile was percolated, in order to reach a concentration of approximately 0.50 µg/l of each compound in water.

This material is destined to calibrate the measurement devices and/or to validate the analytical methods for the determination of herbicides in water.

Preparation of materials

The preparation of materials required the creation of multi compound solutions by consecutive dilutions of a mother solution obtained with pure compounds.

After preparation, the solution is transferred into vials which are immediately sealed under nitrogen.

For the cartridges, 180 l of tap water were taken the same day. Before transferring to a cartridge according to the protocol described below, 5 l of water are spiked with 5 ml of a pesticide solution.

Cartridges are first conditioned with 3 ml of acetonitrile, 3 ml of methanol, 3 ml of water then 250 ml of spiked water are percolated. Water rinsing and nitrogen flow drying steps are then realized before storing them in specific conditions (protected from light and at a temperature lower than –20°C). All these steps were conducted with the ASPEC XL IV robot (GILSON).

There was a prior verification that vial seals would not alter the solution inside and that the reliability of the robot was sufficient to guarantee good cartridge homogenity.

Vial and cartridge preparations required three days and three weeks respectively.

400 sealed vials and 600 cartridges were produced.

Homogenity

The homogenity was verified on two series of ten vials and ten cartridges (Table 1).

Compound	Vials CV (%)	Cartridges CV (%)
Atrazine	1,2	0,9
Deethylatrazine	1,4	0,7
Deisopropylatrazine	1,1	3,2
Simazine	1,5	0,9
Terbutryne	1,3	1,4
Terbutylazine	0,9	1,2
Chlortoluron	1,4	2,9
Diuron	1,4	1,3
Isoproturon	1,2	0,6
Linuron	1,3	1,6

Table 1. *Vial and cartridge homogenity*

Stability

During the life span of the material, analyses are done each month. Until now and in accordance with the feasibility study, no significant evolution of concentrations was detected as is shown for example in Figures 4 and 5.

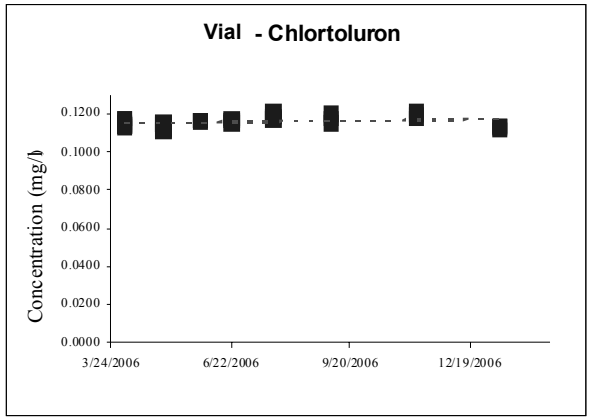


Figure 4. *Stability of chlortoluron in solution in the vials*

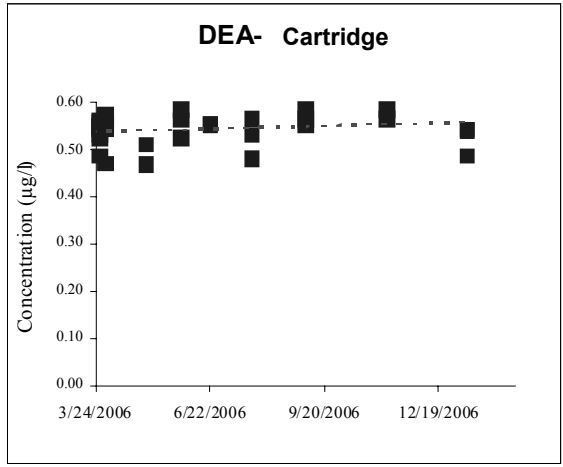


Figure 5. *DEA stability in cartridge*

Inter laboratory testing

16 laboratories with different techniques participated in this test (Table 2).

Samples (3 vials and 3 cartridges) were sent on April 4th 2006 and all results were received on May 20th 2006.

Laboratory	Analytical techniques
Bouisson Bertrand Laboratoires	CPL – SM^2 (67% of participants) CPL – UV (28% of participants) CPG – TSD (5% of participants)
MMA/APO BRGM ORLEANS	
CTCPA	
Département de l'Économie et de la Santé (DES) Service de Protection de la Consommation (SPCo) - Geneva	
IANESCO	
Institut Départemental d'Analyse et de Conseil	
Pôle Organique de l'INERIS	
ISSeP (Institut de Scientifique de Service Public) - Belgium	
LECA / ESPCI	
LDA 22	
Laboratoire Départemental D'Analyses de la Drôme	
Laboratoire national de métrologie et d'essais (LNE)	
Micropolluants Technologie	
Nestlé Waters MT Laboratoire chimie - PTC Waters	
Pôle Analytique des Eaux /LABO/CEE Plouzané	
SGS Multilab – Laboratoire de l'Essonne	

CPG: gas chromatography,
CPL: liquid chromatography,
SM^2: tandem mass spectrometry,
TSD: thermoionic detector,
UV: UV absorption.

Table 2. *Laboratories participing in the certification testing and implemented techniques*

Results

Raw results for cartridges and vials are grouped in Tables 3 and 4 respectively.

labo	Atrazine	DEA	DIA	Simazine	Terburtryne	Terbutylazine	Chlortoluron	Diuron	Isoproturon	Linuron
A	0.093	0.082	0.079	0.093	0.106	0.094	0.118	0.104	0.114	0.116
	0.100	0.092	0.079	0.103	0.116	0.100	0.121	0.100	0.123	0.123
	0.095	0.083	0.084	0.096	0.109	0.099	0.128	0.097	0.120	0.107
B	0.106	0.099	0.100	0.133	0.112	0.097	0.102	0.095	0.109	0.105
	0.110	0.101	0.102	0.137	0.116	0.102	0.109	0.098	0.111	0.110
	0.110	0.101	0.100	0.136	0.116	0.102	0.106	0.097	0.110	0.110
C	0.076	0.054	0.076	0.100	0.066	0.042	0.105	0.114	0.107	0.057
	0.088	0.068	0.084	0.101	0.086	0.056	0.111	0.112	0.117	0.075
	0.096	0.077	0.087	0.103	0.107	0.068	0.112	0.111	0.119	0.084
D	0.093	0.094	0.079	0.105	0.121	0.089	0.095	0.096	0.132	0.106
	0.095	0.103	0.079	0.105	0.123	0.095	0.111	0.096	0.106	0.106
	0.102	0.096	0.067	0.107	0.122	0.091	0.098	0.097	0.114	0.095
E	0.111	0.097	0.094	0.130	0.128	0.105	0.107	0.100	0.121	0.111
	0.111	0.101	0.100	0.120	0.130	0.109	0.111	0.101	0.126	0.131
	0.107	0.093	0.095	0.126	0.125	0.106	0.109	0.096	0.117	0.119
F	0.094	0.084	0.096	0.100	0.102	0.086	0.125	0.126	0.161	0.119
	0.093	0.082	0.093	0.100	0.093	0.085	0.115	0.117	0.145	0.123
	0.096	0.082	0.100	0.100	0.098	0.086	0.132	0.123	0.155	0.124
G	0.102	0.097	0.096	0.116	0.134	0.105	0.119	0.110	0.130	0.119
	0.100	0.094	0.091	0.111	0.130	0.101	0.114	0.108	0.126	0.116
	0.104	0.099	0.097	0.117	0.136	0.106	0.121	0.111	0.132	0.120
H	0.110	0.130	0.115	0.130	0.160	0.110	0.130	0.110	0.145	0.140
	0.110	0.135	0.110	0.130	0.150	0.110	0.130	0.110	0.145	0.145
	0.105	0.125	0.100	0.125	0.150	0.100	0.125	0.105	0.135	0.140
I	0.109	0.098	0.070	0.130	0.138	0.100	0.117	0.105	0.130	0.121
	0.110	0.102	0.070	0.132	0.142	0.102	0.124	0.107	0.132	0.110
	0.113	0.106	0.076	0.142	0.145	0.107	0.134	0.110	0.139	0.126
J	0.108	0.098	0.084	0.119	0.121	0.103	0.107	0.104	0.123	0.113
	0.108	0.098	0.084	0.119	0.123	0.104	0.107	0.104	0.124	0.111
	0.110	0.098	0.084	0.117	0.121	0.102	0.106	0.104	0.122	0.114
K	0.114	0.115	0.103	0.112	0.109	0.091	0.129	0.109	0.133	0.138
	0.117	0.117	0.102	0.111	0.107	0.092	0.122	0.115	0.136	0.126
	0.133	0.140	0.113	0.121	0.138	0.106	0.152	0.117	0.147	0.144
M	0.086	0.090	0.082	0.084	0.083	0.075	0.085	0.093	0.104	0.076
	0.085	0.083	0.081	0.087	0.082	0.072	0.087	0.086	0.099	0.076
	0.087	0.084	0.089	0.098	0.087	0.073	0.090	0.083	0.107	0.087
N	0.108	0.118	0.154	0.116		0.165	0.135	0.115	0.155	0.113
	0.085	0.095	0.129	0.100		0.139	0.112	0.093	0.120	0.098
	0.107	0.103	0.163	0.109		0.154	0.130	0.106	0.140	0.106
O	0.106	0.097	0.078	0.108	0.128	0.104	0.129	0.103	0.125	0.116
	0.105	0.097	0.079	0.108	0.129	0.104	0.133	0.103	0.127	0.116
	0.107	0.099	0.081	0.109	0.133	0.105	0.134	0.101	0.128	0.117
P							0.120	0.099	0.162	0.113
							0.096	0.082	0.125	0.086
							0.115	0.104	0.148	0.112

Table 3. *Laboratory results for cartridges expressed in µg/l of water*

Lab	Atrazine	DEA	DIA	Simazine	Terburtryne	Terbutylazine	Chlortoluron	Diuron	Isoproturon	Linuron
A	0.385	0.451	0.384	0.305	0.342	0.370	0.417	0.250	0.370	0.369
A	0.400	0.466	0.410	0.328	0.395	0.382	0.423	0.226	0.429	0.416
A	0.369	0.429	0.375	0.314	0.369	0.378	0.425	0.328	0.449	0.396
B	0.415	0.503	0.457	0.442	0.393	0.421	0.393	0.400	0.454	0.429
B	0.408	0.488	0.453	0.432	0.371	0.414	0.376	0.378	0.426	0.412
B	0.419	0.522	0.473	0.447	0.379	0.417	0.388	0.387	0.438	0.428
C	0.433	0.603	0.553	0.345	0.391	0.364	0.439	0.454	0.540	0.434
C	0.423	0.511	0.502	0.334	0.383	0.350	0.425	0.441	0.530	0.418
C	0.387	0.468	0.467	0.303	0.386	0.346	0.375	0.388	0.468	0.380
D	0.336	0.424	0.320	0.304	0.392	0.352	0.312	0.372	0.388	0.368
D	0.320	0.440	0.324	0.304	0.416	0.340	0.364	0.360	0.384	0.400
D	0.324		0.368	0.308	0.420	0.336	0.360	0.404	0.428	0.384
E	0.379	0.457	0.370	0.353	0.406	0.389	0.419	0.404	0.479	0.442
E	0.384	0.455	0.414	0.366	0.417	0.399	0.433	0.398	0.492	0.464
E	0.357	0.426	0.380	0.324	0.373	0.353	0.404	0.396	0.450	0.416
F	0.341	0.448	0.469	0.300	0.368	0.312	0.352	0.436	0.556	0.388
F	0.359	0.451	0.464	0.326	0.357	0.334	0.408	0.460	0.620	0.468
F	0.321	0.408	0.427	0.300	0.326	0.307	0.416	0.500	0.624	0.442
G	0.401	0.499	0.474	0.378	0.477	0.449	0.420	0.443	0.514	0.472
G	0.391	0.504	0.485	0.374	0.471	0.430	0.427	0.452	0.530	0.468
G	0.404	0.518	0.497	0.382	0.490	0.448	0.422	0.452	0.530	0.475
H	0.405	0.565	0.395	0.360	0.480	0.450	0.405	0.415	0.525	0.455
H	0.400	0.555	0.390	0.355	0.460	0.420	0.400	0.380	0.520	0.460
H	0.430	0.565	0.390	0.370	0.490	0.435	0.400	0.405	0.520	0.470
I	0.430	0.560	0.310	0.440	0.470	0.460	0.500	0.430	0.580	0.490
I	0.420	0.420	0.258	0.420	0.380	0.460	0.450	0.390	0.450	0.500
I	0.440	0.570	0.303	0.460	0.460	0.460	0.500	0.440	0.590	0.520
J	0.422	0.500	0.400	0.389	0.422	0.434	0.368	0.404	0.481	0.429
J	0.429	0.503	0.406	0.390	0.430	0.442	0.376	0.415	0.486	0.443
J	0.430	0.501	0.392	0.394	0.442	0.447	0.377	0.425	0.491	0.449
K	0.453	0.502	0.424	0.350	0.399	0.371	0.424	0.386	0.493	0.464
K	0.434	0.451	0.440	0.339	0.402	0.348	0.432	0.382	0.502	0.466
K	0.404	0.440	0.491	0.346	0.382	0.336	0.429	0.408	0.507	0.406
L	0.330	0.400	0.390	0.340	0.380	0.360	0.380	0.300	0.460	0.280
L	0.380	0.430	0.440	0.400	0.450	0.410	0.400	0.340	0.520	0.300
L	0.360	0.420	0.470	0.350	0.420	0.410	0.380	0.300	0.500	0.340
M	0.310	0.467	0.325	0.247	0.325	0.293	0.317	0.415	0.461	0.364
M	0.332	0.453	0.351	0.261	0.338	0.308	0.331	0.438	0.484	0.354
M	0.327	0.421	0.347	0.269	0.335	0.310	0.340	0.448	0.482	0.351
N	0.454	0.538	0.485	0.451		0.435	0.426	0.439	0.543	0.424
N	0.411	0.449	0.451	0.361		0.384	0.351	0.356	0.484	0.319
N	0.441	0.462	0.401	0.390		0.396	0.373	0.388	0.501	0.361
O	0.422	0.596	0.411	0.364	0.420	0.397	0.262	0.435	0.570	0.455
O	0.463	0.631	0.428	0.388	0.449	0.443	0.270	0.446	0.599	0.485
O	0.470	0.634	0.439	0.392	0.452	0.450	0.259	0.449	0.591	0.490
P							0.378	0.321	0.508	0.328
P							0.328	0.255	0.461	0.290
P							0.288	0.259	0.391	0.277

Table 4. *Laboratory results for vials expressed in mg/l*

Statistical data

Statistical data analysis

The statistical data analysis was performed based on operation standards for interlaboratory tests NF ISO 5725-2 [6] and NF ISO 5725-5 [7]. The first phase was to detect atypical values from statistical homogenity tests (Grubbs and Cochran). The exclusion of some data was based on statistical and technical conclusions. The second phase, from results obtained, was to quantify the parameters summarizing the RM, average and standard deviation of reproducibility as closely as possible. The results of this process are summarized in Tables 5 and 6.

Compound	A	B	C	D
Atrazine	14	-	-	14
Deethylatrazine	14	-	-	14
Deisopropylatrazine	14	N	-	13
Simazine	14	-	-	14
Terbutryne	13	C, K	-	11
Terbutylazine	14	-	-	14
Chlortoluron	15	-	-	15
Diuron	15	-	-	15
Isoproturon	15	-	-	15
Linuron	15	-	-	15

A: number of laboratories providing results,
B: laboratories eliminated for variance homogenity,
C: laboratories eliminated for average homogenity,
D: number of laboratories retained for certification.

Table 5. *Statistical data for vials*

Compound	A	B	C	D
Atrazine	15	-	-	15
Deethylatrazine	15	-	-	15
Deisopropylatrazine	15	-	-	15
Simazine	15	-	-	15
Terbutryne	14	-	-	14
Terbutylazine	15	-	-	15
Chlortoluron	16	-	-	16
Diuron	16	-	-	16
Isoproturon	16	-	-	16
Linuron	16	-	-	16

A: number of laboratories providing results,
B: laboratories eliminated for variance homogenity,
C: laboratories eliminated for average homogenity,
D: number of laboratories retained for certification.

Table 6. *Statistical data for cartridges*

Certified values

The value assigned to CRM is the average result from laboratories and the standard uncertainty is given by the standard deviation of reproducibility. This uncertainty corresponds to the uncertainty on the result of an analysis of this CRM carried out by a laboratory working under the same conditions as the circuit laboratories.

Tables 7 and 8 group CRM certified values.

Compound	Concentration (mg/l)	Expanded uncertainty k=2 (mg/l)
Atrazine	0.103	0.022
Deethylatrazine	0.098	0.034
Deisopropylatrazine	0.089	0.025
Simazine	0.113	0.029
Terbutryne	0.122	0.039
Terbutylazine	0.099	0.044
Chlortoluron	0.116	0.029
Diuron	0.104	0.019
Isoproturon	0.128	0.032
Linuron	0.112	0.038

Table 7. *Certified values for vials*

Concentrations in cartridges are expressed in µg of compound by liter of water. In each cartridge, 0.25 l of water was percolated.

Compound	Concentration (µg/l)	Expanded uncertainty k=2 (µg/l)
Atrazine	0.395	0.088
Deethylatrazine	0.49	0.12
Deisopropylatrazine	0.41	0.13
Simazine	0.36	0.11
Terbutryne	0.407	0.093
Terbutylazine	0.39	0.10
Chlortoluron	0.39	0.11
Diuron	0.39	0.12
Isoproturon	0.50	0.12
Linuron	0.41	0.12

Table 8. *Certified values for cartridges*

Conclusion

Validation for this experimental approach makes it possible to propose a certified reference material which is now available for the analysis of pesticides in water. It is presented in the form of a kit in order to respond to analysis laboratory requirements:

– sealed vials containing compounds in solution in acetonitrile, which can be used for the verification of calibration or for spiking water matrices;

– cartridges on which water containing compounds was percolated representing a real sample.

From the development to certification of this reference material, the study took five years of study. In addition, this CRM is in compliance with the ISO 34 guide specifying the conditions of production, sample conservation and stability.

References

[1] Etudes et Travaux IFEN, “Les pesticides dans les eaux”, September 2000.

[2] J. Deplagne, J. Vial, V. Pichon*, Béatrice Lalere, G. Hervouet and M.-C. Hennion, Feasibility study of a reference material for water chemistry: Long term stability of triazine and phenylurea residues stored in vials or on polymeric sorbents, Journal of Chromatography A, 1123 (2006) 31-37.[3] K. El Mrabet, M. Poitevin, J. Vial*, V. Pichon, S. Amarouche, G. Hervouet, B. Lalere, An Interlaboratory Study to evaluate potential Matrix Reference Materials for pesticides in Water, Journal of Chromatography A, 1134 (2006) 151-161.

[4] S.A. Senseman, T.C. Mueller, M.B. Riley, R.D. Wauchope, C. Clegg, R.W. Young, L.M. Southwick, H.A. Moye, J.A. Dumas, W. Mersie, J.D. Mattice, R.B. Leyy, Interlaboratory comparison of extraction efficiency of pesticides from surface laboratory water using solid-phase extraction disks, J. Agric. Food Chem., 51 (2003) 3748-3752.

[5] M.B. Riley, J.A. Dumas, E.E. Gbur, J.H. Massey, J.D. Mattice, W. Mersie, T.C. Mueller, T. Potter, S.A. Senseman, E. Watson, Pesticide extraction efficiency of two solid phase disk types after shipping, J. Agric. Food Chem., 53 (2005) 5079-5083.

[6] NF ISO 5725-2 Exactitude (justesse et fidélité) des résultats et méthode de mesure partie 2: Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée.

[7] NF ISO 5725-5 Exactitude (justesse et fidélité) des résultats et méthode de mesure partie 5: Méthodes alternatives pour la détermination de la fidélité d'une méthode de mesure normalisée (analyses robustes).

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Determination of aflatoxin M1 using liquid chromatographic method: comparison between uncertainty obtained from in house validation data and from proficiency test data

C. Focardi, M. Nocentini, S. Ninci, P. Perrella, G.Biancalani.

Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana - Sezione di Firenze, Via di Castelpulci 43, 50010 San Martino alla Palma, Florence, Italy.
Tel +39055721308- Fax +390557311323 – e-mail: claudia.focardi@izslt.it

ABSTRACT: An HPLC method with fluorescence detection has been validated for the determination of aflatoxin M1 in milk samples. Certified Reference Materials and Spiked samples have been used for in house validation. These validation data have been applied for the uncertainty estimation with the bottom-up approach. Results obtained with this method have been compared with the expanded uncertainty determined from proficiency testing FAPAS.

Introduction

Aflatoxins are a group of hepatocarcinogen molecules produced by *Aspergillus flavus* and *Aspergillus parasiticus*. When aflatoxin B1 (AFB1), present in contaminated feed, is ingested by dairy cattle, it is excreted into milk as aflatoxin M_1 (AFM1). Both AFB1 and AFM1 can cause DNA damage, gene mutation, chromosomal anomalies and as a consequence the International Agency for Research on Cancer (IARC) has classified AFB1 and, recently also AFM1, as class 1 (carcinogenic to humans) [1]. Strict regulatory limits for AFM1 are currently in force in the European Community; the Regulation (EC) 1881/2006 [2] set a maximum residue level (MRL) of AFM1 in milk, intended for adults, at 0.050 μg/kg, and at 0.025μg/kg for milk, intended for infants or for baby-food production. The European Commission with the Regulation (EC) 401/2006 [3] fixed the performance criteria for the analytical methods.

The adopted method is a liquid chromatographic one with fluorimetric detection. The sample is purified by using immunoaffinity column. The method is in house fully validated using Certified Reference Materials and spiked samples.

The EN ISO/IEC 17025 [4], as well as the Regulation (EC) 401/2006, requires all the measurements to be accompanied by estimation of expanded uncertainty.

Various methods are available to evaluate measurement uncertainty. One approach, applied by Eurachem [5], is the so called "bottom-up" and consists in separately identifying and quantifying error components that are consider important from in house validation data. Another approach is to use results from proficiency testing, by comparing reproducility variance with the repeteability variance of the laboratory [5].

In our case we use the data obtained with participation of the laboratory at 11 FAPAS during the period 2002-2006.

Experimental

Materials

Aflatoxin M1 Standard Reference Material at a concentration of 500 μg/l in methanol was purchased from Riedel de Haën. The chemical structure of AFM1 is presented in Figure 1.

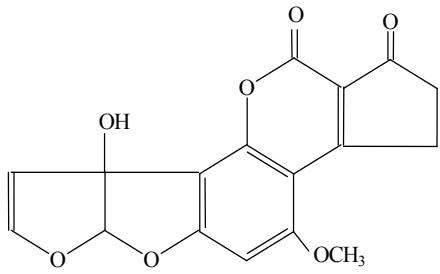


Figure 1. *Chemical structure of Aflatoxin M1*

Deionised water obtained by a Milli-Q water (Millipore) and acetonitrile HPLC grade where used throughout. AflaTest® Immunoaffinity columns containing antibodies against AFM1 were purchased from VICAM. AFM1 working standard solutions at five different concentration levels (0.08, 0.2, 0.4, 0.6 and 0.8 ng/ml) were prepared by dilution of the stock solution (5 ng/ml). All the solutions are dissolved in acetonitrile water in the ratio 10/90 (v/v) (Figure 2).

Sample preparation

The reference materials used for HPLC method validation and uncertainty estimation are reported in Table 1.

Reference Material	*Certified Concentration*	*Certified Uncertainty*
Standard Reference Material	500 µg/l	±50 µg/l
CRM 285	0.76 µg/Kg	± 0.05 µg/Kg
FAPAS round 0477	0.44 µg/Kg	± 0.015 µg/Kg
FAPAS round 0445	0.26 µg/Kg	± 0.01 µg/Kg

Table 1. *Reference materials used for method validation*

These milk powder samples were prepared by dilution with water in the ratio 1:10 (w/w), in the way that the final concentration of AFM1 in the samples lays down in the HPLC calibration curve. Besides spiked samples at the nominal concentration of 0.05, 0.04 and 0.03 µg/Kg were prepared adding, under a gently magnetic stirring, respectively 500 µl, 400 µl and 300 µl of AFM1 stock solution (5ng/ml) to 50 g of homogenized milk, previously tested to demonstrate the absence of AFM1 residues, under a gently magnetic stirring.

Reconstituted Certified Materials (FAPAS and CRM), spiked and blank milk samples were extracted and purified with the procedure used by Tuinistra [6] and reported in the flow diagram (Figure 2).

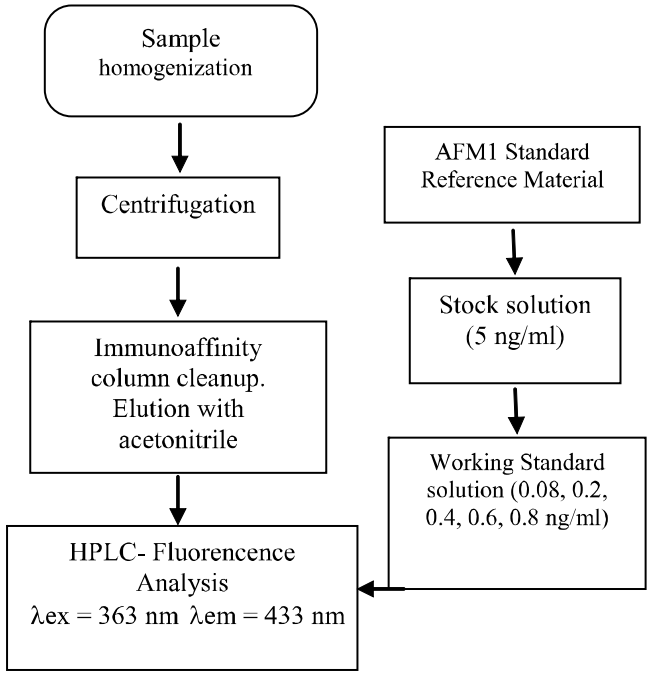


Figure 2. *Flow diagram of sample analysis*

To ensure the homogenity of samples, they were gently magnetically stirred at room temperature for about 15 minutes. After homogenization, 50 g of samples were centrifuged for 25 minutes at 3500 rpm, to eliminate fat components. Skimmed milk was passed through the immunoaffinity column. The column was washed with 10 ml of water and AFM1 eluted with 5 ml of acetonitrile.

For chromatographic analysis the sample was evaporated under a stream of nitrogen to a volume of approximately 0.3 ml and reconstituted to 5 ml with water.

HPLC analysis

Chromatography was performed with a Perkin Elmer (USA) Series 200 system consisting of a quaternary gradient pump, an autosampler, a fluorescence detector and a degassing system using helium. Chromatographic separation was achieved using a Spherisorb ODS2 (250 × 4.6 mm, 5 μm) reversed phase column, with a guard column of the same type. The mobile phase was constituted by water and

acetonitrile in the ratio 75/25 (v/v). An isocratic elution was performed at a flow rate of 1 ml/min for a total run time of 20 minutes. The injection volume was 500 µl. The detector was set at a excitation wavelength of 363 nm with a emission wavelength of 433 nm.

HPLC validation criteria and uncertainty estimation

The method for the determination of Aflatoxin M1 in milk was in house validated by a set of parameters which are in compliance with the Eurachem guide [7] and the Commission Regulation 401/2006 [3].

For evaluating the overall uncertainty, the "bottom up approach" has been adopted, following the statements of Eurachem guide [5].

Internal quality control

The method used for the determination of aflatoxin M1 in milk fulfils the requirements of EN ISO/IEC 17025 [4] and is accredited in the laboratory.

Internal quality control is achieved following the IUPAC harmonized guidelines [8]. A HPLC sequence consists of the analysis of a standard, a blank sample, a Certified Reference Material to check the recovery and finally samples in duplicate.

The Shewart control chart has been applied for the statistical control of the measurement process by using the CRM and an example is reported in Figure 3. The Reference value µ is the concentration level of CRM and σ is the uncertainty associated with the material. The warning limits are considered equal to µ plus and minus 2σ, the action limits are µ plus and minus 3σ below the center line.

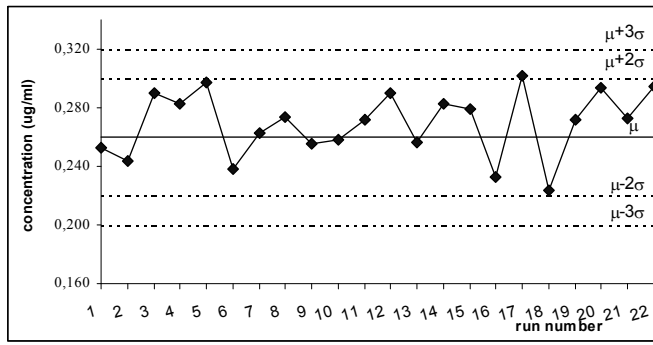


Figure 3. *Shewart control chart related to FAPAS 0445*

Proficiency testing

The performance of the laboratory is periodically checked with the partecipation at Proficiency testing organized by FAPAS®. The results were evaluated in the form of a z-score which is the estimated laboratory bias divided by the target value for the standard deviation [8].

The laboratory has participated in eleven proficiency tests over the years. In Figure 4 we show the value of z-score obtained according to time (omissis); all values are included in the range of ± 1.

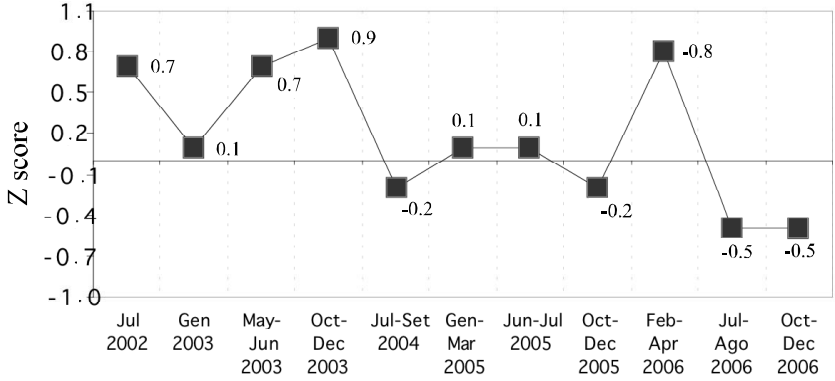


Figure 4. *z-score profile obtained during the period 2002-2006*

Results and discussion

In-house validation of the method

The validation data are obtained by using Certified Reference Materials and spiked samples.

Specificity. The chromatograms of a blank milk sample, CRM and standard solution are previously reported [9]. The blank milk sample is free of interferences at the elution time corresponding to the AFM1 peak, demonstrating a good specificity of the method proposed.

Linearity. The calibration curve of AFM1 in standard solutions (Figure 5) is linear (r = 0.9999) over the concentration range between 0.08 and 0.8 ng/ml. These data are the results of six different series obtained by different operators.

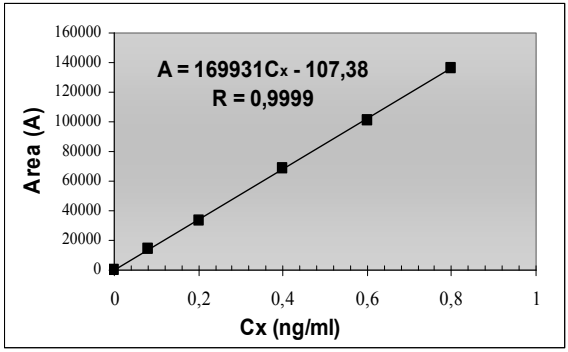


Figure 5. *Calibration curve for aflatoxin M1*

Evaluation of systematic error. The Bartlett test applied both on Certified Reference Material and spiked samples demonstrated the homogenity between the variance and subsequently the absence of systematic errors.

Accuracy and precision. The precision and accuracy determined are listed in Table 2.

According to the EC Regulation 401/2006, the accuracy for all of the samples fell in the range between –20 % to 10%. The precision of the method is expressed as RSD, Relative Standard Deviation, values for all concentrations. In Table 2 are also reported the values of the RSD_{max} which is equal to two thirds of the Horwitz equation:

$$RSD_{max} = 2/3 \times 2^{(1-0.5\log C)}$$

where C is the mass fraction expressed as a power of 10. According to the EC Regulation 401/2006 experimental RSD value is lower than RSD_{max}.

Sample	C (μg/kg)	n	Mean (μg/kg)	Accuracy (Within – 20% to 10%)	RSD (%)	RSD_{max} (%)
CRM 285	0.76	15	0.754	- 0.7	7.8	31.1
FAPAS 0477	0.44	10	0.434	- 1.5	10.7	33.8
FAPAS 0445	0.26	15	0.261	+ 0.5	6.8	36.6
"Spiked"	0.05	9	0.048	-3.6	9.6	46.9
"Spiked"	0.04	5	0.035	-11.5	8.4	48.5
"Spiked	0.03	5	0.026	-14.7	6.5	50.6

Table 2. *Data of accuracy and precision obtained for CRM, FAPAS and "Spiked" samples, where n is number of replicates and Mean is the average concentration obtained*

Limit of detection and limit of quantification. The limit of detection has been calculated from the calibration curve. Taking into account the standard deviation ($s_{y/x}$) of the regression analysis, the limit of detection is found to be C_{LOD}= 0.001 μg/kg. The limit of quantification (LOQ) has been set as 10-fold the LOD giving the value of C_{LOQ}= 0.01 μg/kg.

Uncertainty estimation

The equation of the measurand is as follows:

$$\text{C (μg/kg)} = \frac{C_x \cdot V_{ext}}{M \cdot R} \tag{1}$$

where M (g) is mass weight of the milk sample, V_{ext} is the final volume of the extraction, C_x (ng/ml) is the concentration obtained by calibration curve and R is the recovery rate obtained on suitable reference sample (CRM or FAPAS sample). Every factor of the equation is shown in the cause and effect diagram (Figure 6).

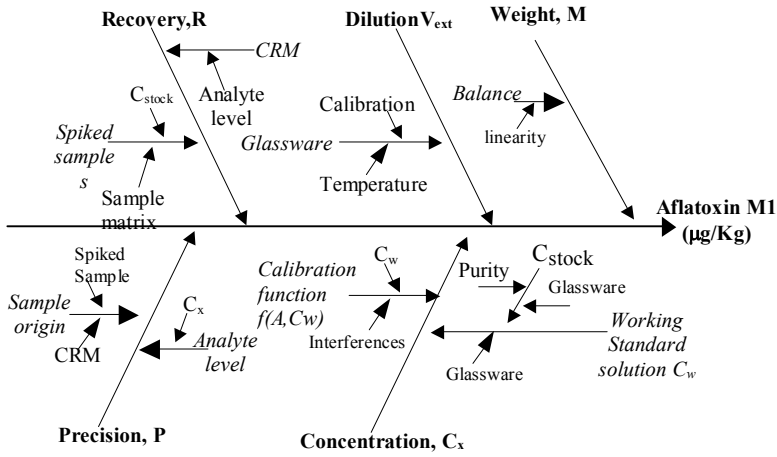


Figure 6. *Cause and effect diagram*

The detail of uncertainty estimation, based on bottom-up approach taking into account the in-house validation data are previously reported [9]. For the standard uncertainty associated to the dilution volume and weight, a triangular distribution has been chosen.

In Figure 7 is reported the Error Budget diagram; the contribution to uncertainty of mass weight (M) and dilution (V_{ext}) are negligible compared to the others, for all samples analysed.

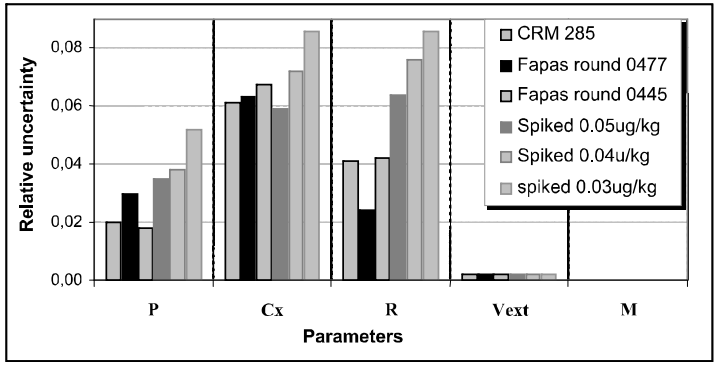


Figure 7. *Error budget diagram*

In Figure 8 are presented the pie charts for the reference materials and for the spiked samples.

If we compare the different components to uncertainty of certified reference materials to those obtain for the spiked samples, it is clear that in the second case increases the contribution of the recovery, due to the homogenity of the sample. At the same time, the contribution due to the concentration Cx decreases. This second effect should be correlated with the value of the chromatographic peak area which is lower in the case of Certified Reference Material, due to the dilution of the sample.

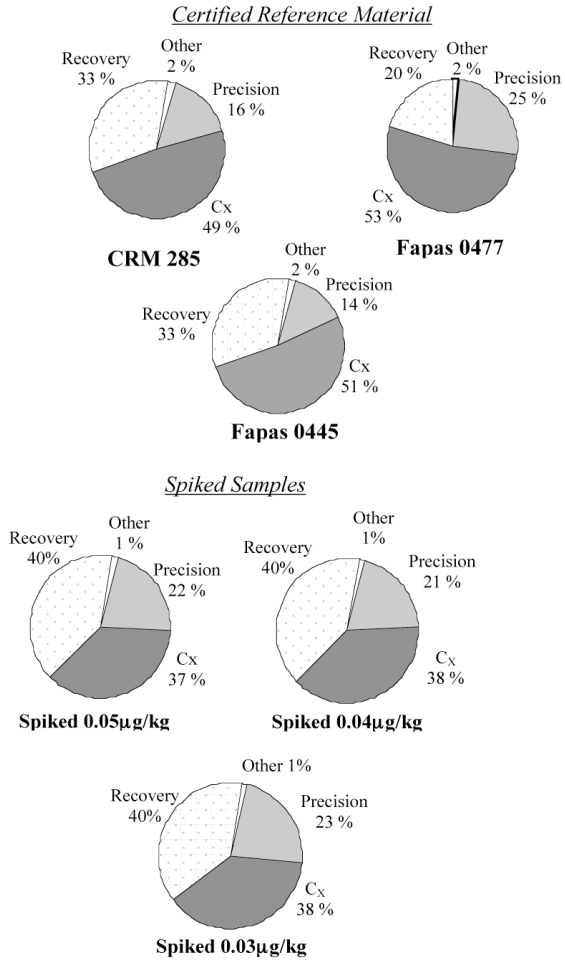


Figure 8. *Contributions to combined standard uncertainty. Charts show the relative sizes of uncertainty associated with precision, bias, calibration and other effects*

According to the bottom up approach the global uncertainty, is given by the addition of all the uncertainty associated with the component that influences the measurand. Taking into account equation (1) the overall uncertainty, in terms of relative uncertainty, is determined by the formula:

$$\text{ů}(C) = \sqrt{(\text{ů}(P))2 + (\text{ů}(Cx))2 + (\text{ů}(M))2 + (\text{ů}(Vext))2 + (\text{ů}(R))2}$$

The standard uncertainty is then obtained by the equation

$$u(C) = \text{ů}(C) \times C \tag{2}$$

In Table 3 we show, for the Certified Reference Materials and for the spiked sample, the relative uncertainty ů(C).

Sample	Concentration (µg/Kg)	ů(C)
CRM 285	0.76	0.08
FAPAS round 0477	0.44	0.07
FAPAS round 0445	0.26	0.08
Spiked	0.05	0.09
Spiked	0.04	0.11
Spiked	0.03	0.13

Table 3. *Relative and standard uncertainty calculated for reference materials and spiked samples*

The relative expanded uncertainty is obtained multiplying the mean value of the relative uncertainty, by a coverage factor, *k*=2

$$\overset{0}{U}(C) = \overset{0}{u}(C) \times 2 = 0.09 \times 2 = 0.18$$

As an alternative method to measure uncertainty for chemical measurements, the Analytical Method Committee of the Royal Society of Chemistry [10] proposed an approach based on precision data assessed in an interlaboratory study.

The uncertainty associated with a mesurand result y is given by the following formula:

$$u_y = u_v = \sqrt{s^2_R + u^2_{ref}}$$

where s^2_R is the reproducibility variance between laboratories and u_{ref} is uncertainty associated with the accepted reference value. A further example of this expression is possible when u^2_{ref} is negligible in comparison to the reproducibility variance.

Some applications of interlaboratory data in the estimation of the uncertainty of chromatographic methods have been described [11,12]. Considering the results obtained by the laboratory, applying the proficiency test FAPAS, we can compare laboratory repeatability with the reproducibility of interlaboratory study (Table 4).

Round	C (μg/Kg)	$\underline{\sigma}$	$\underline{\sigma}_{rel}$
0445	0.26	0.09	0.35
0450	0.059	0.016	0.27
0462	0.058	0.02	0.34
0465	0.084	0.005	0.06
0472	0.069	0.02	0.29
0480	0.052	0.015	0.29
0487	0.218	0.075	0.34
0492	0.53	0.159	0.30
0496	0.068	0.019	0.28

Table 4. *Results of FAPAS proficiency testing: C is Assigned value and $\underline{\sigma}$ is the robust standard deviation, $\underline{\sigma}_{rel}$ is the relative robust standard deviation*

The mean value obtained for the intralaboratory relative repeatability is equal to 0.09, an order of magnitude lower than the interlaboratory reproducibility. This indicates that the method precision of the laboratory is comparable to that of the laboratories which took part in the collaborative trial. It is therefore acceptable to use the reproducibility standard deviation from the collaborative trial in the uncertainty budget of the method. The mean value resulting from the interlaboratory test of relative robust standard deviation is equal to 0.28.

We can calculate the expanded uncertainty at the MRL level by using the relative uncertainty obtained with the two different approaches applying equation (2).

$$U(C) = 0.28 \times 0.05 = 0.014 \text{ μg/kg}$$
$$U(C) = 0.18 \times 0.05 = 0.009 \text{ μg/kg}$$

We can conclude that the expanded uncertainty obtained with the two different methods are of the same order of magnitude.

Conclusions

This chromatographic method for the determination of aflatoxin M1 in milk samples meets compliance with the Regulation (EC) 401/2006. Internal quality controls are achieved by applying the Shewart control chart to the data obtained each round with the Certified Reference Materials. Uncertainty has been evaluated following two different methods; the bottom up approach and the interlaboratory data, in the specific case, FAPAS round materials over the period 2002-2006. Expanded uncertainty obtained with the two different approaches calculated at the Maximum Residue Level are comparable.

References

[1] IARC International Agency for Research on Cancer Monograph on evaluation of carcinogenic risks to humans. IARC Lyon France **2002** 82, 171

[2] *Official Journal of European Commission* L 364/5 **2006** Commission Regulation No 1881/2006 of 19 December 2006. Brussels Belgium

[3] *Official Journal of European Commission* L 70/12 **2006** Commission Regulation No 401/2006 of 23 February 2006. Brussels Belgium

[4] EN ISO/IEC 17025: 2005. General Requirements for the competence of calibration and testing laboratories. **2005** ISO Geneva

[5] Quantifying uncertainty in analytical measurement *Eurachem Guide* **2000**, second edition LGC.

[6] Tuinistra L. et al., *J. AOAC Int*, **1993,** 84, 2. 1248-1254

[7] The Fitness for purpose of analytical methods A laboratory guide to method validation and related topics *Eurachem Guide* **1998**, edition 1.0 LGC.

[8] Thomson M., *Pure Applied Chem*, **1995,** 4, 649-666

[9] M. Nocentini, C. Focardi, M. Vonci, F. Palmerini. *Proceedings at 11th International Metrology Congress*, 20-23 October 2003 Toulon- France

[10] AMC (Analytical Method Committee) *Analyst*, **2005,** 130, 721

[11] Dehouck P., et al., *J. of Chrom A*, **2003,** 1010, 63-74

[12] Dehouck P., et al., *Anal Chim Acta*, **2003,** 481, 261-272

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Purity analysis in gas metrology

F. Dias and G. Baptista

Laboratório de Gases de Referência, Instituto Português da Qualidade
R. António Gião, 2, 2829-513 Caparica, Portugal

ABSTRACT: The Reference Gas Laboratory (LGR) of the Portuguese Institute for Quality (IPQ) is the Primary Laboratory in Portugal in the field of gas metrology. Its main mission is to assure and guarantee the accuracy and traceability [1] of the gas measurements to national and international standards.
LGR is also a consumer of pure gases to be used as raw material in the gravimetric preparation of primary gas mixtures. The laboratory needs to measure and control the purity level of the gases supplied by the industry in order to determine the gas composition with higher accuracy. Several methods are used for purity analysis, namely, Gas Chromatography (GC), Fourier Transform Infrared Spectrometry (FTIR) and Mass Spectrometry (MS). These analytical methods will be briefly described and documented with some examples of gas purity analysis made at LGR: Propane (C_3H_8), Nitrogen Monoxide (NO) and automotive mixtures ($CO+CO_2+C_3H_8$ in Nitrogen).

Introduction

LGR is a consumer of pure gases to be used as raw material in the preparation of primary gas mixtures. The need for the quantification of impurities is of high interest in the calculation of the measurement uncertainty and in the estimation of the result accuracy.

The choice of the more adequate analytical technique is made according to different properties such as, method specificity, cross-interference and matrix effects and also the method detection limits.

At LGR, several methods are being implemented for purity analysis, namely, Gas Chromatography (GC), Fourier Transform Infrared Spectrometry (FTIR), and Mass Spectrometry (MS). In GC technology, the purity of parent gases is obtained by using the two following techniques: Flame Ionization Detection (FID) and Thermal Conductivity Detection (TCD). The FTIR technique is used to measure gas species, which may be difficult for GC analysis. GC-MS technology is used to detect multi-species by mass signatures and identify the unknown impurities.

In order to obtain a better knowledge of the purity of gases involved in gravimetric preparations and/or in dynamic mixtures, our laboratory is now developing a Purity Analysis service. Although not fully implemented these techniques already provide important information concerning raw material control, fundamental to primary gas standards preparation. The quantification of impurities is of high interest in the calculation and estimation of the results. Indeed, one of the major uncertainty contributions is the purity component of the gas.

The qualitative analysis of impurities in pure gases is also important due to the possibility of having interferents and because of the need to ensure that cross-interferent impurities are small enough not to contribute significantly to the measured results.

The gas mixture preparation method according to ISO 6142 accounts for each component (including impurities) as a fraction amount of the total mixture. It then evaluates the uncertainty of each fraction amount. This approach is totally correct, but it might lead to neglecting certain correlations in the uncertainties when the same source of gas is used more than once in the preparation of a mixture.

Gas Chromatography (GC)

Gas Chromatography (GC) [2] is a common analytical method for gas purity analysis. GC is able to analyze almost all gas components, but requires a variety of columns, specific for certain chemical species, and detectors. This technique has low gas consumption and it is possible to connect it to auto-sampler equipment enabling automatic data acquisition. One limitation is the need for reference standards to validate peak retention time. It is a good tool for quantitative analysis.

In the LGR two different detectors are used, namely, Thermal Conductivity Detector (TCD) and Flame Ionization Detector (FID), which will be briefly described below.

Figure 1. *Gas Chromatograph Agilent 5890*

Thermal Conductivity Detector (TCD)

Gas Chromatography by TCD is a non-destructive method in which the detector is sensitive to the thermal conductivity of the carrier gas. Each time a component different from the carrier gas passes through the detector, the TC decreases the given origin to a signal proportional to that component concentration. One important advantage is that it is valid for almost all components.

Flame Ionization Detector (FID)

The flame Ionization Detector (FID) is a destructive method on which the carrier gas is mixed with hydrogen and air, being the final mixture burned. The burning produces ions causing an electric current change. The signal generated is proportional to the ion concentration. It is a very sensitive method allowing the detection limit in GC to be decreased. It applies to almost all organic compounds, but, is not sensitive for common inorganic species (CO, NO_x, SO_2, H_2S, H_2O). One way to change this is using a Nickel catalyst, which converts the CO and CO_2 to CH_4 for further detection.

Pure C_3H_8 analysis

An example is shown of pure Propane (99.95%) analyzes by GC-TCD with different columns.

Equipment: Agilent 5890

Column: Porapack Q, Molecular sieve 5A

Temperature: 160ºC, 150ºC

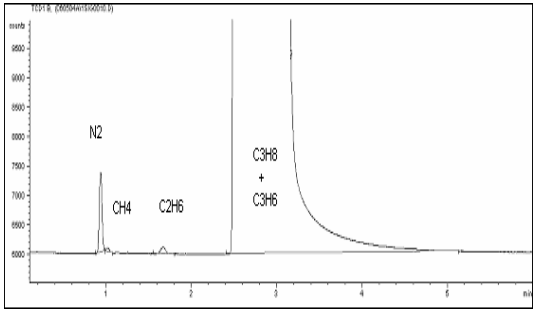


Figure 2. *C_3H_8 pure sample analysis with Porapak col*

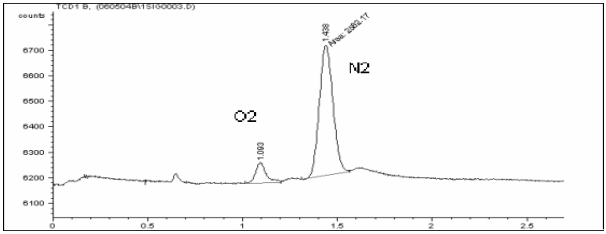


Figure 3. *C_3H_8 pure sample analysis with Mol. Sieve col*

<u>C_3H_8 99.95%</u>
Specification: H_2O (5 ppm), O_2 (10 ppm), CO_2 (5 ppm), N_2 (40 ppm), C_3H_6 (200 ppm), C_nH_m (200 ppm)

Results: H_2O (?), O_2 (8 ppm), CO_2 (2 ppm), N_2 (83 ppm), C_nH_m (42 ppm CH_4 + 21 ppm C_2H_6)

For the C_3H_8 99.95% sample, the results show that the mixture is under the specification, except for N_2 where it was found at a concentration higher than specified. Nevertheless, it was not possible to separate propene from propane, nor to measure H_2O with this technique

Fourier Transform Infrared Spectrometry (FTIR)

FTIR [3] is a spectroscopy method where working principle is based on the infrared (IR) absorption by molecules. Any molecular vibrations, which displace an electric charge, will absorb IR radiation. On FTIR the scanning of IR region allows us to detect several components in a sample. However, it applies only to species absorbing an IR radiation, which means that atomic species (He, Ne, Ar), as well as the homonuclear diatomic species without a permanent dipole (H_2, N_2, O_2) cannot be observed.

This limitation does not imply that the majority of environmental and pollutant gases cannot be observed by this technique. According to NIST, approximately 100 of the hazardous air pollutants listed in the US EPA Clean Air Act can be measured.

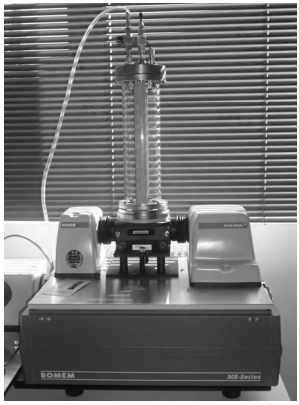


Figure 4. *Fourier Transform Infrared Spectrometer BOMEM MB100*

IR signatures are easily recognized and do not change according to the mixture, showing no matrix effects. The main limitation of this method is the difficulty to eliminate residual H_2O (1325-1900 cm^{-1}: 3550-3950 cm^{-1}) and CO_2 (2295-2385 cm^{-1}) existing in the absorbance spectrum, mainly due to the evolving atmosphere. The large amount of gases consumed during an analysis is also a disadvantage when the sample quantity is a limitation. For all the reasons explained above, FTIR is mainly used to measure gas species difficult to measure in GC. This technique is a useful tool to measure pure NO and its contaminants.

Pure NO analysis

An example of pure NO 99.90% analysis by FTIR is shown.

Equipment: BOMEM MB100

Scanning Region: Mid Infrared (4000-400 cm^{-1})

Resolution: 8 cm^{-1}

Gas Cell: Graseby Specac glass cell

Optical Path: variable path length (1-8 m)

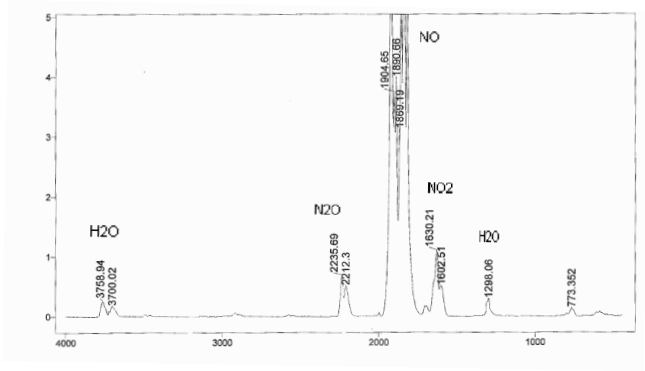


Figure 6. *Absorbance vs wavenumber (cm^{-1}) FTIR analysis*

NO 99.90% specification
Impurities: H_2O (20 ppm), NO_2 (100 ppm), CO_2 (100 ppm),
N_2O (200 ppm), N2 (500 ppm),

In the example, the NO 99.90% pure sample, is not in accordance with the specification, namely for the NO_2 (590 ppm) and N_2O (420 ppm) components which are significantly greater. This difference can be due to the reactivity of NO. However, it is an example of the lack of accuracy in analysis made by some gas suppliers not traceable to the SI.

Gas Chromatography – Mass Spectrometry (GC-MS)

GC-MS [4] is a hyphenated technique, which combines the Chromatographic separation with the spectral information. The GC separates mixtures into their components which will then pass through the MS [5]. Here, each component is fragmented into several ions, by the ion source. The mass filter or quadrupole classifies the different ions into the mass-to-charge ratio (m/Z). The detector will further produce a signal proportional to the ion concentration. GC-MS is very useful for identifying unknown compounds, namely trace contaminants that can be found in pure or balance gases. When there are no clues or reference standards to compare with an unexpected signal appearing in a GC analysis, this analytical technique can give a result, by comparing the mass signature obtained with the mass spectrum library, although the accuracy of quantification is poor.

Figure 7. *Gas Chromatograph – Mass Spectrometer Agilent 6890*

Automotive exhaust gas analysis

An example of an automotive gas analysis of a Certified Reference Material (CRM) provided by a gas supplier is presented. This CRM contained an unexpected impurity.

Equipment: Agilent 6890

Column: HP PlotQ, HP Molesieve

Temperature: Variable (60–150ºC)

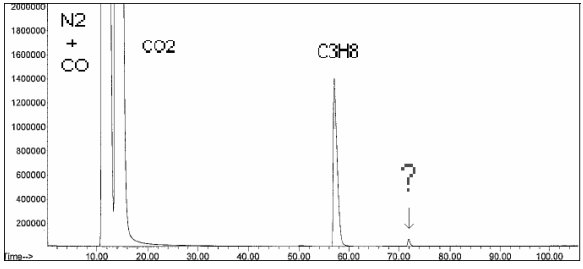


Figure 8. *Chromatogram for an automotive CRM gas sample*

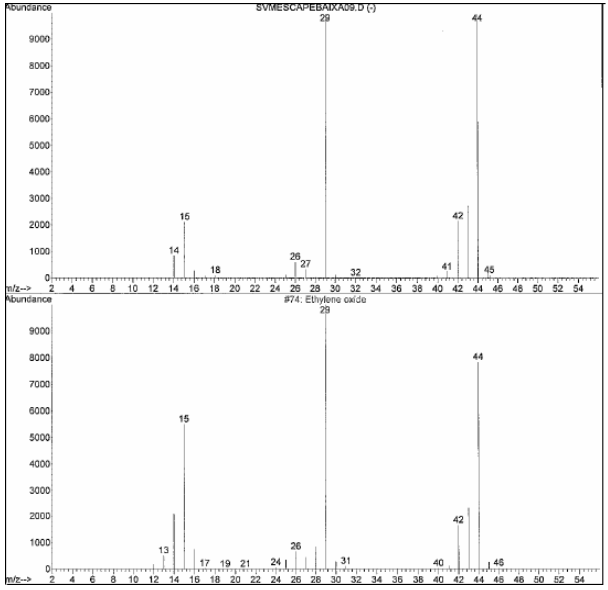


Figure 9. *Mass spectrum comparison: above, unknown component sample; below, ethylene oxide*

Automotive exhaust gas CRM specification
Composition: CO (0.5%), CO_2 (6%), C_3H_8 (100 ppm), N_2 (matrix)

Comparing the mass spectrum of the unknown component sample with the mass spectrum of Ethylene Oxide (C_2H_4O), it was observed that the contaminant corresponds to this component. Besides not interfering with components in the mixture, this molecule could introduce a considerable source of error if it was, as previewed, used as standard material for calibration of automotive gas analysis equipment, since this apparatus reads total hydrocarbons.

Conclusions

In gas purity analysis there is no universal equipment. The maximum information on a sample characterization will be given by combining different techniques.

To obtain a quantitative analysis there is the need to purchase all the correspondent reference standards, which is not always easy due to price or availability restrictions.

Gas purity analysis is crucial on primary standards preparation enabling us to decrease measurement uncertainty and therefore reach more accurate results.

References

[1] BIPM et al., *International Vocabulary of Basic and General Terms in Metrology*, 2ª ed, Geneva, ISO, 1993. 59 p. ISBN 92-67-01075-1.

[2] Agilent 6890 Series Gas Chromatograph Operating Manual, USA, 2000.

[3] The Michelson Series FT-IR Spectrometer – BOMEM User's Guide, Version 1.0, Canada, 1994.

[4] Agilent GC-MSD Chemstation and Instrument Operation "basique" H 8720-90000. USA, 2005.

[5 Agilent Technologies. 5973 Inert Mass Selective Detector Hardware Manual. USA, 2003.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Comparison of two different approaches to evaluate the on line gas chromatography for natural gas

Elcio Cruz de Oliveira

Petrobras Transporte S.A., Av. Presidente Vargas 328,
7º andar – Rio de Janeiro – RJ,
ZIP 20021-060, Brazil; Tel + 55 21 3211 9223; Fax + 55 21 3211 9300
Email: elciooliveira@petrobras.com.br

ABSTRACT: Recently, several approaches to evaluate the uncertainty in measurement have been developed. Within these, we may highlight the following: the guide to the expression of uncertainty in measurement (GUM), which evaluates the uncertainty of a measured result through the combination of each source of uncertainty in the measuring process and the approach derived from control chart techniques. The objective of this article is to determine if these two approaches are equivalent, or if in the case of gas chromatography of natural gas, there are differences between them.
KEYWORDS: Measurement uncertainty; GUM approach; Control charts; Natural gas.

Introduction

The evaluation of uncertainty associated with an analytic result is an essential part of the measurement process. The uncertainty of a measurement is defined as "a parameter associated with the result of a measurement, which characterizes the dispersion of values that can be fundamentally attributed to a measurand" [1]. The result of a measurement is considered as the best estimate of the value of measuring accompanied with all the sources of uncertainty that contribute to its propagation [2]. Consequently, the result of a measurement cannot be correctly interpreted without knowledge of the uncertainty of the result [2].

Several concepts have been developed for the evaluation of uncertainty related to the result of a measurement. The approach most used is the one proposed by GUM [3] for the expression of uncertainty in measurements, which combines the diverse sources of uncertainty, by expansion of the Taylor Series. In the beginning of the 1990s, EURACHEM [5] adopted GUM for analytical chemistry.

The control chart as a graphical means of applying the statistical principles of significance to the control of a production process was first proposed by Dr. Walter Shewhart in 1924 [4,5]. Control chart theory recognizes two types of variability. The first type is random variability due to "chance causes" (also known as "common causes"). This is due to the wide variety of causes that are consistently present and

not readily identifiable, each of which constitutes a very small component of the total variability but none of which contributes any significant amount.

Nevertheless, the sum of the contributions of all of these unidentifiable random causes is measurable and is assumed to be inherent to the process. The elimination or correction of common causes requires a management decision to allocate resources to improve the process and system.

The second type of variability represents a real change in the process. Such a change can be attributed to some identifiable causes that are not an inherent part of the process and which can, at least theoretically, be eliminated. These identifiable causes are referred to as "assignable causes" or "special causes" of variation. They may be attributable to the lack of uniformity in material, a broken tool, workmanship or procedures or to the irregular performance of manufacturing or testing equipment.

When we observe that the second type of variability does not exist, the first one – random errors – can be considered as the process uncertainty.

A standard material or control sample is measured repeatedly over time. The presumption of this practice is that the variation experienced on this material will be indicative of the laboratory total expected variation. Incorporation of specific known or potential sources of variation in the testing program is encouraged.

A control chart is prepared and the results are evaluated to identify short-term variation and longer-term variation. These data can then be used to determine an estimate of uncertainty standard deviation.

Methodology

After calibration, over a period of five days, a certified reference material (CRM) was analyzed by gas chromatography. This sample was a natural gas composition. We obtained 20 results, four replicates per day.

The outlier data were tested by Grubbs' test [6] and the data normality was evaluated using the Anderson-Darling statistic [7].

GUM approach

The GUM approach (also known as the bottom-up approach) consists of the identification and quantification of the relevant sources of uncertainty following the combination of these individual estimates. In the estimation of total uncertainty, it is necessary to deal separately with each source of uncertainty to know its

contribution. Each one of the separate contributions of uncertainty is referred to as a component of uncertainty.

When expressed as a standard deviation, a component of uncertainty is known as a standard uncertainty, *u(x)*. If there is a correlation between some of the components, then we must calculate the co-variance. However, it is frequently possible to evaluate the combined effect of the various components.

For a result of measurement *x*, standard combined uncertainty, $u_c(x)$ is an estimate of combined standard deviation equal to the positive square root of the total variance obtained by the combination of all the components of uncertainty evaluated, using the law of propagation of uncertainty.

For many propositions in analytic chemistry [8], an expanded uncertainty, *U*, may be used. The expanded uncertainty comes from the interval in which the value of the measurand is believed to be with a certain level of confidence. *U* is obtained by multiplying $u_c(x)$, the combined standard uncertainty, by a coverage factor *k*. The choice of the factor *k* is based on the required level of confidence.

In gas chromatography, when a quantitative analysis is carried out, the sources of uncertainty identified have to be taken into consideration, both in the calibration and in the sample. Normally, in gas chromatography, due to the difficulty in obtaining certified reference materials, with different concentrations of components, calibration of a single point is utilized. With a view to minimizing the possible systematic errors, the CRM must present the closest possible concentration to the sample being analyzed. For the test considered here, the approach used follows ISO 6974-2 [9]. The mathematical model used to calculate the non-normalized molar fraction is the following:

$$\overset{*}{x}_i = \frac{x_{CRM,i}}{\hat{x}_{CRM,i}} \hat{x}_i \tag{1}$$

where:

$\overset{*}{x}_i$ is the non-normalized molar fraction of component i;

$x_{CRM,i}$ is the molar fraction of component i in the CRM, found on the certificate;

$\hat{x}_{CRM,i}$ is the molar fraction of component i in the CRM, analyzed;

$\hat{x}_i$ is the molar fraction of component *i*, analyzed.

The uncertainty of the non-normalized molar fraction of component *i*, $\overset{*}{x}_i$, from equation (2), in considering the independent quantities among themselves, is:

$$\left(\frac{u_{\overset{*}{x}_i}}{\overset{*}{x}_i}\right)^2 = \left(\frac{u_{x_{CRM,i}}}{x_{CRM,i}}\right)^2 + \left(\frac{u_{\hat{x}_{CRM,i}}}{\hat{x}_{CRM,i}}\right)^2 + \left(\frac{u_{\hat{x}_i}}{\hat{x}_i}\right)^2 \tag{2}$$

where:

$u_{\overset{*}{x}_i}$ is the standard uncertainty of the non-normalized molar fraction of the component *i*;

$u_{x_{CRM,i}}$ is the standard uncertainty of the molar fraction of component *i* in the CRM, found on the certificate;

$u_{\hat{x}_{CRM,i}}$ is the standard uncertainty of the molar fraction of component *i* in the CRM, analyzed;

$u_{\hat{x}_i}$ is the standard uncertainty of the molar fraction of component *i*, analyzed.

The mathematical model used to calculate the normalized molar fraction is the following:

$$x_i = \frac{\overset{*}{x}_i}{\sum_{w=1}^{q} \overset{*}{x}_w} \tag{3}$$

where:

x_i is the normalized molar fraction of component *i*;

$\overset{*}{x}_w$ is the non-normalized molar fraction of each component *i*.

The uncertainty of the normalized molar fraction of component *i*, x_i, from equation (4) is:

$$u_{x_i} = x_i \sqrt{\frac{1 - 2\overset{*}{x}_i}{\overset{*}{x}_i^2} u_{\overset{*}{x}_i}^2 + \sum_{w=1}^{q} u_{\overset{*}{x}_w}^2} \tag{4}$$

Control chart approach

Preliminary control charts are then prepared and examined. These charts are evaluated to determine if the process is in a state of statistical control. The usual principles of control charting use short-term variability to estimate the limits within which samples of test results should vary. For control sample programs this short-term variability is equivalent to repeatability precision. It is expected, however, that additional contributions to variation will be present over time and therefore additional variation, equivalent to intermediate precision, will be encountered.

Two types of control charting methods are recommended to develop estimates of uncertainty. These include:

– mean and range or standard deviation charts are used when multiple test results are conducted in each time period;

– individual charts are used when single test results are obtained in each time period.

Either a range chart or a standard deviation chart may be used to estimate the short-term variability when multiple assays are conducted under repeatability conditions per time period.

An estimate from the control chart data can be compared to other estimates of repeatability (within a laboratory, short-term variation) if available.

Sample averages are examined and may provide estimates of variation caused by other factors. Such factors may include environmental effects, operator factors, reagents or instruments. In general, these sources of variation will be defined (including acceptable tolerances) by the test method.

A specified number of independent test results are taken during each time period.

Either a range chart or a standard deviation chart is prepared. This is examined for special cause variation. If the variability appears random then an estimate of repeatability is calculated. This may be done by pooling the sums of squares, using the average standard deviation, or using the average range.

A means chart is used to examine variation among time periods. Limits on this chart enable comparison of variation between time periods using repeatability as the estimate of error.

If the control chart shows a state of statistical control then the uncertainty will be assumed approximately equivalent to the repeatability standard deviation.

In most cases it will be expected that the variability between means will show an "out of control" condition indicating that there are "special" causes of variation in addition to repeatability. The between means variation and within means repeatability estimates are then used to calculate an estimate of uncertainty standard deviation.

The following steps are suggested:

1 – Prepare a standard deviation control chart
Calculate the average of the standard deviations (*p* test periods):

$$\bar{s} = \frac{\sum s}{p} \tag{5}$$

If the standard deviations in many of the samples were zero, then we recommend replacing the values of 0 with a value calculated as scale division divided by $2\sqrt{3}$.

2 – Estimate the within sample standard deviation this is an estimate of a single laboratory repeatability standard deviation.

A direct estimate of single laboratory standard deviation, s_r , is calculated based on the "pooled" variances. This is found by: calculating the squares of each standard deviation; summing the squares, $\sum s^2$, dividing by the number of samples; and taking the square root.

$$s_r = \sqrt{\sum s^2 \Big/ p} \tag{6}$$

3 – Estimate the between time or sample variation

Since there is a between sample or between time variation, an estimate of the between time standard deviation is then calculated. First the standard deviation among the sample averages is found.

The s_{time} is then calculated as:

$$s_{time} = \sqrt{s_{\bar{x}}^2 - \frac{s_{within}^2}{n_{within}}} = \sqrt{s_{\bar{x}}^2 - \frac{s_r^2}{n_{within}}} \tag{7}$$

where:

$s_{\bar{x}}$ = the standard deviation of the averages,

n_{within} = number of repeats (4),

s_{within} = standard deviation within groups and is equivalent, to s_r = single laboratory repeatability standard deviation.

Note: If the difference under the radical sign is negative, meaning the estimate of s^2_{time} is negative, then this may be interpreted as indicating that the variation associated with time is negligible and the estimate of s_{time} is set to zero.

4 – The uncertainty standard deviation is estimated from a single time and a single repeat

$$s_u = \sqrt{s^2_{time} + s^2_r} \tag{8}$$

Note: This value is equivalent to an estimate of intermediate precision based on multiple time periods.

Results and discussion

GUM approach

The GUM approach requires the quantification of all the contributions relevant to the uncertainty.

We have followed method B, from ISO 6974-2, which uses only one calibration point.

We have calibrated the chromatograph using a mixture of standard gas and this CRM was analyzed three times as a sample. This implies the use of linear calibration without an interceptor.

The results of this methodology are found in Table 1, based in the equations (1) to (4).

Component	Concentration % mol/mol	Relative standard uncertainty, %	
		Sample	CRM
N_2	1.206	0.31	0.12
CO_2	1.705	0.69	0.28
C1	82.927	0.06	0.28
C2	8.160	0.32	0.18
C3	3.500	0.40	0.27
iC4	0.800	0.39	0.18
C4	1.200	0.41	0.18
iC5	0.300	0.52	0.16
C5	0.202	0.50	0.16

Table 1. *Relative standard uncertainty by the GUM approach*

The effect of the normalization algorithm in the GUM approach minimizes the uncertainty of the component with more than 50% in its composition.

With the exception of methane, where the uncertainty of the CRM is the highest source of contribution, as we increase the uncertainty of the CRM, we observe an increase in the uncertainty of the component.

Control chart approach

No value is considered an outlier using Grubbs' test and the data have normal distribution based on the Anderson-Darling criterion.

The results of this methodology are found in Table 2, based on equations (5) to (8).

Component	$\%s_u$
N_2	0.13
CO_2	0.39
C1	0.01
C2	0.04
C3	0.13
iC4	0.21
C4	0.17
iC5	0.45
C5	0.70

Table 2. *Relative standard uncertainty by the control charts approach*

Finally, we compare both approaches in Table 3. The first approach is separated into two parts: with and without CRM.

	GUM with CRM	GUM without CRM	Control charts
N_2	0.31	0.16	0.13
CO_2	0.69	0.42	0.39
C1	0.06	0.02	0.01
C2	0.32	0.05	0.04
C3	0.40	0.16	0.13
iC4	0.39	0.22	0.21
C4	0.41	0.19	0.17
iC5	0.52	0.44	0.45
C5	0.50	0.41	0.43

Table 3. *Comparison of the approaches*

When we do not consider the CRM, the results become too similar. This occurs because this source is the largest and the control chart approach can not perceive this variation.

Conclusion

The new approach – control charts – presented compatible results in relation to the GUM results.

This methodology showed itself to be simple to use (without differentials as the GUM), as well as using the same statistical of the process control data.

Acknowledgements

Experimental data were collected thanks to Nilson Menezes Silva and his staff from PETROBRAS/UN/SEAL.

References

[1] International Organization for Standardization, *International Vocabulary of Basic and General Terms in Metrology*, second ed., International Organization for Standardization, Geneva, 1993.

[2] Analytical Methods Committee, *Analyst* 120 (1995) 2303–2308.

[3] International Organization for Standardization, *Guide for the Expression of Uncertainty in Measurements*, International Organization for Standardization, Geneva, 1993.

[4] IS0 7870, Control charts – General Guide and Introduction, 1993.

[5] ISO 8258, Shewhart Control Charts, 1993.

[6] Miller, J. N. & Miller, J. C. *Statistics and Chemometrics for Analytical Chemistry*, 2000.

[7] ASTM D-6299, Standard Practice for Applying Statistical Quality Assurance Techniques to Evaluate Analytical Measurement System Performance, 2006.

[8] EURACHEM, *Quantifying Uncertainty in Analytical Measurement*, first ed., EURACHEM, 1995.

[9] ISO 6974: Natural Gas – Determination of Composition with Defined Uncertainty by Gas Chromatography, Part 2, 2001.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Performance characteristics of GF-AAS and some metrological aspects of trace element determination

Steluta Duta

Institute for Reference Materials and Measurements, EC-JRC-IRMM, Retieseweg 111, 2440 Geel, Belgium, steluta.duta@ec.europa.eu/steluta.duta@inm.ro

ABSTRACT: The performance characteristics of atomic absorption spectrometers used for determination of metal pollutants in water are set-up by legal metrology recommendations. As estimation of results is based on the instrumental comparisons between the signal provided by a known reference solution and the sample under investigation, the instrumental performance characteristics of the equipment are important to provide proper quality results. Some metrological characteristics of the instrument are presented and evaluated. From the measurement uncertainty evaluation of metal pollutant determination in water as well as from the metrological verification/calibration activities of atomic absorption spectrometers, some aspects affecting the quality of results are also identified and discussed.
KEYWORDS: GF-AAS, Trace elements in water, Metrology in Chemistry, Measurement uncertainty

1. Introduction

The determination of metal pollutants in water by atomic absorption spectrometry with graphite furnace (GF-AAS) is a common and well established analytical technique for many chemical testing laboratories. There are the standardized procedures [1] as well as other analytical procedures described in the instrument manual from the manufacturers.

The atomic absorption spectrometers with graphite furnace used for metal pollutant determination in water should fulfill some criteria that are set-up by legal metrology recommendations [2]. At the national level, some legal metrological norms [3] and some specific procedures [4] to calibrate the atomic absorption spectrometers are issued and applied on the basis of OIML recommendations.

It is also well known that for accreditation purposes the chemical testing laboratories should fulfill some technical requirements according to ISO/IEC [5]. It is mentioned in this document that 'equipment shall be calibrated or checked to establish that it meets the laboratory's specification requirements and complies with the relevant standard specifications'. The measurement uncertainty evaluation and assuring the traceability of measurement results are also required.

In this context, the paper presents some aspects concerning the metrological performance characteristics of atomic absorption spectrometers with a graphite furnace in connection with their consequence on the quality of the analytical results.

2. Performance characteristics of GF-AAS

The traditional metrological assurance of spectrochemical measurement results is mainly concerned with the qualification of instrumental performance characteristics of atomic absorption spectrometers that includes the verification and calibration of the equipment. The proper use of the certified reference materials to assure the traceability of measurement results as well as the evaluation of measurement uncertainty are also important activities.

Accordingly, with the legal metrology recommendations [2, 3], the atomic absorption spectrometers used for metal pollutant determination in water are subject to the compulsory periodical verification of their metrological performance characteristics. Additionally, the authorized metrological laboratories perform the calibration of these instruments at the environmental request of laboratories.

The instruments are tested in a standard configuration and under standard operating conditions to determine the performance characteristics. This is obviously the case for flame atomic absorption spectrometers. For atomic absorption spectrometers with a graphite furnace the experimental conditions should be carefully selected and optimized.

The main instrumental performances of atomic absorption spectrometers with a graphite furnace that are verified are: (i) the working range; (ii) the short term precision; (iii) the limit of detection and (iv) the characteristic concentration (or sensitivity).

As is well know in the atomic absorption technique, the concentration of an unknown sample is evaluated from the calibration data. Calibration means the set of operations that establish, under specified conditions, the relationship, within a specific range, between values indicated by the instrument and the corresponding values assigned to reference standard solutions. The instrument is usually calibrated as far as possible according to the manufacturer's procedure.

Copper is the recommended chemical element used to evaluate the instrumental performances of atomic absorption spectrometers. A copper standard solution with the nominal concentration (1000±3) mg/L is used as a stock standard solution. The corresponding standard solutions in the working range (2-40) µg/L are prepared by successive dilution steps from working solutions that normally have the concentration in the range (10-100) mg/L.

A spectrometer type SOLAAR 939 with GF 90 furnace and video-auto-sampler SFTV is used. The measurement conditions are wavelengths 324.8 nm; bandpass 0.5 nm; lamp current 80% mA, sample volume 20 µL, D_2 background correction. Table 1 shows the optimized temperature programme.

Step	Temp. ^{0}C	Temp. variation, ^{0}C/s	Time, s	Gas flow, L/min
Dryness	120	10	30	240
Pyrolise	850	50	20	240
Atomization	2100	0	3	0
Cleaning	2500	0	3	300

Table 1. *Temperature programme for copper determination*

An **optimal working range** is considered as the range where the copper concentration varies proportional with the instrumental signal. The copper standard solutions with the concentration in the range (4-40 µg/L) are used to investigate the working range.

Calibration, evaluated as the regression line of concentration and corresponding absorbances of up to five copper standard solutions is performed. In practice there are two possibilities to perform the calibration, concentration versus height signal or concentration versus area signal. The calibration line and the regression coefficient are shown in Figure 1.

The atomization process is well described either by the height, respectively by the area of the instrumental signal. This is because the complete atomization takes place in a short period of time and gives proper information on the linearity of response as well as on the sensitivity of the measuring instrument. In practice, a better linear correlation is obtained (regression coefficient, R=0.9991) if the peak area is considered.

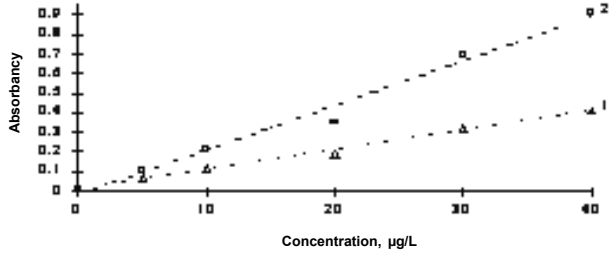


Figure 1. *Linear calibration of SOLAAR 939 spectrometer*

1-peak area: A=0.010c+0.0069; R=0.99911

2-peak height: A=0.0228c+0.0201; R = 0.9926

In addition to this, in the analytical practice the quadratic or cubic althgoritms could be used to calibration the new types of spectrometers. As an example, chromium calibrate graphs are presented in Figure 2 using two different types of instruments. Good regression coefficient is obtained for the quadratic or cubic althgoritms used. In this case, the above definition of the optimal working range is not valid anymore.

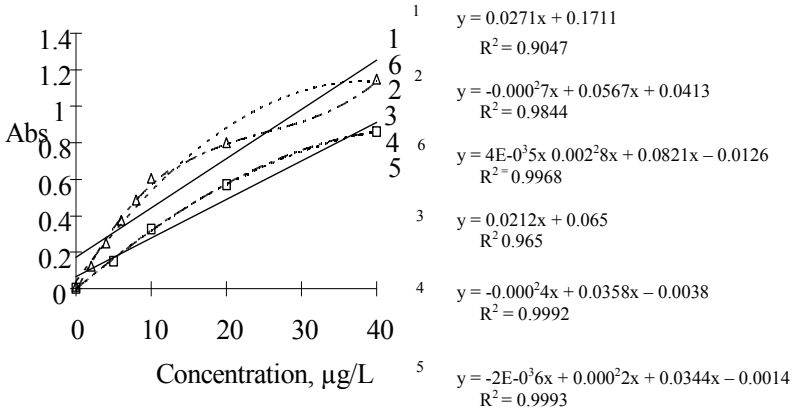


Figure 2. *Linear, quadratic and cubic calibration*

VARIAN instrument: 1- linear; 2- quadratic; 6 - cubic;
PERKIN-ELMER instrument: 3-liniar; 4-quadratic 5-cubic

It is important to mention that the Calibration Certificate issued by metrological authorized laboratories is required in the context of the accreditation process of the laboratories. The procedure [4] describes how the calibration uncertainty is evaluated. One of its components is the associated standard deviation due to the regression line [6].

The standard deviation of the predicted concentration value s_{ca} is expressed by the equation:

$$s^2_{\overline{c_A}} = \frac{s_0^2}{b^2}\left[\frac{1}{p_A} + \frac{1}{n} + \frac{(\overline{c_A} - \overline{c})^2}{\sum_{j=1}^{n}(c_j - \overline{c})^2}\right]$$

where: s_0 - the standard deviation of the regression line

b - the slope of the calibration line

N - the number of standard used for calibration

P - the number of replicates

The calibration uncertainty in the range (4.00–16.00) μg/L represents (6.0-2.5)% as relative standard uncertainty.

Chemical element	Cu		
Calibration linear range, μg/L	1.0 … 20.0		
Slope	0.0150 ± 0.0003		
Intercept	0.0232 ± 0.0046		
Regression coefficient	0.9990		
True concentration, μg/L	4.00	8.00	16.00
Observed value, μg/L	3.82	7.93	16.18
Calibration uncertainty, μg/L	0.23	0.25	0.39
Relative standard uncertainty	6.0%	3.2%	2.4%

Table 2. *Calibration uncertainty for copper determination*

From an analytical point of view it is not clearly defined how the testing laboratories could use in practice the calibration uncertainty information provided by authorized metrological laboratories. This is because the calibration line of each instrument is performed before each measurement.

If it is still the case to evaluate this uncertainty, the long term precision (reproducibility) of the instrument calibration could be included in the calibration uncertainty value. However, in this case an over-estimated measurement uncertainty will be propagated into the laboratory reported results.

The short term precision (repeatability) of spectrometric measurements is usually evaluated as relative standard deviation of repeated measurements.

The bias is defined as the difference between the mean of observed values and the certified value of the standard solutions, a maximum 5% is considered an acceptable limit; 3.5% for relative standard deviation of five repeated measurements can be assumed as acceptable limits for short term precision.

Table 3 presents some results on the performance characteristics of an instrument in the concentration range (5.00-40.00) μg/L. The optimum measurement conditions such as peak area, low noise signal, bandpass etc. were carefully selected. Even thus, it can be noted there are some values that do not comply within the acceptable limits. From a metrological point of view, the instrument should be rejected. It is obvious such a decision should be taken carefully because it is possible that not all factors affecting the results were well documented. At this stage it should be also considered if more effort should be spent to improve the above performance characteristics. Due to the particularities of metal pollutant determination there are some other factors that significantly affect the quality of results, such as matrix effect, the lower detection limit, etc.

Concentration, µg/L	Observed Concentration µg/L	Relative bias, %	RSD, %
5.00	5.31	- 6.2	3.5
10.00	10.23	- 2.3	2.7
20.00	19.35	+ 3.2	4.1
30.00	30.63	- 2.10	3.3
40.00	38.67	+ 3.32	3.9

Table 3. *Results on the performance characteristics of GF-AAS*

The detection limit for atomic absorption spectrometers is an important figure of merit for evaluating the instrumental performances. This is a statistical measure of the smallest concentration of a particular analyte that can be distinguished from the baseline noise signal.

To evaluate the detection limit, the calibration in the concentration domain is usually performed by a linear least square regression using up to five standard solutions. Several times series of blank measurements should be performed and the standard deviation of the mean value is calculated. The signal corresponding to the detection limit is the blank signal plus 3 times standard deviation of the blank signal. Than, the detection limit is quantified based on the calibration line in the concentration domain. Poor precision is obtained in the low concentration range due to significant noise signal, as a consequence it is recommended to use a higher concentration, usually 100 times above the detection limit.

As the detection limit reflects important information on the intrinsic performance of an atomic absorption spectrometer, such as instrumental background, signal to noise ratio etc. the optimal measurement conditions should be carefully set-up. As an example, if the dryness period in the temperature graphite programme is extended to 30 s as well as the pyrolysis temperature is set-up to 850°C and an atomization temperature is set-up at 2100°C, then the suitable limit of detection limit for copper measurements is obtained (Table 4). A careful calibration should be performed in the low concentration range. The uncertainties associated with the standard reference solution should be taken into consideration as well. The use of chemical modifiers also confers some benefits to improve the detection limit even for a simple matrix.

The characteristic concentration (sensibility) in atomic absorption spectrometry is defined as the concentration that gives an absorbance signal of 0.0044. The term sensitivity used in this context is different from VIM definition [7] that is mostly applied to the measuring instruments. It is obvious that different chemical elements have different sensitivity values. Using the slope of the calibration curve, the characteristic concentration is obtained by dividing 0.0044 by the slope. Using different concentration sub-ranges, the characteristic concentration can easily be improved. It is also considered that the characteristic concentration for

metal pollutants in water is easily improved if the peak area is observed instead of the peak height. From a practical point of view the characteristic concentration is an important information for the analyst to predict the measuring range for a specific analyte in an unknown sample.

In Table 4 the detection limit and the characteristic concentration set-up by OIML are mentioned. Some results on the metrological evaluation of several instruments used for measuring the metal pollutant concentration in water are presented as well.

	Limit of detection, μg/L	Characteristic concentration, μg/L
Requirement [2], [3]	0.2	0.4
Instrument 1	0.06	0.06
Instrument 2	0.2	0.08
Instrument 3	0.12	0.09
Instrument 4	0.25	0.18
Instrument 5	0.18	0.45

Table 4. *Limit of detection and characteristic concentration*

From a metrological point of view, if the acceptable detection limit for copper determination is 0.2 μg/L and the acceptable characteristic concentration is 0.4 μg/L, several instruments should be rejected.

However, to take this decision, more information should be considered regarding the confidence interval of the above evaluated performances characteristics, the manufacturer specifications of each instrument as well the analytical aspects of metal pollutant determination in real water samples.

3. Analytical aspects of metal pollutant determination in water

Determination of metal pollutant concentration in water by atomic absorption spectrometry with a graphite furnace is a well accepted technique in many chemical laboratories. When the measurement uncertainty is evaluated, the analytical procedure is better understood. In this respect some aspects on the measurement uncertainty evaluation are presented.

The measurement uncertainty of an analytical result is defined as one important "parameter associated with the result of measurement, that characterizes the dispersion of values that could reasonably be attributed to the measurand". The standardized and well accepted procedure to evaluate the measurement uncertainty is described by [8]. The main steps to evaluate the uncertainty of chemical results are:

(1) define the measurand;

(2) describe the model equation of measurement procedure;

(3) identify all possible sources of uncertainty;

(4) evaluate all input quantities;

(5) evaluate the standard uncertainty of each input quantity; (6) calculate the value of the measurand using the model equation;

(7) calculate the combined standard uncertainty of the result; (8) analyze the uncertainty contribution index;

(9) calculate the expanded uncertainty;

(10) document all steps in a report.

An atomic absorption spectrometer type Ati Unicam SOLAAR 939 with GF 90 Plus furnace and SOLAAR FS95 auto-sampler is used. The sample is collected and dispensed via an inert PTFE capillary with rapidly interchangeable tips. All facilities are programmed from the instrument software and sampling facilities. The D2 background wavelength correction is performed. The optimized parameter, such as, the wavelength 324.8 nm; the bandpass 0.5 nm; the lamp current 80% mA and the sample volume 20 μL are selected.

The analytical steps for copper determination are:

(i) water stabilization with nitric acid;

(ii) preparation of the standard reference solutions;

(iii) calibration of the measurement procedure;

(iv) injection of sub-sample water into the graphite furnace;

(v) determination of instrumental signal for water sample;

(vi) determination of copper concentration in water, in μg/L;

(vii) report the result.

The concentration of copper is calculated using two calibration solutions by bracketing the unknown sample concentration. The final result is further corrected by recovery factor and dilution factor.

According to [8] the reported result for copper concentration in water is:

$Ccu = (2.25 \pm 0.18)$ μg/L, k=2

From the detailed measurement uncertainty evaluation, presented in [9], the main contributors affecting the quality of results are summarized in Figure 3.

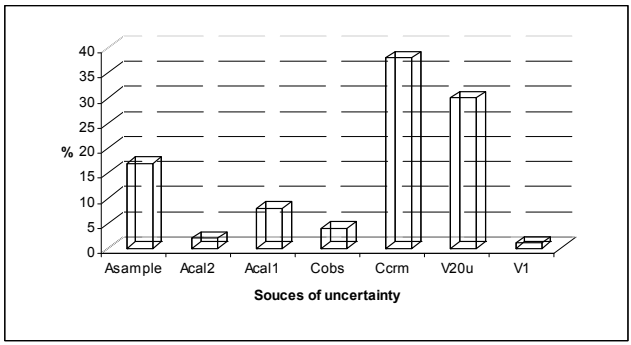


Figure 3. *The main sources of measurement uncertainty*

The ***sample absorbance*** A_{sample} is determined using the two point calibration, by bracketing the signal of an unknown sample between the signal A_{cal1} and A_{cal2} that correspond to copper amounts w_{cal1} and w_{cal2} contained in 20 µL standard solutions.

The absorbance measurement for a water sample has 17% contribution to the overall uncertainty. The contribution of the absorbance measurement of the standard solutions A_{cal1} and A_{cal2} represents 8% and respectively 2%. These are quantified on the measurement repeatability data.

The ***recovery*** of the analytical procedure was evaluated against a certified reference material (CRM), by dividing the observed concentration C_{obs} to the certified concentration C_{crm}. The uncertainty contribution of certified reference material is significant, respectively 38% to the overall uncertainty. The observed concentration of certified reference material C_{obs} evaluated as repeatability of measurements, contributes with 4% to the overall uncertainty.

The ***volume measurement*** of 20 µL sub-sample injected into graphite furnace contribute with 30% to the overall uncertainty. Even the FS95 auto-sampler was used and the sample was collected and dispensed via an inert PTFE capillary with rapidly interchangeable tips; the contribution of this volume measurement is significant. Other volume measurements contribute non-significantly to the overall uncertainty.

From the metrological performance characteristics of atomic absorption spectrometers with a graphite furnace presented above, in this case study the significant contribution to the copper concentration is the short term precision (repeatability) of measurements.

From an analytical point of view the recovery of analytical procedure seems to be the critical issue, even for the simple matrix sample. The proper selection of the suitable certified reference material and its associated uncertainty is also important.

4. Conclusions

The main performance characteristics of atomic absorption spectrometers with a graphite furnace used for metal pollution determination in water are defined. The optimum working range, the short term precision, the detection limit and the characteristic concentration are evaluated for several types of instruments in accordance with the legal metrological recommendations. The linear calibration of the instruments and its uncertainty were examined in accordance with metrology practice as well as in connection with the analytical practice.

From the analytical point of view, the main contributors to the measurement uncertainty of copper determination are discussed, in connection with the instrumental performance characteristics of the spectrometers. The absorbance measurements are the subject of improvement from the repeatability point of view. The proper selection of certified reference materials with smaller uncertainty used for the recovery is pointed out from the uncertainty study case.

5. References

[1] ISO 15586: Water quality: Determination of trace elements using atomic absorption spectrometry with graphite furnace, 2003.

[2] OIML R100, Atomic Absorption Spectrometers for measuring metal pollutants in water, 1991.

[3] NML 9-02-94 Atomic Absorption Spectrometers for measurement of metal pollutants in water, 1994.

[4] PS-02-5.1-INM: Specific Procedure: Calibration of the Flame Atomic Absorption Spectrometers, 2006.

[5] ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories, 2005.

[6] S. Duta, The role of Certified Reference Materials in Romanian Traceability Scheme*, OIML Bulletin*, vol. 38, No. 3, 15-23, July 1997.

[7] International Vocabulary of Basic and General Terms in Metrology, ISO, Geneva, 1993.

[8] *Guide to the Expression of Measurement Uncertainty (GUM)*, second ed. ISO, Geneva, 1995.

[9] S. Duta and all., Evaluation of measurement uncertainty of trace elements in water using atomic absorption spectrometry, *Proceeding of Colloquium Spectroscopium International XXXIV*, 2005.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The influence of water density and conductivity variation in a gravimetric calibration method

E. Batista, E. Filipe

Instituto Português da Qualidade (IPQ) – Rua António Gião, 2 – Caparica – Portugal
ebatista@mail.ipq.pt, efilipe@mail.ipq.pt

ABSTRACT: This paper describes the influence of using water as a calibrating liquid, with different characteristics from those recommended by the ISO 3696 standards in gravimetric calibration of glassware equipment, from 1 ml to 5,000 ml.
The influence of using different density tables/formulae, available in technical literature is also discussed.

Introduction

The volumetric equipment normally used in analytical or chemical laboratories is calibrated using the gravimetric method described in ISO 4787 standard [1], where the mass of liquid (pure water) contained or delivered from the calibrated instrument is converted into volume using an appropriated formula, at a reference temperature of 20ºC.

$$V_{20} = (I_I - I_E) \times \frac{1}{\rho_W - \rho_A} \times \left(1 - \frac{\rho_A}{\rho_B}\right) \times \left[1 - \gamma(t - 20)\right] \quad (1)$$

The density of the water is one of the most important elements in a gravimetric calibration and is determined at the measured temperature using a density table or formula. There are several water density tables/formulae in technical literature (Tanaka, Spieweck, Wagenbreth & Blanke, Patterson & Morris, IAPWS, among others) that can be used for the conversion of mass into volume. The authors present experimental values that are different from each other. These small but important differences may imply meaningful variations in volume determination.

In this gravimetric method the calibration liquid – pure water - should be, according to ISO 3696 standard [2], at least of grade 3, which means that the water should be, at least, distilled and have, among other properties, a conductivity lower than 5 μS/cm.

To compare the importance of using different density tables or a calibration liquid with different characteristics from the one presented in the literature,

several tests were performed with several types of laboratory glassware from 1 ml to 5,000 ml.

Conductivity

The electrolytic conductivity of a solution is a measure of the amount of charge transported by ions. When the source of ions is from impurities, the conductivity becomes a measure of purity, the lower the conductivity the purer the solution.

The use of conductivity to measure water purity can be used in several areas of industry. The pharmaceutical industry use the water as an ingredient and in others processes, analytical laboratories use pure water as a reference material or solvent, the environmental laboratories measure conductivity for monitoring water quality and the calibration laboratories use pure water for volume determination.

The theoretical conductivity of pure water can be determined using individual ionic conductivity of H^+ (349.81 S·cm^2/equiv) and OH^- (198.3 S·cm^2/equiv) using the following equations [3]:

$$\Lambda_{water} = \frac{1000 \times \kappa}{c}$$

$$\Lambda^0_{water} = 349.81 + 198.3 = 549.11 \text{ S·cm}^2\text{/equiv} \quad (2)$$

$$c = 10^{-7} \text{ equiv/l}$$

$$\kappa = \frac{\Lambda c}{1000} = \frac{549 \times 10^{-7}}{1000} \text{ S/cm} = 0.055 \text{ μS/cm}$$

This value will increase in the present of ionic impurities and in the presence of CO_2.

General solution conductivity can be assessed from resistance, R, or conductance, G, measurements in a cell with well known geometry or in a commercial cell with a constant, K_{cell} (cm^{-1}) assessed with the standard solution of known conductivity, κ, taking into account the temperature effect:

$$\kappa = K_{cell} \times G \times \left[1 + \alpha \times (20 - t)\right] \quad (3)$$

where

α – 2 %/ºC;
t – temperature (ºC).

Density

Density is a physical property of matter that describes the degree of compactness of a substance - i.e, how closely packed together the atoms of an element or molecules of a compound are. The more closely packed the individual particles of a substance are, the more dense that substance is. Since different substances have different densities, density measurements are a useful means for identifying substances.

In the case of water and for the same temperature, the increase of purity lowers its density. The density of pure water is a widely used reference for determinations of unknown densities and volumes.

Since the beginning of the 20th century, tables for water density from 0 ºC to 40ºC have been deduced from measurements carried out by Thiesen *et al* [4] and by Chappuis [5]. In 1975 Kell [6] performed a revision on the stated density values of water and converted them into the temperature scale IPTS68. In 1990 Kell's formula and the Wagenbreth and Blanke formula [7] were converted in the current temperature scale ITS90 by Bettin and Spieweck [8].

Recent progresses in science and technology have provided metrologists with several new concepts and new analytical instrumentation for the determination of the water density. The preparation of de-aerated pure water at standard atmospheric pressure has been improved and standardized and includes the definition of "pure" water with a isotopic composition of the Vienna Standard Mean Ocean Water (V-SMOW).

Using semiconductor technology for the production of pure water, and the development of new techniques such as the use of magnetic floating devices (sinkers) with very low uncertainties in the order of 1×10^{-6}, a number of researchers have collected more accurate data such as Patterson and Morris [9] in 1994 or Wagner and Pruss (IAPWS) [10] in 1996.

In 2001 Tanaka [11] combined four sets of results published in the 1990s by different authors and obtained an analytical expression for the density of SMOW water for the conditions recommended by IUPAC [12]. This expression is today the one recommended by the BIPM and has an uncertainty of 1×10^{-3} kg/m^3.

Despite this recommendation, in the metrological volume community, there is no agreement on the table or formula to be used.

Discussion of results

Water characteristics

Conductivity and density of three water samples: taped, distilled and ultra pure, were determined for gravimetric purposes for a reference temperature of 20ºC.

For conductivity determination all three water samples were placed for 24 hours, together with the Metrohm 712 conductivity meter and the Metrohm conductivity cell (previously washed with ultra pure water and dried in ambient air) in the same room kept at a constant temperature of (20 ± 0.5)ºC. The cell was then connected to the conductivity meter and immersed in each sample placed in a glass recipient until a steady measure of the conductivity was obtained. The instrument has an internal current frequency of 1 kHz.

Temperature was then recorded and conductivity readings were later corrected according to equation (3). The procedure was repeated 5 times and an average value was calculated.

The density of the three samples was determined using a digital density meter, a direct reading of Anton Paar DMA 5000. The samples were also placed for 24 h together with the density meter in the same room kept at constant temperature of (20 ± 0.5)ºC. The results are presented in the following table.

Water Type	**Condutivity (μS/cm)**	**Density (g/ml)**
Ultra Pure	0.487	0.998192
Distilled	7.873	0.998220
Tap	462.4	0.998376
Literature	0.0055 [3]	0.998207 [11]

Table 1. *Conductivity and density measurements*

The tap water sample being the most impure has the larger conductivity and density, the distilled water type has an intermediate value and the ultra pure has the lowest value and the closest to the literature value.

The results for conductivity and density are correlated as it can be seen in Figure 1.

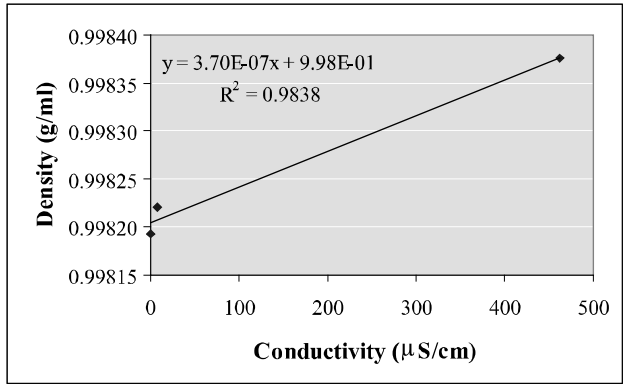


Figure 1. *Correlation between conductivity and density*

Volume determination

With the determination of each sample density the value was then used to determine the volume of several instruments using the gravimetric method. The results are presented in the following tables, where the theoretical volume is determined with the literature density (Tanaka formula) and the real volume based on the density measurements (Table 1).

1000 ml Flask			
Water Type	**Theoretic Volume (ml)**	**Real Volume (ml)**	**Expanded Uncertainty (ml)**
Ultra pure	999.974	999.989	0.044
Distilled	999.974	999.960	0.044
Tap	1000.145	999.976	0.043

Table 2. *Volume determination for a 1000 ml Flask*

100 ml Pycnometer			
Water Type	**Theoretic Volume (ml)**	**Real Volume (ml)**	**Expanded Uncertainty (ml)**
Ultra pure	99.981	99.982	0.005
Distilled	99.982	99.981	0.003
Tap	99.999	99.982	0.004

Table 3. *Volume determination for a 100 ml Pycnometer*

25 ml Cylinder			
Water Type	**Theoretic Volume (ml)**	**Real Volume (ml)**	**Expanded Uncertainty (ml)**
Ultra pure	25.059	25.059	0.014
Distilled	25.069	25.068	0.013
Tap	25.069	25.064	0.011

Table 4. *Volume determination for a 25 ml cylinder*

5 ml Pipette			
Water Type	**Theoretic Volume (ml)**	**Real Volume (ml)**	**Expanded Uncertainty (ml)**
Ultra pure	5.030	5.030	0.004
Distilled	5.032	5.032	0.004
Tap	5.029	5.028	0.004

Table 5. *Volume determination for a 5 ml pipette*

From the presented results it can be seen that the theoretical volume for tap and distilled water are in most cases larger than the real volume because the used density value in the table is lower that the real one. Applying the correct and real density value in the calibration formula (1) the volume for the same instrument using different water samples became similar within the range of the uncertainty.

For the ultra pure water the results for real volume and theoretic volume are similar because there are no significant differences between table density and determined density.

Also the larger the volume of the instrument is the larger the difference between theoretical and real volume.

Water density tables

In order to verify possible variations in volume measurements due to small differences between several density formulae presented in current literature, a set of volumetric flasks of different capacity were calibrated for a temperature of 20ºC. The obtained volume results are presented in an increasing order in the table below:

Density Formula	Volume Flask (ml)			
	5000	**1000**	**500**	**100**
Wagner & Pruss	5000.215	999.876	500.083	99.968
Tanaka	5000.218	999.876	500.083	99.968
Patterson & Morris	5000.223	999.877	500.084	99.968
Bettin & Spieweck	5000.236	999.880	500.085	99.968
Wagenbreth & Blanke	5000.248	999.882	500.086	99.968
Maximum variation (ml)	0.033	0.006	0.003	0

Table 6. *Volume calibration of a volumetric flask using different density formulae*

As can be seen in Table 6, for small amounts of liquid the used density formula is not relevant, but as we increase the capacity of the flask the differences became more significant. It can also be verified in Table 6 that the major differences are for the Wagenbreth and Blanke formula.

In order to see if there is a correlation between the volume capacity and the maximum variation between the formulae a new figure is presented.

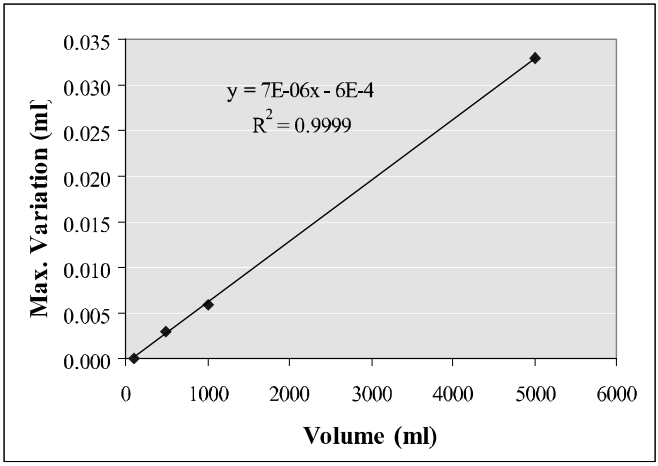


Figure 2. *Linear correlation between maximum volume variation and volume capacity*

This figure shows that there is a linear correlation between the maximum volume variation and the increase of the flask volume capacity, which means that this variation can be extrapolated to larger volumes.

Concluding remarks

The use of a suitable calibration liquid (water) with well defined characteristics is very important in the gravimetric calibration of glassware equipment. Density and conductivity should be controlled in order to obtain the correct volume results if the calibration liquid has different characteristics from those recommended by the standards.

Also through the analysis of the presented results it can be seen that there are small differences in volume calibration when using different density formulae, these differences increase with the increase of the instrument volume capacity. This variation is important and when comparisons between National Metrological Institutes in volume calibration are performed is fundamental to establishing an agreement on which density formula is to be used in order to have comparable results.

The Tanaka density formula was determined based on recent density results from different authors from 0ºC to 40ºC. As this formula is the one recommended by BIPM, and the one with the lowest uncertainty, the authors recommend its use by all the metrological volume community.

References

[1] ISO 4787:1984, “Laboratory glassware - Volumetric glassware - Methods for use and testing of capacity”.

[2] ISO 3696:1987, “Water for analytical laboratory use – Specification and test method”.

[3] Shreiner R.H., Pratt, K.W., *Nist Special Publication,* 2004, 260-142.

[4] Thisenm., Scheel K., Diesselholst H., *Phusik. Techn. Reichsanst. Wiss. Abho.,* 1900, **3**, 1-70.

[5] Chappuis P., *Trav. Mém. Bur. Int. Poids et Measures*, 1907, **13**, D1.

[6] Kell G. S., *J. Chem. Eng. Data,* 1967, **12**, 66-69; ibid., 1975, **20,** 97-105.

[7] Wagenbreth H., Blanke W., *PTB Mitteilungen*, 1971, **6**, 412-415.

[8] Bettin H., Spieweck F., *PTB Mitteilungen*, 1990, **100**, 195-196.

[9] Patterson J. B., Morris R. C., *Metrologia*, 1995/96, **32**, 333-362.

[10] Wagner W., Puss A., "The IAPWS formulation 1995 for thermodynamic properties of ordinary water substance for general and scientific use", *J. Phys. Chem. Ref. Data,* 1996, **25**.

[11] Tanaka M., Girard G., Davies R., Peut A., Bignel N., *Metrologia*, 2001, **38**, 301-309.

[12] IUPAC, Pure and Applied Chem., 1976, **45**, 1-9; Girard G., In *IUPAC Recommended Reference Materials for the Realization of Physicochemical Properties*, Oxford, Blackwell, 1987, 5-3.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Hydraulic Quantities

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Uncertainty estimation of free surface flow rates determined by ADCP or current meter gauging

Audrey Olivier[1], Bertrand Blanquart[2], Gilles Pierrefeu[1], Mattia Scotti[1]

[1]Laboratoire d'hydraulique et mesures de la Compagnie Nationale du Rhône, Lyon
[2]CETIAT, Division Métrologie, Villeurbanne

Abstract: The need to know the values of the flow rates measured with ever less uncertainty led the Compagnie Nationale du Rhône, with the help of CETIAT, to carry out an in-depth study on the uncertainty of the flow rates measured by two main methods, i.e. current meter and ADCP (Acoustic Doppler Current Profiler).
Standards NF EN ISO 748 and NF ENV 13005 are applied to the current meter. Then the uncertainty of the flow rate measured by ADCP is estimated by a comparative study between ADCP and the current meter (ISO 5725).
This study has led CNR to choose an uncertainty of ± 7% for current meter gauging and ± 5% for ADCP gauging.

Introduction

Knowledge of the Rhône flow rate and its affluents is of vital concern for Compagnie Nationale du Rhône (CNR) which manages it. CNR ensures its own flow measurements by current meter or ADCP. Gauging operations are the basis of the establishment of a relationship between height and water flow. This relationship is then used to calculate the water flow from a height measurement made continuously in a hydrometric station.

Several methods are used to measure the flow rate of rivers. Both methods examined in this document are the traditional "current meter" method and the more recent ADCP (Acoustic Doppler Current Profiler) method based on the Doppler effect. The first one has been used for a long time and is considered as the reference method, but its implementation is time expensive (up to half a day for one measurement). The second method is faster and can give data in trickier conditions, but is not as well known. Until now, the estimation of uncertainties was only carried out for the first method and was based in the interpretation of the dispersion of measurements obtained under reproducibility conditions in reference sites.

Estimation of measurement uncertainty by an analytical method is intended for the determination of predominant sources of uncertainties and the value of uncertainty that can be obtained for each of these two methods.

The measurement uncertainty associated with the flow rate measured by current meter can be estimated by the application of the NF ENV 13005 standard (August 1999) [1], usual reference in metrology, or by the application of the NF EN ISO 748 standard (January 2002) [2], specifically describing the estimation method of the measurement uncertainty for current meter gauging. The recommendations of these two standards will be applied and results will be compared.

For the ADCP method, there is no specific standard. The estimation of uncertainties will be based on comparisons between the current meter and ADCP achieved by Compagnie Nationale du Rhône since 1994. These comparison data, obtained under reproducibility conditions, will mainly be analyzed by following the recommendations of the NF ISO 5725 standard [3].

Current meter gauging

Determination of total flow rate

The total flow rate is determined from measurements of depth, width and water speed in different points of the section (Figure 1, from [2]).

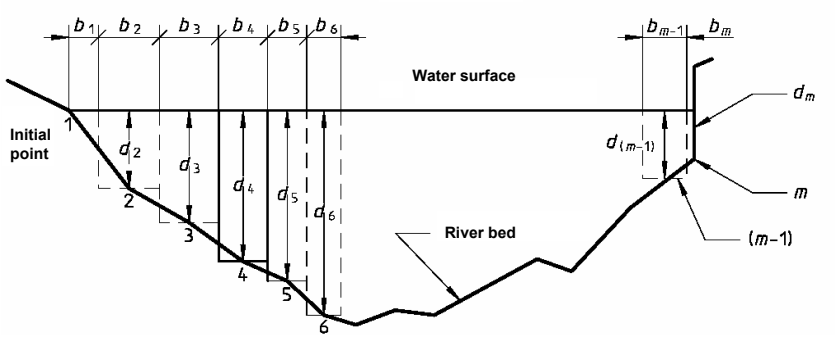


Figure 1. *Current meter gauging diagram*

Flow rate is calculated by relation (1):

$$Q = \sum_{i=1}^{N} Q_i = \sum_{i=1}^{N} b_i \cdot d_i \cdot V_i \tag{1}$$

where:

- N is the number of verticals;

- Q is total flow rate;

- Q_i is the elementary flow linked to vertical i;

- b_i is the width linked to vertical i;

- d_i is the depth linked to vertical i;

- V_i is the average speed associated with vertical i.

$$V_i = \frac{1}{d_i} \cdot \begin{bmatrix} (0.1 \cdot d_i + 0.1) \cdot V_{0.20} + (0.2 \cdot d_i - 0.1) \cdot V_{0.2d_i} + 0.2 \cdot d_i \cdot V_{0.4d_i} \\ +0.2 \cdot d_i \cdot V_{0.6d_i} + 0.15 \cdot d_i \cdot V_{0.8d_i} + (0.1 \cdot d_i + \frac{0.05}{3} d_i) \cdot V_{0.9d_i} \end{bmatrix} \quad (2)$$

This formulation of the average speed linked to vertical i (Figure 2) is inherent to CNR (formula used as long as the depth is higher than 1 m). It is however compliant with the NF EN ISO 748 standard. It leads to the application of a weight factor of one third of the speed measured at 0.9 d_i to the bottom layer (for a thickness of 0.05 d_i at the bottom).

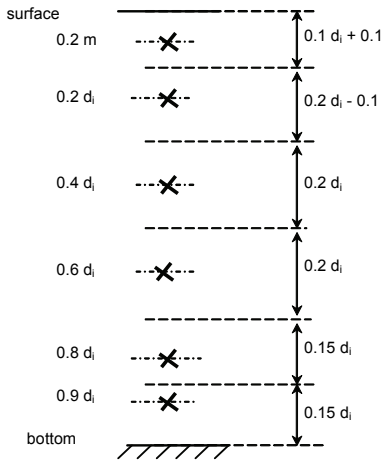


Figure 2. *Distribution of speeds in a vertical during standard current meter gauging*

The uncertainty for total flow rate is estimated from measurement uncertainties for the different quantities measured (width, depth, speed). In order to do this, two normative documents propose a method of estimation:

- standard NF ENV 13005, "*Guide pour l'expression de l'incertitude de mesure*", August 1999 (universal standard which can be applied to all measurement processes);
- standard NF EN ISO 748, "*Mesure de débit des liquides dans les canaux découverts. Méthode d'exploration des vitesses*", January 2002.

Application of standard NF EN ISO 748

Propagation of measurement uncertainties

Although standard NF EN ISO 748 is more recent than NF ENV 13005, it does not integrate the defined concepts and notations. The metrologist will therefore be surprised to find such terms as "random uncertainty" and "systematic uncertainty". For clarity purposes, definitions and mathematical notations used in this section are from the NF EN ISO 748 standard.

Flow uncertainty is defined by the following relation (3):

$$X_Q = \pm\sqrt{X'^2_Q + X''^2_Q} \tag{3}$$

where:

- X_Q is the total flow rate uncertainty;
- X'_Q is the "random" uncertainty;
- X''_Q is the "systematic" uncertainty.

- X'_Q is estimated by the following relation (4):

$$X'_Q = \pm\sqrt{\left[X'^2_N + \frac{1}{N}\left(X'^2_b + X'^2_d + X'^2_e + X'^2_p + X'^2_c\right)\right]} \tag{4}$$

This formula can be applied to gauging including at least ten verticals, if the elementary flow rates per vertical are not too far apart (elementary flow rate on a vertical cannot exceed 10% of total flow rate). The terms of the equation are:

- X'_N: random uncertainty linked to the number of N verticals;
- X'_b: random uncertainty associated with the determination of a vertical width;
- X'_d: random uncertainty associated with the determination of a vertical depth;
- X'_e: random uncertainty linked to the current meter exposition time;
- X'_p: random uncertainty linked to the number of points on each vertical;
- X'_c: random uncertainty associated with current meter calibration.

– X''_Q is estimated by:

$$X''^2_Q = X''^2_b + X''^2_d + X''^2_c \tag{5}$$

with:

- X''_b: systematic uncertainty on width;
- X''_d: systematic uncertainty on depth;
- X''_c: systematic uncertainty on current meter calibration.

Relations (4) and (5) combine uncertainties on measured parameters (current meter calibration) and on parameters linked to the measurement method (number of verticals, number of points per vertical, etc.).

X values are expressed in percentage, with a level of confidence of 95%. By analogy, these values represent expanded uncertainties (U) of the NF ENV 13005 standard.

Estimation of uncertainty components

The standard does not provide any recommendation on the practical determination of components X' and X'' values. These values should be estimated for each gauging site by the implementation of specific experimental methods. This is particularly difficult to achieve for rivers, contrary to flows confined to well known sections (channels, pipelines).

By default, the calculation can be based on the values given in Appendix E of the NF EN ISO 748 standard, even though these values were probably obtained under different conditions from those encountered by CNR on the Rhône and its affluents. However, even with default parameters, results obtained on flow uncertainty converge with forecasts of the experts.

A "standard" current meter gauging completed by CNR includes a maximum of 30 verticals in order to maintain a reasonable gauging time period while closely examining the measurement section, i.e. under three hours. In this way, in the case of CNR application using default values of the NF EN ISO748 standard, the minimum random uncertainty for 30 verticals is:

$$X'_Q = \pm 3.3\%$$

Values that will determine systematic uncertainty are given in Appendix E of the standard, in the context of an example, where $X''_b = X''_d = X''_c = 0.5\%$. In the absence of other information, these values are used and result in a value of:

$$X''_Q = \pm 0.87\%.$$

Based on relation (4), the combination of both components for N=30 results in:

$$X_Q = \pm 3.4\%$$

Since systematic uncertainty X''_Q is very low, uncertainty on total flow rate X_Q has the same general value as X'_Q and X'_N (this last component is equal to ± 3% for 30 verticals).

For 20 verticals, the same calculation leads to a value $X_Q = \pm 5.4\%$ for a value of $X'_N = \pm 5\%$, thus the same range.

These calculations show that the predominant uncertainty component in X_Q is the number of verticals (Figure 3), on the condition that elementary flow rates are correctly distributed "each of them must remain close to the total flow rate by 5%, not exceeding 10% of the total flow rate".

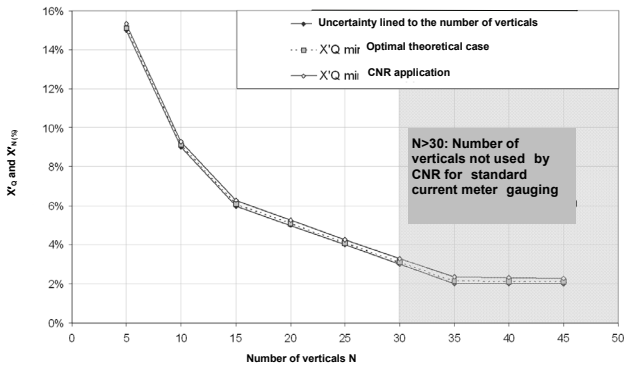


Figure 3. *Random uncertainty X'_Q according to the number of verticals, in the best conditions*

Conclusion on the use of NF EN ISO 748 standard

Predominant parameters in uncertainty calculation are the number and distribution of verticals. The standard indicates that the formula used to estimate X'_Q is valid only if the elementary flow rates are slightly equals and if the elementary flow rates do not exceeds 10% of the total flow rate. This condition can only be verified when gauging is completed.

Uncertainty value of ± 5% (for 20 verticals for example) can only be considered as a minimal value, for gauging completed in good measurement conditions. It appears that uncertainty of ± 7% is more realistic because of general gauging conditions.

Application of standard NF ENV 13005

The method described in NF ENV 13005 standard is based on the measurement process description and its modeling to determine the relation between input quantities (speed, width, etc.) and the output quantity (flow rate).

Propagation of measurement uncertainties

The main hypothesis for the establishment of uncertainty propagation formula is that quantities estimations are not correlated. This hypothesis is realistic if the random error sources are predominant compared to systematic error sources. The application of the law of propagation of uncertainty results in writing:

$$u^2(Q) = \sum_{i=1}^{N} Q_i^2 \left[\frac{u^2(b_i)}{b_i^2} + \frac{u^2(d_i)}{d_i^2} + \frac{u^2(V_i)}{V_i^2} \right] \tag{6}$$

with:

- $u(Q)$: standard uncertainty of standard total flow rate;
- Q_i: elementary flow linked to vertical i;
- N: the number of verticals;
- $u(b_i)$: standard uncertainty of width b_i associated with vertical i;
- $u(d_i)$: standard uncertainty of depth d_i associated with vertical i;
- $u(V_i)$: standard uncertainty of average speed V_i of vertical i.

Estimation of uncertainty components

The calculations are carried out without taking into account the correlations on measured values.

The "one point" speed measurement uncertainty is due to the current meter (trueness, drift, resolution and repeatability) and its position in the flow rate. The representativity of this punctual measurement on a given surface is linked to the number of verticals (elementary surface width affected) and to the number of points per vertical (elementary surface height). Representativity is mathematically known on classic sections (in channel for example), which is not the case for flows in a natural section. This component is often insignificant for medium to strong flows and can be quantified *a posteriori* from the analysis of the discredited field of speeds if this latter can be known with a large number of values.

The uncertainties on width and depth, which have repercussions on wet surface calculation, are estimated by considering that the maximum error linked to the instrument is mainly due to its implementation (for example verticality of the current meter support finger line or pole and horizontal position of the cable for width measurement). Measurement uncertainty components for width and depth associated with the instrumentation itself (such as calibration, resolution, etc.) are then insignificant compared to the implementation. The estimation of maximum errors on width and depth are based on the experience of operators and the analysis of several measurement realised with different instruments.

Conclusion on the use of NF ENV 13005 standard

The calculation of the uncertainty under different conditions leads us to conclude that the number of verticals, as well as their position, is the predominant factor in uncertainty on total flow rate measured by standard current meter gauging.

For all the gaugings under study, the minimum uncertainty is about ± 5% and is obtained for a number of verticals higher than 20.

Comparison of the NF EN ISO 748 and NF ENV 13005 standards

Regardless of the method used, uncertainty estimation methods presented in these two standards presume a permanent flow system. Simplifying hypotheses must be accomplished for the treatment of correlations, which are ignored here insofar as random error sources are predominant compared to systematic error sources.

With a correct distribution of partial flows (between 5 and 10%), implying a sufficient number of verticals (between 20 and 30), the minimum uncertainty values obtained for both standards converge toward a value close to ± 5%.

Because of real gauging conditions with a distribution of partial flows that is not always correct (even with enough verticals), a ± 7% uncertainty will be retained for standard current meter gauging carried out under good conditions.

ADCP gauging

Flow determination

The ADCP flow measurement requires the analysis of the speed field for the whole section. This analysis is done by the movement of a boat equipped with ADCP, going from one shore to the other (Figure 4). During this movement, the device calculates flow at its vertical.

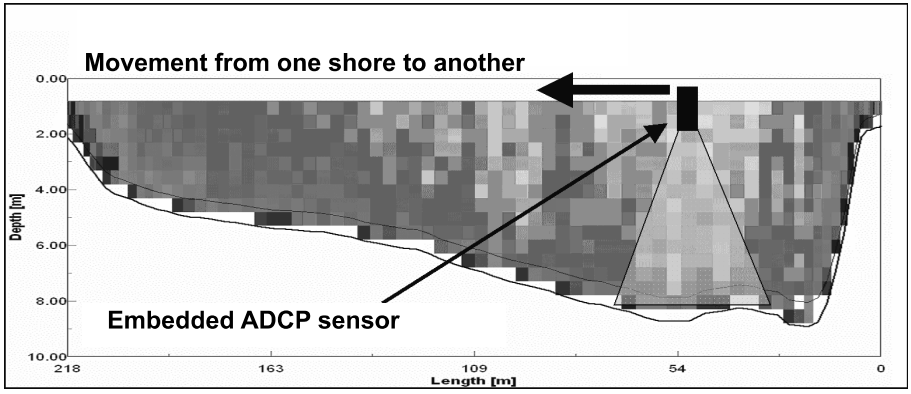


Figure 4. *Across profile of the river to gauge*

If the ADCP manufacturer provides intrinsic characteristics of the device (uncertainty for depth measurement, speed measurement, etc.), it gives no indication on uncertainties associated with flow measurement. In addition, no formalized method exists for the determination of this uncertainty or for making the metrological connection of the instrument to "flow" quantity. Uncertainty estimation is then achieved in two ways: the application of the methods used for current meter gauging and a comparison method with current meter gauging.

Application to the ADCP of the methods used to determine current meter gauging uncertainty

In the absence of an ADCP specific method, a first way is to apply the methods previously seen for current meter gauging.

The establishment of formulae for the estimation of measurement uncertainty, in the case of NF EN ISO 748 and NF ENV 13005 standards, is based on the hypothesis of absence of correlation between measures. This hypothesis is possible for standard current meter gauging insofar as uncertainties linked to systematic errors are insignificant compared to uncertainties linked to random errors ($X''_Q{}^2 << X'_Q{}^2$). For the application of these formulae to the ADCP method, this hypothesis must also be verified.

As ADCP instantaneously examines the complete flow section with a number of elementary measurements much higher than the current meter method, systematic error of instrument can be of importance. The number of elementary measurements with the ADCP is much higher than the current meter since the section is more discredited compared to current meter gauging. Multiplication of measurements increases the effect of systematic errors and consequently correlations. As we can not know the exact impact of these correlations, it is not relevant to apply methods of NF EN ISO 748 or NF ENV 13005 standards to the case of ADCP gauging.

Experimental determination of uncertainty by comparing ADCP measurement to current meter measurements

This estimation of the uncertainty consists of comparing the flow measured with ADCP and measured with a method considered as a reference method, in this case current meter gauging. CNR has carried out these comparative measurements between ADCP and with the current meter methods since acquisition of the ADCP instrument in 1994.

Comparative measurements retained

Values retained for this study were determined in a steady flow regime. They were obtained with two teams carrying out measurement in parallel, one equipped with ADCP, the other with current meter. 19 comparative measurements with ADCP Broadband 1200kHz are used (Figure 5). They were carried out on 11 different sites along the Rhône and its affluents, for flows from 11 m^3/s to 3500 m^3/s, with speeds of 0.1 m/s to 2.5 m/s, over a period of ten years. For each gauging, river bed stability was checked in order for the ADCP measurement not to be biased.

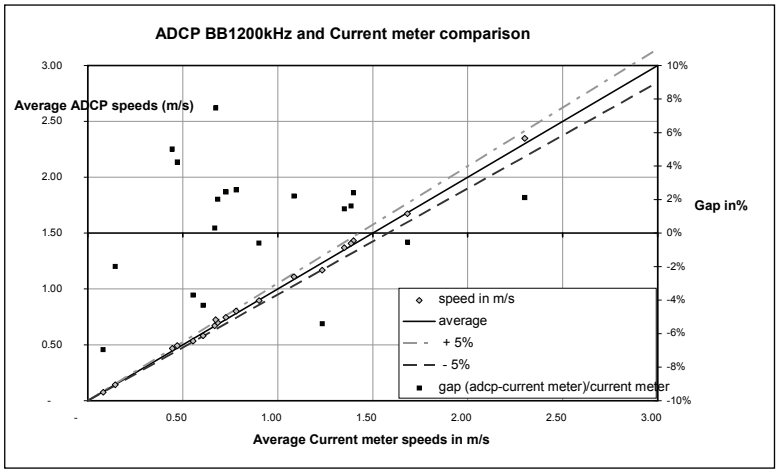


Figure 5. *Comparison between ADCP Broadband 1,200 kHz (BB1,200 kHz) and current meter*[1]

Comparative measurement analysis

The number of analyzed measurements and conditions of obtaining these measurements enable the evaluation of accuracy and reproducibility of the ADCP Broadband 1200 kHz / Current meter comparison, by using recommendations of the NF ISO 5725 standard [1]. Due to its historic anteriority the current meter method is considered as the reference.

The bias between the two methods is equal to the average of deviation between ADCP measurements and current meter measurements, or 0.5% for the considered sample. This value is much lower than the minimal flow uncertainty obtained from current meter gauging (± 5% to ± 7%). As the different teams used different current meters on the period of ten years while they used only one ADCP instrument, this bias should be considered as a systematic error linked to the ADCP. However, it is possible that a residual bias is linked to the interpolation method of current meter gauging. The small bias value could then be explained by an ADCP interpolation bias in the same direction. Interpolation method sensitivity can be studied in post-processing. In equation (7), this bias is considered as a correction that is not applied and is added linearly in expanded uncertainty .

Reproducibility is equal to the standard deviation between ADCP measurements and current meter measurements, or 3.7% for the considered sample. Measurement

1 ADCP and current meter average speeds are very close according to the ratio between total flow and wet section.

uncertainty of a flow obtained with current meter gauging (minimum ± 5%) is significant compared to this value of reproducibility. It is not possible to conclude that the latter is only linked to the reproducibility uncertainty of the ADCP method only.

Uncertainty estimation of ADCP measurement

The application of NF ISO 5725 standard makes it possible to combine bias and reproducibility to estimate expanded measurement uncertainty U, associated with the average deviation between both measurement methods (ADCP and current meter):

$$U = \text{bias} + k \text{ x reproducibility} \tag{7}$$

For $k = 2$, the result $U = 7.9\%$.

This uncertainty is also equal to the quadratic sum of uncertainty components of each of these two methods (ADCP and current meter), or:

$$U = k\sqrt{\left(\frac{U_{ADCP}}{k}\right)^2 + \left(\frac{U_{Currentmeter}}{k}\right)^2} = 7.9\% \text{ for k = 2} \tag{8}$$

where:

– U_{ADCP} is the expanded uncertainty of measurements achieved by the ADCP method;

– $U_{Currentmeter}$ is the expanded uncertainty of current meter gauging.

This combination can be achieved because both methods are completely independent from one another. Knowing the expanded uncertainty $U_{Currentmeter}$, it is then possible to determine the expanded uncertainty $U_{ADCP.}$

Figure 6 represents U_{ADCP} according to $U_{Currentmeter}$ so that the quadratic sum of these two components corresponds to a constant expanded uncertainty to ± 7.9% (k = 2). The curve is therefore drawn in order to respect equation (8).

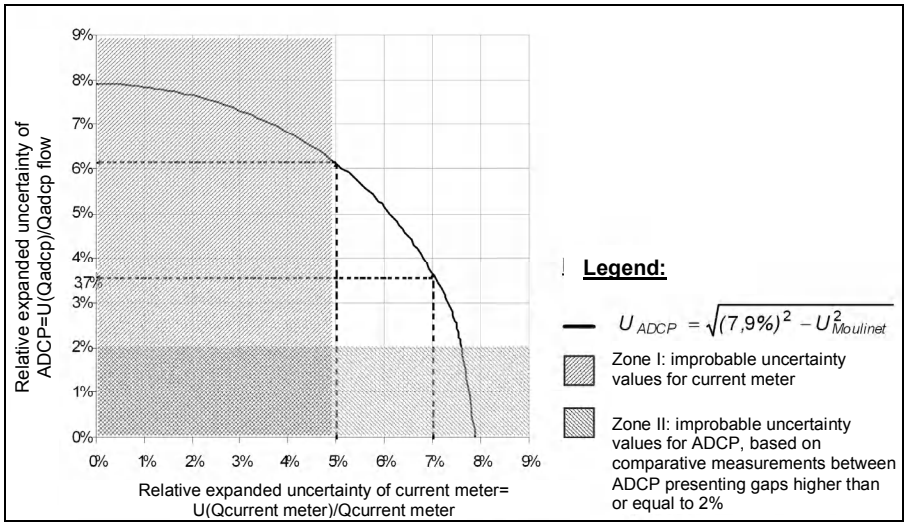


Figure 6. *Uncertainty U_{ADCP} based on uncertainty $U_{currentmeter}$*

Figure 6 indicates that current meter gauging uncertainty (carried out in CNR application conditions) is necessarily lower than ± 7.9%. Similarly, ADCP gauging uncertainty is necessarily lower than ± 7.9%.

From this graphic, we can deduce minimum and maximum uncertainties of ADCP gaugings. From minimum and maximum uncertainties of current meters gaugings, the minimum uncertainty for current meter gauging is between ± 5% and ± 7%. In this way, maximum uncertainty of ADCP Broadband 1200 kHz gauging is between ± 3.7% (for $U_{currentmeter}$= 7%) and ± 6% (for $U_{currentmeter}$= 5%).

We will observe that ADCP gauging uncertainty is the same magnitude of current meter gauging uncertainty, and even lower.

This analysis was carried out on the basis of a sample of 19 comparative measurements. Further measurements will lead to refine the measurement uncertainty ADCP gauging. A similar analysis is in progress to extend these results to other ADCP instruments, notably the 600 kHz ADCP.

Conclusion

For current meter gaugings, the application of NF EN ISO 748 and NF ENV 13005 standards confirms the uncertainty value deduced from the analysis

of experimental data. Expanded uncertainty is about ± 7% for standard current meter gauging carried out under excellent conditions.

It is difficult to determine which standard is the most relevant since they both lead to similar uncertainty values while presenting different hypotheses and gaps for the consideration of certain parameters. To offset these gaps, the exploration of the speed field must be optimized since the number and distribution of verticals are fundamental elements.

For ADCP gaugings, CNR has carried out an experimental comparison with current meter over several years. The application of the ISO 5725 standard, with the current meter method as the reference, helps in the estimation of the flow measurements uncertainty by ADCP to a value between ± 4% and ± 6%. Taking care of quality of the sample used in the comparison (number of measurements and representativity of conditions), this approach leads to more reliable information than the estimations that one could obtain by applying the methods of NF EN ISO 748 and NF ENV 13005 standards.

As a practical conclusion, the operation mode of current meter gauging guaranteeing a ± 7% uncertainty is more difficult to implement than the ADCP operation mode leading to better uncertainty: when it can be used, the ADCP will be chosen in priority over other methods.

References

[1] NF ISO 5725-1 to 6, *Application de la statistique. Exactitude (justesse et fidélité) des résultats et méthodes de mesure*, December 1994

[2] NF EN ISO 748, *Mesure de débit des liquides dans les canaux découverts. Méthode d'exploration des vitesses*, January 2002

[3] NF ENV 13005, *Guide pour l'expression de l'incertitude de mesure*, August 1999

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Influence of the insertion depth on the response of a hot film anemometer in different wind tunnels

Mathias Rohm*, Yvan Cordier-Duperray**

*E+E Austria, **CETIAT France
mathias.rohm@epluse.at - yvan.cordier@cetiat.fr

ABSTRACT: In calibration intercomparisons of hot-film, hot-wire or vane anemometers, significant deviations between different calibration laboratories are observed.
The anemometer's probes influence the flow profile with which velocity should be measured.
Our experiments give an example, how the geometric installation in a wind tunnel can affect the calibration results of anemometers.
A typical hot-film anemometer has been used as test unit and the reference air velocity has been determined by a Laser Doppler Anemometer.
The results of calibrations, according to the insertion depth of the probe in two different wind tunnels, are compared in this paper.

Introduction

To define the sensor characteristic of an anemometer, its output signal is compared to a reference value obtained by a Laser Doppler Anemometer (LDA), Pitot tube etc. in a wind tunnel.

With the worldwide growing importance of quality assurance, the periodic comparison of the characteristics of the instruments to the respective national laboratory or a traceable standard has become very important.

The EUROMET (European Collaboration in Measurement Standards) Project 388 "Intercomparison of anemometers" [1] in 1999 showed that the low uncertainties of the national laboratories are not implicitly assignable to the uncertainties of the calibrated anemometer. In that round-robin test, 3 vane anemometers of different design were calibrated in 15 calibration laboratories to the respective national standard.

Even between the institutes using a LDA as their standard, air velocity deviations up to 10% of reading were reported.

As no commonly unified standard procedure for calibrating anemometers has been achieved, was assumed to be the main reason for these deviations the different procedures for referencing the air velocity were assumed.

The anemometer is influencing the flow ("blockage effect"). Bad positioning of the point of reference can then cause deviations.

Müller *et al*. [2] characterizes this specific problem in detail.

For an 8 mm probe the maximum deviation caused was around 2%.

The larger the anemometer's blocking surface, the more the flow and the calibration should be influenced.

However, comparing the maximum deviations at the EUROMET 388 experiments of the three tested anemometers with their blocking surfaces, the more a negative correlation can be noted.

	type 1	type 2	type 3
Blockage Area [cm^2]	39	26	24
Deviations [% o.r.]	~2	~5	~10

As the wind tunnels used are of a different size and shape, deviations in the geometric placement of the anemometer in the flow profile necessarily occur.

Figure 1 shows the average deviations of the calibrated velocities of each laboratory versus the calibration of the PTB Germany (using LDA as their standard) during the EUROMET-round-robin tests. Requiring that the point of measurement of all calibrations is placed in the centre of the flow-profile, the deviations are plotted over the insertion depth of the anemometer (the distance of the measuring point to the mounted boundary layer) (Figure 1).

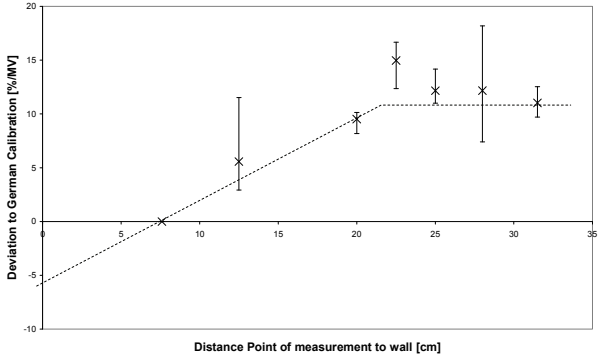


Figure 1. *Deviations of calibrations of different laboratories to German calibration with used insertion-depth of the instrument (Type 3)*

The German calibration was performed in a wind tunnel using an open test section with a diameter of only 15 cm. The graph shows that increasing the insertion depth up to about 22 cm, the deviations to the referenced calibration are increasing approximately linear. From that point on the calibrations remain constant. It is interesting to note that from this point the whole probe is situated in the flow.

Experiments

With the following experiments, we want to show the influence due to different installations of an anemometer in different wind tunnels on the calibration results.

For these tests a commercial hot-film anemometer is used.

The sensor is fixed on a round probe, and covered by a head-construction to protect the sensor. The probe has a diameter of 8 mm so that the influence of the blockage effect is low.

The influences of atmospheric pressure and temperature on the sensor are well known and corrected.

Calibrations are carried out in two totally different wind tunnels, to observe the dependence on installation-effects on the calibration result.

CETIAT France
Type: closed-loop with **closed** test section;
Flow section **square 0.5 x 0.5 m**
Reference: **LDA**

E+E Austria
Type: closed-loop with **open** test section
Flow section **diameter = 0.25 m**
Reference: **LDA**

Both wind tunnels offer a very good parallel uniform flow in the test section with a turbulence factor smaller than 0.4%.

A Laser Doppler Anemometer is used as a non-flow influencing reference.

In order to reduce the influence of the blockage of the probe on the results, a certain calibration method is used for all the tests.

The LDA-reference is measured at the same time, at a certain position in front of the probe, where the flow is not influenced by the probe. The difference of the

uninfluenced flow between the reference positions to the position of sensor is taken into account by a correction-factor.

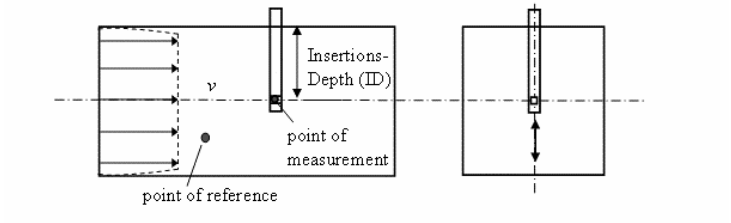


Figure 2. *Definition of insertion depth of an anemometer*

Additionally, calibrations varying the position (insertion depth (ID)) of the anemometer are performed. The position of the anemometer in the flow-profile is changed by an automatic positioning system, so that uncertainties on installations such as the angle between the anemometer and the flow direction are constant.

Calibration points are 1, 5, 15 and 20 m/s of air velocity.

Results

Figures 3 to 6 show the deviations in percent to the measurement-value resulting from the calibrations carried out in E+E's and CETIAT's wind tunnels varying the insertion depth from 2.5 to 25 cm.

Uncertainty estimations are based on the standard procedures of each institute. Not including uncertainties caused by different mounting of the probe (e.g. angle between anemometer and flow direction, uncertainties of insertion depth, etc).

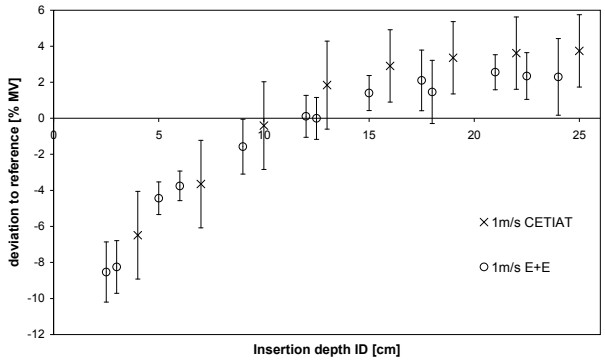


Figure 3. *Calibration of the hotfilm anemometer at 1 m/s*

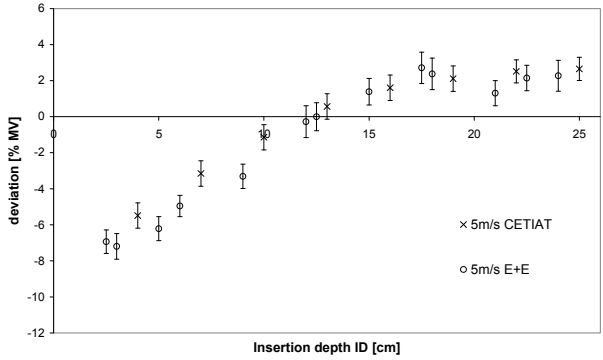


Figure 4. *Calibration of the hotfilm anemometer at 5 m/s*

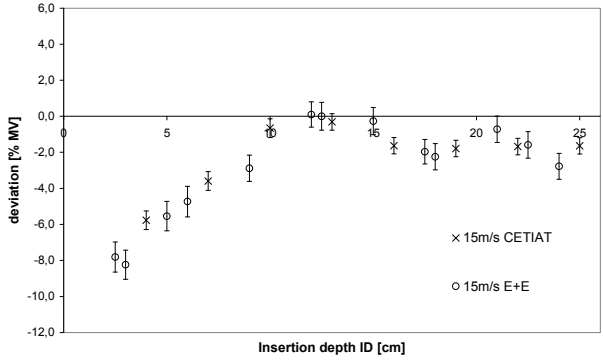


Figure 5. *Calibration of the hotfilm anemometer at 15 m/s*

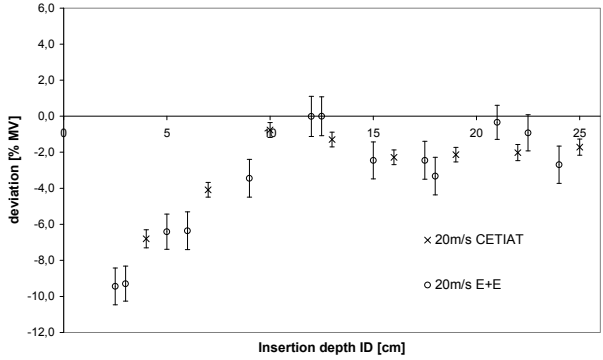


Figure 6. *Calibration of the hotfilm anemometer at 20 m/s*

At all tested velocities, the deviations of the anemometer show a similar characteristic when varying the insertion depth.

A linear increase of the deviations up to a certain insertion depth, of approximately 16 at 1 and 5 m/s and approximately 12 at 15 and 20 m/s.

This insertion depth is defined as "critical insertion depth" ID_{crit}.

From that point, the deviations keep constant and seem to be independent of the insertion depth.

Comparing the results of the two different wind tunnels there is no significant systematic error detectable. Both show the same characteristics relative to the insertion depth. This shows that the relating effect does not depend on the geometry of the wind tunnel. It only depends on the insertion depth of the probe.

Furthermore, the differences of the boundary layer, wall or open test section do not cause significant differences.

Discussion

The discussion of the results needs some theoretical knowledge about the flow situation around the anemometer.

For this, the construction of the anemometer is simplified to a finite circular cylinder on a flat plate.

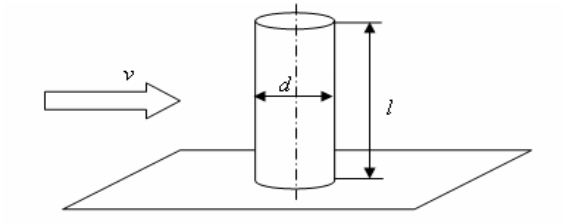


Figure 7. *Circular cylinder with diameter d and an finite length l on a flat plate, in a parallel flow with a velocity of v*

The flow around an infinite cylinder is more or less 2-dimensional, characterized by the Reynolds number Re consisting of the diameter d and the fluid parameters, density ρ and dynamic viscosity μ.

$$\mathrm{Re} = \frac{v \cdot d \cdot \rho}{\mu}$$

The flow around a finite cylinder has, compared to the infinite cylinder, an essential change: the flow is dominated by 3-dimensional effects caused by the flow around the free end and interactions with the boundary layer developed on the flat plate. Additional to the diameter another geometric parameter, the length *l* of the finite cylinder, has an effect on the flow structure.

Kawamura *et al*. [3] characterized the flow around a finite cylinder with varying *l/d* ratio.

Their experiments were performed in a range of *l/d* from 1 to 8 with a Reynolds number of 32,000 (Figure 8).

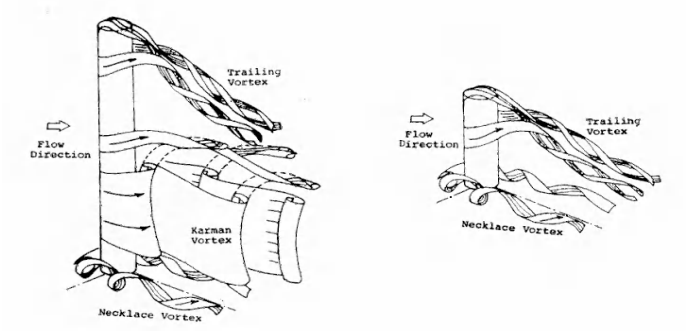


Figure 8. *Flow around a short and a long finite cylinder [3]*

At low l/d ratios, the flow is complex and 3 dimensional, as also shown by Richter and Leder [4] using numerical simulations for *l/d=2* and $Re= 2 \cdot 10^5$ (Figure 9).

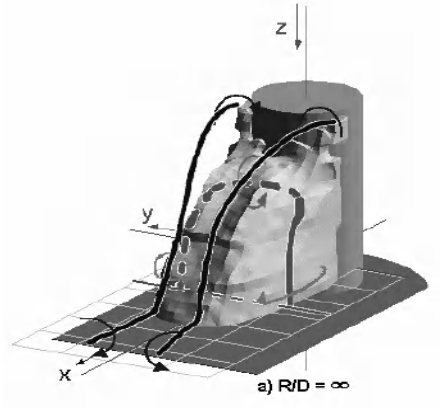


Figure 9. *Characteristic of the flow around a finite cylinder with l/d=2 for Reynolds number* $2 \cdot 10^5$ *[4]*

Figure 9 shows the characteristic formation of the reverse flow area and the interaction of the wall and the flow around the cylinder-head. The structure is dominated by two head-vortices. That pair of vortices are following the back flow area. At the end they are redirected into flow direction, and merged with the vortex system on the ground. This ground vortex system is built by the stagnation point flow at the root of the cylinder and the boundary layer of the wall.

Experiments by Fox and West [5] show that the flow around the free end is influenced up to an *l/d*-ratio of 20.

For higher *l/d*-ratios in the middle of the cylinder, a flow structure similar to the 2D-consideration (Figure 8) occurs. This area breaks the interaction between wall and cylinder-head, the flow structure around the cylinder-head is in this case no longer influenced by the boundary layer.

The knowledge about the flow around a finite cylinder can be used to describe the characteristic with the insertion depth of the anemometer in a wind tunnel (Figure 10).

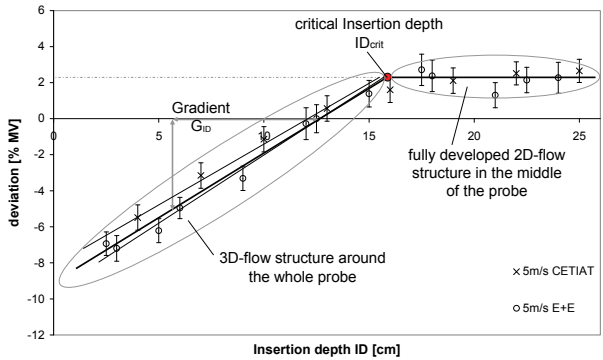


Figure 10. *Approximate specification of the characteristic of the anemometer in a wind tunnel by varying insertion depth*

The characteristic in Figure 10 can be divided into two main parts.

Below the critical insertion depth ID_{crit}, the flow around the whole probe is 3-dimensional. The flow around the cylinder head and the calibration result depends on the insertion depth, showing an approximately linear increase with a gradient G_{ID}.

Above an insertion depth of ID_{crit} a quasi-2D-flow region is developed in the middle of the probe. The result of the anemometer calibration is independent of the insertion depth.

In summary, the characteristic can mainly be described with two parameters: ID_{crit} and $G_{ID.}$

The main parameters of ID_{crit} have already been mentioned for the flow around a cylinder of finite length:

– the length l and the diameter d, representing the geometry of the anemometer;

– δ, the thickness of the boundary layer;

– Re, the Reynolds number, describing the flow structure around the probe:

$$ID_{crit} = \mathrm{f}\left(l, d, \delta, \mathrm{Re}\right)$$

In order to obtain a parameter independent from the geometry of the probe, ID_{crit} is transformed to a dimensionless parameter.

$$\left(l/d\right)_{crit} = \frac{ID_{crit}}{d} = \mathrm{f}\left(\delta/l, \mathrm{Re}\right)$$

The bigger the diameter d of the probe, the deeper the probe has to be inserted into the flow to find a result independent of the insertion depth.

$(l/d)_{crit}$ can be specified for each wind tunnel with its specific boundary layer δ depending only on the Reynolds number Re of the probe.

The influence of the boundary layer could not be investigated yet. Comparing the calibrations of E+E and CETIAT (Figures 3, 4, 5 and 6) (opened and closed test section), the differences between these wind tunnels are low.

Relating to the calibrations carried out by E+E and CETIAT with the 8 mm probe, both wind tunnels offer an $(l/d)_{crit}$ ratio of about 20 for 1 m/s and 5 m/s (Figures 3 and 4), and of about 15 for 15 m/s and 20 m/s (Figures 5 and 6). Most probably the differences of $(l/d)_{crit}$ with the velocities are caused by the Reynolds number Re.

G_{ID} seems to be dominated by the head construction of the anemometer, which depends on the extent to which the flow around the sensor is more or less influenced by the flow around the probe.

Differences may also be caused by the boundary layer and the Reynolds number.

Velocity [m/s]	1	5	15	20
Re []	500	2,600	7,800	$10.4 \cdot 10^3$
G_{ID} CETIAT [%/cm]	~1.1	~0.7	~0.75	~1
G_{ID} E+E [%/cm]	~0.95	~0.8	~0.9	~1
$(l/d)_{krit}$ []	20	20	15	15

Conclusion

In this paper we have shown, that the response of a flow disturbing anemometer depends on its geometric installation in a wind tunnel.

Calibrations of a commercial hot-film anemometer in two geometrically totally different wind tunnels at E+E and CETIAT show excellent agreement if the anemometer is placed at the same insertion depth in the flow profile.

The plot of the deviations of the response against the insertion depth shows the same characteristic in both wind tunnels.

Thus, the type of wind tunnel and the different dimensions of their test sections have no significant influence on the calibration result. The observed characteristic – deviation versus insertion depth – can be interpreted based on the model description of the flow structure around a finite cylinder. A non-dimensional parameter $(l/d)_{crit}$ defines the critical insertion depth for a given wind tunnel. For bigger l/d ratios the response of the anemometer is independent of the distance to the boundary layer (insertion depth).

Certainly the Laser Doppler Anemometer is the state-of-the-art air velocity standard and with a good placement of its measurement point relative to the distance of the unit being tested [2], it allows a highly accurate independent determination of a reference air velocity.

However for comparing calibrations of flow disturbing anemometers, the variations in the geometric setup caused by different wind tunnels and mounting procedures have to be considered.

The described experiments finally give an example of how geometric parameters affect calibrations of anemometers in wind tunnels.

For future intercomparisons of flow disturbing anemometers, it is necessary to consider or preferably to equalize the geometric setup of each calibration. Then it will be possible to obtain lower deviations, and to find a convenient interpretation of the results.

References

[1] H. Lerch: "Intercomparison of Anemometers", *EUROMET Project No. 388*, Draft Final Report, Part 1, 28. January 1999.

[2] H. Müller, Pape, Sodomann, Kampe, Dopheide,"Einsatz der Laser-Doppler-Anemomterie für die Kalibrierung von Anemometern in Windkanälen", 14. Fachtagung, "*Lasermethoden in der Strömungsmesstechnik"- GALA 2006*, Braunschweig, 5-7 September 2006, article 17.

[3] Kawamura, Hrwadam Hibino, Mabuchi, Kamuda, "Flow around a finite circular cylinder on a flat plate", *Bulletin of the JSME 27*, 1984, S. 2430-2439.

[4] F. Richter, A. Leder, "Dreidimensionale Strömungs- und Turbulenzstrukturen im Nachlauf eines Kreiszylinderstumpfes", *12. Fachtagung, "Lasermethoden in der Strömungsmesstechnik"- GALA 2004*, Karlsruhe 7-9 September 2004, article 14.

[5] Fox, T.A.; West, G.S., "The aerodynamic disturbance caused by the free-ends of a circular cylinder immersed in a uniform flow", *Journal of Fluid Mechanics*, Vol. 122, pp.109-111, 1982.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Recent Development of the ITS

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The characterization of relative humidity and temperature probes for use in interlaboratory comparisons

R. Farley[a], W. Rütti[b], M. Stevens[c]

[a] Rotronic Instruments (UK) Ltd., United Kingdom
Crawley, West Sussex RH10 9EE

[b] Rotornic AG, Switzerland

[c] National Physical Laboratory, United Kingdom

ABSTRACT: Evaluating the level of agreement and checking the quality of results to a level coherent with the accredited scope best measurement capability (BMC) is a matter of growing interest, as the claimed specifications and thus the customers' calibration needs are becoming more difficult to substantiate. The technical specifications of high-quality relative humidity sensors are now of the same order as the claimed BMCs.

Establishing the specification for a relative humidity sensor is an almost impossible task without defining the limits of the environment it is to be subjected to. However, in the context of use in the environment of a calibration chamber or generator, the strictest application of the specification can be applied and determined based on the percentage relative humidity (%rh) and temperature cycles involved, and without undue influence related to contamination and degradation issues.

This paper presents and discusses the results of a preliminary investigation on the use of relative humidity and temperature probes in order to establish their level of reproducibility and suitability for use in inter-laboratory comparisons within their specified range of operation. A total of six HygroClip devices with nominal relative humidity ranges of (0 to 100)%rh were tested in two groups. The first group comprises two sensors of different models: "S" for general use and "S3", for meteorology applications, with temperature ranges of (-40 to 85)ºC and (-40 to 60)ºC, respectively. The second group comprises four devices: two as in the first group, plus two model "IC-1", designed for high-temperature applications, with nominal temperature range (-50 to 200)°C.

The instruments of the first group were characterized in the range from 10%rh to 90%rh at a temperature of 23ºC. The second group were tested in the same relative humidity range, but at temperatures of 5ºC, 23ºC and 50ºC. The results of the initial characterization performed both at the manufacturer's own calibration laboratory and at an independent calibration laboratory are presented and discussed in the context of confirming their use as suitable transfer standards.

KEYWORDS: relative humidity, comparison.

1. Introduction

The procedures and results discussed represent the first phase of a series of calibrations of standard Rotronic%rh probes against National Measurements Institutes (NMI) world-wide. This paper provides an overview of the work so far and a initial summary of some of the results obtained. A further series of calibrations are planned over an extended period, with two objectives:

1.1. The primary objective is to establish the best measurement capability of the%rh probes over the temperature range 5 to 50°C and 10 to 90%rh. Of particular interest is the long term stability or drift of the probes, and the data compared between probes that are only used at ambient temperatures (~23°C) and probes used over extended temperature ranges (5 to 50°C).

1.2. A secondary objective is the opportunity for an inter-comparison between the various%rh calibration laboratories. The majority of the international inter-comparisons are dew-point based, so this work may provide a useful means of assessing the equivalence between the respective laboratory%rh standards.

2. Overview of the work so far

2.1. *Factory preparation and first calibration*

The six%rh probes were produced and adjusted according to the Rotronic factory procedures. Each probe was equipped with a connecting cable compatible with a Rotronic HygroLab 2 display instrument. The digital output of each probe could then be displayed, and the output transmitted as an RS232 signal so that measurements could be displayed and logged on a PC running Rotronic HW3 software. The probes were then calibrated in the Rotronic Swiss Calibration Service (SCS) calibration rig, comprising a Thunder 2500ST and MBW chilled mirror hygrometer. The calibration data is shown later marked ‘ROT’.

2.2. *First NMI – Instituto Nacional De Tecnica Aerospacial, Madrid, Spain (INTA)*

Dr Robert Benyon of INTA has provided significant guidance in the definition of the calibration procedures used, and ran the first NMI calibrations between June and November 2005. It is planned that the final calibration will also be performed at INTA, to provide a definitive and independent end to the intercomparison ‘loop’. Calibration data is marked ‘INTA’.

2.3. *Second NMI – National Physical Laboratory, London, UK (NPL)*

Mark Stevens and Richard Gee at the NPL in London performed the third calibration in January and February 2006, fitting in the work despite short notice as a result of a change in the planned ‘route’ for the instruments!

2.4 *Third NMI – Physikalisch Technische Bundesanstalt, Braunschweig, Germany (PTB)*

Dr Nobert Boese must be thanked for the fast handling of our requirements to carry out the fourth calibration in March 2006, providing us with the data in time to include an additional calibration comparison in this initial presentation of data.

3. Procedures

Both the Rotronic SCS laboratory and INTA performed an initial characterization at 23°C, with the%rh cycled 50, 10, 50, 90, 50 five times. This was to provide both an indication of repeatability.

The calibration points performed by all the laboratories are shown in the table below. Two probes were only subjected to calibration at 23°C, with 11%rh points. Four probes were subjected to all 44 calibration points as shown.

Temperature	23°	5°	50°	23°
%rh values	50	50	50	50
	10	10	10	10
	35	35	35	35
	50	50	50	50
	75	75	75	75
	90	90	90	90
	75	75	75	75
	50	50	50	50
	35	35	35	35
	10	10	10	10
	50	50	50	50

The total number of 324%rh calibration points were performed by Rotronic and INTA, NPL and PTB reported a total of approximately 198%rh points.

4. Provisional results

Based on the measurements so far, it has been encouraging to note the agreement between the various laboratories.

Clearly there is a significant amount of data available; some example data plots are included as Appendices. A summary of provisional conclusions are detailed:

– probe repeatability at 23°C can be shown to be within ±0.2%rh over the range 10 to 90%rh (Appendix 1).

– probe hysteresis at 23°C can be shown to be within ±0.35%rh over the range 10 to 90%rh (Appendix 2)

– probe hysteresis at 5°C can be shown to be within ±0.4%rh over the range 10 to 90%rh (Appendix 3)

– probe hysteresis at 50°C can be shown to be within ±0.2%rh over the range 10 to 90%rh (Appendix 4)

The actual values on each graph in the Appendices have been removed to ensure that future calibration tests within the intercomparison are not influenced.

5. Degrees of equivalence

In comparing the calibration data from the various laboratories, Mark Stevens of the NPL carried out a limited analysis of the degree of equivalence at 3 randomly selected calibration points.

50°C 50%rh	**Rotronic**	**INTA**	**NPL**	**PTB**
Rotronic	-	-0.44 ± 1.28	-0.50 ± 1.12	-0.82 ± 1.12
INTA	0.44 ± 1.28	-	-0.06 ± 0.94	-0.38 ± 0.94
NPL	0.50 ± 1.12	0.06 ± 0.94	-	-0.32 ± 0.71
PTB	0.82 ± 1.12	0.38 ± 0.94	0.32 ± 0.71	-
Reference value	0.44 ± 1.00	-0.01 ± 0.80	-0.06 ± 0.50	-0.38 ± 0.50
5°C 50%rh	**Rotronic**	**INTA**	**NPL**	**PTB**
Rotronic	-	0.02 ± 1.80	-0.11 ± 1.58	-0.57 ± 1.58
INTA	-0.02 ± 1.80	-	-0.13 ± 1.12	-0.59 ± 1.12
NPL	0.11 ± 1.58	0.13 ± 1.12	-	-0.46 ± 0.71
PTB	0.57 ± 1.58	0.59 ± 1.12	0.46 ± 0.71	-
Reference value	0.16 ± 1.50	0.19 ± 1.00	0.06 ± 0.50	-0.40 ± 0.50
23°C 50%rh	**Rotronic**	**INTA**	**NPL**	**PTB**
Rotronic	-	0.50 ± 1.08	0.48 ± 0.78	0.36 ± 0.78
INTA	-0.50 ± 1.08	-	-0.02 ± 1.03	-0.14 ± 1.03
NPL	-0.48 ± 0.78	0.02 ± 1.03	-	-0.12 ± 0.71
PTB	-0.36 ± 0.78	0.14 ± 1.03	0.12 ± 0.71	-
Reference value	-0.34 ± 0.60	0.17 ± 0.90	0.15 ± 0.50	0.03 ± 0.50

By treating each institute's measurements as an independent distribution, simple probability calculus can quantify the bilateral demonstration of equivalence, and the confidence associated with equivalence to approximately 95% confidence level. In layman's terms if the degree of equivalence is within the estimated confidence, the uncertainties quoted by the institute can be justified. The uncertainties in this evaluation are assumed to be uncorrelated.

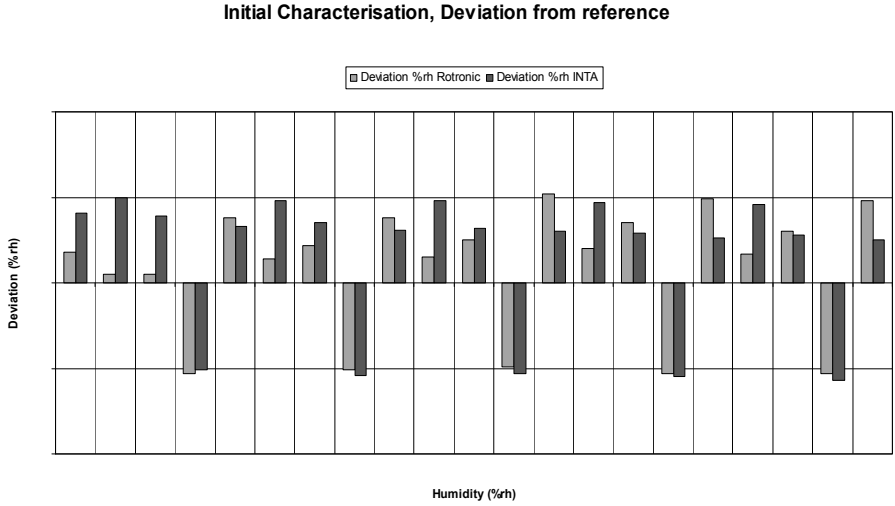
6. Conclusions

It is difficult to draw definitive conclusions based on the limited number of calibrations and the limited time available so far to carry out a full analysis of the data. However, the reasonable agreement between the respective calibration laboratories and the consistent response of the probes encourages continuation of the project.

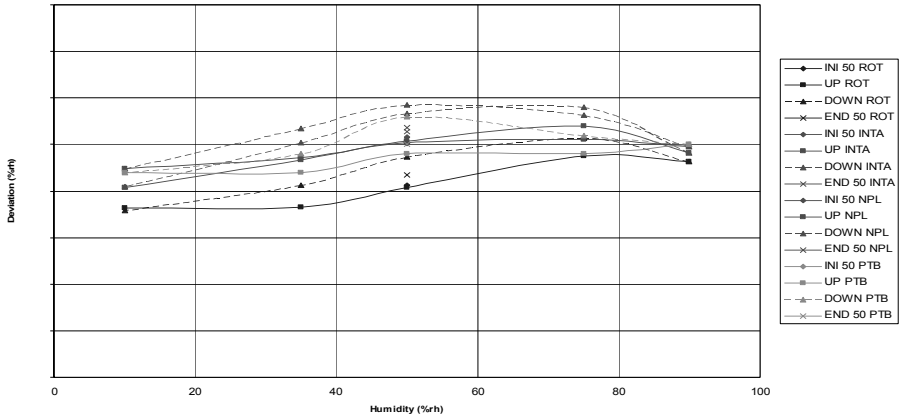
The probes have so far been exposed to a large number of calibration points, over a reasonable range of temperatures, over a period of approximately one year. Drift is most notable at high humidity and temperatures, and as can be seen in Appendix 4, comparing the INTA data from summer 2005, with the PTB data from spring 2006, shows stability within ±0.5%rh for probes used over a temperature range of 5…50°C.

During spring/summer 2006, further calibrations will be performed, with the final set carried out at INTA to close the first 'loop'. A final analysis of all the data will be performed and the results presented at a future symposium.

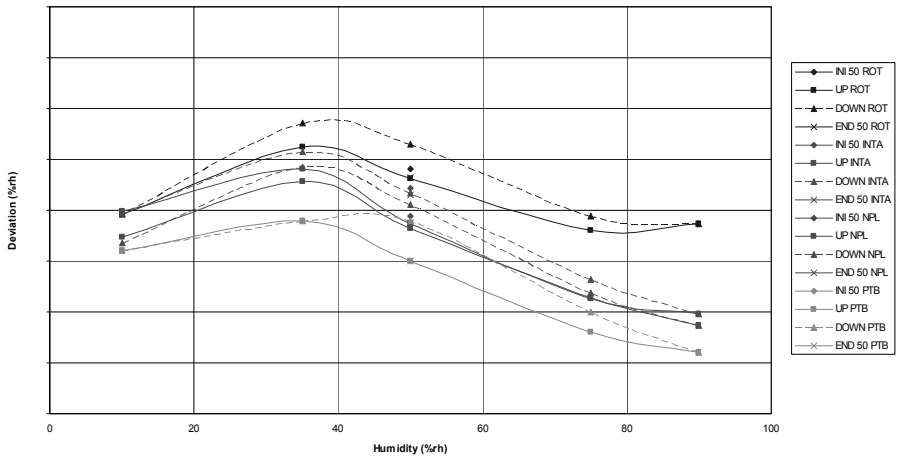
Appendix 1 – Sample plot demonstrating repeatability at 23°C (Y Axis range 4%rh)



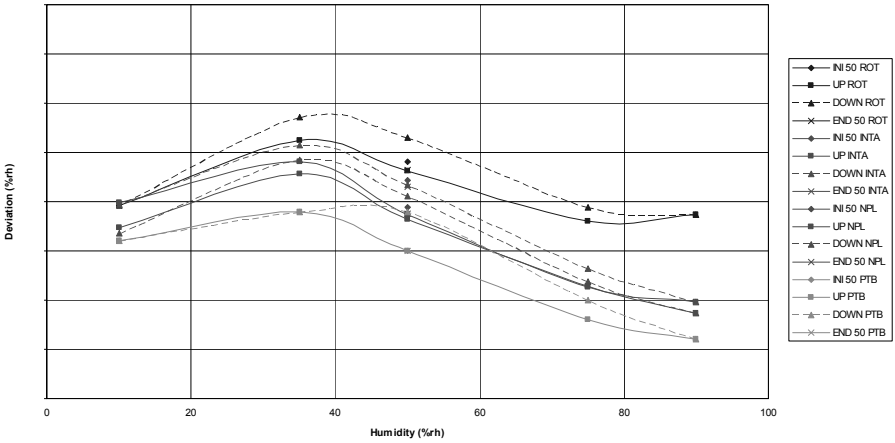
Appendix 2 – Sample plot demonstrating hysteresis and stability at 23°C (Y Axis Range 4%rh)



Appendix 3 – Sample plot demonstrating hysteresis and stability at 5°C (Y Axis Range 4%rh)



Appendix 4 – Sample plot showing probe hysteresis and stability at 50°C (Y Axis Range 4%rh)



Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

High temperatures: new fixed points and thermodynamic temperature

M. Sadli, B. Rougié, F. Bourson, S. Briaudeau

Conservatoire National de Arts et Métiers, Institut National de Métrologie
LNE-INM/Cnam, 61, rue du Landy 93210 La Plaine Saint-Denis, France

ABSTRACT: ITS-90 temperature measurements above the copper point are performed by extrapolating the spectral radiance from a fixed point (Ag 961.78°C, Au 1064.18°C or Cu 1084.62°C) using the Planck law for radiation. This extrapolation will mathematically induce an increase of the uncertainty and will also be affected by the base parameters of the extrapolating instrument: the pyrometer.
To help avoid this inconvenience in a future temperature scale or a different manner for putting the kelvin into practice, our institute has worked for several years now on two parallel paths: the thermodynamic temperature measurement with a direct traceability to the cryogenic radiometer, and the study of high-temperature fixed points based on metal(carbide)-carbon eutectics. These two paths are coming today to a crossroads where we should contribute to the international effort aiming to assign thermodynamic temperatures to a set of high-temperature fixed points by 2011-2012 [1] toward a new international temperature scale.
In this paper, we will review the up-to-date progress of our work in the field of high temperature fixed points, thermal optimization of the furnace used for implementing these fixed points and the new means devoted to thermodynamic temperature measurement in our institute.

Introduction

Studies conducted for several years on high-temperature fixed points [2-4] tend today to concentrate on the most promising points from the half a dozen points between 1153 and 2474°C: Co-C 1324°C, Pt-C 1738°C and Re-C 2474°C. The first objective is the development and study of the best possible cells: the materials used are very pure, the design of crucibles has evolved and the filling method was improved to minimize the risks of furnace contamination. Indeed, the quality of the melting plateau (small melting range, long duration, good repeatability and good definition of phase transition) of eutectic points depends on the purity of the material (metal, carbon). It also depends on the thermal conditions in which the cell is implemented. That is why improving the temperature uniformity of the HTBB 3200PG furnace is one of the important tasks of our work.

Simultaneously, the development and improvement of a measurement method for the thermodynamic temperature [5] enables us today to hope for a typical uncertainty of about 0.3 K and 2000 K. A tuneable Ti-Sa laser source, operating in the spectral range from 700 to 950 nm will allow the characterization of the spectral responsivity of a linear pyrometer and, when associated to a silicon trap detector

traceable to the cryogenic radiometer, to determine thermodynamic temperature in conditions not seen to this day in our institute.

High-temperature fixed points

The "radiation thermometry" working group of the Comité Consultatif de Thermométrie has been working for two years to coordinate studies involving high-temperature fixed point with the goal of completing a research plan gathering many national metrology institues. These studies should possibly lead to the definition of a new international temperature scale above 1,000°C within five years, or at least, to the possibility of implementing these high-temperature fixed points as a means of disseminating thermodynamic temperatures or for the p of the kelvin.

CCT-WG5 project on high-temperature fixed points

This project is coordinated by the NPL [1] and is divided into six distinct workpackages:

– Workpackage 1: assessing long-term stability and robustness of high-temperature fixed points (WP1 – LNE-INM France) – 2008-2010.

– Workpackage 2: defining best cell construction methodologies to obtain a repeatability of 100 mK and construction of a series of "best-achieved" cells (WP2 – NMIJ Japan) – 2008-2011.

– Workpackage 3: understanding the interactions furnace/fixed point/pyrometer and quantifying the associated uncertainties (WP3 – NMIJ Japan) – 2008-2011.

– Workpackage 4: circulating cells and measuring their thermodynamic temperatures to compare and Improve methods for absolute thermometry (WP4 – PTB Germany) – 2009-2010.

– Workpackage 5: assigning, in a multilateral way, absolute temperatures to a series of fixed points of reference based on cells of best possible quality (WP5 – NPL United Kingdom) – 2010-2011.

– Workpackage 6: recommending to the CCT a way of revising ITS-90 or a new mise-en-pratique in order to improve the scale for at high temperatures (WP6 – NPL United Kingdom) – 2011.

Fourth generation of eutectic cells and new fill method

In the CCT-WG5 project's first workpackage, led by LNE-INM, the long-term stability of eutectic cells of Co-C, Pt-C and Re-C is evaluated. "Ageing" cycles of approximately 50 hours are planned for certain cells which, compared before and

after ageing with pilot cells should make it possible to quantify the drift of these cells, if any. LNE-INM must in particular develop a batch of the 4 Co-C cells. This task requires refining of the design and characterization of our cells, in order to optimize the quality of the cells (quality of the plateaux, reproducibility, reliability, robustness etc.). Accordingly, we developed a 4th generation of cells (Figure 1) for considerably reducing the number of filling steps (melting – return to ambient temperature) ranging from 5 to 10 so far.

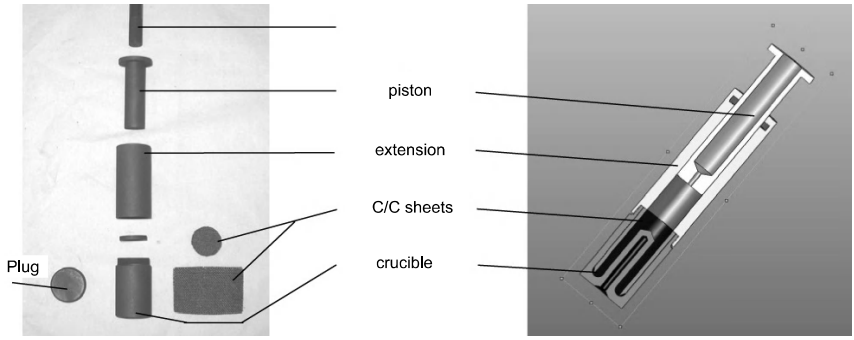


Figure 1. *Details of the fourth generation eutectic cells and associated filling series*

This new generation of cells has the following advantages:

– limitation of pollution risk during the filling by the complete closing of the cell during melting-freezing cycles;

– reproducible filling rate, by the removal of meniscus formed on the metal surface;

– visual monitoring of melting temperature by piston horizontal displacement (which limits the risk of overheating the eutectic mixture);

– better temperature uniformity inside the cell due to C/C sheets (melting plateaux have smaller melting ranges);

– sample for chemical analysis of the impurity content is obtained during filling.

This method was implemented and validated on several Co-C cells, but unfortunately the graphite used for crucible development turned out to be too fragile and could not absorb the strong thermal expansion of Co-C complicated by the tendency of the alloy to stick to the graphite walls. Several cell breakages occurred

during the implementation of the cells. The use of high expansion factor graphite should solve this problem.

One of the innovations of this cell generation is the use of carbon-carbon composite sheets, which are rolled around the ingot. Their efficiency was proven by the improvement of the melting plateaux [6] by locally uniformizing the temperature. This was observed on Co-C cells with and without C/C sheets as shown in Figure 2. The first plateau was obtained in June 2006, with the third generation of cells (cell ume1 developed with our Turkish institute colleagues [7]), before improvement of the HTBB temperature profile. The second was performed with the fourth generation of cells (with C/C sheets) after the optimization of the temperature profile.

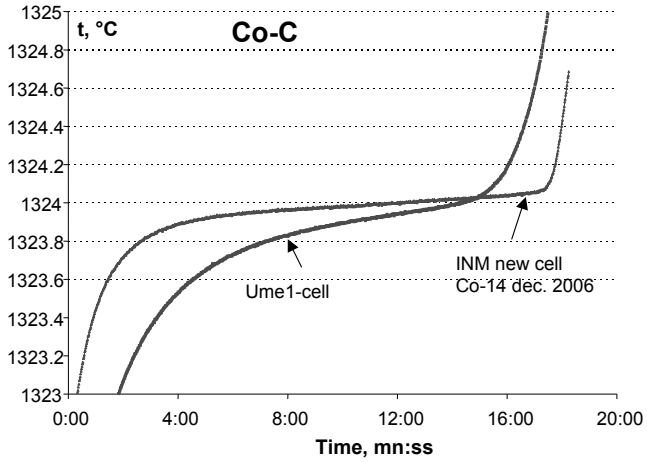


Figure 2. *Evolution of a standard melting stage of two Co-C cells*

In parallel to the study completed on the cells, the temperature distribution of the HTBB furnace, used for implementation of the cells, was improved.

Thermal optimization of the Vega HTBB 3200 PG furnace

The HTBB furnace was not originally designed for the implementation of eutectic points, but as a variable temperature blackbody source. This furnace, open to one end and cooled at the other, presents a bell-shaped temperature distribution. This implies that the cell in any case is in an large temperature gradient and that its positioning will affect the direction and amplitude of this gradient.

However, this furnace temperature uniformity can be improved by efficiently arranging the resistive rings which form it.

That is why, in cooperation with the furnace manufacturer, we developed a relative measurement device for the resistance of the rings (Figure 3). This device allows to individually measure the "resistivity" of rings and to classify them according to their thermal dissipation and to "model" the temperature profile by increasing dissipation at the edges compared to the center.

Figure 3. *Resistance measurement device for HTBB furnace rings*

The studies carried out in collaboration with the PTB [8] show that it is possible to significantly improve the temperature profile and to observe the positive effects on the melting plateau quality.

The temperature profile measurement is performed using a tilted pyrometer with an angle of 5-6° with a frontal lens equipped with a diaphragm of 8-10 mm (Figure 4). Thus, by avoiding vignetting at the furnace entrance, a central furnace zone of approximately 20 cm can be scanned to measure axial temperature distribution on the radiating cavity or the cell carrier.

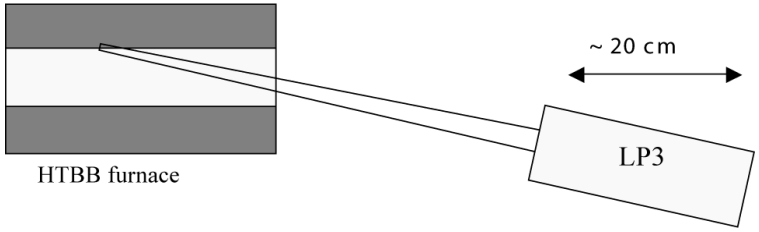


Figure 4. *Layout diagram of the pyrometer to measure the temperature distribution in the HTBB furnace*

It is thus possible to modify the rings positioning and to measure the consequences in terms of temperature profile. This makes it possible to know beforehand the temperature distribution in the furnace and to deduce, as can be seen in Figure 5 the ideal place to position the cell during its implementation in the furnace for the best thermal conditions.

We observe that temperature distribution worsens and that this maximum distribution is pushed to the back of the furnace when the temperature increases.

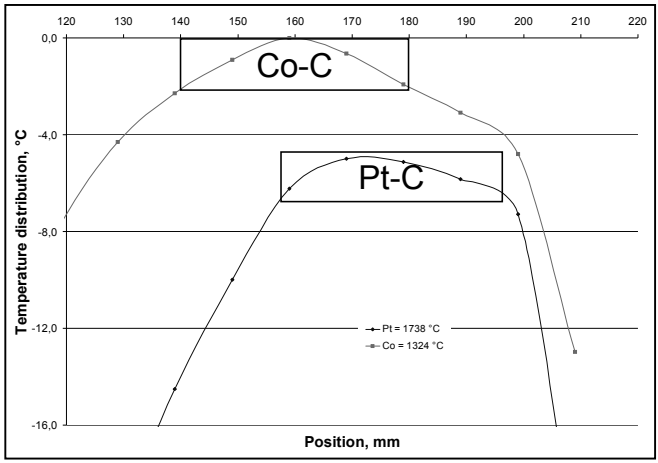


Figure 5. *Temperature profile in the HTBB furnace after ring rearrangement. A 2°C gradient can be obtained along the 40 mm of cell length*

Lastly, in Figure 6, we can observe the effect of furnace ring rearrangement, and the improvement of the temperature gradient on the length and quality of the Pt-C point melting plateaux.

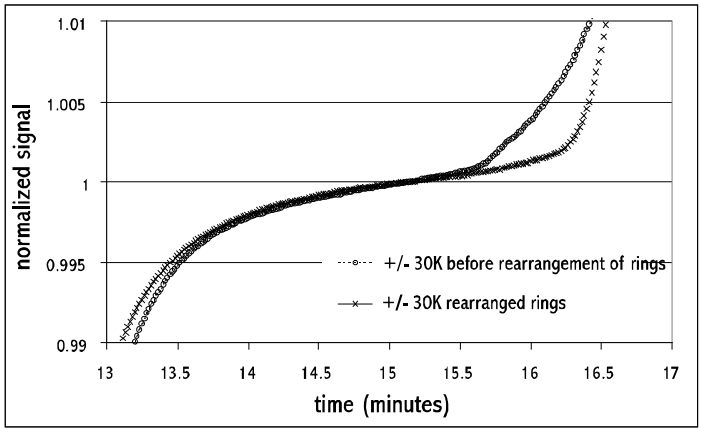


Figure 6. *Effect of rearranging HTBB rings on the duration and flatness of the plateaux [8]*

It is currently possible to completely characterize our HTBB heaters and prepare ring arrangement adapted to the eutectic points studied. The temperature profile will be measured at each temperature used to position the cell where temperature distribution does not noticeably affect measurement results of the melting points.

This *a priori* characterization, although long and tiresome, has a real advantage for obtaining better repeatability and plateaux with small melting ranges.

Thermodynamic temperature measurement

The interest of a thermodynamic measurement

The Planck's law for radiation allows to obtain by extrapolation from an ITS-90 fixed point, the temperature of a blackbody. However uncertainty associated with the fixed point T_{pf} value is propagated at any other temperature with a coefficient $(T/T_{pf})^2$. We must add uncertainties of the comparison instrument. This is the main reason motivating the development of a thermodynamic temperature measurement.

The radiometric method

We can distinguish three fundamental steps in pyrometric measurement of temperature by radiometric method: measuring the radiance, then its spectral density and finally establishing the link with the temperature.

The methodology developed at LNE-INM

The measurement principle has already been implemented at the institute and presented in 2004 at Tempmeko [5]. The radiance of a homogenous and monochromatic source (integrating sphere lit by a laser) is measured with a detector with a responsivity that is directly calibrated by comparison with the national responsivity reference. The radiance of a blackbody is measured by comparing this source to the blackbody through a spectro-radiometer which has very finely measured spectral responsivity.

The Planck's law establishes the connection with temperature by taking into account effects of source dimension and characteristics of the blackbody (emissivity and temperature drop in the walls of the crucible). This measurement principle follows the three-step radiometric measurement.

The current process

Differences

In the current process, the spectro-radiometer is replaced by a filter pyrometer. To measure the spectral responsivity of this pyrometer the radiance source wavelength must explore the field of pyrometer responsivity. The laser lighting the source was replaced by a titanium sapphire laser, simultaneously more powerful (700 mW) and tuneable between 700 nm and 950 nm.

The pyrometer

The pyrometer (LP3 IKE brand) has excellent qualities such as spectral selection, small size-of-source effect, and stability, so that the measured responsivity should remain stable for several days or even several weeks.

It is equipped with 750 nm, 650 nm, 850 nm and 950 nm filters. The 850 nm filter is used for thermodynamic temperature measurement.

The monochromatic source

The laser is injected through a multimode fiber in a 50-mm diameter Spectralon sphere. The exit port is delimited by an aperture of 6 mm. Radiance of the sphere is controlled by a detector. The power of the laser can then be stabilized. Due to the high power of the laser, an acousto-optical crystal accepting a power higher than the one with liquid crystals had to be used as the active control element. The test of the new control system shows a stability of 5.10^{-5}.

It also should be noted that two meters of fiber are dipped into an ultrasound bath to mix its propagation modes and thus, by the averaging effect, reduce the "speckle" noise at the sphere's exit.

The radiancemeter

Since radiance is the quotient of a flux by a geometric extent, the radiancemeter consists of a detector measuring a flux and of two 6-mm apertures (one is the exit port of the sphere) maintained at a fixed distance thus determining the geometrical conditions. Aperture surface and the distance separating them are measured. Detector sensitivity (A/W) is also calibrated by comparison to the cryogenic radiometer which is the national standard of detector sensitivity.

In the new radiancemeter version, the distance between the two diaphragms was doubled (600 mm). Inter-reflections between the two apertures are still under test, but the hope is that their effect is negligible.

The cycle of measurement

The first stage of the cycle consists of measuring the pyrometer responsivity. In order to do this, the laser wavelength explores pyrometer responsivity one wavelength at a time. For each wavelength, sphere radiance is measured with the radiancemeter and the pyrometer signal placed in front of the sphere is recorded.

In a second stage the radiancemeter measures the radiance of a blackbody.

Uncertainties

Uncertainties presented are the result on one hand of the evaluation made with the current process, and on the other hand of the LP3 pyrometer characterization [9]. They are expressed as standard uncertainties. Temperature values were calculated for a blackbody at 2000 K and for a wavelength of 850 nm.

	$10^{-4}.\ \sigma_X/X$	σ_T (mK)
Homogeneity of the source	5	100
Stability of the source	0.5	10
Aperture surface	7	140
Distance source-detector	5	100
Detector responsivity	10	200
Radiancemeter interreflections	2	40
Size-of-source effect	1	20
Blackbody emissivity	2	40
Global		288

Table 1. *Uncertainties on thermodynamic temperature measurement*

Conclusion

The LNE-INM is committed in the field of high temperatures with the goal of implementing new references which should, in the long term, considerably reduce uncertainty for the highest temperatures.

We have presented the current studies on the development of metal(carbide)-carbon eutectic cells and the optimization of the existing thermal means as well as the recent developments on thermodynamic temperature measurement by radiometric method.

Our implication in European and international projects in these fields enables us to improve our expertise and to actively contribute to the changes emerging in the field of thermometry, particularly at the highest temperatures.

REFERENCES

[1] G. Machin, P. Bloembergen, J. Hartmann, M. Sadli, Y. Yamada "A concerted international project to establish high-temperature fixed points for primary thermometry", *Tempmeko 2007 Proceedings*, to be published.

[2] M. Sadli, M. Fanjeaux and G. Bonnier "Construction and implementation of a set of metal–carbon eutectic fixed points" *Proc. Temperature: Its Measurement and Control in Science and Industry*, 7 (Chicago, USA) AIP Conf. Proc. 684, 267–72 (2002).

[3] M. Sadli, F. Bourson, M. Fanjeaux, S. Briaudeau, B. Rougié B and G. Bonnier "Study of metal–carbon eutectic points: from construction to temperature determination" *Proc. 9th Int. Symp. on Temperature and Thermal Measurements in Industry and Science* (Tempmeko) (Dubrovnik) (Zagreb, Croatia: Laboratory for Process Measurement, Faculty of Mechanical Engineering and Naval Architecture) pp 611–6 (2004).

[4] M. Sadli, J. Fischer, Y. Yamada, V.I. Sapritsky, D.H. Lowe and G. Machin "Review of metal–carbon eutectic temperatures: proposal for new ITS-90 secondary points", *Proc. 9th Int. Symp. on Temperature and Thermal Measurements in Industry and Science* (Tempmeko) (Dubrovnik) (Zagreb, Croatia: Laboratory for Process Measurement, Faculty of Mechanical Engineering and Naval Architecture,) pp 341–7 (2004).

[5] S. Briaudeau, B. Rougié, M. Fanjeaux, M. Sadli, G. Bonnier "Thermodynamic temperature determination in high temperature range at BNM-INM", *Proc. 9th Int. Symp. on Temperature and Thermal Measurements in Industry and Science* (Tempmeko) (Dubrovnik) (Zagreb, Croatia: Laboratory for Process Measurement, Faculty of Mechanical Engineering and Naval Architecture), pp 119-125 (2004).

[6] Yamada Y, Khlevnoy B, Wang Y, Wang T and Anhalt K, "Application of metal (carbide)–carbon eutectic fixed points in radiometry" *Metrologia* 43, S140–4 (2006).

[7] M. Sadli, Ö. Pehlivan, F. Bourson, A. Diril, K. Ozcan, "Collaboration on Co-C Eutectic Fixed Point Construction and Characterisation between UME and LNE-INM", *Tempmeko 2007 Proceedings*, Int. J. Thermophysics, DOI 10.1007/s10765-008-0470-5.

[8] M. Sadli, K. Anhalt, , F. Bourson, S. Schiller, J. Hartmann "Thermal Effects in HTBB-3200pg Furnace on Metal-Carbon Eutectic Point Implementation", *Tempmeko 2007 Proceedings*, Int. J. Thermophysics, DOI 10.1007/s10765-008-0479-9.

[9] F. Bourson, M. Fanjeaux, B. Rougié et M. Sadli "Caractérisation et étalonnage d'un pyromètre de référence pour la mesure des hautes températures" *Revue Française de Métrologie* n° 6, Volume 2006-2, p 3-8.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Implementation of a polyvalent measurement set-up for temperature and humidity used to characterize thermal and climate chambers: presentation of results

Roland Platteau

Laboratoire Belge de thermométrie (LBT)
Rue du bosquet 7, B-1348 Belgium

ABSTRACT: In order to guarantee a good level of uncertainty, a low long term drift and a great flexibility of use, LBT developed its own multi purpose data acquisition systems (20 temperature and/or relative humidity sensors etc.) used daily in different fields of application like thermal or climate chambers mappings, autoclave validation, thermocycler (PCR) characterization, etc.
We will consider the reasons for this development, the requirements for the hardware and the software, the specifications of the selected hardware and software, the results of measured experimental and field data recorded over more than 5 years.

Development

Why such a development?

Customers have been asking us to control incubators with ±0.5°C specifications.

Like any good calibration laboratory, we wanted to respect the metrological 4 to 10 ratio but for us it is always the 10 ratio! So we needed to find on the market a piece of measurement equipment (instrument and sensors) able to guarantee a combined uncertainty around ±0.05°C, including long term drift and residual error after correction.

In addition to that we also thought about a multipurpose instrument able to measure PT100 and thermocouples.

We could actually not find on the market such a standard instrument.

Equipment requirements

Quality of the measurement

We focused on 3 items important to us:

– residual error after correction: it should be low thanks to an efficient correction feature of the equipment;

– uncertainty: measurement equipment should be very stable to minimize type A uncertainty, resolution should be low to minimize type B uncertainty, the calibration laboratory in charge of the calibration should show very low uncertainties;

– long term drift: instruments and sensors should remain stable over years in order to minimize type B uncertainty.

NF X 15 140 standard

Our data acquisition equipment has to match the NF X 15 140 requirements, so we looked at the following items:

– the equipment should at least have 15 temperature channels for chambers > 2 m^3. We will choose 20 temperature channels to get some spares;

– the equipment also has to measure ambient conditions (temperature and relative humidity), power supply voltage and power supply frequency;

– the equipment will also have to measure dewpoint.

Hardware requirements

Here is the list of the main items we were looking for during the equipment selection.

Liability of the electronics:

– short term stability;

– long term stability.

Temperature inputs – direct input:

– PT25;

– PT100 (4 wires for optimal measurements and no concern if we want to extend the cables, test current selectable);

– thermocouples with internal or external cold junction compensation;

– thermistors.

Humidity input – RS232 input:

– cooled mirror dewpoint hygrometer.

Other inputs:

– 4-20 mA (mainly pressure);

– mV.

Sampling frequency:

– target: 1 scan of all the inputs per minute;

– maximum sampling frequency available;

– what will be the loss in accuracy when increasing the sampling frequency?

Software requirements

Recorded data structure:

The recorded data should allow the user to control manually the raw and processed data. We should find the following in the recorded data files:

– date and time;

– number of the record;

– raw data (ohm, μV, mA,...);

– scaled data (°C, Pa, F0,…);

– set-up of every input channel;

– set-up of the data acquisition equipment.

Error correction feature

In order to minimize the residual error and depending of the input to process we should have the possibility to apply one of those different corrections curves:

– pair of points;

– linear;

– polynomial;

– IEC-751;

– ITS-90.

Carefree software requirements:

– easy to install (simple sub-directory copy);

– no shared DLLs;

– compact, small disk space;

– very basic but user friendly interface (what matters is the quality of the measurement, not animation on the display!). One mouse click to choose the set-up, one mouse click to choose the data file name and one mouse click to start recording;

– ASCII files for set-up, *.ini and data files. Thus we can easily read those files with a single text editor.

Selected equipment

In order to reach the 0.05°C combined uncertainty we choose the following equipment:

Hardware

7 ½ digit Keithley digital multimeter

The important specifications for us are:

– resistance input:

4 wire measurement.

1 kOhm resolution: 100 μohm.

1 year accuracy: 50 ppm of rdg+ 2 ppm of range.

– DC voltage input:

100 mV resolution: 10 nV.

1 year accuracy: 37 ppm of rdg+ 9 ppm of range.

Keithley scanner

The important specifications for us are:

– 4 wire measurements.

– 20 channels minimum.

PT100

– Working range: -40 +230°C.

– 1/10 din.

– Alpha = 0.00385 Ohm/Ohm/°C.

We cycled sorted PT100 to release their mechanical stress and to age them.

RH transmitter

This sensor is used to monitor ambient conditions.

Cooled mirror dewpoint hygrometer

This sensor is used for climate chamber characterization. We measure both temperature and dewpoint and then we calculate the relative humidity for every temperature sensor location.

Software

We developed the data acquisition software in house. The software development language is DELPHI. We chose DELPHI because LBT had a TURBO PASCAL function library available and also because this language matched our previously mentioned requirements.

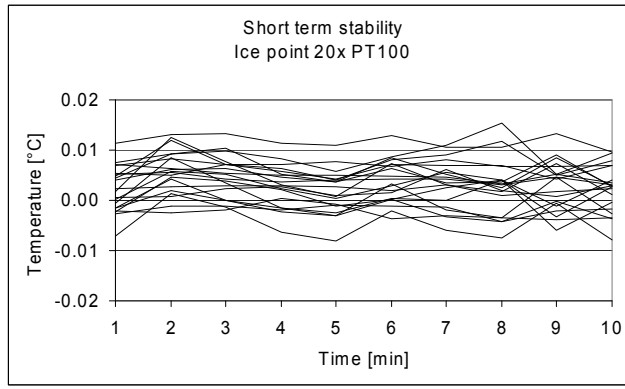
Actual data

PT100 calibration

We will only consider the PT100 calibration since it will better show the performances of the LBT measurement equipment.

Short term stability

The best way to show short term stability is to measure the 20 PT100 in an ice point so that measurements are not influenced by the bath stability. As you can see on the graph below, the 20 PT100 are within a 0.03°C range.



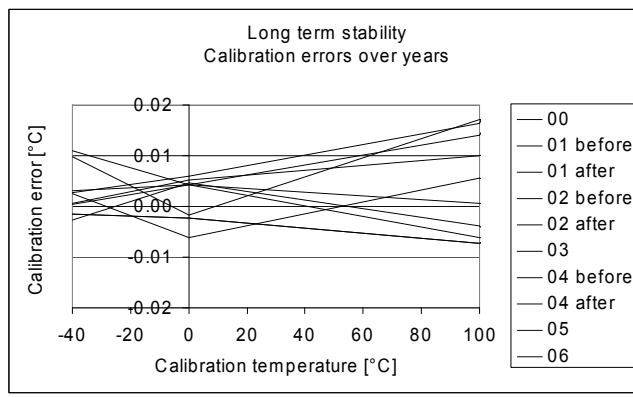
Below are short term stability figures of the recorded data:

Worst stability:	0.006°C
Worst homogenity:	0.011°C
Worst standard deviation:	0.004°C

The stability shows the quality of the electronics, the residual noise. The homogenity shows the quality of the correction algorithm and the polynomial applied to the measured raw data. It also includes the homogenity of the ice bath.

Long term stability

The graph below shows calibration data of 1 PT100 out of the 20 PT100 sensors over the range –40 +100°C from 2000 to 2006. “Before” and “after” next to the year are data before and after adjustment.



Below are long term stability figures of the recorded data (7 years):

Max drift @ -40°C:	0.012°C
Max drift @ 0°C:	0.007°C
Max drift @ +100°C:	0.021°C

Uncertainties

The calibration certificate gives us the residual error and the uncertainty of the PT100 sensor.

We have to combine the certificate uncertainty with resolution, drift and self heating to obtain the actual uncertainty of the PT100 sensor.

Uc from the certificate of calibration (k=2)	0.012°C
Resolution	0.003°C
Drift	0.021°C
Self heating	0.015°C

The new combined uncertainty will be:

Uc = 0.032°C (k=2)

However, it remains a residual error after calibration on the certificate:

maximum residual error: 0.023°C

We will combine this residual error with the uncertainty to avoid having to take the error into account during measurements. This new uncertainty will be:

Uc = 0.040°C (k=2)

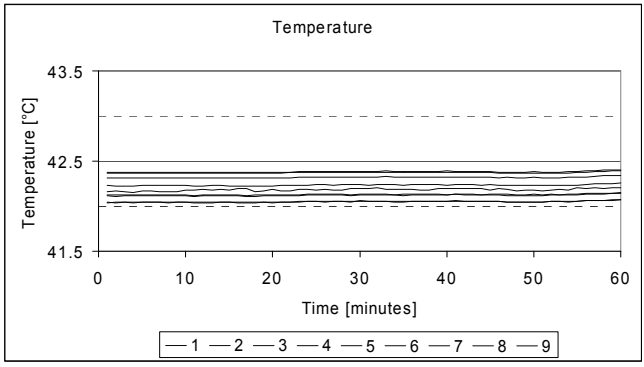
This working uncertainty is different for each PT100 sensor. It fluctuates between 0.040 and 0.053°C.

The target of our project was achieved!

PT100 mapping results

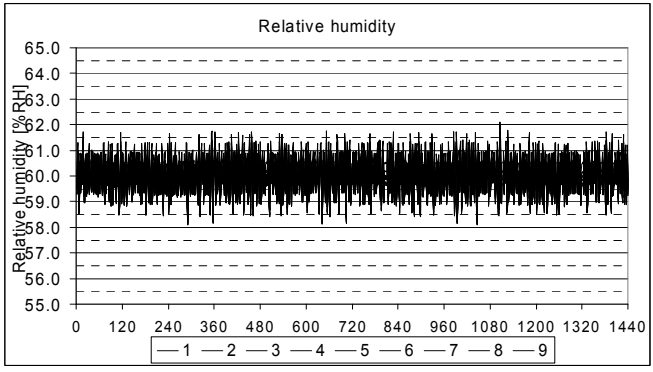
Thermal chamber

Here under a 42.5±0.5°C application, 1 hour after stabilisation, 9 PT100.



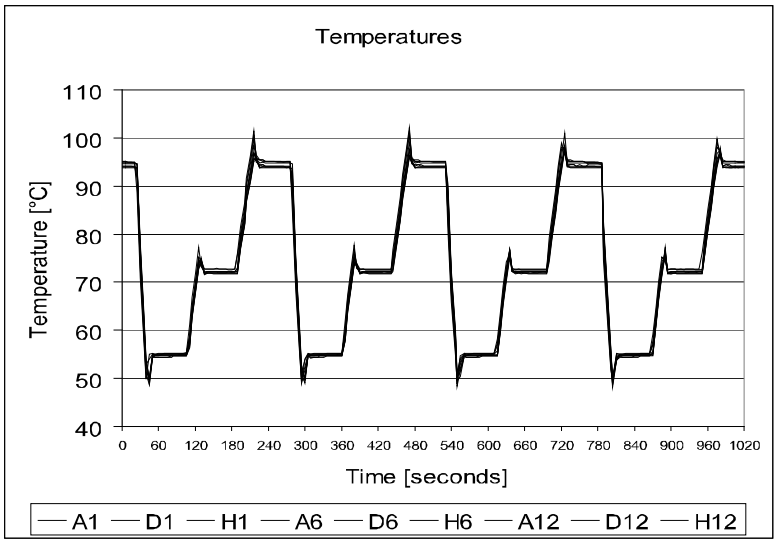
Climate chamber

Here under a relative humidity application, 24 hours after stabilisation, 9 PT100 + 1 cooled mirror dewpoint hygrometer (only RH data).



PCR

The same equipment is also able to characterise Polymerase Chain Reaction equipments (PCR) but it needs a special TC T and a very specific mechanical interface.



Conclusions

We have been using 3 pieces of equipment for more than 5 years and are very satisfied at the metrological point of view (high accuracy, low drift and good uncertainties) which is the most important for a calibration laboratory. Some specific items are discussed more in detail hereunder:

Cost:

This equipment is a high budget compare to existing measurement equipment's (up to 20x more)

Weight:

It is not very convenient to go on site with such an equipment.

High accuracy and liability:

Actually one can buy measurement equipment with a very nice user graphical interface but with a very cheap measurement electronics. In our case we focused

on the best possible measurement electronics, a very basic graphical user interface and the most efficient error correction possibilities.

Multipurpose:

One can switch from PT100 to TC with the same accurate hardware and make use of any type of existing sensors.

References

[1] Norme NF X 15-140, October 2002.

[2] Norme IEC-751, amendement 2, 1995-07.

[3] *Guide pour l'expression de l'incertitude de mesure*, first edition, Organisation internationale de normalisation (ISO), 1995.

[4] B. Crétinon and J. Mérigoux, *La mesure de l'humidité dans les gaz*, BNM, Monographie 18, June 2000.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The SI, Now and Tomorrow

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Determination of the Avogadro constant: a contribution to the new definition of mass

I. Busch[1], S. Valkiers[2] and P. Becker[1]

[1]Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany
[2]Institute for Reference Materials and Measurements, Retieseweg, B-2440 Geel, Belgium
Ingo.Busch@ptb.de

ABSTRACT: An improved attempt of several national metrology institutes to replace the present definition of the kilogram with the mass of a certain number of ^{12}C atoms is described. This requires the determination of the Avogadro constant, N_A, via the silicon route with a relative uncertainty better than $2 \cdot 10^{-8}$. Previously, the limiting factor is the measurement of the average molar mass. Consequently, a world-wide collaboration has been set up, to produce, approximately, 5 kg of ^{28}Si single-crystal with an enrichment factor greater than 99.985% to be used for an improved determination of N_A. The first successful tests of all technological steps (enrichment of SiF4, purification and synthesis of silane, deposition of polycrystalline ^{28}Si, single crystal growth) for the production of high-purity ^{28}Si are reported.

Introduction

Today the kilogram is the last base unit in the International System of Units (SI) which is not traced back to an invariant of nature. A few drawbacks are associated with this definition upon a prototype, which are summarized as follows:

– it can be damaged or even destroyed;

– it is not well defined (it accumulates foreign material and has to be cleaned with unknown effect);

– it ages at an unknown rate (perhaps 50 µg change within the last 100 years). This is because of possible instabilities of the international prototype based on observations of drift between the international and national prototypes over about 100 years;

– it limits the stability of the kilogram relative to fundamental constants to a few parts in 10^8 at best.

In 1870 James Clerk Maxwell [1] defined the well-known requirement, that the physical units should not be founded on macroscopic quantities but on the properties of these “imperishable and unalterable molecules”. Up to now this requirement has not been realized in case of the mass definition.

One of the possibilities discussed among metrologists today is to trace back the unit of mass to an atomic mass. The kilogram would then be defined as the mass of a fixed number N_{kg} of atoms ^{12}C. Since the factor to connect atomic mass standards with the macroscopic mass is the Avogadro constant, in a first step it has to be determined within the framework of the SI with an relative uncertainty of $2 \cdot 10^{-8}$. Thereafter the numerical value would become a defined constant. The new definition of the kilogram could be like: "the kilogram is the mass of N_{kg} free ^{12}C atoms at rest in their ground state". With

$$N_{kg} = \frac{N_A}{A_r(^{12}C)} \times 10^3 \tag{1}$$

where $A_r(^{12}C)$ is the relative atomic mass of ^{12}C with $A_r(^{12}C) = m(^{12}C)/m_u$, m_u being the unified atomic mass constant.

To realise a redefinition of the kilogram on the basis of Si atoms, it is necessary to determine the ratio of the corresponding nuclide masses to ^{12}C, which is the present atomic mass standard, with an accuracy exceeding that of most atomic masses. At present, Penning traps are the most accurate mass spectrometers [2], with an accuracy achieved with these devices is now approaching 10^{-10} (rel.). This is due to the long storage time and the confinement of ions in a small volume in well-defined electric and magnetic fields.

The presupposed final measuring uncertainty of about $2 \cdot 10^{-8}$ (rel.) to the application of the Avogadro constant in a new definition of the mass unit [3], is a challenge for the experimental determination of all quantities involved, macroscopic density, isotopic composition, and unit cell volume of a silicon crystal. A value of the Avogadro constant with a relative measurement uncertainty of $3 \cdot 10^{-7}$ has been obtained by using single crystals of natural Si. However the uncertainty attained is close to a practical limit [4]. The combination of data from several independent measurements of the unit cell and the molar volumes leads to a value for the Avogadro constant of $N_A = 6.022\ 1353\ (18) \cdot 10^{23}\ mol^{-1}$ [5].

Determination of N_A

The desired number N_{kg} can be obtained by counting all silicon atoms in a mole. A practically perfect single crystal is required, to achieve the demanded high accuracy. The international consortium has chosen a silicon crystal, which are widely available and which are almost perfect in structure and purity.

The Avogadro constant N_A can be derived from

$$N_A = \frac{M_{mol}}{m_{Si}} = \frac{n M_{mol} v}{V_0 m} \tag{2}$$

with v, volume, m, mass of a sample, M_{mol}, silicon molar mass, and V_0, volume of the unit cell (with n atoms). From equation (2) it can be deduced that the Avogadro constant relates quantities on the atomic scale and the macroscopic scale: N_A is nothing other than the ratio of the molar volume V_m and the atomic volume V_a. To acquire the value practically the following quantities must be measured:

1)the volume occupied by a single Si atom, i.e. the lattice spacing of an almost perfect, highly pure silicon crystal. This include the precise determinations of the content of impurity atoms and Si self-point defects,

2) the macroscopic density of the same crystal, and

3) the isotopic composition of the Si crystal, with the three stable isotopes ^{28}Si, ^{29}Si, ^{30}Si.

The first two tasks have been attempted by several national laboratories, the National Institute of Standards and Technology (NIST) in the USA, the Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig (Germany), the Istituto Nazionale di Ricerca Metrologia (INRIM) in Turin (Italy), the National Measurement Laboratory (NMI-A) of Australia, and the National Metrology Institute of Japan (NMIJ) in Tsukuba.

The determination of the molar mass, the third point stated above, has been actively pursued at the Institute for Reference Materials and Measurements (IRMM) of the European Community in Geel (Belgium), in a cooperative project under the umbrella of the Bureau International des Poids et Mesures (BIPM) [6].

Fortunately, the lattice defects in semiconductor-grade silicon are reduced to nearly zero dimensional point defects, which include

1) impurity atoms occurring on regular lattice sites by substitution of silicon atoms,

2) impurities on interstitial lattice sites increasing the average number of atoms per unit cell,

3) Si-vacancies and Si self-interstitials favored by the relatively small packing density of the lattice.

The crystal quality is determined by the starting material and the growth process. Fortunately, the number density of the defects is quite low in undoped silicon and therefore in many cases close to the detection limits of the analyzing methods. Only 10^{-8} of the atom quantity are not silicon atoms [7]. This negligibly small amount is achieved by the highly efficient purification process during the floating zone (FZ) refining of crystal growth. This means that the number of atoms per unit cell, *n,* in equation (2) is no longer an integral number and has to be slightly modified by $n = (N_o \pm \delta)$, with $N_o = 8$ and δ in the order of 10^{-8} as a correction due to the impurity content in the crystal.

Molar mass

The molar mass M(Si):

$$M(Si) = \Sigma\, f(^{i}Si)M(^{i}Si) = [\Sigma\, M(^{i}Si)\, R_{i/28}] / \Sigma R_{i/28} \qquad (3)$$

is obtained from the measurement of the isotope abundance ratios $R_{i/28}$ of the three stable Si isotopes ^{28}Si, ^{29}Si and ^{30}Si which are then combined with the molar mass values M (^{i}Si), all available with an uncertainty $< 10^{-8}$, for ^{28}Si with $< 10^{-9}$. The condition $[\Sigma f(^{i}Si)] \equiv 1$ is always fulfilled. However, it is sufficient to measure only two ratios and to calculate the third ratio. Nevertheless all three ratios are measured directly, since the combination of the single values via $(R_{29/28}/R_{30/28}/R_{29/30}) = 1$ can be used as a check. A deviation from 1 indicates systematic errors in the measurement of at least one isotope, which would show up in two ratio amount but not in the third one. The experimentally acquired value is 0.999927 (57). To calibrate the measurement of $R_{i/28}$, IRMM used gravimetrically prepared synthetic mixtures from highly enriched ^{30}Si, ^{29}Si, and ^{28}Si isotopes.

To achieve a further reduction of the uncertainty of N_A, an improvement in the molar mass determination is demanded. This can be realised by fabricating an isotopically pure Si single crystal sphere of ≥ 99.99% enriched ^{28}Si, which means ^{29}Si and ^{30}Si abundances of the order of 0.005% [8]. A relative combined uncertainty of ≤1% on each of these abundance-value corrections contributes a relative uncertainty on the molar mass value of the highly enriched ^{28}Si of $\leq 3 \cdot 10^{-8}$. Using this concept, the task is now to measure the (very) small ^{29}Si and ^{30}Si abundances in the highly enriched ^{28}Si, instead of the calibrated determination of the isotope ratio amount in Si of natural isotopic composition using synthetic isotope mixtures. That results in (very) small corrections to the molar mass value of ^{28}Si, known to a relative combined uncertainty of $\leq 10^{-9}$. Only the uncertainty of these small corrections will be included in the uncertainty budget of the molar mass of the highly enriched ^{28}Si in the Avogadro crystal [9]. These corrections will be measured directly.

This concept has become feasible because of a source of very highly enriched Si isotopes in Nizhny-Novgorod. The Institute of Chemistry of High-Purity Substances RAS (IChHPS RAS) has demonstrated in a first step its capability to produce ^{28}Si with the desired level of enrichment. A second step was due to an up-scale of the production facility to produce a sufficient amount of enriched ^{28}Si material to grow a 5 kg single crystal. The isotopic homogenity has to be assured during the complete process, starting with the transformation of gaseous SiF_4 (the form in which the Si is enriched in the isotope ^{28}Si), through several wet-chemical steps to elemental Si, which serves as a source to the growth process. Hence any isotopic contamination must be minimised. In addition to the requirements for isotopic purity, a very high chemical purity of this enriched ^{28}Si material is also requested. However, isotope enrichment usually acts as a chemical purification process. Therefore the delivered material at the end of the enrichment process is also chemically pure.

Isotope enrichment

Some technological steps have to be carried out on the production method of isotopically enriched silicon [10]. The principle workflow form natural silicon to highly enriched, chemically pure Si crystal is displayed in Figure 1.

The process starts with the gas centrifugation for the isotope separation in the form of silicon tetrafluoride (SiF_4), the transformation of SiF_4 in the monosilane gas SiH_4 followed by low-temperature rectification and then the deposition of polycrystalline silicon (by thermal decomposition of silane) in form of silicon rods used as starting material for FZ single crystal growth. The operations and products necessary to grow the perfect Si single crystal are described in part 1 of the Technical Road Map for the production of ^{28}Si single crystal precursors (TRM28) [11].

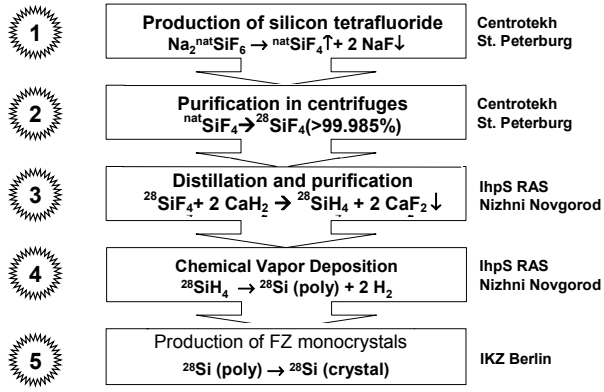


Figure 1. *Workflow of the enrichment process of ^{28}Si, starting with natural silicon and resulting in the production of the float zone (FZ) crystal*

The second part of this roadmap is addressed to the analytical monitoring of the isotope enrichment, chemical purity and measurements of the crystal perfection. Finally, the third part is for quality management purposes and installs a sample identification system, to trace back the course of the Si material through the whole technological chain back to its origin.

Silicon tetrafluoride

The separation of ^{28}Si is initiated with SiF_4, prepared from Na_2SiF_6 powder as follows:

$$Na_2SiF_6 \rightarrow 2\ NaF\downarrow + SiF_4\uparrow \tag{4}$$

The lightest fraction, $^{28}SiF_4$, was separated by ultra-centrifugation in two stages using cascades of gas centrifuges [12], [13]. In the first stage the technology of enrichment of the ^{28}Si isotope was optimized and the separation unit was updated to improve the chemical purity of $^{28}SiF_4$, followed by the second stage of increasing the separation capacity by up-scaling.

Silane

The hydride method was used for the production of high-purity silicon. Therefore, $^{28}SiF_4$ was first transformed into monosilane $^{28}SiH_4$ using the chemical reaction

$$^{28}SiF_4 + 2CaH_2 = {}^{28}SiH_4 + 2CaF_2 \tag{5}$$

in the flow-through mode. The mixture of hydrogen with $^{28}SiF_4$ was passed through the reactor with fine-dispersed calcium hydride. The conversion efficiency of this synthesis exceeded 90%. Processing without organic solvents was chosen to avoid carbon contamination of the silane. The synthesized monosilane was mainly contaminated by fluorine-containing compounds (fluorine-siloxanes) and light hydrocarbons $C_1 - C_4$. The monosilane was additionally ultra-purified by subsequent cryofiltration and rectification.

Hydrocarbons were removed by rectification in a stainless steel column with the feeding reservoir placed in the center of the column in the periodic-mode operation. The fractions containing compounds with boiling points lower or higher than that of silane, were removed simultaneously from the top and bottom of the column, respectively. The concentration of $C_1 - C_4$ hydrocarbons in the selected fractions was monitored by gas chromatography. Due to the ultra-purification, the concentration of hydrocarbons in the monosilane was reduced by more than factor 100.

Polycrystalline silicon

For the production of the polycrystalline silicon crystal a vertically arranged single-rod set-up was specially designed for the thermal decomposition of monosilane. To start the silane deposition process, a ^{28}Si slim rod (7 mm diameter) with an enrichment of 99.99% was used. The seed crystal was grown by pedestal technique from another ^{28}Si crystal.

The slim rod was heated by a direct current with axial movable molybdenum contacts, to compensate for the linear thermal expansion of the rod. The operating temperature of the rod surface was 800°C and was kept constant by increasing the heating power during the deposition process. Due to an increase of the diameter, the supply rate of silane was continuously increased during the process, to stabilize the radial deposition rate at 0.05 mm/h.

In a separate experiment using silane of natural composition, the unwanted, but expected, separation of the three isotopes was investigated. In case of an enrichment of more than 99.985% ^{28}Si for the Avogadro project, the isotopic inhomogenity of about 10^{-10} is negligibly small.

Single crystal growth

In autumn 2005 approximately 400 g ^{28}Si with an enrichment of more than 99.99% was delivered to the PTB. One part of this material was used for fundamental research with enriched silicon, the second part served as feed-rod to the growth of the final polycrystal.

The deposited polycrystalline ^{28}Si rod was transformed into a dislocation free single crystal by FZ crystal growth. This method is very suitable for the single crystal growth of enriched silicon, since the silicon has no contact during melting with other materials, contrary to the Czochralski process, and additional contamination sources are avoided. The crystal growth, performed in Ar atmosphere and with conventional growth conditions, confirmed the quality of the polycrystalline material.

To improve the impurity content down to the level required for the Avogadro crystal, seven additional growth runs were carried out, with five runs under vacuum conditions. On the basis of the measured impurity contents after the first run and the well known carbon distribution coefficients (k_0=0.07), the final concentration was estimated to be lower than $2 \cdot 10^{15}$ cm^{-3} in about 2/3 of the crystal volume.

Results of the enrichment

The investigations carried out due to the quality control of the small ^{28}Si crystal, produced in the first phase of the enrichment process, have proven the level of isotopic enrichment and the chemical purity requested by the International Avogadro Coordination (IAC).

The analysis of isotopic composition of the crystal was performed by laser mass spectrometry (LIMS) at the IChHPS RAS and Electron Impact GAS Mass Spectrometry (EIGSM) and by Thermal Ionization Mass Spectrometry (TIMS), both at IRMM. The results of isotopic analysis (Figure 2) show that the content of the ^{28}Si in the single crystal exceeded 99.99%. The enrichments for the minor isotopes ^{29}Si and ^{30}Si were also measured for the different production steps [14]. Single crystal IR-spectroscopy (IChHPS RAS and PTB) and laser ionization time-of-flight tandem mass reflection at IChHPS proved that the grown crystal is the purest with respect to electro-active impurities.

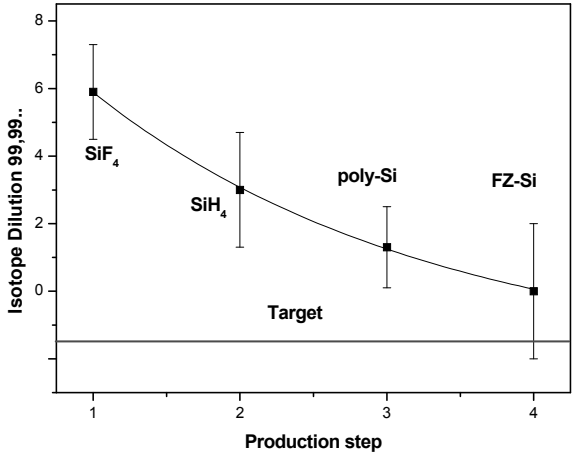


Figure 2. *Isotope dilution of ^{28}Si during the entire production steps*

After the successful finish of the first phase, demonstrating the enrichment process to a degree of more than 99.99% for ^{28}Si, the second part of up-scaling the production is also successfully finished. At the end of 2006 a polycrystalline ^{28}Si rod of approximately 800 mm length, 60 mm diameter and a mass of more than 6 kg was delivered to the IAC. The enrichment is 99.994%, proven by the IRMM at Geel and the impurity content, measured by IR spectroscopy in the PTB, is $3 \cdot 10^{15}$ cm^{-3} for carbon, $4 \cdot 10^{15}$ cm^{-3} for oxygen, and $3 \cdot 10^{13}$ cm^{-3} for boron.

The first FZ run under vacuum conditions to transform the polycrystal into a dislocation free single crystal has already been performed successfully at the "Institut für Kristallzüchtung" (IKZ) at Berlin.

Atomic volume

The determination of the lattice parameter will be performed with an X-ray interferometer. This method, developed by Bonse and Hart [15] in 1965, measures the lattice constant of Si independent to the X-ray wavelength and is traceable to the unit of length (meter).

Simultaneous measurement of the travel distance of a Si lamella with an optical laser interferometer, and counting the X-ray intensity maxima provides the calibrated – mean – silicon lattice parameter *a*, as an average over the cross-section of the X-ray beam. Assuming the unit cell to be cubic in shape and of no more and no less than eight atoms, the atomic volume is $V_o = a^3/8$.

Up to a resolution of several femtometers (1 fm = 10^{-15} m) the purity of semiconductor-grade silicon is high enough to use the X-ray interferometer without any correction due to impurities. For resolutions better than fm, the actual content of impurity atoms, Si self-point defects and, of minor importance, the actual isotopic composition have to be taken into account. Whereas the element specific analysis of the impurities is made spectroscopically, their influence on the lattice homogenity and the average interatomic distance can be measured with special X-ray crystal arrangements.

Volume and Density

The use of a spherical single Si crystal serves to an accurate measurement of its volume. Cylinders or cubes are not applicable due to the sharp corners and edges, which are susceptible to damage. The diameter of the sphere of 1 kg (approximately 100 mm) has to be measured by optical interferometry with a sub-nanometer uncertainty. As a logical consequence the surface quality has to be checked with the same accuracy. With an additional mass determination the density of the sphere is traced back to the base units of the SI, mass and length. An alternative approach is the pressure-of-flotation [16], which compares unknown specimens to density standards, where an almost ideal shape is not necessary.

The grinding and polishing process is the primary influence on the roundness of the sphere. The ultimate roundness achievable on a 100 mm diameter sphere made from single-crystal silicon is in the order of 30 nm, with an surface roughness of about 0.5 nm (peak to valley) [17]. During the fabrication of the sphere, reliable

roundness and roughness measurements have to be applied at every step of the polishing process.

For the determination of the Avogadro constant the volume of the sphere will be determined by measuring diameters either with a Fizeau interferometer that measures about 16,000 diameters simultaneously or with a large number of single measurements in a Fabry-Perot etalon [18]. Two different approaches are used to fit the wave fronts to the curvature of the sphere: light beams with a very small diameter, or expanded light beams with spherical wave fronts to obtain a complete topographical map of the sphere. Reproducibility in diameter measurements of about 1 nm is achieved.

Conclusion

The determination of the Avogadro constant using single crystals of natural silicon has attained a limit in the uncertainty of $3 \cdot 10^{-7}$. To overcome this limit and realise a further reduction of the uncertainty down to $2 \cdot 10^{-8}$, a sphere of highly enriched ^{28}Si (99.99%) has to be produced.

Within a cooperation of eight national institutes, international research institutes and two Russian R&D institutes, about 6 kg of highly enriched ^{28}Si material have been produced and delivered to the International Avogadro Coordination (IAC) at the end of 2006.

In the first step, the Russian partners have produced a prototype polycrystal to demonstrate all technology steps and the desired enrichment of >99.99% ^{28}Si in the final single crystal. This step was completed by verification of the enrichment and chemical purity of the prototype in autumn 2005.

The subsequent up-scaling of the process was successful and the cooperation was finalized with step two of the project and the delivery of about 6 kg of highly enriched ^{28}Si polycrystal. Again the isotopic and chemical composition of the material met the requirements of the IAC.

The first FZ run has successfully finished at the IKZ and the sphere should be available in autumn 2007.

In the determination of density and lattice parameter, analogous to those on Si of natural isotopic composition as described above, no new problems are expected. A structured task distribution between the participating laboratories will be implemented on a world-wide level amongst participants under the aegis of the working group on the Avogadro constant of the Comité Consultatif pour la Masse of the Comité International des Poids at Mesures (CCM-CIPM). It “integrates” – rather

than "federates" – the efforts at various Metrology Institutes on a fundamental constant.

Acknowledgement

We would like to thank all contributors to the determination of the Avogadro constant, especially our partner institutes in St. Petersburg and Nizhny-Novgorod, represented by A.K. Kaliteevski, project leader.

References

[1] J.C. Maxwell, 1890 Report Brit: Mathematics and Physics, Association Adv. Science XL, *Math. Phys. Sec.*, Notices and Abstracts, pp. 1-9.

[2] F. DiFilippo, V.Natarajan, K.R. Boyce, D.E. Pritchard: "Accurate Atomic Masses for Fundamental Metrology", *Phys. Rev. Lett. 73*, 1481-1484, 1994.

[3] P. Becker *Rep. Prog. Phys*. **64** (2001) pp.1945-2008

[4] P. Becker, H. Bettin, H.-U. Danzebrink, M. Gläser, U. Kuetgens, A. Nicolaus, D. Schiel, P. De Bièvre, S. Valkiers, P. Taylor, *Metrologia* **40** (2003) pp.271-287

[5] K. Fujii, A. Waseda, N. Kuramoto, S. Mizushima, P. Becker, H. Bettin, A. Nicolaus, U. Kuetgens, S. Valkiers, P. Taylor, P. De Bièvre, G. Mana, E. Massa, Jr. E.G. Kessler, M. Hanke, R. Matyi *IEEE Trans I&M* (2005) pp.854-859

[6] P. Becker: History and progress in the accurate determination of the Avogadro constant, *Rep. Prog. Phys.* **64** (2001) pp. 1945-2008

[7] W. Zulehner, B. Neuer, G. Rau: "Silicon", in: *Ullmann´s Encyclopedia of Industrial Chemistry*, Vol. **A23**, VHC Publishers Weinheim, 1993, pp. 721-748

[8] H. Friedrich, "Isotopically enriched Silicon: the atomic path to the kilogram?", Laboratory Report, Physikalisch-Technische Bundesanstalt, Braunschweig, 2002

[9] P. De Bièvre, R.Werz "How enriched must enriched isotopes be to measure synthetic isotope mixtures", *Int. Jour. Mass. Spect. and Ion Physics* **48,** 365-368, 1983

[10] P. Becker et al., *MST* **17** (2006), pp. 1854-1860

[11] H.-J. Pohl and P. De Bièvre *Technical Road Map for the Production of ^{28}Si Single,* 2005

[12] E.I. Abakumov, A.K. Kaliteevski, V.I. Sergeev et al. *Atomnaya Energiya* **67**(4) (2005) 255 [in Russian]

[13] Y.V. Tarbeyev, A.K. Kaliteyevski, V.I. Sergeyev, R.D. Smirnov, O.N. Godisov *Metrologia* **31** (1994) pp. 269-273

[14] K.M. Itoh, J. Kato, A.K. Kaliteevski, O.H. Godisov, D. Bulanov, G.G. Devyatych, A.V. Gusev, P.G. Sennikov, H.-J. Pohl, R. Riemann, N. Abrosimov *Jpn. J. Appl. Phys.* **42** (2003) pp. 6248-6251

[15] U. Bonse, M. Hart: "An x-ray interferometer", *Appl. Phys. Lett.* **6**, (1965) pp. 155-156

[16] H. Bettin, H. Toth; *MST* **17** (2006), pp. 2567-2573

[17] A.J. Leistner, W.J. Giardini: "Fabrication and sphericity measurements of single-crystal silicon spheres", *Metrologia* **31**, (1995) pp. 231-243

[18] R. A. Nicolaus and G. Bönsch: "A novel interferometer for dimensional measurement of a silicon sphere", *IEEE Trans. Instrum. Meas.* **46** (1997) pp. 563-565

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Development and application of the femtosecond frequency comb

S.A. van den Berg[a], M. Cui[b], N. Bhattacharya[b] and J.J.M. Braat[b]

[a] NMi van Swinden Laboratory,
Thijsseweg 11, 2629 JA Delft, The Netherlands
svdberg@nmi.nl

[b] Optics Research Group, Department of Applied Sciences,
Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

ABSTRACT: An optical frequency comb has been built at NMi van Swinden Laboratory. The first application is the realization of the coupling from the meter to the second. This is accomplished by measuring the frequency of an iodine-stabilized helium-neon laser (our primary length standard) directly with respect to our time standard. The second application that will be discussed is the use of the frequency comb as an advanced white light source for measurement of long distances.

Introduction

The invention of the stabilized femtosecond frequency comb has ushered in a revolution in frequency metrology. Frequency comb technology now is generally applied in the fields of optical frequency metrology, ultrashort pulse generation, high resolution spectroscopy, including the work on fundamental constants, and forms the basis for optical clocks [1,2]. The synergy between ultrashort laser technology and the recent development of highly nonlinear fibers has paved the way for octave-spanning frequency comb generation. Reviews on the development of frequency comb technology are provided by Cundiff and Ye [3] and Udem [4].

In this paper we will first summarize the principle of operation of the frequency comb and its experimental realization. Then the measurement of optical frequencies with the comb will be discussed and illustrated with some experimental results. Another application of the femtosecond frequency comb lies in the field of distance measurement. The scheme will be discussed, followed by some simulations of the effect of frequency noise on the measurement results.

Frequency measurement

The frequency comb is generated with a mode-locked femtosecond laser. The periodic train of pulses emitted from the laser has its counterpart in the frequency

domain as a comb of equidistant modes with a mutual separation equal to the repetition frequency fr. The difference between the group velocity and phase velocity gives rise to a pulse to pulse phase shift between the carrier wave and the envelope. This phase shift results in an offset frequency f0, generally referred to as the carrier-envelope offset frequency. Therefore, a frequency comb has two degrees of freedom which are the repetition frequency fr and the offset frequency f0, with f0 < fr. This is illustrated in Figure 1.

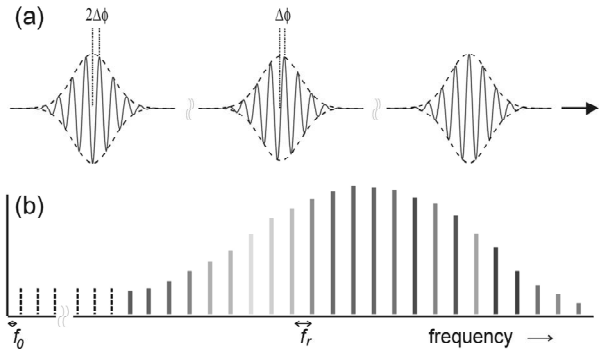


Figure 1. ***a*** *Illustration of the carrier-envelope phase shift in the time domain. A pulse to pulse phase shift* $\Delta\phi$ *is observed;* ***b*** *corresponding optical frequency spectrum with* f_0 *the offset frequency due to the carrier-envelope phase shift, and* f_r *the pulse repetition rate. In reality the comb consists of several hundreds of thousand individual lines*

The frequency of a single line is given by $f_n = f_0 + n f_r$, with n an integer number. Both f_r and f_0 need to be stabilized to have a stable frequency comb. The development of broadband femtosecond lasers and techniques to further broaden the spectrum using self phase modulation in a micro-structured optical fiber, provide an experimental handle to measure and stabilize f_0. Such a stabilized frequency comb laser provides absolute optical frequencies, spanning over an octave.

Setup

The optical frequency comb is generated with a Kerr-lens modelocked Ti:Sapphire laser (GigaOptics), emitting pulses of about 40 fs (after compression) at a repetition rate of 1 GHz. The oscillator is pumped at 5.5 W (Coherent Verdi V6), resulting in an output power of 830 mW. The pulse train is sent through a 1 m

microstructured fiber with a core diameter of 2 μm. This results in an octave-spanning spectrum that largely covers the optical domain, and extends into the infrared. The green and infrared part of the spectrum are selected and are sent into a f-2f interferometer to measure f_0. Here the frequency of the infrared spectral part (around 1064 nm) is frequency doubled in a 7 mm BBO crystal. The frequency doubled output is mixed in an interferometer with the green light (532 nm) that is generated in the fiber as well. The lowest beat frequency that is measured represents f_0 (signal >40 dB above noise floor). The offset frequency is stabilized to a reference oscillator by controlling the intensity of the pump beam with an acousto-optic modulator. The repetition frequency f_{rep} is measured directly and is stabilized to a second reference oscillator by applying feedback to the cavity length with a piezo-mounted mirror. Both reference oscillators are directly linked to our primary time standard, a commercial high-performance cesium clock. The optical frequencies of the comb are thus directly referenced to the time standard as well.

Measurement results

In order to measure the frequency of the iodine-stabilized helium-neon laser NMi-5, the laser output is mixed with the red part of the frequency comb on an avalanche photodiode. The frequency f_{beat} (S/N ratio of about 30 dB), is selected with a tunable RF filter and measured with a frequency counter (Agilent 53132A) that is referenced to the cesium clock. The laser frequency can be written as $f_{laser} = n f_r \pm f_0 \pm f_{beat}$.

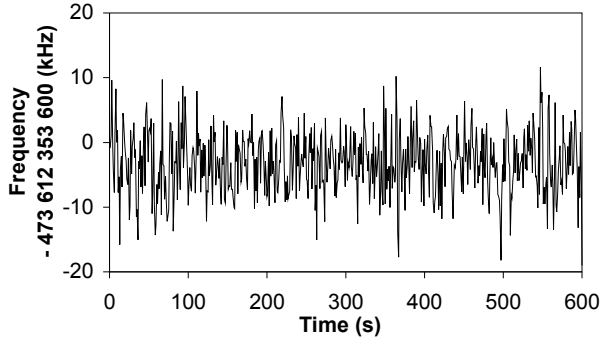


Figure 2. *Measured laser frequency of NMi-5 (f-component) over a period of 10 min. The noise is fully dominated by the frequency fluctuations of NMi-5*

The value of the integer n and the signs of f_0 and f_{beat} are determined from the nominal value of the laser frequency. The result of a measurement of the laser frequency over a period of 600 s is shown in Figure 2. The measured average laser frequency of the f-component of NMi-5 is 473612353596.9 kHz. This result agrees

well with a previous measurement of the same frequency with the BIPM frequency comb that was performed about one year earlier (473612353597.6 kHz).

Distance measurements

The active stabilization of f_{rep} and f_0 does not only lead to frequency stability in the optical domain, it also fixes the pulse-to-pulse phase shift $\Delta\phi$ (see Figure 1). The phase relationship between subsequent pulses is thus conserved, which allows for interference between different pulses from the femtosecond laser. In the frequency domain this may be viewed as an extension of multiwavelength interferometry, with the mode-locked laser as the multi-wavelength source. In this case, the synthetic wavelength is determined by the mode spacing (equal to f_{rep}), which is typically of the order of 100-1000 MHz for pulsed lasers, like Ti:Sapphire lasers of fiber lasers. This leads to a required accuracy of the initial value of the distance of the order of 1 m. Interference between two pulses can be observed if the path-length difference is a multiple of the distance between the pulses. This can be accomplished by changing the cavity length, and thus the interpulse distance [5]. Alternatively, we propose to apply a delay line to tune the interferometer to zero path-length difference. This method is especially suitable when the distance to be measured is relatively short with respect to the cavity length. This setup is shown in Figure 3.

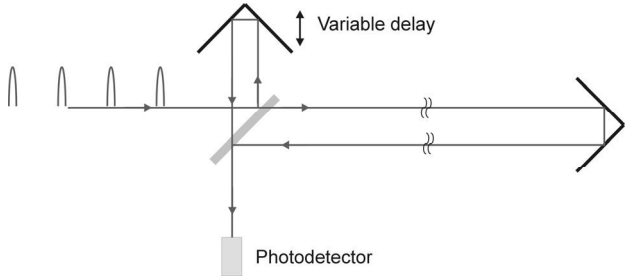


Figure 3. *Scheme of the experimental setup. A pulse train is sent into an interferometer with a long measurement arm and a short reference arm that can be adjusted such that pulse overlap at the detector is established*

The pulses generated by the frequency comb source are split into two at the beam splitter (BS). One part of the pulse goes into the measurement arm and the other is sent to a calibrated scanning delay line, which is the reference arm. The range of the delay line equals half of the optical cavity length of the frequency comb laser. Since the distance between successive pulses is equal to the cavity length, this range suffices to obtain spatial overlap between two pulses. The optical path length in the measurement arm can be written as $L_m = m\ l_c + x$. Here m is an integer number and x a distance shorter than the cavity length ($x < l_c$). For the reference arm, the optical path is set such that $L_r = x$. This leads to a path length difference ΔL

$= L_m - L_r = m\ l_c$. The cross-correlation function can be measured by scanning the delay line around the position $L_r = x$.

The effect of noise on the correlation function

The ultimate distance that can be measured with such techniques will be limited by the noise or jitter that is present in the pulse train. A minimum stability is required to obtain sufficient fringe contrast. The jitter between the two pulses that interfere should be less than an optical period (2.7 fs@ 800 nm wavelength) for fringes to be visible. In order to estimate the effect of timing jitter on the fringe contrast of the cross-correlation function, we simulated the correlation measurement for various distances. We consider only the effect of noise on f_r. The power spectral density (PSD) of the noise plays an important role, since the spectral distribution of the noise determines the time scale on which the fluctuations take place. For the simulations presented below, a typical PSD of the noise is used, that approximates measured noise features [6,7] (see Figure 4).

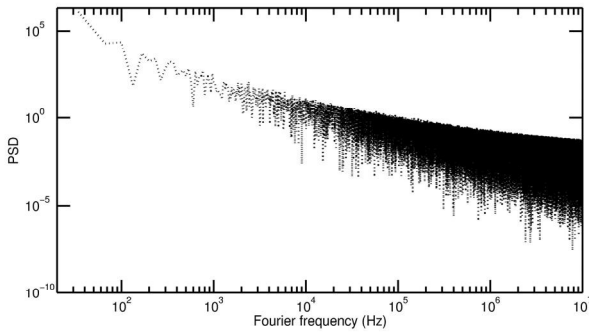


Figure 4. *Power spectral density of the timing phase noise used for the simulations*

Timing jitter noise with the PSD shown in Figure 4 was added to the pulse train arising from the frequency comb source. The cross correlation function is then calculated for various distances between the pulses, as can be seen from Figure 5.

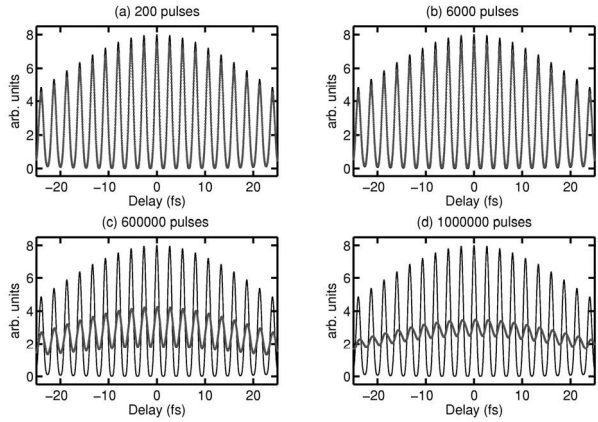


Figure 5. *Cross correlation signals between pulses in the train separated by 200, 6000, $6 \cdot 10^5$ and 10^6 pulses, respectively. For a 100 MHz repetition rate of the pulsed source this translates to distances of 300 m, 9 km, 900 km and 1500 km. The fringe contrast drops to 22% for the longest distance*

The presence of noise leads to a decreasing fringe contrast with increasing distance. This is easily understood since low frequencies (long time scales) have a larger amplitude in the frequency noise spectrum. Note that we have not taken into account the noise of the reference clock here. In practice, the maximum distance that can be measured will be limited by the quality of the clock. Commercial atomic clocks with an accuracy on the 10^{-13} level allow distance measurements with optical interferometry to distances over 10^6 m.

Conclusion

The femtosecond frequency comb has been applied for measurement of optical frequencies. By measuring the frequency of the primary length standard NMi-5, we have obtained traceability from the meter to the second in the Netherlands. As a second application of the frequency comb we have presented an interferometric scheme for absolute determination of long distances and demonstrated how the fringe contrast diminishes as a result of timing jitter. The noise power spectral distribution has been chosen such that experimentally observed noise distribution is approximated. If the timing jitter has a small enough amplitude (i.e. f_0 and f_{rep} are tighly locked to a reference clock), the fringe contrast does not vanish completely for long distances.

References

[1] S.A. Diddams *et al*., "An optical clock based on a single trapped $^{199}Hg^{+}$ ion", *Science*, vol. 293, pp. 825-828, Aug. 2001.

[2] D.J. Jones *et al.*, "Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical freqency synthesis", *Science*, vol. 288, pp. 635-639, April 2000.

[3] S.T. Cundiff and J. Ye, "Colloquium: Femtosecond optical frequency combs", *Rev. Mod. Phys*., vol. 75, pp. 325-342, Jan. 2003.

[4] Th. Udem *et al*., "Optical frequency metrology", *Nature*, 416, pp. 233-236, March 2002.

[5] J.Ye, "Absolute measurement of a long, arbitrary distance to less than an optical fringe", *Opt. Lett.*, Vol. 29, pp. 1153-1155, 2004.

[6] R.K. Shelton *et al*, "Subfemtosecond timing jitter between two independent, actively sunchronized, mode-locked laser", *Opt. Lett.*, Vol 27, pp. 312-314, 2002.

[7] E.N. Ivanov, S.A. Diddams and L. Hollberg, "Experimental study of noise properties of a Ti:Sapphire femtosecond laser", *IEEE Trans. on Ultr., Ferr. and Freq. Contr.*, Vol 50, pp. 355-360, April 2003.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Recent development in time metrology

J. Achkar and D. Valat

Observatoire de Paris, LNE-SYRTE
61 avenue de l'Observatoire 75014 Paris, France

ABSTRACT: The recent development in time metrology done at LNE-SYRTE, allows the laboratory to improve its measurement uncertainties, especially on i) the determination of the deviation between the coordinated universal time UTC and its national realization UTC(OP), evaluated within ±1.5 ns, ii) the equipment of remote atomic clocks comparison, through microwave links with communications satellites (TWSTFT technique) or radio-navigation satellites (GPS P3 technique) and iii) the frequency accuracy of the French atomic time TA(F) permitting the realization of the SI second at better than $\pm 1.10^{-14}$ and making an excellent standard for the laboratories accredited by the COFRAC.

National time scales and contribution to coordinated universal time

Two time and frequency references are established by LNE-SYRTE (also described as OP in this document) and made available to users. They contribute to the development of the international atomic time TAI and coordinated universal time UTC maintained by the Bureau International des Poids et Mesures BIPM and connected together by the following relation: TAI – UTC = 33 s, since January 1st 2006. The deviation between these two scales is caused by leap seconds announced by the international earth rotation service IERS located at the Observatoire de Paris.

UTC(OP) as national time reference

The Temps Universel Coordonné de l'Observatoire de Paris UTC(OP) is materialized by second pulses from a high performance cesium commercial atomic standard driven in frequency by a micro phase stepper operating in continuous mode. The UTC(OP) time scale is the physical realization of the international UTC reference, basis for the French legal time according to the decree of 1978. Compared to the coordinated universal time, time deviation |UTC-UTC(OP)| published by the BIPM is maintained below 50 ns with a 1.5 ns standard type uncertainty (Figure 1). This uncertainty value positions the laboratory at the highest international level, with very few laboratories. This performance was reached with the use of the 2 way microwave link in the French contribution to TAI [1].

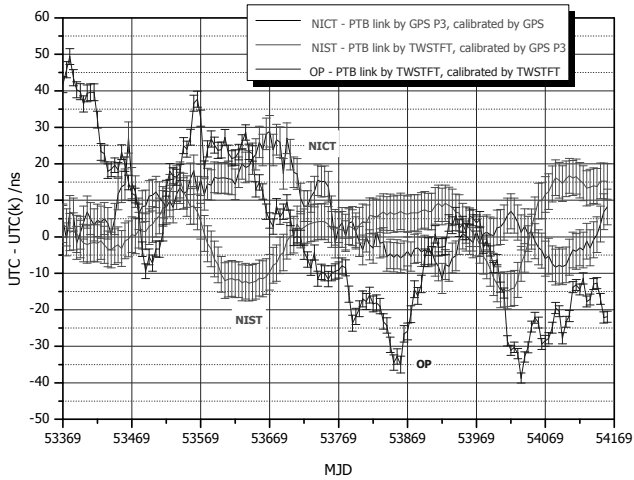


Figure 1. *Comparison of a specific national time scale sample in relation to UTC, over the period between December 2004 and February 2007 [1]*

Because of this, measurement data carried out by the LNE-SYRTE station are integrated in the TAI calculation from January 1st, 2005 in the OP-PTB link calibrated with a 1 ns accuracy, which is five times lower than with the same link carried out by the GPS C/A 1-way single frequency microwave link. In addition, in the same figure, two specific links are also reported, showing deterioration of the GPS calibration uncertainty of a link, whether it is based on the TWSTFT or GPS P3 method.

TA(F) as national frequency reference

The French atomic time scale TA(F) is a scientific time scale aimed at frequency stability and accuracy qualities. It is calculated monthly at the LNE-SYRTE by using daily readings of a series of commercial clocks (23 in 2006) operating in 8 laboratories in France. The clocks in the 8 laboratories are compared several times a day by the common view method of GPS C/A signals, with the master LNE-SYRTE clock, UTC(OP). Clock related data is introduced in an algorithm intended to ensure frequency stability in the medium term (30 days) of the time scale TA(F). The TA(F) frequency stability is typically of 3.10^{-15} over a 30 day integration period.

By the creation of a weighted average of clock data, the TA(F) algorithm produces an intermediate time scale with a better frequency than the clocks contributing to it. It is the free FEAL atomic scale. The term “free” comes from the fact that we only integrate into the calculation clocks with a frequency that can freely evolve to avoid any bias at the algorithm output.

The weight of each clock is inversely proportional to the error resulting from the comparison between actual clock data and a prediction by a statistic ARIMA (Auto Regressive Integrated Moving Average) model [2]. The parameters of model ARIMA, commonly used in the analysis of time series were optimized for high performance cesium commercial type 5071A atomic clocks.

The FEAL frequency reported at SI was approximately $-2.4.10^{-13}$ in January 2007. This dimension represents FEAL frequency accuracy which is not calibrated.

The primary laboratory standards create the second SI with a frequency accuracy from 2.10^{-15} (thermal jet standard) to $4\text{-}5.10^{-16}$ (cold atom fountain) [3]. In order to obtain TA(F), the accuracy of these standards is transferred to FEAL in the form of a monthly frequency correction. In 2006, the TA(F) frequency was calibrated on the basis of a thermal jet standard (JPO) relayed in the second semester by cold atom fountains (FO1 and FO2). In the second semester of 2006, TA(F) accuracy was maintained at 3.10^{-15}, which makes it one of the most exact atomic time scales in the world (Figure 2).

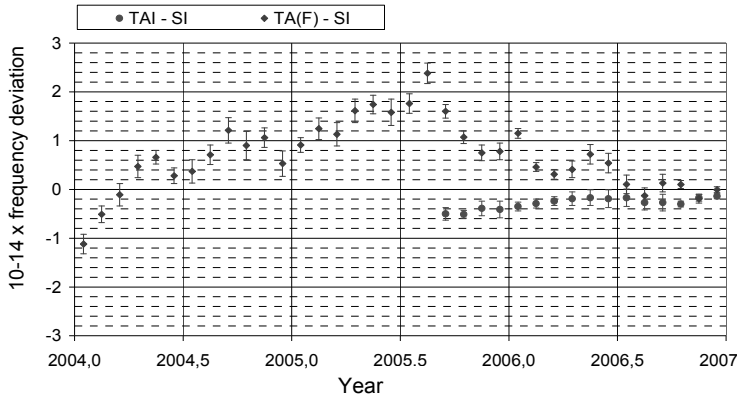


Figure 2. *Average monthly TAI and TA(F) frequencies related to the SI second*

In this way the TA(F) constitutes an excellent reference for COFRAC accredited laboratories. The clocks of these laboratories are matched in frequency with TA(F) during the monthly calculation before the 5th of each month. The monthly frequency link of the TA(F) at second SI is based on the BIPM circular T database and the result is published with a one month delay in the monthly report from the LNE-SYRTE (Bulletin H).

The TA(F) frequency accuracy also makes it possible to periodically evaluate the frequency drift of hydrogen masers. The frequency drift stability is a significant criterion in the choice of local oscillator that will be used in the context of studies on frequency transfer and for TAI frequency calibration.

In addition, the TA(F) clocks contribute to the International Atomic Time TAI calculated by the Time section of the BIPM.

Atomic clock comparisons by satellite systems

The laboratory reinforced these remote atomic clock comparison systems, in particular with the implementation of geodetic receivers for an ionospheric delay compensation, and a 2 way station time calibrated at ± 1 ns.

Transfer of time and frequencies by common GPS views

Common view principle: when 2 ground stations (S_A and S_B) equipped with GPS receivers simultaneously collect the signals transmitted by the same satellite, they simultaneously carry out a measurement of pseudo-distance. It is the distance measurement between the satellite (transmission of the signal in GPS Time) and the receiver (reception of the signal dated by the ground clock). If we cut off the real geometric distance covered by the signal transmitted at pseudo-distance, we obtain a residual distance from which we deduce the time variation which separates the TGPS from the ground clock at moment T, $(H_A\text{-TGPS})_T$ and $(H_B\text{-TGPS})_T$. The pooling of these results makes it possible to easily obtain $(H_A - H_B)_T$.

With a single channel receiver and according to the BIPM schedule, we obtain up to 48 measurements of this type by 24 hour periods. The coordinates of the ground station antennas must be known with an uncertainty of a few centimeters. For the GPS P3 reference antenna from the LNE-SYRTE, the IGS (International GNSS Service) provides the coordinates within the framework of our contribution to the Permanent GPS Network (RGP).

The GPS satellites transmit signals over two frequency bands L1 and L2. These signals are modulated by two types of pseudo random codes of noise: codes Coarse/Acquisition C/A on L1 (1575.42 MHz) and Precision/Secure P/Y on L1 and L2 (1227.60 MHz).

Certain “geodetic” type receivers have recently given access to code P of the GPS. Moreover, these multichannel receivers make it possible to observe several satellites simultaneously and thus increase the number of measurements.

One of the main advantages of the GPS P3 is the compensation of the ionospheric delay by a linear combination of measurements, called P3, since code P is transmitted simultaneously on both L1 and L2 GPS signal carriers, as the effect of the ionosphere crossing is proportional to the inverse of the signal frequency square.

A new software program developed at LNE-SYRTE makes it possible to make a selection among the simultaneous common views between two laboratories: application to the PTB-OP link [4].

With this original method of filtering we obtain a noticeable improvement of the short-term stability of these measurements, a major element in the frequency comparisons of the local oscillators of the primary frequency standards. This method was applied to the link data between OP and PTB in 2004 (Figure 3).

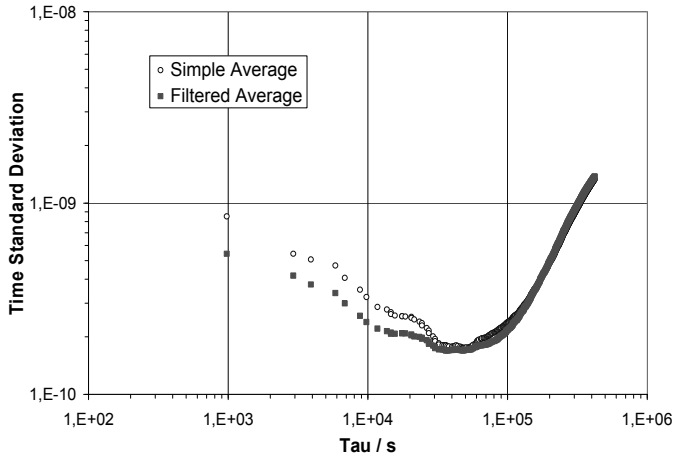


Figure 3. *Time stability of the link between Maser H4 (PTB) and Maser 816 (OP) by the average of simultaneous GPS P3 common views*

In Figure 3, the hollow circles represent the stability obtained from a simple average, whereas the full squares represent stability after filtering. We note a slight improvement of stability for periods of analysis around 10000 S, which particularly enables a better visualization of the daily periodic term around 40000 S. The best stability obtained is of 170 PS over a period of analysis of 36000 S.

Time calibration: all the delays due to equipment must be measured in order to determine the true value of $(H_A - H_B)_T$. It is the delay introduced by cables and internal receiver delay. The latter is determined by comparing the receiver to calibrate with the BIPM receiver, calibrated in absolute, when they are supplied by the same clock. The LNE-SYRTE is the reference laboratory for the calibrations of BIPM GNSS receivers.

The method can be improved with the integration of IGS products in this type of measurements: precise ephemeris, tropospherical maps, satellite clock variation errors in relation to the GPS Time for transcontinental link processing.

Moreover and with the GPS P3 data, the LNE-SYRTE takes part in the installation by the BIPM of a new method of clock comparisons by GPS called "All in View". This method will, with the help of IGS products, make it possible to use all satellite trackings, including those not in common view of the two stations involved.

In the near future, the techniques mentioned above will be extended and adapted to the Russian GLONASS and European GALILEO satellite data.

Transfer of time and frequencies by the 2 way method (TWSTFT)

The laboratory developed a system of comparison of ground atomic clocks by a 2 way microwave link (in Ku band) with a telecommunications satellite. This very powerful system makes it possible to compare two remote clocks, by direct link – the geostationary satellite being used as microwave relay, without the help of an intermediate clock as with the GPS satellite radio navigation system. It is with this goal in mind that a 2 way station was developed (Figure 4), equipped with a SATRE modem generating a spread spectrum carrier, modulated by a pseudo-random code of noise with 2.5 MChips/s, containing the clock signal 1 PS. The main goal is twofold: to contribute to the development of TAI calculated by the BIPM by using an efficient technique independent of the GPS, and to take part in the comparison of primary remote frequency standards [5].

The signal delays vary with distance, ionosphere, troposphere, weather, ground conductivity, etc. Thus, for two given remote 2 way stations, time signals are transmitted at the same moment, with each station receiving the signal from the other station. The difference between the two clocks is then calculated: the comparison accuracy depends on the non-reciprocal residual effects, which must be determined.

Figure 4. *TWSTFT station of LNE-SYRTE*

The performances reached by this technique have a frequency stability of the link of 1.10^{-15} at one day of sampling and a level of noise of 2.10^{-16} is reached on average over only 5 days [6].

Comparisons of the two systems TWSTFT and GPS

An effective comparison of the two systems involving three technical measurements (TWSTFT, GPS P3 and GPS C/A) is illustrated here. The variations observed on the PTB – OP link, over a 27 day analysis period for the TWSTFT and 19 days for the GPS, are 4.5 ns with the TWSTFT method and 9 ns with the GPS P3 method in common view, with excellent recovery however. As for the corresponding time Allan standard deviation (Figure 5), it is clearly observed that the masers are reached by the TWSTFT link at 0.6 day (date of angle change) with a $\sigma_x(\tau) = 80$ ps, and at the end of 2 days by the GPS P3. The comparison of these two techniques compared to GPS C/A (typical curve of Figure 3) clearly shows the improvement made by the laboratory in the comparison of clocks.

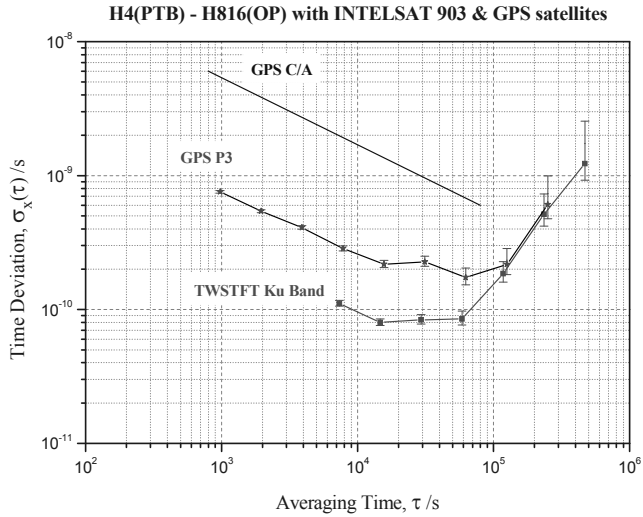


Figure 5. *Time stability of three microwave links used in the comparison of remote masers*

The implementation at the laboratory of various transfer links of time and frequencies makes it possible:

– to control various techniques in this field and to compare them with each other to better adapt the requirements in time metrology;

– to ensure an active redundancy, high performances and lower cost, of the main TWSTFT link in the TAI French contribution. This is the case with GPS P3;

– to ensure the permanence of the existing links in relation to the French metrology missions entrusted to the LNE-SYRTE.

Conclusion and perspectives

With the recent developments carried out in time metrology, the laboratory is positioned at the highest international level in this field along with some powerful worldwide laboratories. The prospects for evolution will continue mainly in the following fields.

Creation of a new time scale

UTC(k)s are generated by either a single primary frequency standard (PTB Germany), or by an algorithm (NIST and USNO United States, NICT Japan), by a maser with hydrogen controlled in frequency (NPL UK), by a cesium clock average (ROA Spain), or by a single cesium clock (INRIM Italy, OP France and ORB Belgium). Performances from the last three UTC (K)s are probably limited by cesium clock instabilities. A study is carried out in order to improve the short term stability of UTC (OP) and to improve real time prediction of UTC (Figure 1).

This study is motivated by new goals: participation in key BIPM comparisons, space projects (EGNOS, Galileo) with the goal of improving real time UTC prediction and in metrology for the improvement of short term UTC (OP) stability used as single source of remote comparisons.

The improvement of UTC (OP) performances will be done as with the NIST [7], by using a hydrogen maser for short term stability and an atomic scale “E” based on a series of clocks for long term stability. UTC prediction will be improved by the calibration of the frequency of this time scale and long term steering to UTC (OP) maser on UTC (Figure 6).

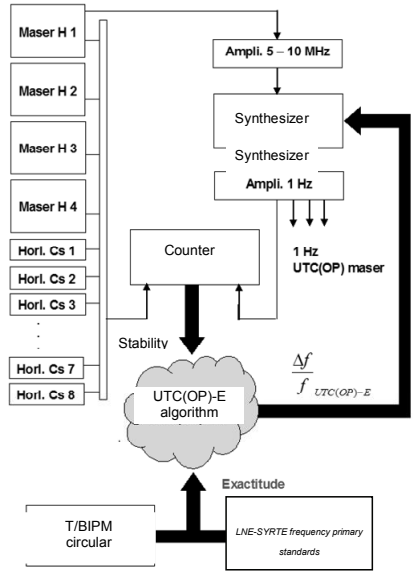


Figure 6. *New chain of UTC (OP) maser development*

The first studies involved the implementation of the atomic time scale "E" algorithm, weighted average of a series of cesium clocks. At the end of the optimization studies of the algorithm parameters in "prediction" mode done in 2005, time scale "E" reaches a frequency stability between 3.10^{-15} and 30 d in Allan standard deviation.

The real conditions of the application of such an algorithm demand the development of management modules of clock input/outputs to ensure phase continuity of the scale and to make the algorithm available in "operational" mode.

The development of an automatic detection module for phase skips and/or missing data enabled the development of an automatic clock elimination module. In 2006, a second version of the software made it possible to test a weighting mode based on the Allan variance and to adapt the parameters to the various types of clocks encountered (cesium, hydrogen) in order to prevent a clock from becoming dominant in calculation (Figure 7)

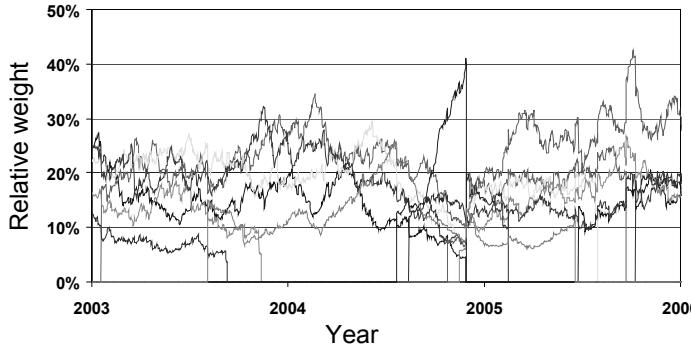


Figure 7. *Relative weight of clocks in the calculation of time scale "E" (extracted from data 2003-2006)*

The experience gained during studies on TA(F) accuracy will be used to optimize the accuracy of "E". A very long term UTC control will be useful in avoiding a phase accumulation that is too high over time. Tests of various control modes are in progress.

The use of a maser with hydrogen as a source of UTC (OP) will reduce uncertainties on the transfers of remote time and the control of its frequency on that of "E" will have to allow for a gain in factor 2 on the real time prediction of UTC.

Implementation of a satellite simulator

An original absolute calibrating device of internal delays of the station installed on the site of the Observatoire de Paris was recently developed (Figure 8).

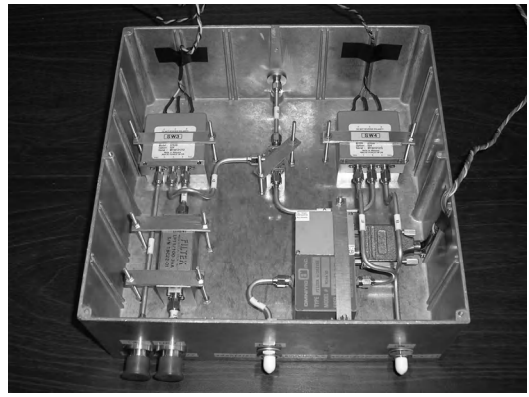


Figure 8. *Microwave part of the satellite simulator developed at LNE-SYRTE*

This device, based on a satellite simulator, makes it possible:

i) to calibrate in absolute mode the differential delay of the station, which is part of the time equation describing the difference between two remote atomic clocks. The knowledge of this delay is vital in the determination of the variation of absolute time between the two clocks in comparison (study of accuracy);

ii) to follow-up over time the differential delay based on atmospheric conditions (temperature, pressure, moisture), in order to highlight certain systematic effects (in particular daily and seasonal variations of the delay) degrading accuracy. The consideration of these effects will make it possible to improve the level of accuracy (stability study).

The current performance of the equipment including the station and simulator show an exceptional time stability, lower than 50 PS over a sampling period ranging from 0.3 to 3 days [8].

Such a device is increasingly recommended by the CCTF working group on TWSTFT in order to better evaluate the component of uncertainty related to the variation of the differential delay of the station between two calibration campaigns per portable station. Along with some simulators present on the market manufactured from the VSL design, the LNE-SYRTE developed a new calibration device based on two different designs: a faster calibration process of station delays and a complete characterization of its internal delays with the help of a vectorial network analyzer.

Contribution to the use of Galileo in time metrology

The laboratory is a member of the Fidelity consortium which obtained the contract with Galileo Joint Undertaking in June 2005 for the development of the Galileo Time Service Provider prototype. Fidelity started construction and will operate the GTSP [9] prototype in order to connect Galileo System Time GST to the universal time UTC coordinated for its validation in orbit planned for 2008. The role of the GTSP is to provide the parameters for the GST control, such as it is carried out in Precise Timing Facility, on UTC (modulo 1 S).

The laboratory is mainly involved in the following activities: definition and design of the prototype, responsible for GPS link calibration and participation in the calibration campaigns of TWSTFT links. The laboratory also takes part in the studies on the use of Galileo in time metrology and the relation with EGNOS.

References

[1] BIPM, *Circular T*, 2004 to 2007.

[2] C. Andreucci, "A new algorithm for the French atomic time scale", *Metrologia*, Vol. 37, pp. 1-6, 2000.

[3] S. Bize *et al*, "Cold atom clocks and applications", *J. Phys. B: At. Mol. Opt. Phys.*, Vol. 38, S449, 2005.

[4] J. Achkar, P. Uhrich, P. Merck and D. Valat, "Recent time and frequency transfer activities at the Observatoire de Paris", *Joint IEEE IFCS and PTTI Systems and Applications Meeting*, 2005, pp. 247-253.

[5] J. Achkar and P. Merck, "Comparaisons d'horloges atomiques au sol par liaisons micro-ondes deux voies avec un satellite de télécommunications", *Revue française de métrologie*, Vol. 2, No. 6, pp. 9-24, 2006.

[6] A. Bauch *et al*, "Comparison between frequency standards in Europe and the USA at the 10^{-15} uncertainty level", *Metrologia*, Vol. 43, pp. 109-120, 2006.

[7] M. A. Weiss, D. W. Allan, T. K. Peppler, "A Study of the NBS Time Scale Algorithm", *IEEE Trans. on Instrumentation and Measurement*, Vol. 38, No. 2, pp. 631-635, 1989.

[8] P. Merck and J. Achkar, "Stability study of the TWSTFT microwave satellite simulator developed at LNE-SYRTE", *Conf. Digest CPEM*, 2006, pp.600-601.

[9] R. Jones *et al*, "Fidelity – progress report on delivering the prototype Galileo time service provider", *EFTF-IFCS 2007*.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Original monolithic device of 4-crossed axis with flexure pivots to be used for the French watt balance experiment

Patrick Pinot[1], Stéphane Macé[1], Gérard Genevès[2],Darine Haddad[2], François Villar[2]

1 - LNE-INM/Cnam 61, rue du Landy F-93210 La Plaine Saint-Denis (France)

2 - LNE 29, avenue Roger Hennequin F-78197 Trappes Cedex (France)

ABSTRACT: In the French watt balance experiment, in order to have a single point of application of a weight and an electromagnetic force balancing this weight, a monolithic system of double gimbals with flexure pivots was made out of copper-beryllium alloy. It also allows a rotation according to four coplanar axes which intersect in one point. The parameters of stiffness of the flexure pivots have been determined by the way of a simple dynamometric method. The study of the dynamic behavior in air of this device has shown a very good decoupling between each gimbal.

Introduction

The kilogram is the only basic unit of measure of the Système International d'unités (SI) still defined by a material artifact. Considering the past evolutions of SI and the lack of knowledge of the stability of the international prototype of the kilogram, the definition is not satisfactory. We are attempting to substitute a definition either based on atomic properties or fundamental constants.

Among the different research studies currently taking place in the national metrology institutes, one of the most promising paths seems to be "watt balance". Its principle consists of comparing a mechanical and an electromagnetic power. It is the result of a measure carried out in two steps:

– a static phase during which the Laplace force occurring on a conductor driven by an electrical current and subject to a magnetic induction field is compared to the weight of a mass standard;

– a dynamic phase where we determine the voltage induced at the terminals of the same conductor when it is moved through the same induction field at a known velocity. The determination of electric quantities by comparison with the Josephson effect and with the quantum Hall effect then makes it possible to connect the mass unit to the Planck constant. Even though the principle of the experiment remains simple and direct, obtaining a relative uncertainty that is low enough (1×10^{-8}) supposes the development, at the best level, of devices from different fields in physics.

In the static measurement phase, the coil on which the electromagnetic force acts and the load pan are suspended at the same end of the beam, and the other end is loaded with a counterweight.

One of the challenges with this experiment is to reduce the misalignment effects to a negligible level. In order for the balance of moments of force to be equivalent to the balance of applied forces at within 10^{-9} in a relative value for a beam arm length of 100 mm, the gap between both arms must be constant to within 0.1 nm, including for a mass off-centering of 1 mm on the load pan.

To reduce the effects of off-center loading and to ensure a constant length of beam arms, a monolithic device with flexure pivots axes was designed. This original device forms two identical gimbal systems each having two orthogonal, coplanar rotation axes, intersecting at a single point on the vertical symmetric axis of the device.

Experimental studies of static and dynamic mechanical behaviors of this device were carried out to characterize the stiffness of its flexure pivots and its dynamic behavior in terms of coupling between the two suspension systems.

We were able to determine the mechanical characteristics of this device for its application in the French watt balance experiment of which a demonstrator is under construction at the Laboratoire national de métrologie et d'essais in Trappes [1-3].

Watt balance experiment

Comparison of force

The watt balance experiment is based on the principle of virtual power to indirectly compare a mechanical power with an electric power. The idea was proposed in 1976 by the Englishman B.P. Kibble [4]. The experiment is conducted in two step, one is dynamic and the other static, the principle of the latter is the only one interesting us here.

Static phase: Figure 1 presents the diagram of the principle of the static phase. A horizontal wire conductor of length L, in which a current I is flowing, is placed in a magnetic field B homogenous and uniform so that the Ampère-Laplace F_e force acting upon it is vertical. The direction and amount of current I are adjusted so that this force, acting on the conductor suspended to one of the ends of a beam, compensates the force due to the weight P of a mass m suspended at the other end of the beam and subject to the acceleration due to gravity g. With hypotheses of orthogonality and perfect geometric configuration, the equilibrium of the equal arm beam balance corresponds to the following equation:

$$mg = BLI \tag{1}$$

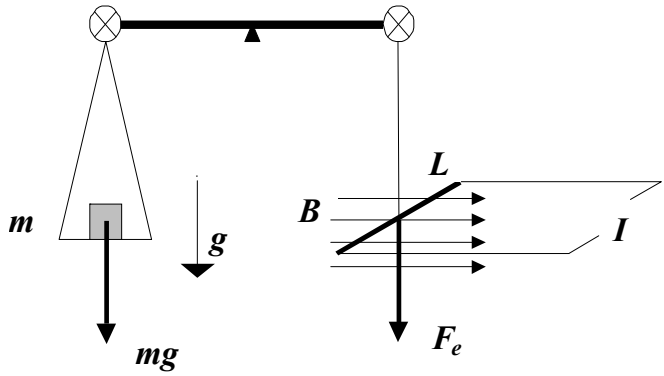


Figure 1. *Diagram of the static phase principle*

Application point and forces alignment: For this phase of static equilibrium, the conductor is actually suspended at the same end of the beam as the load pan carrying the transfer mass m (in the French experiment, mass m is 500 g). In this way, the electromagnetic force F_e and the weight P can then be applied at the same distance from the center of rotation of the beam. In addition, mass off- centering effects on the load pan must be insignificant and the electromagnetic force must be vertical, because a deviation of a few dozen micro-radians of the mass suspension line or of the coil in relation to the vertical would cause a significant gap in the length of the beam arm. And in this case, the balance of forces F_e and P moments would no longer be equivalent to the balance of these forces.

Load pan and coil suspensions

4-crossed axes system: In order to ensure the same application point of the electromagnetic and gravitational forces, we have designed an monolithic element made from copper-beryllium alloy (material chosen for its elastic properties). It consists of a double 2-crossed axis system (double gimbals) where its rotation axes intersect in one point within a micrometer (the four axes are coplanar at the allowed precision available by the machining process). Each rotation axis is materialized by two symmetric flexure pivots.

These rotation pivots aremachined in a monolithic part by spark machining. Their profile is concave circular with a 3 mm radius of curvature. The width of these pivots is 1.5 mm and their minimal thickness at the center is 40 μm. Figure 2 shows the perspective view of this complex piece.

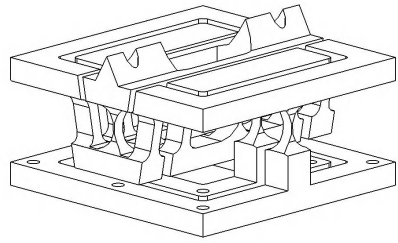


Figure 2. *Perspective view of the 4-crossed axis system*

Double suspension line: The stiffness of the 4-crossed axis system pivots is not negligible; this leads to a non-zero bending resistance of the pivots supporting the coil and transfer mass suspension lines, which leads to a coupling effect on the beam in the case of misalignment of forces.

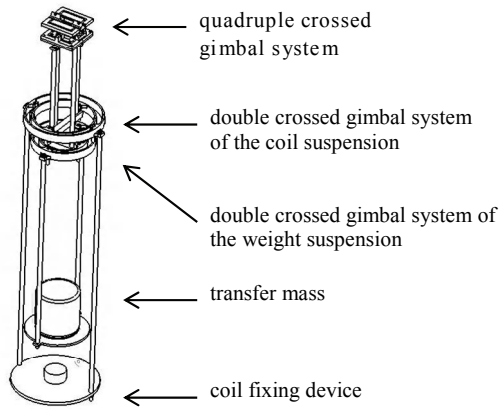


Figure 3. *Double suspension line of the mass pan and coil*

In order to reduce the static and dynamic couplings of the suspended elements (i.e. mass pan and coil), each suspension line is made up of two sections tiltable around point pivots offering a great mobility for this type of pivot [5]. This technique is widely used for mass comparators and balances in order to lessen the effects caused by the off-center loading [6-8]. Figure 3 presents the double suspension line fixed under the double gimbal designed for the mass pan (internal moving part) and coil (external moving part).

Flexure pivot system

Theoretical considerations

Flexure pivots: The flexure theory of a loaded pivot is now well known [9-14] and the advantages of using flexure pivots instead of knifes/ball bearings for mass comparators have been studied for many years [6, 9, 15, 16].

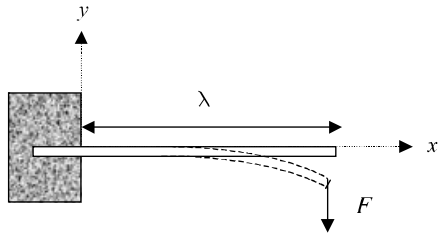


Figure 4. *Diagram of cantilever beam flexure under the effect of force*

In Figure 4, we consider a vertical force F applied at the free end of a cantilever beam, made from homogenous material with a Young's modulus E, a constant rectangular section (width w and thickness t), a length λ and a moment of inertia I.

For conditions with given limits, we obtain the expression of deflection f for $x = \lambda$, we have $y = f$ such that:

$$f = \frac{F\lambda^3}{3EI} \tag{2}$$

We deduct the stiffness k (where C is the stiffness coefficient of the pivot):

$$k = \frac{EI}{w} = \frac{F\lambda^3}{3fw} = \frac{C\lambda^3}{3w} \tag{3}$$

Stiffness measure

Measurement method: The measurement technique consists of horizontally setting one of the pivot ends (for which we want to measure stiffness) to the integral support of a micrometer screw lift table. This method is similar to the one used for the stiffness measurement of AFM tips [17]. The free end of the pivot rests either on a knife if it is a thin strip without end reinforcement or on a point in the

case of a reinforced end strip. This knife (or point) is laid on the load pan of a balance. The vertical translation of the pivot resting on the knife (or the point) enables it to bend if possible under the effect of the supporting force F applied by the knife (or the point).

From its horizontal position, by increasing the supporting force by vertical translation and without exceeding the pivot's elasticity limits, conventional values of mass m indicated by the balance can be directly converted into force by multiplying them by the conventional value of the acceleration due to gravity. We assume that the horizontal component of this force is negligible for very low deflection values.

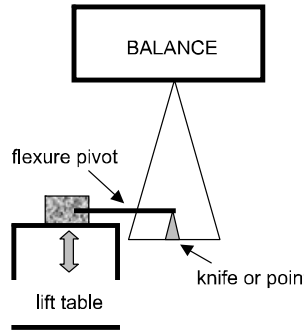


Figure 5. *Principle diagram of the stiffness coefficient measurement of a thin strip*

Deflection f is measured by the micrometer screw of the lift table and the distance λ between the fixed end of the pivot, and the point of application of force F is measured when the pivot is supported horizontally. For a thin constant section pivot, we observe a linear relation of the mass variation Δm indicated by the balance in relation with the position variation Δd given by the micrometer screw. In this case, we directly obtain a constant stiffness coefficient C given by:

$$C = g\frac{\Delta m}{\Delta d} = \frac{\delta F}{\delta f} \tag{4}$$

If the dimensional parameters of the pivot are constant and known, we can deduce from expression (3) the stiffness k, as well as the Young's modulus E of the pivot.

For non-constant section pivots, as is the case with the 4-crossed axes system pivots, the stiffness coefficient varies in relation to the deflection. We would then need to use analysis of data issued from finite elements modeling techniques to obtain the stiffness coefficient value for a given deflection. Nevertheless, for very

low flexure values, we can accept that the stiffness coefficient leans toward zero flexure.

Experimental results: In order to validate this method, a first measurement experiment was performed on a thin constant section strip made from stainless steel with the following characteristics:

mass *m*:	approximately 20 mg
thickness *t*:	10 µm
width *w*:	13 mm
length λ:	9 mm.

This same type of pivot is set up as central and end pivots on the force comparator prototype beam of the French watt balance experiment, hence the interest for this study.

Figure 6 presents curves of a series 1 of increasing flexure measurements and a series 2 corresponding to the return to zero flexure.

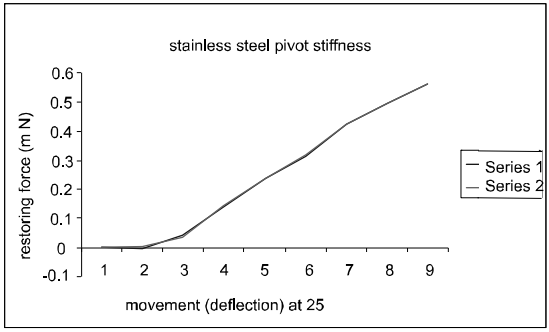


Figure 6. *Curves of restoring force applied by a 10 µm thick stainless steel pivot in relation to its deflection*

Once the transition corresponding to pivot weight compensation is passed, we obtain a quasi-linear variation of the restoring force in relation to the deflection. We observe no hysteresis effect, which means that flexure occurred only in the purely elastic region of the pivot. Calculation results in a stiffness coefficient *C* of 0.33 $N.m^{-1}$ and a stiffness *k* of 19×10^{-6} $N.m.rad^{-1}$. The relative measurement uncertainty is lower than 20%.

By knowing the pivot dimensions, we can express the inertia moment *I* for a constant rectangular section pivot and from expression (3), we can determine Young's modulus *E*. With value *k* determined above, we obtain an *E* value of

222,000 MPa. It is a value that is quite in accordance with the Young's modulus of stainless steel (approximately 200,000 MPa), to within uncertainties.

Because of error sources (pivot fixing constraints, directional constraints caused by throttling, etc.), the range of theoretical and experimental results is sufficiently coherent to validate the method.

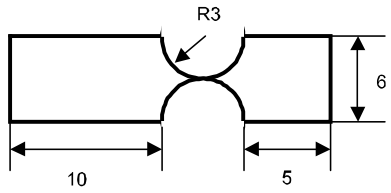


Figure 7. *Profile of the copper-beryllium alloy pivot of central thickness 40 µm and of width 34 mm used for the stiffness measurement*

This method was then applied to a pivot made with the same copper-beryllium alloy and profile as 4-crossed axis system pivots. Figure 7 shows the profile of the pivot used.

Figure 8 presents curves of the restoring force in relation to the deflection obtained from measurements. 1 and 2 downward were performed with decreasing flexure, whereas 2 upward corresponds to increasing flexure measurements.

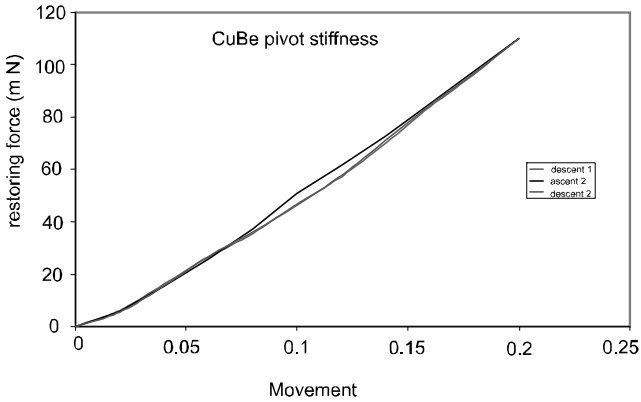


Figure 8. *Curves of the restoring force applied by a pivot in Cu-Be according to its deflection*

We observe a slight hysteresis effect which can be linked either to the pivot elasticity limit or to lift table problems or to a non-zero horizontal component of the

applied force. Curves do not seem to be linear. By carrying out polynomial regressions at order 2, we obtain expressions such as:

$$F = af^2 + bf + c \tag{5}$$

By choosing among them the regression where c has the lowest value ($c \approx -0.8$ mN) and by deriving expression (5), we obtain a stiffness coefficient C in $N.m^{-1}$ with f in mm such as:

$$C = \frac{dF}{df} = 2af + b \approx 1656f + 390 \tag{6}$$

Limit C value when f leans toward 0 is 390 $N.m^{-1}$. A regression at order 3 would have given a similar result to 370 $N.m^{-1}$. For very low flexure angles (lower than 100

μrad), we can have an estimation of the stiffness moment k from the expression (3) with limit C value. Width w of the pivot in the test is 34 mm. But the flexure here is only applied to the central part of the pivot. We determine a value of λ equivalent at approximately 6 mm which is different than the effective length of the flexure pivot. This results in a minimum value of the intrinsic stiffness moment of $k \approx 2.35 \times 10^{-3}$ $N.m.rad^{-1}$ for a 34 mm wide pivot. For flexure angles of up to 1 mrad, we would obtain a k value increased by approximately 3%, much lower than the relative measurement uncertainty which is 10%.

To verify if the stiffness moment range is correct, we calculate the Young's modulus from expressions (2) and (3), we find 146,250 MPa instead of the commonly accepted value of 135,000 MPa. This gap is not really significant in relation to the relative standard uncertainty of C which is 10% at best and according to approximations used.

For an important angular flexure, the stiffness coefficient C increases: in this case we can no longer accept that the pivot behavior is equivalent to the one from a constant section strip, but instead to the one of short strips. In particular, if the λ/t ratio becomes lower than 10, this leads to a behavior change of the pivot [18].

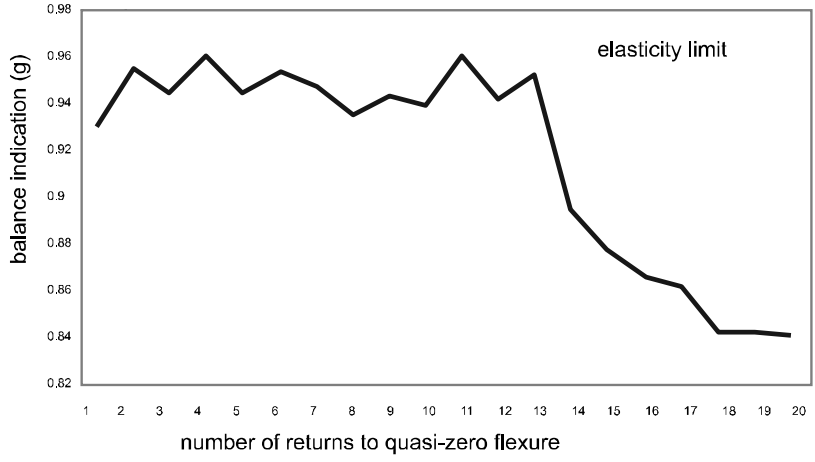


Figure 9. *Determination test of the elasticity limit of a copper-beryllium alloy pivot*

Figure 9 presents a determination test of the elasticity limit of a copper-beryllium alloy pivot. We carry out increasing flexures with a 20 µm deflection step. Between each deflection increase, we come back to an initial flexure as close to zero as possible and we log the balance indication. We observe a large decrease of this force once the pivot incurs a 240 µm deflection corresponding to an angular limit of elasticity of approximately 40 mrad.

Dynamic behavior study

Method of study: The principle of this method consists of the study of the behavior of an oscillating beam hung from the flexure pivot involved. The set is placed in a climate controlled room inside a glass cage to protect the oscillating beam from drafts and on a granite block to isolate it from vibrations. A oscillation triggering device is necessary to start the oscillating beam. We log oscillation signals from two perpendicular directions with an optical device made up of a fixed laser diode with a beam reflected by an integral mirror of the oscillating beam is captured by a 2D photodiode.

First test on simple pivot: We carry out a first experimental assembly with a beam where a central pivot is made up of a simple stainless steel strip of 10 µm wide. The experimental assembly principle diagram is given by Figure 10.

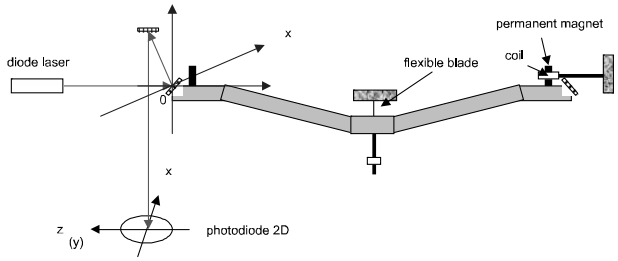


Figure 10. *Assembly for the dynamic study of a stainless steel stript used as central pivot of a beam*

Strainless steel strip characteristics are as follows:

strip thickness *t*:	10 µm
strip width *w*:	10 mm
strip length λ:	10 mm
Beam mass:	200 g
Beam length:	200 mm

The beam's center of gravity is a few millimeters below the beam pivot fixing.

We excite the beam with an electromagnetic device and we log low amplitude oscillations (≤ 10 mrad) in a vertical plane (xOz) of one of the beam's ends. Figure 11 shows a typical example log revealing an elliptical movement turning in the plane (xoz) of the beam extremity. It is a movement coupled in flexure and torsion of the strip. The oscillation frequency of the pseudo-carrier is approximately 0.6 Hz whereas the pseudo-modulation frequency for each mode, obtained by spectral analysis FFT, is 6 mHz.

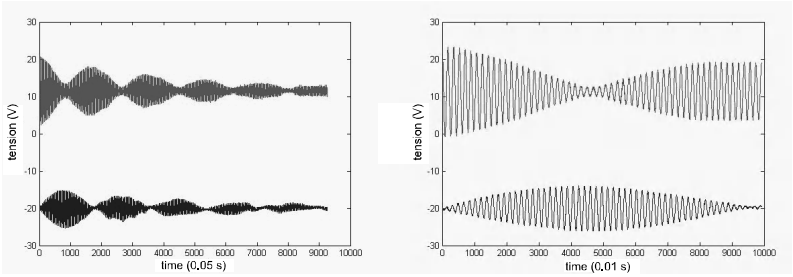


Figure 11. *Logs (left over 500 s and right over 100 s) exit signals of the 2D photodiode after beam excitation:- top according to the z axis (stript flexure)- bottom according to the x axis (strip torsion)*

The origin of this flexure-torsion coupling can be caused either by a component x of the electromagnetic excitation pulse, or by a horizontal position problem x of beam clamps pinching the strip, or even, by asymmetric constraints in the flexure pivot. The temporal representation shows damping which can come from viscous friction with air, internal flexure pivot constraints or Foucault 's currents which could be caused by the oscillating magnet system.

4-crossed axis system test: The four-crossed axis system is fixed to the fixed bed, we hook under each double gimbal a simple pendulum with a mass of 500 g and 20 cm long. As with the previous assembly, the set is placed in a climate controlled room in a glass cage and a granite block.

A mechanical device makes it possible to move the internal pendulum from its resting position based on the axis of rotation yy' of the internal elastic connection and to release the pendulum based on the axis xx' without initial force. Figure 12 presents a schematic view of the assembly.

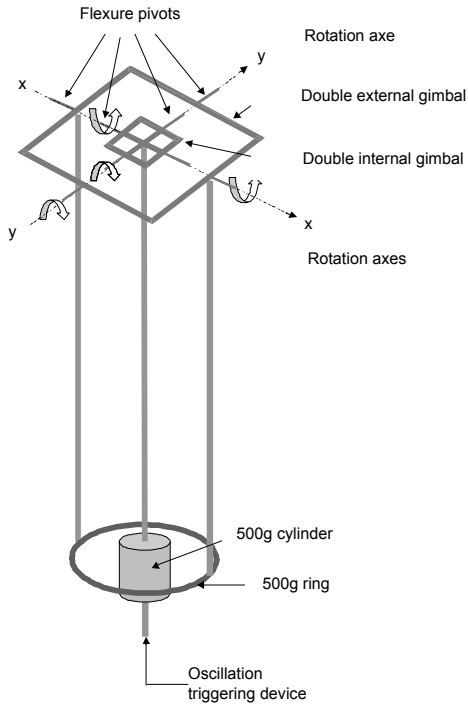


Figure 12. *Schematic view of the 4-crossed axis system assembly to which two simple pendulums are attached (length: 20 cm; mass: 500 g)*

A mirror fixed at the end of each pendulum reflects the laser diode beam on a 2D photodiode making it possible to follow oscillations. This device has a sensitivity of 125 µrad/V.

The theoretical period of this type of simple pendulum is 0.9 s with a frequency of 1.1 Hz. Even though the assembly is installed in a block of granite in a climate controlled room, vibration frequencies of approximately 1 Hz cannot be filtered by the block of granite. We observe permanent residual oscillations with very low amplitude (≤ 40 µrad) caused by noise.

The example in Figure 13 shows almost perfect decoupling between both gimbals. We note that oscillations based on x and y are almost in quadrature, which could be explained by a horizontal position problem of the 4-crossed axis system or by a faulty mechanical device providing an initial component based on the yy' axis.

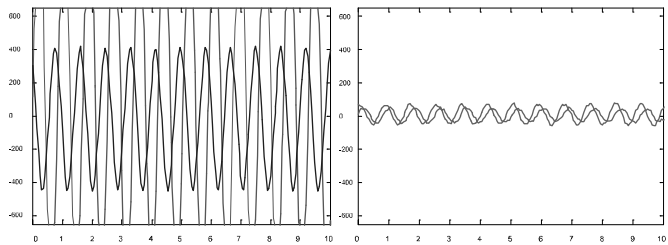


Figure 13. *Oscillation log example (max ± 700 µrad amplitudes) over 10 seconds:- left: of the internal pendulum (started) based on xx' (continuous line) and based on yy' (dotted line);- right: of the external pendulum based on xx' (continuous line) and based on yy' (dotted line)*

The spectral FFT log analysis of the example given in Figure 13 shows that the four oscillations present the same main frequency of 1.2 Hz practically corresponding to the theoretical frequency of simple pendulums. It also shows secondary frequencies with significant amplitudes between 3 and 5 Hz as well as a low amplitude frequency of approximately 40 Hz which does not seem to come from the measurement device noise.

Conclusion

A dynamometric method was developed to measure the stiffness coefficient of a thin flexure pivot. It makes it possible to measure an intrinsic stiffness of 19×10^{-6} N.m.rad^{-1} with a relative uncertainty of approximately 20% for a stainless steel pivot that is 10 µm thick, similar to the one used on the beam from the French watt balance experiment.

This same technique was applied to a copper-beryllium alloy pivot of the same type as the 4-crossed axis system pivots developed for the French watt balance

experiment. It made it possible to determine a pivot stiffness of 2.35×10^{-3} N.m.rad^{-1} for a 34 mm wide pivot with a relative uncertainty of below 20% over an angular range of ± 100 µrad. We determined an elasticity angular field of ±40 mrad for this pivot.

The dynamic behavior study of the 4-crossed axis system in copper-beryllium alloy has shown that decoupling between both gimbals is satisfactory. With pendulum systems that have a theoretical inherent frequency similar to 1 Hz, there is an experimental main frequency of 1.2 Hz, as well as secondary frequencies between 3 and 5 Hz and a low amplitude frequency of 40 Hz.

Even though the environmental conditions of this study are not optimal, this study will be continued by placing the assembly in vacuum, in a non-vibration system that is heavier and better isolated and by fixing it to a sturdier "hexapode" type fixed bed.

This device could be implemented in other fields of metrology, and also find other technological applications in high precision mechanisms.

References

[1] M. Lecollinet, Ch. Bordé, M. Chambon, A.Clairon, T. Coorevits, N. Feltin, G. Genevès, A. Gosset, P. Juncar, P. Pinot, F. Platel, "Vers une balance du watt française", *Congrès International de Métrologie conference papers*, 2001, St Louis, France.

[2] F.Alves, M. Besbes, A. Clairon, L. Chassagne, J. David, A. Gosset, P.Gournay, D. Holleville, P. Juncar, A. Landragin, M. Lecollinet, F.Pereira Dos Santos, P. Pinot, S. Topçu, F. Villar, G. Genevès, "La balance du watt du BNM: état d'avancement", *Congrès International de Métrologie conference papers*, 2003, Toulon, France.

[3] G.Genevès, P.Gournay, A.Gosset, M.Lecollinet, F.Villar, P.Pinot, P.Juncar, A.Clairon, A.Landragin, D.Holleville, F.Pereira Dos Santos, J.David, M.Besbes, F.Alves, L.Chassagne and S.Topçu, "The BNM watt balance project", *IEEE Trans. Instrum. Meas.* Vol. 54 no. 2 - Special issue on CPEM 2004, pp 850-853, 2005.

[4] B. P. Kibble, *Atomic Masses and Fundamental constants* **5**, ed. J.H. Sanders and A.H. Wapstra, New-York Plenum, 1976, pp 545-551.

[5] P. Pinot, G. Genevès, D. Haddad, J. David, P. Juncar, M. Lecollinet, S. Macé and F. Villar, "Theoretical analysis for the design of the French watt balance experiment force comparator", *Rev. Sci. Instrum.*, **78**, 095108, 2007, 11p.

[6] A. Picard, "The BIPM flexure-strip balance FB-2", *Metrologia* Vol. 41, pp 319-329, 2004.

[7] H. A. Bowman and L. B. Macurdy, "Gimbal device to minimize the effects of off centre loading on balance pan", *J. Res. NBS* Vol. 64 C, pp 277-9, 1960.

[8] H. A. Bowman and H. E. Almer, "Minimization of the arrestment error in one-pan, two-knife balance systems", *J. Res. NBS* Vol. 67 C, pp 227-35, 1963.

[9] F. S. Eastman, "Flexure pivots to replace knife edges and ball bearings", *Engineering Experiment Station Series*, Bulletin no. **86**, University of Washington, Seattle, 1935.

[10] F. S. Eastman, "The Design of Flexure Pivots", *J. Aeronautical Science* Vol. 5, pp 16-21, 1937.

[11] W. H. Wittrick, "The Theory of Symmetrical Crossed Flexure Pivots", *Aust. J. Sci. Res. A*, Vol. 1, pp 121-134, 1948.

[12] W. H. Wittrick, "The Properties of Crossed Flexure Pivots, and the Influence of the Point at which the Strips Cross", *Aeronautical Quarterly*, Vol. **11**, pp 272-292, 1951.

[13] W. D. Weinstein, "Flexure-Pivot Bearings", *Machine Design*, Vol. 37, pp 136-145, 1965.

[14] P. H. Sydenham, "Elastic design of fine mechanism in instruments", *J. Phys. E: Sci. Instrum*. Vol. 17, pp 922-930, 1984.

[15] T. J. Quinn, C. C. Speake and R. S. Davis,"A 1 kg Mass Comparator Using Flexure-Strip Suspensions: Preliminary Results", *Metrologia,* Vol. 23, pp 87-100, 1986/87.

[16] T. J. Quinn, "The beam balance as an instrument for very precise weighing", *Meas. Sci. Technol.*, Vol. 3, 141- 159, 1992.

[17] M.-S. Kim, J.-H. Choi, Y.-K. Park and J.-O. Kim, "Atomic force microscope cantilever calibration device for quantified force metrology at micro- or nano- scale regime: the nanoforce calibrator (NFC)", *Metrologia,* Vol. 43, pp 389-395, 2006.

[18] S. P. Salisbury, R. Ben Mrad, "Analytical stiffness estimation for short flexures", *Mechatronics*, Vol. 16, pp 399-403, 2006.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The practical realization of the meter in Portugal

F. Saraiva, S. Gentil

Laboratório Central de Metrologia, Instituto Português da Qualidade
Rua António Gião 2, 2829 – 513 Caparica, Portugal; e-mail: fsaraiva@mail.ipq.pt; sgentil@mail.ipq.pt

ABSTRACT: This paper reports the measured values of the laser systems IPQ2 and IPQ3 performed during recent years at IPQ for the evaluation of the output frequency and the sensitivity coefficients of the lasers: the iodine temperature and pressure, the modulation width and the power coefficients. These parameters are required when the lasers are used as a standard in applications such as frequency calibrations.

The laser systems IPQ2 and IPQ3 are He-Ne lasers stabilized on the saturated absorption of $^{127}I_2$ at $\lambda \cong 633$ nm which are able to satisfy requirements set by the Comité International des Poids et Measures (CIPM) [1] and they represent one of the two possibilities to perform the practical realization of the definition of the meter.

At IPQ, the other possibility to perform the practical realization of the definition of the meter is by means of the optical comb frequency synthesizer (OFS) a femtosecond Kerr-lens mode-locked laser with an optical fiber based on photonic crystals. This device links the optical frequencies to the frequency of the Cs microwave standard (the national standard for the second).

This paper also describes the validation process of the frequency measurements performed at IPQ, through the comparison of its own frequency values with the ones measured by absolute frequency using the OFS and with the values obtained in the international key comparisons participations at the BIPM (BIPM.L-K10 and BIPM.L-K11).

Introduction

The He-Ne laser, operating at 633 nm with frequency stabilized by saturated-absorption in an iodine cell to hyperfine components in the R(127) 11-5 transition, has been widely used for the practical realization of the definition of the meter following the specifications defined in the list of recommended radiations. As the frequency of the He-Ne/I_2 laser operating at 633 nm is given for specified values of the laser parameters, it is necessary to evaluate the laser frequency dependence on parameters such as the iodine temperature and pressure, the modulation width and the power.

The results of sensitivity coefficient measurements of the two He-Ne laser systems, IPQ2 and IPQ3, are presented as well as the frequency stability measurements.

Since the redefinition of the SI unit of time, defined in terms of the frequency of the caesium clock transition (≅ 9.2 GHz), and the unit of length, defined in terms of

a wavelength, the possibility of connecting the two base units of length and time through the relationship $c = \lambda f$, where λ is the wavelength and f is the frequency has been analyzed. That became possible with the development of the stabilization of femtosecond laser frequency combs.

The OFS, recently acquired by the Primary Length Laboratory (LCO) at IPQ, based on a femtosecond laser frequency comb, allows the possibility of realization of a primary wavelength by direct measurement of the laser frequency and the links between an optical frequency in the visible region of the spectrum and the primary atomic caesium clock.

Following the measurement procedures, we present the installation of the OFS at LCO.

Practical realization of the definition of the meter

At LCO, the practical realization of the definition of the meter can be performed using two methods:

1. by means of the wavelength in vacuum λ of a plane electromagnetic wave of frequency f; this wavelength is obtained from the measured frequency f using the relation $\lambda = c_0 / f$ and the value of the speed of light in vacuum $c_0 = 299{,}792{,}458$ m/s;
2. by means of one of the list of radiations recommended by the CIPM, whose stated wavelength in vacuum or whose stated frequency can be used with the uncertainty showed, provided that the given specifications and accepted good practices are followed.

Primary length standard

The primary length standards are:

– for method 1: OFS (Figures 1 and 2) constituted by: an Nb:YVO4 laser; an optical acoustic modulator, Kerr-lens mode-locked femtosecond Ti:S laser and an optical fiber based on photonic crystals and an optical and electronic set that allows us to determine the two frequencies that characterize the OFS, the offset frequency, f_0, and the repetition frequency, f_{rep}.

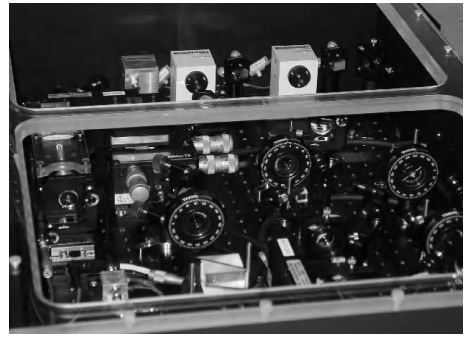


Figure 1. *Opto-mechanical setup of the OFS*

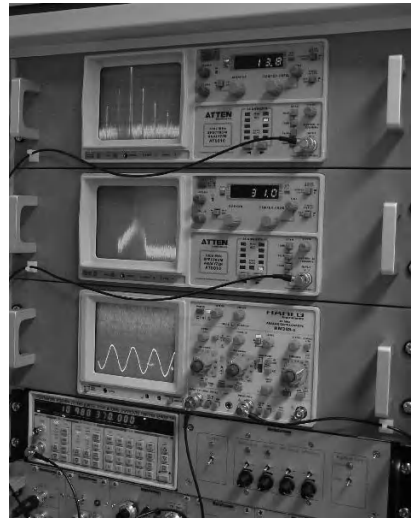


Figure 2. *Control unit of the OFS*

This system is in line with the rule for the practical realization of the definition of the meter by means of the wavelength in vacuum λ of the plane electromagnetic wave of frequency f and with the value of the light in vacuum c_0 = 299,792,458 m/s;

– for method 2: two He-Ne laser stabilized by molecular absorption in hyperfine iodine transition, the IPQ2 and IPQ3, with the following parameters.

Laser standard/ parameter	IPQ2 (Thomson)	IPQ3 (Winters)
Modulation frequency	1.092 kHz	8.33 kHz
Modulation width peak-to-peak	6.01 MHz	6 MHz
Laser cavity length	39.5 cm	26 cm
Mirror curvature *R*1 (tube side)	60 cm	30 cm
Mirror curvature *R*2 (cell side)/cm	∞ (plane mirror)	∞ (plane mirror)
Mirror transmission T1 (tube side)	0.9%	0.7%
Mirror transmission T2 (cell side)	0.9%	0.25%
Iodine cell	BIPM#255	BIPM #445
Cell length/flat windows/origin	10 cm, Brewster, BIPM	10 cm, Brewster, BIPM

Table 1. *Laser system parameters*

The operation of the lasers under the conditions specified in the recommendations is a sufficient condition to ensure that their absolute frequency lies inside the given uncertainty. However, it is important to know the sensitivity coefficients of each laser to evaluate the changes in frequency and/or uncertainty.

Beat-frequency measurement technique at IPQ

The frequency measurements between IPQ2 and IPQ3 were performed using the beat-frequency technique, in which the beam of the two lasers was directed onto an avalanche photodiode (Figure 3). The frequency intervals were measured for all combinations of the components d, e, f, g of the R(127) 11-5 of $^{127}I_2$. The frequency differences between the two lasers, Δf, were measured using the matrix measurement technique [3].

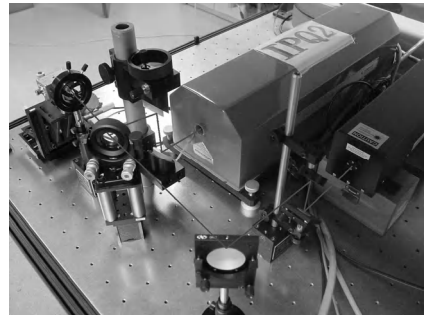


Figure 3. *Beat-frequency between IPQ2 and IPQ3*

Sensitivity coefficients

The sensitivity coefficients for the modulation width, $\Delta f / f_w$ (kHz/MHz), pressure effect $\Delta f / \Delta p_{I_2}$ (kHz/Pa) and power effect, $\Delta f / \Delta p$ (kHz/μW) were obtained by linear regression analysis.

The sensitivity coefficients results obtained for the IPQ3 are shown in Figures 4 to 6:

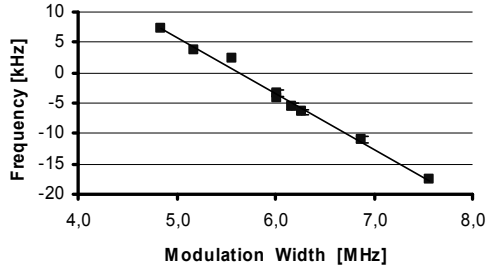


Figure 4. *Modulation width,* $\Delta f / f_w$

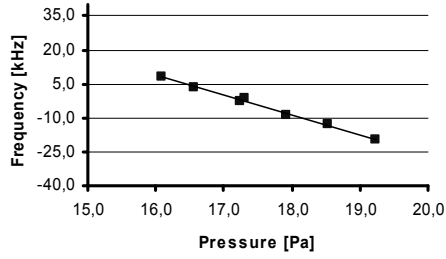


Figure 5. *Pressure effect,* $\Delta f / \Delta p_{I_2}$

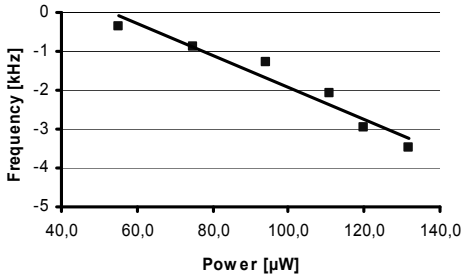


Figure 6. *Power effect,* $\Delta f / \Delta p$

These results are presented in Table 2.

Laser	IPQ2	IPQ3
Parameter	Sens. Coeff.	Sens. Coeff
Modulation width (kHz/MHz)	-8.3	-9.25
Iodine pressure (kHz/Pa)	-8.5	-8.78
Power (kHz/µW)	-0.15	-0.041

Table 2. *Summary of the values of the sensitivity coefficients for the two laser systems*

Output frequency of IPQ3

Using the frequency value of the laser IPQ2 obtained from the BIPM Certificate [2] and used as reference, we performed several matrix measurements with the pair of lasers. The results obtained for the output frequency of IPQ3 were:

$$f_{IPQ3} - f_{IPQ2} = 0.0 \text{ kHz}; \qquad u_c = 4.9 \text{ kHz}$$

$$f_{f(IPQ3)} = 473{,}612{,}353{,}602.7 \text{ kHz} \qquad u_c = 4.9 \text{ kHz}$$

This result shows that the frequency of the lasers fluctuates within 1 part in 10^{11}, which is less than the 1 σ estimate standard uncertainty given in the *Mise en Pratique*.

Traceability

Another point is the traceability of these primary standards. They are achieved through international comparisons made regularly or by absolute frequency measurement.

The participations in two international key comparisons at the BIPM, the BIPM.L-K10 (2005) [2] and BIPM.L-K11 (2006) ongoing comparison, gave us the value of the f component of the R(127)11-5 transition of IPQ2 and IPQ3 by absolute measurement frequency.

The results were:

$f_{f(IPQ2)} = 473, 612, 353, 602.7$ kHz $\qquad u_c$=0.9 kHz

$f_{f(IPQ3)} = 473, 612, 353, 601.5$ kHz $\qquad u_c$=0.9 kHz

These two values are the standard frequency reference at LCO, for the dissemination of the unit.

Absolute frequency measurement at IPQ

The LCO acquired the Optical Frequency Synthesizer MenloSystems FC 8004 which is based on a femtosecond laser frequency comb and the installation of this system is currently being performed at the LCO [4].

The output of a mode-locked laser consists of a train of consecutive pulses showing phase shift between the carrier and envelope. This shift evolves from pulse to pulse by an amount proportional to an offset frequency (f_0). Pulses are emitted at the repetition rate (f_{rep}). Both f_0 and f_{rep} require stabilization. To measure the offset frequency, a mode of frequency $nf_{rep} + f_0$ on the red side of the comb is doubled to a frequency $2 \times (nf_{rep} + f_0)$. As the comb contains more than an optical octave frequency range, there will be a mode with the mode number $2n$ oscillating at frequency $2nf_{rep} + f_0$. The repetition rate and the offset frequency are both phase-locked to an atomic clock.

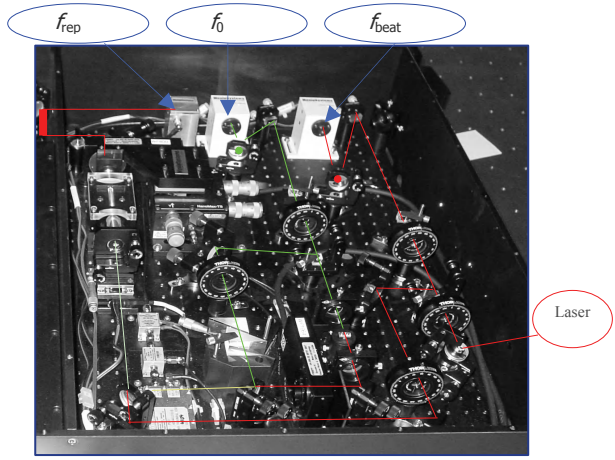


Figure 7. *Scheme of frequency measurements at the OFS*

The microwave frequency, f_{rep}, is measured via photodetection of the output pulse train from the femtosecond laser. The offset frequency, f_0, resulting from differential group and phase delay due to the dispersion in the laser cavity is determined with the self-referencing technique [6]. The output frequency of an external laser is measured by beat detection, f_{beat} (Figure 7).

First results with the OFS at IPQ

The first result obtained by absolute measurements frequency, at IPQ, for the output frequency of IPQ3 was [4]:

$f_{f(IPQ3)} = 473, 612, 353, 607.0$ kHz $\qquad u_c$=0.3 kHz

With the following parameter of the OFS:

f_0 = 20 MHz; f_{rep} = 202, 200.022 kHz, n = 2 342 296

Note: The CCL_{2001} value of the a_{16} (f) of $^{127}I_2$, is [1]:

f = 473, 612, 353, 604(10) kHz

Conclusion

International comparisons and collaborations allow improved calibration methods. The LCO has currently as its priority the continuous measurements of the absolute frequency of the two laser systems, IPQ2 and IPQ3 using the OFS. This work will validate the laser frequency measurements performed at IPQ by beat frequency and absolute frequency measurement with the Optical Comb Frequency Synthesizer, when we compare these values with those obtained in the international Key Comparison.

Acknowledgments

The authors would like to thank the contributions of various individuals to this work. In particular they acknowledge Rui Saraiva, Rui Gomes and colleagues (Laboratório Central de Metrologia) for their contributions for stimulating the work in this field.

References

[1] Quinn, T.J. "Pratical realization of the definition of the meter, including recommended radiations of optical frequency standards (2001)", *Metrologia*, Vol. 40, pp.103-133, 2003.

[2] Robertsson L., Zucco M., Saraiva F., Gentil S., et al,"Results from the CI-2004 campaign at the BIPM of the BIPM.L-K11 ongoing key comparison", *Metrologia*, Vol.42, 04002, Technical supplement 2005.

[3] Prieto E., Saraiva F., Chartier J.-M. "International comparison of stabilized He-Ne lasers by the saturated absorption of 127I2 at l 633 nm involving the CEM (Spain), the IPQ (Portugal) and BIPM", *Metrologia*, Vol. 32, pp 379-384, 1995/96.

[4] Saraiva F., Gentil S. " Primary Length Labotatory and the Length Traceability Chain in Portugal", *Proceedings of the 12th International Congress of Metrology*, 2005.

[5] Saraiva F., Ferreira J. "A realização prática da definição do metro no Laboratório de Comprimento através do Optical Frequency Synthesizer", *Actas da 1ª Conferência Nacional da SPMet*, 2005.

[6] Hollberg L., et al, "The measurement of optical frequencies", *Metrologia*, Vol. 42, pp S105-S124, 2005.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Toward an important evolution of the international system of units: what practical impact?

Marc Himbert

LNE-INM, Conservatoire national des arts et métiers
Case I 361, 61 rue du Landy F 93210 La Plaine Saint Denis, France
himbert@cnam.fr

ABSTRACT: Among the international community of metrologists, several committees have emphasized the usefulness of an evolution in the definitions of the units in the international system of units (SI). The aim is to take into account the most advanced progress in physical theories, and the best experimental determinations of the fundamental physical constants. The present paper presents these projects and gives insights on the expected impact of this evolution.

Introduction

Measuring means comparing the unknown property of a system to a reference with the help of an instrumental chain. The confident use of the result presumes that the confidence granted in particular to the reference is established with no ambiguity: that is how we can be sure of the traceability of measuring results. In most industrialized nations, references used for measurements, controls, tests or analyses are the subject of documented connections to standards materializing units of the international system of units (SI), a conceptual and practical reference system implemented and "maintained" (in the active sense) under the control of an international treaty, the *Meter Convention*.

Technological developments of the last 50 years and diversification of measurement fields have resulted in the multiplication of reference material or systems, better adapted to implement the tools for ensuring the first phase for bringing about traceability of results, out of proximity concern. It is nonetheless imperative to ensure connections to references implemented in the context of SI as often as possible.

The international system of units (SI)

The international system of units must satisfy operations as well as universality requirements: what would be the use of unusable references in practice, or unrecognized references by the scientific, technical, economical and social community? During the last decades, putting units into practice (and sometimes their definition) evolved in the sense of a varied consideration for modern physics, depending on the fields, but always increasing. Model progress underlying the manifestation of quantum effects, progress achieved in the experimental current determination of fundamental physical constants leads to considering important evolutions of unit definitions in order to make them more permanent, more uniform and more precise in their practice. We need to understand their impact.

Disciplinary fields covered

The object of SI is to offer the community a system of units covering the magnitudes involved in the field of electrodynamics, thermodynamics, chemistry, biochemistry and from living sciences from a traditional choice of definitions of a few "base" units; a system covering, if needed, magnitudes linked to human perception; a coherent system, making it possible to accept by a single suite of physical laws any magnitude from "bases"; a rationalized system, where these laws do not include untimely numerical factor. The SI today achieves this goal from definitions given to units of seven basic magnitudes: length, time, mass, electric current intensity, thermodynamic temperature, product quantity, light intensity.

Current definition of units

These definitions are very varied in age and type: the kilogram definition is still based on the choice of an international prototype (kg 1889), the meter prototype corresponds to the numerical value freeze (in meters per second) of light velocity c in space (m 1983); the second definition involves a cesium atomic clock operating in microwave (s 1967/68); the ampere is obtained from the Laplace electro-mechanical law, actually setting the magnetic permeability value of space (A 1948); kelvin reports the thermodynamic temperature (measurable) to the triple point of pure water, where conventional value ensures proximity with older scales best (K 1967/68); the mole ensures connection with chemistry by mass listing of a collection of carbon atoms (mol 1971); candela enables the radio-metric measurement transposition of radiation in the photometric field, where the Commission internationale de l'éclairage validates main aspects (cd 1979).

The evolution of definitions

Competent organs

Each of these definitions presents "pathological" aspects, suggesting improvement paths. The evolution of unit definitions are under the control of the Conférence générale des poids et mesures (CGPM) mandated by its Comité international (CIPM) and is taught within Comités consultatifs (CCs) thématiques of CIPM, and its Comité consultatif des unités (CCU).

The meter example

The meter example explains which trends have been followed long term: with universal vocation and linked to the terrestrial meridian length in the Republic year III, it is in relation to an "international prototype" in 1889; then we must wait until 1960, well after the appearance of the spectroscopic and quantum revolution, for the meter to be linked to the wavelength of a specific transition in a defined isotope from the krypton atom; and 20 years later, because of the limitation imposed by the implementation of this definition in 1960 on performances of c measurement experiments, the meter becomes the length of the path traveled by light in space during a certain amount of time, which sets the numeric value of c in SI units.

Fundamental constants

The evolution of units is firmly considered within modern physics through fundamental physical constants that it introduces in its models. These are permanent and uniform constants in the current state of knowledge that can be verified experimentally; the combined adjustment of numeric values of constants, which are dependent on one another, is proposed by the international group Codata which regularly collates the results published in scientific literature.

Instrumentation

Evolutions in definitions of units, or in recommendations for their implementation, is not neutral for measurement instrumentation: they translate the advanced technological emergence on which, long term, new measurement or reference establishment principles are based on. We can for example mention positioning measurements via signals from slave clock constellations, which use the definition of the meter; we can also mention the "Josephson chip" electric tension references, resulting from recommendations made in 1990.

What perspectives?

Research is being carried out in several directions to contribute to SI improvement, with – and this is just a few decades old – strong implication from national metrology laboratories. They are done at the same time as a prospective scientific reflection brought to a level that has rarely been reached (Académie des sciences, etc.). This is liable to create an international system of units with greatly modified definitions in the sense of much stronger anchoring into fundamental physics and of conceptual simplification of definition and traceability diagrams in a timeline that is to be determined.

The second

For the second, the use of clocks working directly in the optical field (at frequencies 50,000 times higher than with today's cesium) seems promising. All other units associated magnitudes known today as "basic" can experience deeply modified definitions, or can even be replaced by units of other magnitudes.

Kilogram, mole and Avogadro constant

Modifying the definition of mass unit – far from scientific criteria and yet so close – constitutes one of the first challenges. By nature not necessarily permanent, the international prototype today can guarantee traceability of measurements at the per billion scale part in relative value, i.e. at microgram scale, thanks to a century of in-depth studies. Two main research directions were carried out to attempt to equal such performances.

Counting atoms or ions

The first group of experiments is based on counting atoms, ions, individual particles of atomic size. By counting, we can create a link between the mass of an atom or a basic particle (field in which we could choose a permanent mass as reference and for which relative mass comparison can be done, in mass spectrometers, at better than 10^{-9}) and a macroscopic mass. It is actually measurements, by counting, of what we call today the Avogadro (or Loschmidt) constant N_A. Among the experimental leads explored in this path, the construction of a mono-crystalline silicon sphere by a predominantly European consortium is one of the most advanced projects: the measurement of volume, mass, surface roughness, chain parameter of the crystalline network, rate of gaps, impurities or dislocations, etc. today leads to a relative certainty result in the range of 10^{-7}. Another counting experiment with heavy ions in a jet, although conceptually promising, is far from being exact.

Results

The silicon sphere leads to a result that is incoherent today, in the range of 10^{-6} in relative value with the approach explained below. One of the problems being studied during the experiment seems to be the availability of silicon with controlled isotopic purity. In any case, no SI evolution would be considered before this problem is analyzed, understood and solved.

Clearly, any new N_A measurement at high precision level comes down to directly clarifying what we mean by "mole", and opens the door to a definition of the unit of matter quantity that is more compliant to its "object counting" nature for a macroscopic expression of phenomena occurring at molecular level.

Kilogram, electric references and Planck constant

Electric references

The second group of experiments takes advantage of advances from the last 20 years in electric metrology references which should be recalled. The first two advances are already integrated within the SI as conservation standards: their reproducibility and stability are established; their precision is limited by the (difficult) experimental comparison to SI units; metrologists introduced experimental constants in connection with these effects (K_J, R_K) which have had recommended numeric values since 1990. These standards involve the macroscopic manifestation of quantum effects: Josephson effect, which by a *simple* frequency measurement leads to uniform and permanent electric tension references, reproducible at much better than 10^{-9}; quantum Hall effect, leading to the development of resistance references to analog qualities; with the later addition of the mono-electronic effect, where we can command and control load movement, one by one, between the parts of a circuit, which results in simple frequency measurement current etc. Experiments are currently in process to validate the mutual coherence of these three effects, by closing the "quantum electric metrology triangle" where current can be accessed either directly or with the Ohm law. Today the connection could be included in a relative performance in the range of a few 10^{-8}.

Underlying theories

The theoretical description of these effects involves models explaining values of K_J and R_K, models where fundamental physical constants e (basic load) and h (Planck constant leading to the concept of energy quantum of photon hν) occur. More precisely, the Josephson effect comes down to "counting" tensions in units of h/2e. The quantum Hall effect is associated with resistance levels with sub-multiple values of h/e^2. Mono-electronic devices give access to e by temporal counting. One of the limitations encountered today is to be able to appreciate the relevance of these models, where the theoretical approach is not in question, but where the

experimental validity is based on experiments with a relative precision in the range of 10^{-7}. That is particularly the case for the "volt balance" experiment, during which we balance by weighing the force applied to capacitor reinforcing with a load tension measured by the Josephson effect. The experimental property of a capacitor with capacity that can be calculated (with the help of the Lampard-Thomson theorem) from dimensional measurements and current meter and ampere definitions, today leads to somewhat better results. Finally, the experimental implementation of the quantum Hall effect in this theoretical context can lead to a determination of the finite structure constant α, constant without dimension which controls the construction of atomic buildings; this determination is in accordance with the other determinations of α (*ab-initio* calculations, atomic physics measurements) at the scale described above.

The watt balance experiment

Five watt balance experiments are currently in process: at NIST, NPL, LNE, METAS and BIPM. They rely on the following principle: by weighing, on one hand, the force applied to an electric circuit placed in an active magnetic field, and by measuring on the other hand (or simultaneously) the electromotive force appearing at the limits of this circuit when it moves in the same field, we can learn the geometric imperfections of the circuit and field and directly connect the mass used by the weight, with the help of kinematics measures, to electric power etc. i.e., using quantum electric standards to validate behavior models, with only the Planck constant.

Almost all aspects of these experiments are technically critical; for the French experiment, we can cite permanent magnet and thermal dissipation, transfer mass standard (in gold), going from empty to air for this standard, flexible suspension balance(!), alignments, g measurement by atomic interferometry, command and movement speed measurement at sub-nanometric scale by optical interferometry.

The watt balance experiment can be interpreted as a measurement of the Planck constant, as a stability study for mass standards, or as a validation experiment for electric quantum effect models involved in its operation.

Possible evolutions

We can conclude that if uncertainties emerging from the experiments improve, the SI evolution would occur with the grouping made up of the definition of the mass unit and the unit chosen to access electric magnitude units. A consensus seems to have formed to direct the evolution of the mass unit toward setting the numeric value of the Planck constant; the debate remains very open for two important elements however:

- the first one is actually minor and involves the statement of a new definition corresponding to this freeze;

- the second one is more fundamental and involves the choice of the "other" redefined unit belonging to the electric field: set the electron e load (it is the pragmatic viewpoint of electrician metrologists, for immediately storing in this field precision improvements), or setting the "characteristic impedance of empty space", which comes down to, because of the meter definition maintenance, not changing anything to the current situation where the definition of the ampere freezes μ_0 – maybe even the statement of the definition.

The debate is not completely clear-cut today. It is also possible that the redefinition of the mass unit is perceptible to the highest class of mass reference users.

Kelvin and the Boltzmann constant

The realization of the statistical thermodynamic could lead to consider, through a numeric value freeze of the Boltzmann k_B constant, a definition of kelvin which would not be hindered by the triple point of a specific pure matter, with an implementation that is always difficult (the CIPM recently clarified the recommended composition of water involved in the implementation of the fixed point).

Two wide experimental paths to access k_B are explored today.

UHR spectroscopy

In an extremely high resolution spectroscopy experiment, we can deduct k_B from the determination of the form of absorbed spectral line and of the evaluation of its width, which depends on the quantum of thermal excitation k_BT. The first results are very promising.

Speed of sound in cavity

The speed of sound in gas translates the competition between schedules movement and the thermal agitation of the environment. A slightly warped spherical cavity study is carried out to access a sufficient precision; the principle is based on the comparison of light sound and light speeds in the same geometry; speeds are obtained by the measurement of modes resonance linked to micro-warps, we can, in large measure, ignore dimensional properties. These perspectives are also promising.

Measurement methods

These evolutions of the International system currently only constitute decade projects. It is difficult to simply forecast in what measure they will have an impact beyond the implementation of national references for different countries on daily instrumentation. Although it would be illusory to presume that this impact could be insignificant. In fact, the evolution of technical methods precedes and follows the

adjustment of unit definitions. As with today's positioning by satellite, which directly involves the definition of the meter, we can consider that principles of electric measurement or mass measurement will be affected, and even more today using these modifications. Better established on the bases of modern physics, references will also be better adapted to new technologies based on nano-sciences.

Conclusion

The international system of units could be reorganized based on advances in physics theories, or rather based on the improvement of the experimental knowledge of fundamental physical constants involved in validated models in quantum or statistical physics. The challenge is important and corresponds to a historical period of the system of units evolution. Thinking that such conceptual evolutions have no influence on measurement methods implemented by users would be a mistake, because long term, it is established that instrumental technologies broadcast through traceability chains.

References

[1] C.J.Bordé, Base units of the SI, fundamental constants and modern quantum physics, *Phil. Trans. Roy. Soc.*, vol. A 363, p. 2177-2201, 2005.

[2] I.M.Mills, P.J.Mohr, T.J.Quinn, B.N.Taylor, E.R.Williams, Redefinition of the kilogram: a decision whose time has come, *Metrologia*, vol.42, p.71-80, 2005.

[3] I.M.Mills, P.J.Mohr, T.J.Quinn, B.N.Taylor, E.R.Williams, Redefinition of the kilogram, ampere, kelvin and mole: a proposed approach to implementing CIPM recommendation 1 (CI-2005), *Metrologia*, vol.43, p.227-246, 2006.

[4] M.Stock and T.J.Witt, CPEM 2006 round table discussion 'Proposed changes to the SI', *Metrologia*, vol. 43, p.583-587, 2006.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Health and Safety

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Using proficiency testing results latest alternative method for uncertainty evaluation: application in clinical chemistry

Marc Priel*, Soraya Amarouche*,
Michèle Désenfant*, Jacques De Graeve**

*Laboratoire national de métrologie et d'essais
1, rue Gaston Boissier 75024 Paris Cedex 15, France
**Laboratoire de Biochimie Groupe Hospitalier Rangueil – Larrey Toulouse
1 avenue du Pr Jean-Poulhès 31403 Toulouse Cedex 4, France

Summary

There has been a strong increase in proficiency testing (PT) as a powerful tool to the quality assessment of testing and measurement activities. Many providers propose Interlaboratory Comparisons in various domains such as environment agri-food and health. In the classical former methods exploitation of results enabled the evaluation of each participant comparing their measurement results to their peers, calculating a performance statistic such as the "z-score".

Since the advent of the Guide to the expression of Uncertainty in Measurement (GUM) in 1995 founding the principles of uncertainty evaluation, numerous projects have been carried out to develop alternate practical methods, easier to implement notably when it is impossible to model the measurement process for technical or economical aspects.

Besides the method described in Chapter 8 of the GUM so-called "modeling" exploitation of results to the "validation of method" was presented during the Metrology Congress of 2003.

In 2005 the "performance of method" approach was presented and in 2007 another alternate method is presented in this paper based on the exploitation of proficiency testing and internal quality assessments.

Further considerations drawn from activities in the domain of health for the evaluation of measurement uncertainty will be presented in this paper. The example

of the determination of total concentration of glucose in human serum will be described. GUM and Proficiency Testing approaches will be compared.

Typology of evaluation methods for measurement uncertainties

The GUM presents the concepts necessary to the evaluation of uncertainties (precise definition of measurand, list of influence factors…).

It also details an uncertainty evaluation method called a "modeling" approach which is the reference method in uncertainty estimation. Alternative [1] uncertainty quantification methods have been developed, but they still respect the basic concepts described in GUM.

A typology of these methods is presented. It distinguishes between intralaboratory approaches, the case where there is only one laboratory and it uses its own data to evaluate the uncertainty of its results and the interlaboratory approach characterized by a cooperative study between several laboratories.

The intralaboratory approach is split into:

– use of the law of propagation of uncertainty or of propagation of distributions (Monte Carlo Simulation);

– use of the method validation data.

The interlaboratory approach is split into:

– use of the method performance data (NF ISO 5725[3] and ISO TS 21748[4]);

– use of proficiency testing data (ISO 43[5] and /NF ISO 13528[6] guide).

The following figure summarizes these different approaches.

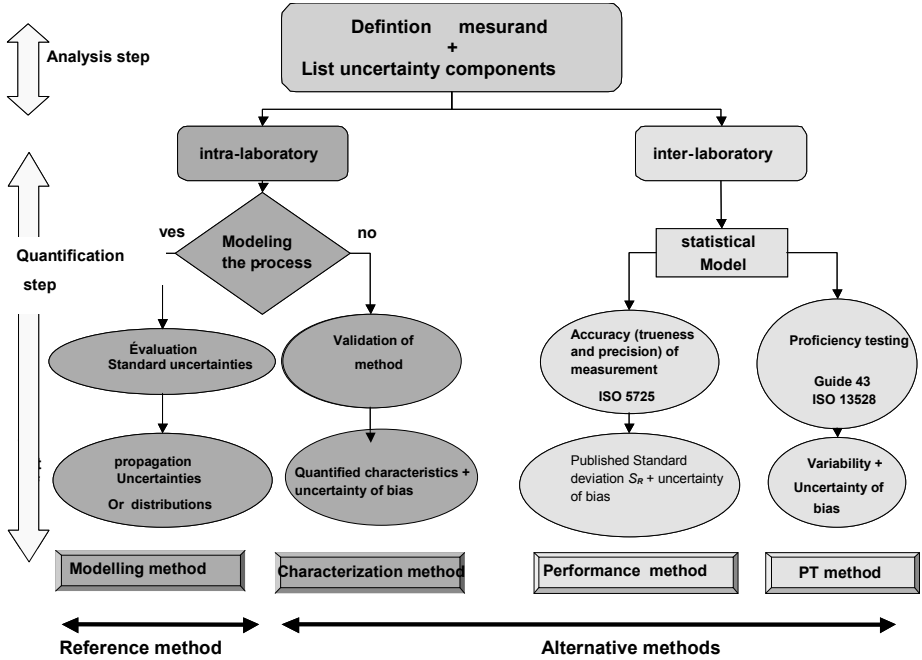


Figure 1. *Road Map of evaluation methods for measurement uncertainties*

GUM and alternative methods

For economic or technical reasons, it is sometimes impossible to establish the mathematical model describing the measurement process. In such cases, alternative methods must be developed. In 2005 at the Congrès de Lyon we presented the "method performance" approach in the conference *"Whatever the uncertainty measurement quantification method, measurand definition is the necessary first step"* [2]. In 2003 at the Congrès de Toulon, we developed the "method validation" approach in the conference *"From the validation of analysis methods to the uncertainty evaluation of measurement results"* [3] and in this paper, we present the latest evaluation method: the use of proficiency testing results.

Justifications for the use of alternative methods

The EN ISO/CEI 17025 general requirements standard concerning the skill of calibration and testing laboratories explicitly clears paths for the use of alternative uncertainty evaluation methods of testing and measurement results:

[…] In certain cases the nature of the test method may preclude rigorous, metrologically and statistically valid, calculation of uncertainty of measurement. In these cases the laboratory shall at least attempt to identify all the components of uncertainty and make a reasonable estimation, and shall ensure that the form of reporting of the result does not give a wrong impression of the uncertainty. Reasonable estimation shall be based on knowledge of the performance of the method and on the measurement scope and shall some cases, the nature of the test method excludes a rigorous calculation that is metrologically and statistically valid from the measurement uncertainty. In such cases; the laboratory must at least attempt to identify all uncertainty components and offer a reasonable estimation, while ensuring that the method of accounting for it does not give an incorrect impression of the uncertainty. A reasonable estimation must be based on knowledge of the method performance and on the measurement field and use the experience acquired and from previous validation data […]

ILAC (International Laboratory Accreditation Cooperation) which is the organization grouping international accreditation organizations published a document ILAC G17/2002 introducing the concept of uncertainty in testing in association with the application of the ISO / CEI 17025 standard which precisely sets the requirements in terms of uncertainty evaluation.

[…] The basis for the estimation of uncertainty of measurement is to use existing knowledge. Existing experimental data should be used (quality control chart, validation, round robin tests, PT,CRM, handbooks etc.[…]

The examination of these two texts clearly shows that we can rely on existing data and the data cited is method validation data; data from quality control such as control charts; interlaboratory comparisons results; and proficiency testing results. These two texts fully justify the development of alternative methods from the reference method described in GUM.

Use of proficiency test results

There are several measurement and testing sectors where the proficiency testing organization constitutes a preferred way to evaluate laboratory skills.

Currently, the common use of proficiency testing consists of calculating the difference between a laboratory result and an assigned value, most of the time established by using the arithmetic, weighed or robust average of the series of results and then to compare this difference to the dispersion of results.

It seems obvious that in the study of deviation between values obtained by a laboratory and the value assigned for the comparison, we can obtain interesting

information on the bias of the laboratory. This information is all the more relevant as the comparisons are regularly organized during a year.

We observe that proficiency testing can help understand the systematic measurement error for a laboratory and for a type of measurement when there are no reference materials. If we have ways to evaluate the "internal deviation" in the form of an internal reproducibility evaluation and by combining bias and internal reproducibility we will be able to evaluate uncertainty.

The approach proposed in clinical chemistry: methods coupling

In the field of medical biology, studies carried out at CHU of Rangueil in Toulouse and LNE have led to proposing to Cofrac a method serving as a basis for the publication of a guide.

The principle retained is based on the use of existing data by taking into account the results of the validation method (method property) and interlaboratory comparisons (proficiency testing).

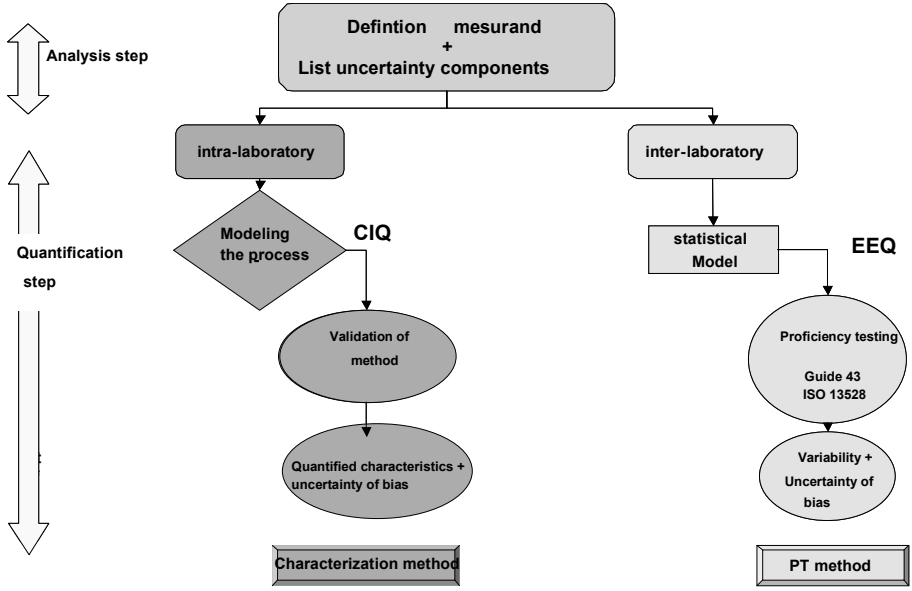


Figure 2. *Coupling of evaluation methods for measurement uncertainties*
CIQ; internal quality control EEQ: external quality evaluation

Application of the evaluation of glycemia measurement uncertainty

In order to validate this uncertainty evaluation approach by "proficiency testing" and "method characteristic" coupling methods, CHU of Rangueil and LNE treated a practical case of glucose testing in human plasma.

The complete definition of the measurand as well as the most exhaustive list of uncertainty components are at the basis of an objective evaluation of the measurement result uncertainty. This step common to all uncertainty evaluation approaches requires pooling of all skills (engineers, technicians, statisticians, etc.).

In the following example, it is the concentration of glucose *Cg* in a human plasma specimen in mmol/L. The measurement process can be summarized as an UV enzymatic method (hexokinase test) after calibration in two points on an Olympus AU 2700.

GUM REFERENCE METHOD (Chap 8)

The definition of the measurand and the process analysis are presented in the 5 M diagram of Figure 3:

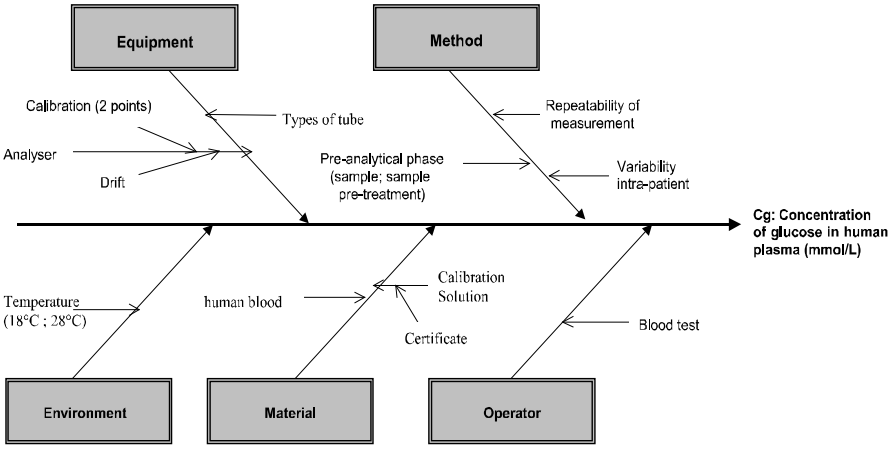


Figure 3. *Diagram of 5 M of the measurement process of total glucose concentration in a blood sample analysis*

The measurement process after analysis of the different supports and experience in the field can be represented by the model below. It is made up of the main sources of uncertainty.

$$C_g = \left[\left(C_0 + \frac{A_s}{A_{Cal}} (C_{cal} - C_0) \right) \times \left(\frac{V_1 + V_2}{V_2} \right) \times f_{matrix} \times f_{derived} + k_{pre} \right]$$

with*:*

C_g*:* total glucose concentration in the human plasma specimen [mmol/L], and for input variables:

C_0*:* total glucose concentration of a solution (pure water) establishing zero point of the calibration curve [mmol/L],

A_s*:* corrected white absorbance of the specimen [AU: unit of absorbance],

A_{cal}*:* the corrected absorbance of calibration solutions in the bucket [AU: unit of absorbance],

C_{cal}*:* total glucose concentration during calibration [mmol/L],

V_1*:* specimen volume (μL),

V_2*:* volume of reactants + water (μL),

f_{matrix} *:* factor caused by the contribution of possible matrix effects (=1),

f_{drift}*:* factor caused by the contribution of a permitted drift in instrument sensitivity (=1),

k_{pre}*:* correction caused by the contribution of the pre-analytical phase [mmol/L] (=0).

By applying the different steps of the GUM reference method after defining the mathematical model of the measurement process:

$y = f(x_1, x_2, \ldots, x_n)$

– Standard uncertainty evaluation of input variables and evaluation of existing covariance*:*

$$u(X_1), u(X_2), \cdots, u(X_n)$$

$$u(X_i; X_j)$$

– Propagation of standard uncertainties*:*

$$u^2(y) = \sum \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum \sum \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i; x_j)$$

– Expression of the result

$$y \pm U \quad (k = 2) \qquad \textit{(measurement unit)}$$

Table 1 summarizes the different steps with an application at concentration of 7.06 mmol/L.

Measurand: total concentration of glucose in a human blood specimen

$$: C_g = \left(\left[C_0 + \frac{A_S}{A_{Cal}} \left(C_{cal} - C_0 \right) \right] \times \left(\frac{V1+V2}{V2} \right) \times f_{matrix} \times f_{derivative} + K_{pre} \right) \text{mmol/L}$$

Code X_i	components	Value x_i	unit	Information provided for evaluation of u(Xi)		type	law	U(Xi)	[Ci]	%
				value	source					
A_S	Adjusted raw absorbance of the measurement	0.3	[AU]	0.06	Reproducibility Laboratory evaluated CQI 2%	A		0.0060	23.5249	44.56 %
A_{cal}	Adjusted raw aborbance of the ground calibration	0.3445	[AU]	0.0046	Experimental dispersion over 6 months	A		0.0046	-20.486	19.86 %
C_0	Total concentration of glucose at pt 0 of the calibration (pure water)	0	mmol/L	0	By hypothesis not uncertainty on pure water	-	-	0.0000	0.13021	0.00%
C_{cal}	Total concentration of glucose during the calibration	8.04	mmol/L	0.1437	By hypothesis 3 times uncertainty of MR NIST 965	B	K=2	0.1437	0.87779	35.58 %
V1	Sampling volume (auto)	1.6	µL	Accuracy: 0.03 Repeatability : 0.014	Uncertainty about pipettes uncertain manufacturer	B	K=2	0.0223	0.03501	0.00%
V2	Volume of 2 reactants + water (auto)	200	µL	Accuracy: 1.6 Repeatability : 0.3	Uncertain about pipettes uncertain manufacturer	B	K=2	0.9700	-0.0003	0.00%
f_{matrix}	Factor due to the effect of the matrix	1	-	0	No information	-	Uncertainty on these factors not taken into account (waiting for information)			
$f_{derivative}$	Factor due to the derivative (sensitivity of the equipment)	1	-	0	Manufacturer info or included in repeatability	-				
K_{pre}	Correction due to the pre-analysis phase	0	mmol/L	0	Experimental design	-				
	Cg = 7.06 mmol/L						*u^2 = 4.47E-02* *u = 2.11E-01* *U = 0.423 (k=2) 5.99%* *U = 0.43 (k=2) 6%*			

Table 1. *Uncertainty calculation on glucose concentration in human blood specimen*

C_g[1] = *7.06 ± 0.43* mmol/L *(*k=*2)*

The uncertainty evaluated for this measurement result is 6%, which corresponds a priori to an expected level by field experts. This first uncertainty quantification will make it possible to obtain a comparison value, more globally to validate the uncertainty evaluation approach using proficiency testing, widely practiced in analysis laboratories in medical biology.

ALTERNATIVE METHOD: "Method characteristics" + "Proficiency testing" coupling

In medical biology, intralaboratory/interlaboratory terminology is replaced by CIQ (internal quality control) and EEQ (external quality evaluation).

The measurement uncertainty evaluation method requires writing a model regardless of the approach. Nevertheless, in this new approach, the model remains very simplified.

The model is written in the following way:

$$C = C_{lue} + K_{EEQ} \tag{1}$$

with,
C: analyte concentration,
C_{lue}: read analyte concentration (analyzer result),
K_{EEQ}: correction *systematic measurement error.*

We consider that in the model exposed above, there are two uncertainty components evaluated from the internal quality control (CIQ) and the external quality evaluation (EEQ). By applying the law of propagation of uncertainties, we obtain:

$$u(C) = \sqrt{u^2(CIQ) + u^2(K_{EEQ})} \tag{2}$$

with, $u^2(CIQ) = u^2(C_{lue})$, which represents the variance of all internal quality control (CIQ) results (C_{lue}) and with, $u^2(K_{EEQ})$, the variance linked to the trueness bias, estimated from the external quality evaluation (EEQ).

This simple and realistic approach based on internal and external control data must still be based on values used for the evaluation of precision and trueness which

1 For information this equals a concentration in g /l of C_g = 1.271 ± 0.078 *g/l* (*k*=2).

take into consideration all factors of influence liable to occur in the measurement process.

Evaluation of internal reproducibility – IQC operation

During these CIQs, medecine laboratories quantify variability of the measurement process by a reproducibility variation factor calculated from standard deviation *s* of reproducibility and from average *m* of results of an CIQ:

$$CV\% = \frac{s}{m} \times 100 = \frac{s}{C_{lue}} \times 100$$

We then have:

$$u^2(CIQ) = (CV_{CIQ})^2 x C_{lue}{}^2$$

for an individual result.

Evaluation of trueness and its uncertainty – EEQ operation

At each interlaboratory comparison *i,* the laboratory obtains its deviation $E_i = (x_{lab} - x_{ref})_i$

During *n* comparisons, we deduce the average deviation and its standard-deviation,

$$\overline{E} = \frac{\sum_i (x_{lab} - x_{ref})_i}{n} \qquad S_E = \sqrt{\frac{\Sigma (E_i - \overline{E})^2}{n-1}}$$

Two procedures are used. The laboratory corrects its results with the average systematic measurement error and integrates the uncertainty of this correction or, the laboratory does not implement the correction and totally integrates the systematic measurement error in its uncertainty, with the consequence of increasing the uncertainty in the end.

Data in mmol/L:			Calculations in mmol/L:	
			Correction	**No correction**
Concentration :	**6.00**	*u(CIQ)*=	0.12	0.12 $K_{EEQ} = 0$
			$K_{EEQ} = -\overline{E}$ $u(K_{EEQ}) = S_E$	$u(K_{EEQ}) = \frac{MAX\ \lvert E_i \rvert}{\sqrt{3}}$
Internal quality control (CIQ)				
s =	0.12	$u(K_{EEQ})$ =	0.1648	0.329089653
External quality evalutation (EEQ)			$u(C) = \sqrt{u^2(CIQ) + u^2(K_{EEQ})}$	
$\overline{E}$ =	0.1242			
S_E =	0.1648	*u(c)* =	0.203860344	0.350285597
max |Ei| =	0.57	*U (k=2)* =	0.41	0.70

	Results mmol/L (k=2):	
Absolute	**5.88 ± 0.41**	**6 ± 0.70**
Relative	**5.88 ± 7%**	**6 ± 12%**

For the expanded uncertainty, results are 7% when we decide to apply a correction and 12% without correction.

By applying a correction, there is equivalence with the 6% obtained by the GUM reference method, even while knowing that using the proficiency testing approach the evaluation of uncertainties is necessarily slightly higher since it takes all factors of influence into account.

Conclusion

The expanded uncertainty evaluated on the measurement of glucose concentration in a human plasma specimen is ± 6% with the reference approach (GUM) and ± 7% with the new approach.

The similarity with both results provides a first validation of the proficiency testing approach coupled with intralaboratory data.

This new approach opens new perspectives and enables us:

- to doubly use proficiency testing, with references currently in revision;

- to diffuse widely the concept of quality of the measurement result through uncertainty calculation;
- to bring a decision support tool notably in diagnostics at the medical level.

The next step will be to distribute this approach to all proficiency testing users, well beyond medical biology laboratories.

References

[1] BIPM, IEC, IFCC, IUPAC, IUPAP, OIML – Guide to the expression of uncertainty in measurement ISO 1995.

[2] Michèle Désenfant, Marc Priel, Road map for measurement uncertainty evaluation, *Measurement* 39 (2006) 841 – 848.

[3] NF ISO 5725, Accuracy (trueness and precision) of measurement methods and results, ISO 1994.

[4] ISO/TS 21748, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation.

[5] Guide ISO 43, Proficiency testing by interlaboratory comparisons.

[6] ISO 13528, Statistical methods for use in proficiency testing by interlaboratory comparisons ISO 2004.

[7] Michèle Désenfant, Marc Priel, Tout se joue dans la définition du mesurande quelles que soient les méthodes de quantification de l'incertitude, *Congrès International de Métrologie*, Lyon 2005.

[8] Michèle Désenfant, Marc Priel, Cédric Rivier, De la validation des méthodes d'analyse à l'évaluation de l'incertitude des résultats de mesure, *Congrès International de Métrologie*, Toulon 2003.

[9] Cofrac LAB GTA November 14 2006, Guide d'évaluation des incertitudes de mesures des analyses de biologie médicale.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Facilitating reliability and traceability of measurement results in laboratory medicine by certified reference materials

B. Toussaint, H. Schimmel, H. Emons

Institute for Reference Materials and Measurements, Joint Research Centre,
European Commission (EC-JRC-IRMM),
Retieseweg 111, 2440 Geel, Belgium

ABSTRACT: Comparability of results from different analytical procedures and commercial diagnostic kits is a key issue in laboratory medicine. In order to achieve this, reference procedures and Certified Reference Materials (CRMs) as well as quality control materials are required. In this presentation, the crucial aspects of the development and certification of CRMs will be described and illustrated by recent examples. In particular, establishing the metrological traceability of pure substance CRMs, developing matrix CRMs and investigating the commutability of these materials will be discussed. Finally, a distinction will also be made between Quality Control Materials and Certified Reference Materials in the way they facilitate reliability and traceability of measurement results.

Introduction

In recent years, the need for standardization of assays in laboratory medicine has been widely recognized. The European Directive 98/79/EC is especially aiming to ensure the quality of In-Vitro Diagnostic (IVD) medical devices through the use of higher order reference materials and calibration methods. In this context, the Institute for Reference Materials and Measurements (IRMM) is developing and certifying higher order reference materials for laboratory medicine. IRMM co-operates with other international organisations (IFCC, JCTLM, CCQM, ILAC) to provide guidance on internationally recognized and accepted equivalence of measurements in laboratory medicine and traceability to appropriate measurement standards.

The need for standardization of assays in in-vitro diagnostics is, for instance, illustrated in Figure 1. A well-defined material containing endogenous progesterone in human serum was characterized using a reference procedure (Isotope Dilution-Gas Chromatography-Mass Spectrometry) and 8 different immunoassays [1]. The comparison of the results considering the reference procedure as giving 100% trueness shows a spread of the results between 80 and 180% of the value obtained by the reference method (5 immunoassays between 80 and 120%, 1 immunoassay between 90 and 130%, 2 immunoassays between 130 and 180%).

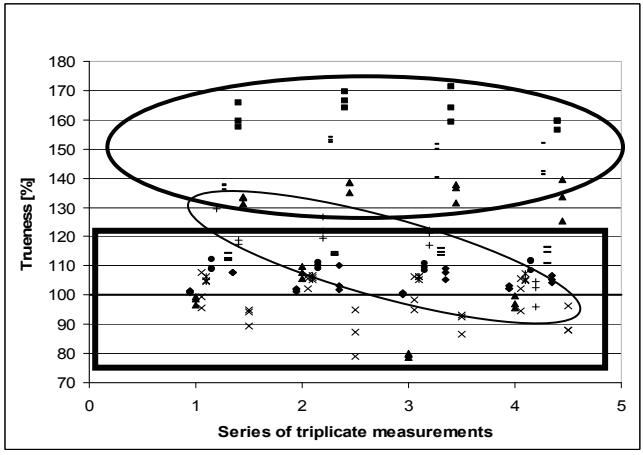


Figure 1. *Progesterone in human serum: trueness of 8 immunoassays compared to the ID-GC-MS reference method (♦)*

Discussion

Reference Measurement Systems

The establishment of a reference measurement system, necessary to achieve the comparability of the results from different assays, is based on three pillars (Figure 2):

– Reference Measurement Procedures, validated for the selected quantity in real samples;

– Reference Materials, suitable for the selected measurand;

– Reference Laboratories, with successful participation in both lab-directed and method-directed proficiency testing.

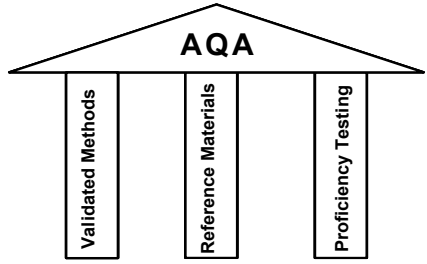


Figure 2. *The three pillars of the Analytical Quality Assurance in a Reference Measurement System*

Metrological traceability

Traceability is a key concept in the establishment of a Reference Measurement System [2]. The definition of traceability, according to the International Vocabulary of Basic and General Terms in Metrology (VIM) (ISO, 1993), reads "property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties". This means that a chain of comparisons must be established between the routine standard used in the laboratory, the working calibrator of this laboratory and the international conventional calibrator.

Certified Reference Materials

Certified Reference Materials represent one of the pillars of a Reference Measurement System. They can be defined as "Reference Materials characterised by a metrologically valid procedure for one or more specified properties, accompanied by a certificate that provides the value of the specified property, its associated uncertainty and a statement of metrological traceability" [3].

As illustrated in Figure 3, CRMs (pure substance and matrix materials) play a key role in the quality assurance of the different steps of a measurement procedure, including qualitative and quantitative aspects [4-5].

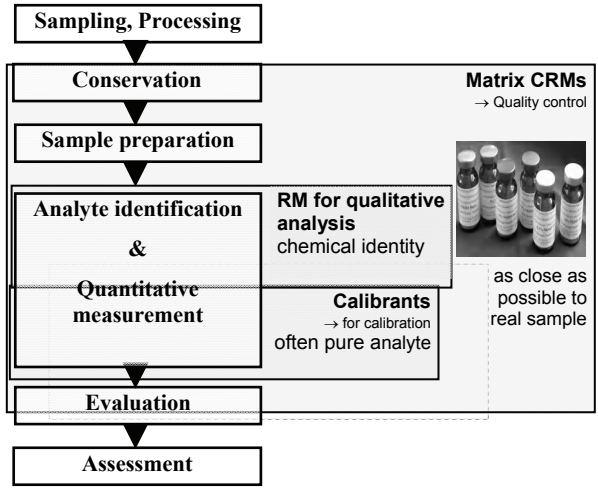


Figure 3. *Materials for Quality Assurance at different steps of the measurement procedure*

The development and certification of a CRM includes homogenity, short-term and long-term stability studies [6-7], characterization for value assignment and often a commutability study. Moreover, after certification, CRMs are regularly monitored for stability, ensuring the user a permanent high quality material. Finally, the certified value of the material property is accompanied by a certified uncertainty which takes into account the uncertainty contributions of the different certification studies [8-12].

Use of Certified Reference Materials

Due to their particularly high quality and characteristics, CRMs can be used for several purposes:

– development and validation of methods in order to evaluate trueness and measurement uncertainty;

– calibration;

– proof of method performance;

– proficiency testing of laboratories.

In each case, the selection of a CRM, with suitable characteristics (matrix, concentration, uncertainty) for the targeted application is essential.

Example

As an example, Table 1 illustrates the characterization of IRMM-468, a CRM for pure thyroxine [13]. In order to determine the purity of that pure substance CRM, impurities detectable by HPLC, inorganic residues and solvents were quantified. The characterization was performed in a ring trial by expert laboratories (A-F) using different validated methods. A combined uncertainty is calculated from the different impurity determinations.

Thyroxine content taking into account HPLC-detectable impurities: **Method (laboratory)**	**mass fraction** **[%]**
HPLC-UV (A)	99.61
HPLC-UV (B)	99.72
HPLC-UV (C)	99.00
HPLC-UV-MS (D)	99.57
HPLC-MS (E)	98.98
Average [%]	99.38
Sum of HPLC-non-detectable impurities:	0.78
Total thyroxine:	**98.60**

Table 1. *Determination of the mass fraction of thyroxine in IRMM-468*

Conclusions

CRMs are higher order reference materials if they are certified by using high quality analytical procedures in networks of expert laboratories and in accordance with ISO Guides 34 and 35 [3,14]. The use of CRMs for the validation of a method provides the user not only with information about the precision of the method, but also about its accuracy and ensures the user traceability to an international Reference Measurement System.

Further information can be found at the following address: http://www.irmm.jrc.be/html/homepage.htm

References

[1] B. Toussaint, C.L. Klein, F. Franchini, H. Schimmel, H. Emons, "Certification of the mass concentration of progesterone at a level of approximately 8 µg/L in human serum, BCR-348R", Report EUR 22299 EN.

[2] ISO 17511:2003, In vitro diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials.

[3] ISO Guide 35:2006, Reference Materials – General and statistical principles for certification.

[4] H. Emons, A. Fajgelj, A. M.H. van der Veen, R. Watters, "New definitions on Reference Materials", *Accred. Qual. Assur.*, Vol. 10, pp. 576-578, 2006.

[5] H. Emons, "The 'RM Family' - Identification of all its members", *Accred. Qual. Assur.*, Vol. 10, pp. 690-691, 2006.

[6] A. Lamberty, H. Schimmel, J. Pauwels, "The study of the stability of reference materials by isochronous measurements", *Fresenius J. Anal. Chem.*, Vol. 360, pp. 359-361, 1998.

[7] T.P.J. Linsinger, A.M.H. van der Veen, B. Gawlik, J. Pauwels, A. Lamberty, "Planning and combining of isochronous studies of CRMs", *Accred. Qual. Assur.*, Vol. 9, pp. 464-472, 2004.

[8] A. Van der Veen, J. Pauwels, "Uncertainty calculations in the certification of reference materials.2. Homogenity study", *Accred. Qual. Assur.*, Vol. 6, pp. 26-30, 2001.

[9] A. Van der Veen, Th. Linsinger, A. Lamberty, J. Pauwels, "Uncertainty calculations in the certification of reference materials.3. Stability study", *Accred. Qual. Assur.*, Vol. 6, pp. 257-263 , 2000.

[10] ISO-IEC-BIPM-IFCC-IUPAC-IUPAP-OIML. *Guide to the Expression of Uncertainty in Measurement (GUM)*, Geneva, Switzerland.

[11] A.M.H. van der Veen, T.P.J. Linsinger, H. Schimmel, A. Lamberty, J. Pauwels, "Uncertainty calculations in the certification of reference materials.4. Characterisation and certification", *Accred. Qual. Assur.*, Vol. 6, pp. 290-294, 2001.

[12] T.P.J. Linsinger, J. Pauwels, A.M.H. van der Veen, H. Schimmel, A. Lamberty, "Homogenity and stability of reference materials", *Accred. Qual. Assur.,* Vol. 6, pp. 20-25, 2001.

[13] B. Toussaint, C.L. Klein, M. Wiergowski, "The certification of the mass fraction of thyroxine in a CRM intended for calibration, IRMM-468", Report EUR 21872 EN.

[14] ISO Guide 34:2000, General requirements for the competence of reference material producers.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Dosimetric standard for continuous x-rays radiation fields of low and medium-energies (< 300 kV)

W. Ksouri, M. Denozière, J. Gouriou and J-M. Bordy

LNE-LNHB, Laboratoire National Henri Becquerel
CEA de Saclay, 91191 Gif-sur-Yvette Cedex, France
wassim.ksouri@cea.fr

ABSTRACT: This work presents the methodology used for the setup of the dosimetric standard at the Laboratoire National Henri Becquerel (LNHB) for the continuous X-rays irradiation field of medium-energies (100 kV to 250 kV) in term of air-kerma rate. Principle of measurements and characteristics of the standard are described. Additionally, a first assessment of global uncertainty budget on the value of air-kerma rate taking into account several correction factors and a first estimate of that of a calibration are developed.

Introduction

X-ray radiations induce erythemas and leukaemias. The ICRP (International Commission of Radiological Protection) enacted the first recommendations to protect against the effects of the ionizing radiations. The recommendations were modified along the years and are today the roots of the **96/29 Euratom Directive** which ensures the protection of the public and workers [1]. The **97/43 Euratom Directive** is particularly applied in the medical field for irradiation of patients to optimize benefit-to-risk ratios [2].

The application of these directives transcribed in the French legislation, set rules for maintenance, internal and external quality control of the machines, written procedures and a metrological traceability.

The LNHB, the French national laboratory for metrology of ionizing radiations, has to deliver to secondary accredited laboratories, or directly to users, dosimetric standards for continuous low and medium-energy X-rays.

This paper gives a description of these X-rays beams, as well as the air-kerma rate standard.

I Materials and method

I.1 *Radiation qualities*

The continuous X-ray beams of low and medium-energies are produced by an industrial generator with its X-ray tube and can go up to 320 kV. The beams are defined by the generating potential of the tube and by the total filtration. The quality beams is characterized by

The Half-Value Layer (HVL) [3] [4] and the air-kerma rate: this HVL is the thickness of an additional copper or aluminum filter (depending of the beam) of high purity (less than 0.01% of impurities) which decreases the air-kerma rate by a factor of 2. The uncertainty on the HVL is around 0.80% (one standard deviation).

The reference point is at 1.20 m of the focus tube. At this distance the beam diameter is 14 cm with less than 0.6% of inhomogenity. The irradiation beam dose rate has a stability of about 0.1% over several hours (≈3h). The measuring device is placed in a temperature controlled room (20°C ± 0.5°C and 50% HR).

I.2 *Principle of the detector (absolute instrument)*

The low and medium-energy X-ray beams are characterized in terms of air-kerma rate and measured with a free-air ionization chamber [5].

The kerma (**K**inetic **E**nergy **R**eleased in **MA**tter) is the quotient (1) of the sum of the initial kinetic energies of all the charged ionizing particles released by non-charged particles dE_{tr} in the air mass dm .

$$K_{air} = \frac{dE_{tr}}{dm} \tag{1}$$

It is defined at point *P* located at the center of the inner side of the diaphragm (Figure 1) and is expressed in Gray ($1Gy = 1J.kg^{-1}$).

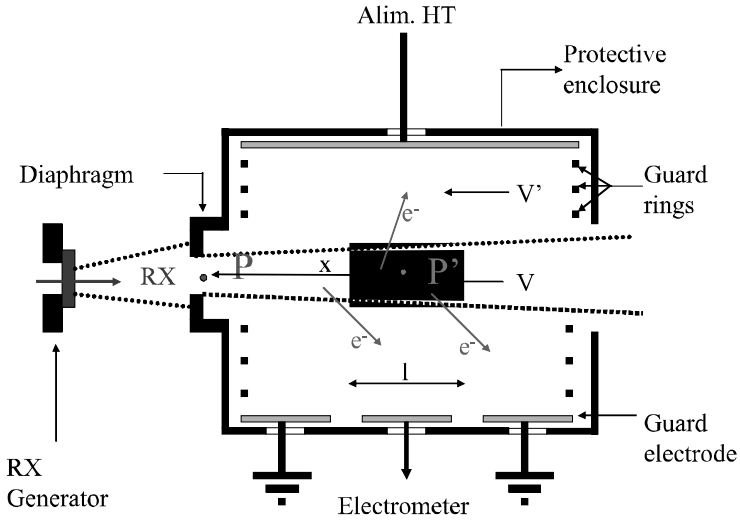


Figure 1. *Schematic plan of a free-air ionization chamber standard*

The free-air ionization chamber is an absolute instrument (known collection volume *V*).

The application of a potential difference between the electrodes makes it possible to collect the charges created by the interaction of the beam with the air. The guard rings connected to a resistances bridge guarantee the uniformity of the electric field over the interaction volume *V*.

This chamber is placed in a lead enclosure to protect it from scattered radiation.

I.3 *Determination of the air-kerma rate*

The equation to calculate the air-kerma rate is:

$$\dot{K}_{air} = \frac{I}{\rho_{air} \cdot V} \cdot \frac{W_{air}}{e} \cdot \frac{1}{1-g} \cdot \prod k_i \qquad (2)$$

with:

I: ionization current measured with the free-air ionization chamber and expressed in A;

ρ_{air} = 1.20479 kg.m^{-3}, density of dry air (20°C and 1013.25 hPa);

V: detection volume in cm^3;

W_{air}/e = 33.97 J.C^{-1}, average energy necessary to create an electron-ion pair in dry air, per unit of the electron charge [6];

g: is the fraction of the initial electron energy lost by Bremsstrahlung production in air. This value is negligible for low and medium-energy X-rays [7];

$\prod k_i$: product of correction factors taking into account the characteristics of the detector (electrodes polarization, scattered photons, recombination of the charges, etc.) and the environmental conditions of the measurement (temperature, pressure and hygrometry).

I.4 *Determination of correction factors*

As indicated by equation (2), the air-kerma rate is corrected by several correction factors. Some are determined experimentally; and the others by using simulations.

– The experimentally determined factors are:

- **recombination factor k_s**: corrects for the loss between the number of produced charges and the number of collected charges due to the recombination of the charges;
- **attenuation factor in the air k_a**: corrects for air attenuation between P (located at the inner side of the diaphragm) and the P' (located in the middle of the volume V), (Figure 1);
- **polarization factor k_{pol}**: takes into account the dissymmetry of ion collection;
- **transmission factor k_l**: corrects for the transmission through the matter around the diaphragm;
- **transmission factor k_p**: corrects for the transmission through the walls.

–The factors determined by software simulation based on the finite elements method are:

- **distortion factor k_d**: takes into account the possible distortion of the electric field inside the chamber which can induce an error in the determination of the volume *V* [8].

– The factors determined by Monte-Carlo simulations are:

- **factor for electron loss k_e**: to evaluate the loss of ionization due to electrons which dissipate part their energy out side the volume *V'*;
- **factor for scattered photons k_{sc}**: corrects for the charges due to secondary photons in *V'*.

II Results

II.1 *Standard radiation qualities (100 kV and 250 kV)*

Table 1 gives two examples of radiation qualities (100 kV and 250 kV) according to the recommendations of the CCEMRI [4]. The X-ray tube has a tungsten anode and an inherent aluminium filtration of 3.2 mm.

Generating potential (kV)	100	250
Additional Al filtration (mm)	0.318	-
Additional Cu filtration (mm)	-	1.552
Al HVL (mm)	4.044	-
Cu HVL (mm)	-	2.487
$(\mu/\rho)_{air}$ (cm^2/g)	0.256	0.141
Air-kerma rate (mGy/s)	0.52	0.53

Table 1. *Characteristics of LNHB standard radiation qualities 100 kV and 250 kV*

II.2 *Details of the free-air ionization chamber WK06*

The LNHB standard (WK06) is suitable for X-ray beams between 100 kV to 250 kV (Figure 2). Table 2 describes some of the principal characteristics.

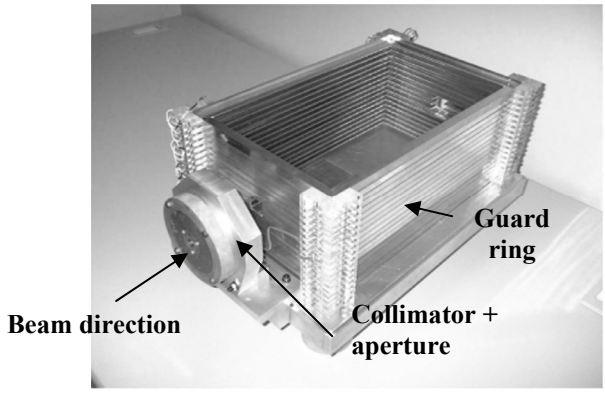


Figure 2. *View of the WK06 free-air ionization chamber*

Standard WK06	
Aperture diameter (cm)	1.0074
Air path length PP' (cm)	**31.8**
Collecting width (cm)	23.1
Electrode separation (cm)	18.0
Collecting length (cm)	6.0004
Measuring volume (cm^3)	**4.7827**
bias voltage (V)	5000

Table 2. *Main characteristics of the standard WK06*

The collecting volume (V) is geometrically defined by the diameter of the beam, itself dependent on the diameter of the aperture diaphragm, and by the collecting electrode width (l).

It is necessary to add two half-spaces corresponding to half of the intersection with the beam of the volume perpendicular to the electrodes and situated between the collecting electrode and the guard electrode. This volume was measured with an uncertainty of 0.05%.

The PP' distance (Figure 1) is selected to ensure the electronic equilibrium.

II.3 *Simulation by the Monte-Carlo method*

Computer codes simulating the radiation-matter interactions and based on the Monte-Carlo method were used to create a numerical model of WK06.

MCNP code version 4C [9], was used to check the detector design (fluorescence inside volume V, collimation of the aperture, etc.) and to consider simplifications with the numerical model which allows us to increase simulation speed.

Another code PENELOPE [10] was used to calculate the correction factors k_e (electron loss) and k_{sc} (scattered photons).

II.4 *Correction factors and their associated uncertainties for 100 kV and 250 kV*

The characteristic correction factors of the detector are given for each radiation quality. These values and their associated uncertainties are listed in Table 3.

The correction factor of hygrometry k_h is taken equal to 0.998 as long as the Relative Humidity (RH) is included between 20 and 80% at 20°C [11] [12].

However, the signal-to-noise ratio is around 1000 for the weakest currents of ionization (pA).

The product of all these factors results in a total correction of around 1% of the value of the air-kerma rate.

Correction Factors	Generating potential		Relative uncertainty	
	100kV	250kV	Type A	Type B
k_{sc}	0,9936	0,9965	-	0,07%
k_e	1.0001	1.0065	-	0.10%
k_s	1.0003	1.0003	0.08%	-
k_{pol}	1.0001	1.0001	0.05%	-
k_a	1.0092	1.0051	-	0.10%
k_d	1.0000	1.0000	-	0.10%
k_l	0.9999	0.9999	-	0.01%
k_p	0.9998	0.9998	-	0.01%
k_h	0.9980	0.9980	-	0.03%
Total	1.0009	1.0061	**0.09%**	**0.19%**

Table 3. *Correction factors for the LNHB standard for 2 quality beams and their associated uncertainties*

II.5 *Air-kerma rate*

Section I.3 describes the relation to calculate the air-kerma rate. For two radiation qualities (100 kV and 250 kV), the values are respectively 0.52 and 0.53 $mGy.s^{-1}$ (Table 1) for ionization currents $I_{100kV} \approx 55$ pA and $I_{250kV} \approx 88$ pA.

Measurement of the ionization current, dimensional measurements (volume of detection), physical constants (W_{air}/e, ρ_{air}…) and various correction factors quoted previously, take part in the assessment of the uncertainty budget.

Table 4 gives the uncertainty budget of the air-kerma rate at one standard deviation. It is equal to 0.31% (at one standard deviation).

	Relative uncertainty	
	Type A	Type B
Volume		0.05%
Positioning	-	0.01%
Ionization current	0.02%	0.15%
Correction factor k_T	-	0.04%
Correction factor k_P	-	0.02%
Correction factors k_i	0.09%	0.19%
Physical constants (W_{air}/e and ρ_{air})	-	0.15%
Air-kerma rate	0.10%	0.29%
Quadratic sum	**0.31%**	

Table 4. *Uncertainties on the air-kerma rate*

III Application

III.1 *Calibration of a transfer instruments*

The calibration coefficient in air-kerma denoted (N_K) is given by the following relation:

$$N_K = \frac{\dot{K}}{I_{tr}} \tag{3}$$

where $\dot{K}$ is the air-kerma rate determined by the standard according to the relation (2) and I_{tr} is the ionization current measured by the transfer instrument with its associated electronics. This current is corrected with k_T and k_P to tally with the standard conditions of measurements (T=293.15 K, P=1013.25 and 50% HR).

III.2 *Uncertainty associated with the calibration coefficient: example*

Table 5 gives the uncertainty established for the calibration of an ionization chamber (NE2571 type). It is equal to 0.35% (at one standard deviation).

	Relative uncertainty	
	Type A	Type B
Standard kerma rate	0.10%	0.29%
Positioning of transfer chamber	-	0.01%
Ionization current I_{tr}	0.02%	0.15%
Correction factor k_T	-	0.04%
Correction factor k_P	-	0.02%
$\mathbf{N_K}$	0.10%	0.33%
Quadratic sum	**0.35%**	

Table 5. *Uncertainties for the calibration of a transfer chamber*

IV Discussion and prospects

The primary air-kerma rate values of the LNHB for the continuous medium-energy X-ray beams have uncertainty levels similar to the foreign counterpart laboratories.

In order to reduce the uncertainty, it would be necessary to improve the magnitude of these components. The air-kerma rate uncertainty comes mainly from W_{air}/e, measurement current and the product of correction factors.

To reduce uncertainty on W_{air}/e (0.15%) requires a fundamental study on the physical parameter such as W_{air}. In the same way, the limit on the measuring accuracy of the current (0.15%) depends on the capacitor uncertainty values.

The level of precision reached at the laboratory is ten times lower than the user's needs (for radiation protection) [13]. Nevertheless, to measure the air-kerma rate of curietherapy radionuclide which requires a calibration in X-rays beams, it is important to reduce these uncertainties.

In the continuity of this work, the LNHB develops a second standard which will be adapted to the continuous X-ray beams for low-energies from 10 kV to 60 kV.

Conclusion

The LNHB has established a primary standard for air-kerma rate in the continuous medium energy X-ray radiation. Forthcoming international comparisons are planned in the current of 2007.

References

[1] Directive 96/29/Euratom du conseil du 13 mai 1996 fixant les normes de base relative à la protection sanitaire de la population et des travailleurs contre les dangers résultant des rayonnements ionisants.

[2] Directive 97/43/Euratom du conseil du 30 juin 1997 relative à la protection sanitaire des personnes lors d'expositions aux rayonnements ionisants à des fins médicales.

[3] Norme Française Homologuée NF ISO 4037, Rayonnement X et gamma de référence pour l'étalonnage des dosimètres et débitmètres et pour la détermination de leur réponse en fonction de l'énergie des photons, décembre 1998.

[4] Qualité de rayonnements, in *BIPM Com. Cons. Etalons Mes. Ray. Ionisants* (section I), 1972, 2, R15.

[5] F. H. Attix, *Introduction to Radiological Physics and Radiation Dosimetry*, New York: A Wiley-Interscience Publication, 1986, ch.12, pp 292-345.

[6] Boutillon M., Perroche A.-M., Re-evaluation of the W value for electron in dry air, *Phys. Med. Biol.*, 1987, 32, 213-219.

[7] Constante physique pour les étalons de mesures, *in BIPM Com. Cons. Etalons Mes. Ray.Ionisants* (section I), 1985, 11, R45.

[8] Tera Analysis Ltd, Quikfield: Finite Element Analysis System, Version 5.1, Knasterhoovej 21, DK-5700 Svendborg, Denmark.

[9] J.F. Briesmeister, Ed., MCNP – A General Monte-Carlo N-Particle Transport Code, Version 4C, Manual LA-13709-M (April 2000).

[10] Salvat F., Fernandez-Varea J. M., Sempau J., PENELOPE-A Code System for Monte Carlo Simulation of Electron and Photon Transport, *NEA Data Bank* ISBN 92-64-02145-0, (July 2003).

[11] Boutillon M., BIPM-96/1 report: Measuring conditions used for the calibration of ionization chambers at the BIPM (Bureau International des poids et mesures); 1996.

[12] ICRU (International Commission on Radiation Unit and measurements), ICRU n°60: Fundamental Quantities and Units for Ionization Radiation; 1998.

[13] Wagner L.K., Fontenla D.P., Kimme-Smith C., Rothenberg L.N., Shepard J. and Boone J.M., Recommandations on performance characteristics of diagnostic exposure meter, report of AAPM diagnostic X-ray imaging task group N°6, Med. Phys., 19, 1992, 231-241.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Metrological Tools and Means

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The international vocabulary of metrology, 3rd edition: basic and general concepts and associated terms. Why? How?

R. Köhler

Bureau International des Poids et Mesures (BIPM)
Pavillon de Breteuil
92312 Sèvres Cedex
France

ABSTRACT: In general, a vocabulary is a "terminological dictionary that contains designations and definitions from one or more specific subject fields" (ISO 1087-1:2000, subclause 3.7.2). The present Vocabulary pertains to metrology, the "science of measurement and its application". It also covers the basic principles governing quantities and units.
In this Vocabulary, it is taken for granted that there is no fundamental difference in the basic principles of measurement in physics, chemistry, laboratory medicine, biology or engineering. Furthermore, an attempt has been made to meet conceptual needs of measurement in fields such as biochemistry, food science, forensic science and molecular biology.
Development of this 3rd edition of the VIM has raised some fundamental questions about different current philosophies and descriptions of measurement.

The international vocabulary of metrology 3rd edition. Basic and general concepts and associated terms. Why? How?

In 1997 the Joint Committee for Guides in Metrology (JCGM), chaired by the Director of the BIPM, was formed by the seven International Organizations that had prepared the original versions of the Guide to the Expression of Uncertainty in Measurement (GUM) and the International Vocabulary of Basic and General Terms in Metrology (VIM). In 2004, a first draft of this 3rd edition of the VIM was submitted for comments and proposals to the eight organizations represented in the JCGM, which in most cases consulted their members or affiliates, including numerous National Metrology Institutes. Comments were studied and discussed, taken into account when appropriate, and replied to by JCGM/WG 2. A final draft of the 3rd edition was submitted in 2006 to the eight organizations for comment and approval.

The 2nd edition of the *International Vocabulary of Basic and General Terms in Metrology* was published in 1993. The need to cover measurements in chemistry and laboratory medicine for the first time, as well as to incorporate concepts such as those that relate to metrological traceability, measurement uncertainty and nominal properties, led to this 3rd edition. Its title is now *International Vocabulary of*

Metrology – Basic and General Concepts and Associated Terms, (VIM), in order to emphasize the primary role of concepts in developing a vocabulary.

The main part of the 3rd edition of the VIM consists of the chapters Quantities and units, Measurement, Devices for measurement, Properties of measuring devices and Measurement standards (Etalons). To facilitate the understanding of the different relations between the various concepts given in this Vocabulary, concept diagrams have been introduced. They are given in Appendix A (informative).

In this Vocabulary, it is taken for granted that there is no fundamental difference in the basic principles of measurement in physics, chemistry, laboratory medicine, biology or engineering. Furthermore, an attempt has been made to meet conceptual needs of measurement in fields such as biochemistry, food science, forensic science, and molecular biology.

Several concepts that appeared in the 2nd edition of the VIM do not appear in this 3rd edition because they are no longer considered to be basic or general. For example, the concept 'response time', used in describing the temporal behavior of a measuring system, is not included. For concepts related to measurement devices that are not covered by this 3rd edition of the VIM, the reader should consult other vocabularies such as IEC 60050, *International Electrotechnical Vocabulary*, IEV. For concepts concerned with quality management, mutual recognition arrangements pertaining to metrology, or legal metrology, the reader is referred to documents given in the bibliography.

The major change of the treatment of measurement uncertainty from an Error Approach (sometimes called the Traditional Approach or True Value Approach) to an Uncertainty Approach necessitated reconsideration of some of the related concepts appearing in the 2nd edition of the VIM. The objective of measurement in the Error Approach is to determine an estimate of the true value that is as close as possible to that single true value. The deviation from the true value is composed of random and systematic errors. The two kinds of error, assumed to always be distinguishable, have to be treated differently. No rule can be derived on how they combine to form the total error of any given measurement result, usually taken as the estimate. Usually, only an upper limit of the absolute value of the total error is estimated, sometimes loosely named "uncertainty".

The objective of measurement in the Uncertainty Approach is not to determine a true value as closely as possible. Rather, it is assumed that the information from measurement only permits assignment of an interval of reasonable values to the measurand, based on the assumption that no mistakes have been made in performing the measurement. Additional relevant information may reduce the range of the interval of values that can reasonably be attributed to the measurand. However, even the most refined measurement cannot reduce the interval to a single value because of the finite amount of detail in the definition of a measurand. The definitional

uncertainty, therefore, sets a minimum limit to any measurement uncertainty. The interval can be represented by one of its values, called a “measured quantity value”.

In the GUM, the definitional uncertainty is considered to be negligible with respect to the other components of measurement uncertainty. The objective of measurement is then to establish a probability that this essentially unique value lies within an interval of measured quantity values, based on the information available from measurement. In the GUM, the concept of true value is kept for describing the objective of measurement, but the adjective “true” is considered to be redundant. The IEC does not use the concept to describe this objective. In this Vocabulary the concept and term are retained because of common usage and the importance of the concept.

In this Vocabulary, a set of definitions and associated terms is given, in English and French, for a system of basic and general concepts used in metrology, together with concept diagrams to demonstrate their relations. Additional information is given in the form of examples and notes under many definitions.

This Vocabulary is meant to be a common reference for scientists and engineers – including physicists, chemists, medical scientists, as well as for both teachers and practitioners – involved in planning or performing measurements, irrespective of the level of measurement uncertainty and irrespective of the field of application. It is also meant to be a reference for governmental and inter-governmental bodies, trade associations, accreditation bodies, regulators and professional societies.

Concepts used in different approaches for describing measurement are presented together. The member organizations of the JCGM can select the concepts and definitions in accordance with their respective terminologies. Nevertheless, this Vocabulary is intended to promote global harmonization of terminology used in metrology.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Current problems of independent laboratories in Europe

Pavel Klenovsky

General Director
Czech Metrology Institute
Okruzni 31
638 00 Brno
Czech Republic
e-mail: pklenovsky@cmi.cz

ABSTRACT: In the context of this paper, by independent we mean those labs which are not associated with manufacturers of goods or providers of services (largely different from those associated with calibration and testing), inclusive e.g. national metrology institutes (NMIs) as well. It can be argued that the rise of independent labs can be traced back to the advent of quality management systems in companies resulting in high demand for measurement operations such as calibration and testing. At the same time it gave rise to accreditation in 1990s. As a result, metrology labs (calibration and testing) have had ever since to demonstrate formally their competence to generate correct measurement results – these results have far reaching consequences for any economy. The core issue here is how to ensure traceability of measurements in a coherent and recognizable way which is closely connected with the matter of technical competence of those providing this traceability. Nearly at the same time the effects of globalization and liberalization have set in to change the background for their operation working towards elimination of technical barriers to trade as a main priority. In the course of these developments various stakeholders in this area, where the labs themselves, regulatory bodies, manufacturers, accreditors, certification bodies and service providers have produced measures to maximize their benefits and minimize the associated costs. The interplay of all the involved conflicting interests can be demonstrated on the recent amendment of ISO 17025 standard setting down requirements for laboratories. The most challenging changes have already been in place for some time and this resulting "post modern" situation can therefore be subject to more detailed scrutiny. The aim of the paper is to discuss the character and the implications of those changes and to propose some proposals to improve the situation from the view point of independent labs especially.

1. Past developments

World-wide traceability of results of measurements to the *International System of Units* (SI) is a prerequisite for research, industry and international trade. It also underpins many regulated areas of modern society such as trade, security, health, the environment and quality of life generally. It is worthwhile to point out here that traceability is not a property of an instrument or a laboratory, but it is a property of the outcome of a process which involves instruments and labs – it is a property of the result of a measurement. The historical development of the definition of this term is described in [1]. The original widely used definition was given in the 1993 VIM [2] requiring traceability to stated references, in most cases to national measurement standards maintained by an NMI. Variations of this definition

introduced the important additional requirement that a quality assurance system be in place. The most comprehensive and up-to-date definition of traceability, widely used in practice, is given in ISO/IEC 17025 standard [3] laying down detailed requirements for competence of labs in its par. 5.6. This concept stresses a link to a primary realisation of the SI independently of where this realisation takes place. Since all instruments and standards are subject to changes, however slight, over time, measures involving time relations (metrological timelines) must be carefully designed to get traceability right (see [4]).

In most industrialized countries the demand for traceable calibrations increased dramatically after 1970, so that over the years the workload by far exceeded what could be provided solely by the National Metrology Institutes (NMIs), originally endowed with this mission. The advantage of this arrangement was that the corresponding technical competence could be kept under a tight control on the part of responsible Governments but a dramatic increase in the scope and frequency of individual traceability operations required, in the core of them lying calibrations of measuring instruments, made this arrangement untenable.

To solve this problem and to maintain the mission of their NMIs, some countries established national calibration services in charge of disseminating the SI to industrial and other customers. To assure confidence in calibrations provided by this new infrastructure as well as in their traceability to the SI, these laboratories were, in general, periodically assessed and formally recognized – nowadays called accredited – by the respective NMI. This explains the early and close involvement of many NMIs in laboratory accreditation activities. In the course of time Governments arrived at the conclusion that this process could not be reasonably governed by legislation (anyhow, this vast variety of necessary operations escapes any attempts to subject it completely to a regulation). Governments therefore confined themselves to a protection of recognized public interests, nowadays a matter of legal metrology, and to establishment and supervision of a body responsible for these assessments. Considerations on regulation by way of legislation here are not that far-fetched as it might appear: differences in calibration procedures and corresponding uncertainties have far-reaching consequences to competitive positions of various labs resulting in a non-level playing field, e.g. in the 1990s there was a dispute between a German and a Spanish calibration labs about different uncertainties on calibration of pressure sensors misused to attract customers. The background behind the EA (now EUROMET) Guideline on non-automatic weighing instruments (NAWIs) is another example: it was found out that different labs arrived at different uncertainties evaluating the same set of raw data.

On the other hand, technological progress brought to the foreground, in the 1980s, the issue of quality in modern manufacture and service provision – it gave rise to harmonized standards on quality management systems and their independent assessment by third party bodies (certification of QMS). These quality–related standards have always kept measuring instrumentation under close control laying

down specific requirements for its handling to maintain the metrological parameters of measuring instruments within required limits – the main tool here being their regular calibration. Certification of QMS has become a hot issue in competitiveness – such self-regulation has eventually made any attempts in any regulation here irrelevant. Calibration of measuring instruments not touching on a public interest has ever been since considered in industrialized countries to be a non-regulated area of metrology.

At one stage in the course of these developments a higher scrutiny was given to the operation of accredited calibration labs in terms of technical competence than to those of NMIs – it was eventually pointed out in the 1990s that a higher attention, at least on the formal side, should be paid to quality issues concerning national standards being at the top of the traceability pyramid. At the same time, confidence among different realisations of the same units was not established, leading regulators in some countries to require traceability to a local NMI (the most notable example is the USA – traceability to NIST). This eventually led to the CIPM Mutual Recognition Arrangement (MRA) drawn up by CIPM (International Committee of Weights and Measures) and signed by members (NMIs) of the Metre Convention in 1999 designed to overcome these problems. The arrangement is technically based on regular participation by signatories in a network of interlaboratory comparisons (ILCs) and on the existence of a quality assurance system based on ISO/IEC 17025. In the case of national standards it is more suitable to speak about a degree of their equivalence than about a simple vertical line of traceability. Until recently it was not absolutely clear what the attitude of accreditors towards the CIMP MRA might be – there was surely a temptation on their part to stick to a requirement that NMIs be accredited to achieve recognition by them. This impasse has been hopefully settled by the Joint Statement by CIPM and ILAC on the roles and responsibilities of NMIs and nationally recognized accreditation bodies (NABs) issued in November 2005 which could be considered, at least implicitly, as a mutual recognition of the respective MRAs (ILAC MLA, CIPM MRA). The second problem has been overcome by the fact that calibration certificates of all the signatories are considered equivalent to the degree given by their respective calibration measurement capabilities (CMCs) specified in a special database (http://kcdb.bipm.org, Appendix C) giving the traceability to the SI. This however does not include traceability to its own national standards when required by regulators. The CIPM MRA provides the solution: regulators can now require traceability to the SI and not to national standards of a particular NMI, a process in progress at present e.g. in the USA.

The signatories of the Metre Convention have an obligation to ensure the dissemination of the SI units to the users in their countries. In most countries, the national measurement system requires the NMI to provide the national measurement standards against which they calibrate the standards held by accredited calibration laboratories, which in turn provide calibrations to industrial users with accuracy sufficient for its intended use. This calibration chain can only be as strong as its weakest link, so that many countries believe that the link between the NMI, the

accreditation body and accredited calibration laboratories should be as close as possible in order to make the knowledge of the NMI experts, being undisputably technically competent and ideally impartial, available to all concerned to the greatest possible extent. The above mentioned statement links together both recognition arrangements, CIPM MRA and ILAC MRA, being complementary in assuring an unbroken traceability chain from day-to-day measurements made by users to the SI. It is widely believed that accreditation will be able to deal with all the technical issues in labs on a self-regulatory basis to provide correct and coherent measurements and to generate a level playing field on the calibration market.

2. Current situation

It can be argued that, with accreditation as a last major change being around for more than 10 years now, the situation in the metrology business is relatively stabilized and it is therefore worthwhile to analyze the current trends and approaches, especially from the viewpoint of independent labs.

Firstly, the situation in metrology in general is greatly influenced by underlying (paper) standards under the auspices of world standards bodies (especially ISO). Companies these days are increasingly involved in a multitude of activities starting from various conformity assessment activities (CAAs) over calibration and testing up to provision of other products. It could be expected that this complex situation is addressed, in terms of requirements, by writing the corresponding standards, like e.g. ISO 17000 series for CAAs, on the platform of ISO 9001:2000 having a generic nature to achieve a desired compatibility of all the standards greatly facilitating their subsequent assessments of whatever kind. ISO CASCO has a policy how to use ISO 9000 standards in the 17000 series but it concentrates only on technicalities how to intertwine the standards. After all, all the CAAs are about delivering a product (service) to customers and all the management requirements are aimed at supporting this delivery – they have therefore the character of quality management requirements in terms of ISO 9001:2000. It offers itself to write the 17000 series of standards, wherever possible, in terms of quality management requirements of ISO 9001:2000 as a common platform – as a result, companies could prepare an integrated quality management system covering all the activities possessing the character of provision of goods in a broad sense whereby greatly reducing duplicities in its assessments (by way of accreditation or certification) and the associated costs (it could potentially lead to a one-stop assessment). This is further supported by the on-going process of alignment between different management systems (quality, environmental, OH&S). Unfortunately, due to vested interests aimed at protecting one´s own business, this approach is not pursued in the practice of ISO CASCO – instead, artificial differences are rather amplified or maintained as is the case of ISO 17025 (see below).

Secondly, it has to be noted that the CIPM MRA has created a competition in calibration services among NMIs themselves and repercussions from it still need addressing. The econo-organizational basis and environment of various NMIs (ranging from full Government bodies up to private bodies) is extremely diversified at the moment, there is a danger that various forms of subsidies being involved here might negatively influence the situation. At present when Government resources are becoming scarce (globalization, ageing of population, welfare state) the time will come for all NMIs when a greater involvement in various kinds of relevant highly qualified services will be a matter of survival. Such a development would, however, generate another sensitive problem: a competition between NMIs and independent labs. Currently, large NMIs, being basically research-oriented organizations, are strictly forbidden to enter any competition with private labs – it is given by the fact that the share of Government subsidies to cover their costs is overwhelming (around 90%). On the other hand, some smaller NMIs are much more involved in various services which inevitably leads to overlaps with private labs. In such a case it is imperative that such NMIs have a transparent accounting system being able to assign costs and revenues to individual units and tasks to be able to prove that public funds are not used to violate competition rules. In Europe there is a substantial infrastructure and legislation watching over fair competition which might act upon any established cases of misdemeanor inclusive of those regarding NMIs (the Treaty establishing the European Community, Rules on Competition, art. 81 – 86 + EC Regulations) – a universal ban on entering competition of the type mentioned above is therefore not necessary.

Whereas a competition among NMIs is a matter of fact (even highlighted in ISO/IEC 17025 – see Note 6 after par. 5.6.2.1.1), competition among NABs is strictly forbidden, arguably as a result of long-term lobbying by accreditors themselves. The reasons for this concept are either securing a necessary impartiality in carrying out this type of conformity assessment (an argument going rather to an extreme – there are other situations, e.g. in legal metrology, where decisions have to be made under even more conflicting and pitched-up interests) or that it is simply a requirement of regulatory bodies (with which e.g. calibration has nothing in common). Whatever the reasons, accreditation bodies are now set up as national monopolies in all the ILAC member countries with the exception of the USA. Here it is more or less a competitive business, in Europe at least cross-border accreditation has been, after much haggling, eventually conditionally accepted to be finally effectively banned in the draft revision of the new approach (see draft EU Regulation 2007/0029 (COM)).

Great care is therefore taken not to expose NABs to any commercial motivation – arguably nothing, even their monopoly position, cannot beat the negative influence of the fact that this service is charged to applicants. Clearly, there is a natural tendency toward a cozy relationship between accreditors (and to that effect, certifiers as well) and their customers – granting the accreditation satisfies both sides, a declined accreditation is a nearly non-existent case. Even the non-profit

character cherished by regulators (European Commission in case of the EU) cannot save much here: NABs do make a profit and these can easily be transposed into various forms of remuneration of their own employees, unlike profit-making bodies where the profit is a property of the owners. In such a way it can easily occur that employees of a non-profit company are paradoxically more commercially motivated than employees of profit-making companies which is exactly what the arrangement is seeking to avoid. There are other negative effects as well: a general tendency to overcharging and to save up as much as possible on external contracting which has a profoundly negative effect on technical activities associated with accreditation (see below). Especially the former matter is quite sensitive for independent labs which cannot use any internal cross-subsidy to compensate the costs.

Apart from apparent disadvantages for their customers a monopoly position at least gives accreditation bodies an ideal position for impartial decision-making which should subsequently imply major improvements in real technical competence of laboratories. This effect can be assessed by an evaluation of results in interlaboratory comparisons (ILCs) in the calibration area or proficiency testing (PT) in the testing area achieved in the long-term. Unfortunately, no major breakthrough has been achieved here as reported on several occasions in the past (see e.g. [5]).

As far as regulators are concerned, they have obviously had a tendency to promote the use of accreditation in any type of technical assessments required by regulations, probably due to a lack of any other tool – they have put all their eggs into one accreditation basket and occasionally have tried to make corrections. Whereas accreditation conceived in such a way is surely acceptable in the regulated area, it is surprising that e.g. the European Commission goes in the revision of the new approach that far to forcefully promote a monopoly of accreditation in the non-regulated area (accreditation of calibration and testing labs) – the fact that well over 10,000 bodies, being mostly small and medium enterprises, get their access to the market by way of such a monopolistic structure is against very basic principles of the EU – a free movement of goods, services in this case. The revision of the new approach is the last twitch in this matter: a legislative guarantee to monopoly is being granted to NABs represented by the EA. In the harmonized regulated area it might be acceptable, especially when it virtually establishes a level playing field in all the EU member countries. We would expect that such a move is preceded by a study to analyze to what extent accreditation really guarantees technical competence, especially when an extensive source of such information has been piled up by one of the EU's own research institutes, IRMM (the IMEP program of proficiency testing in the critical chemical area which unfortunately shows rather the other way – see below). What is surprising here is that the European Commission (CEC) goes that far to legalize monopoly of accreditation in non-regulated area as well. Apart from it being legally correct by some unintentional magic it is just at the time when the new directive on services aimed at establishing a single market in services takes effect. From the broader viewpoint of laboratory community (accreditation of labs: 80% of the NAB workload) it is not clear why such situation in non-regulated area should be

beneficial for labs but the whole matter goes largely unnoticed by the laboratory community – independent labs are in majority small and medium size enterprises (SMEs) so that this is a extremely sensitive matter for them.

Another problem is that a customary close cooperation between NMIs and NABs has caused some in the accreditation community to raise concerns about the independence of the NABs within NMIs. As mentioned above such close ties have historical origins and do not lack a lot of good justification: an easy access to impartial technical expertise is guaranteed under such arrangements which should therefore be rather cherished by NABs instead of being criticized – after all, it is the accreditation community that has maintained that the main difference between accreditation and certification is an involvement of the former in technical underlying matters. This issue has caused a spilt e.g. in Europe a lot of bad blood but with the advent of ISO/IEC 17011:2004 the dispute will hopefully be ended: a NAB can be part of a NMI when proper firewalls are established between accreditation and other activities. One of the prerequisites here is a withdrawal of those NMIs from a competition with private calibration labs – NMIs with a NAB inside have therefore been set on a different path of development than the other ones trying to attract more services of a highly qualified nature (see above).

With an expansion of accreditation to other areas (now even to manufacturers of CRMs) and with a huge number of labs exposed to accreditation (in Europe over 10,000) the lucrativeness of this business attracts more and more attention of certifiers of QMS, this certification being a similar type of business. We sometimes get the impression that behind the scene a fierce struggle between accreditors and certifiers for the so to speak "laboratory market" is going on. The unsubstantiated differences between ISO 17025 and ISO 9001 and the harassment on the part of accreditors against certifiers of QMS described below can potentially lead to a drive on their part to take over the whole issue of management systems in laboratories from accreditors (attempts in this direction have already been made, e.g. in France – the AFAQ case going up to the minister of industry). Indeed, they are very well positioned for that as their qualification for management matters is the same as that of accreditors and for technical matters they can contract external experts exactly as most NABs do. A successful strategy on the part of certifiers here has to be based on shifting the stress in auditing more to technical competence and on starting to use suitably qualified technical experts in the field being subject to auditing.

It cannot escape the attention that certification of QMS, to the dismay of big industry, has grown to a huge business as well. As accreditation now basically governs an access to the market for laboratories the same is applicable to certification/registration in relation to product provision. A lack of expertise in the product provision under assessment often leads to a certain formalism in handling metrology issues by certifiers: all the equipment items measuring something are put under mandatory calibration independently of their application. Metrological services, especially calibration, are among those most exposed to outsourcing,

nowadays quite popular with big globalized industry, with the aim to cut the costs of needed calibrations. Internet-based auctioning of calibration services is one of the most controversial approaches when we take into account the inherently non-tangible output of this service whose quality cannot be easily checked by the user. A provision of a complete service together with necessary adjustments and at least basic repairs and maintenance is an indispensable prerequisite for a success here. All these trends generate a "low-end" market in calibrations where no lows in calibration fees offered are low enough. Foreign direct investors (FDIs) moving with their factories around the globe, shopping for low costs, are the biggest proponents of this approach. When this is combined with a cut-throat competition in the calibration business, like e.g. in the Czech Republic (nearly 100 accredited labs in a nation with a population of 10 million in comparison to 72 in the Netherlands, 300 in Germany, etc.), then it would take a lot to maintain a standard of quality (technical competence) at an acceptable level – a demanding role for accreditation. This is currently the situation in countries with a lot of FDIs (new EU members) but it is bound to spread to traditional industrialized countries (old EU members) in the long term as well.

Matters are further impaired here by a discernable tendency in Europe for the accreditation bodies to get less involved in the technical issues of laboratories. Instead, they concentrate more on what they consider their core competence, which is lead assessments with their focus on the quality management system, virtually leaving the technical assessments and other technical activities, including writing of guidelines, to subcontractors. Under such conditions it cannot be expected that great care is given to get those crucial matters right, resulting in unfair competition. Moreover, some observers feel that the competence profile of the personnel in the accreditation body is starting to resemble the competence profile of a certification body. There is a danger if accreditation starts to look more like certification that it will then be more difficult to distinguish between the two in the marketplace. A calibration laboratory may then consider, as an alternative, certification towards ISO 9001:2000 with the addition of ISO 10012.

When quality issues are at stake metrology is lucky enough to possess a tool providing nearly complete feedback on technical performance of measurements of any kind – it is proficiency testing (PT), in physical metrology wrongly called interlaboratory comparisons (ILCs) – see [6] (sometimes also called horizontal traceability). First introduced in everyday practice by ISO/IEC 17025:1999 PTs have taken off the ground rather slowly due to 2 reasons:

1) it is only one but the most challenging one out of a number of options in ISO 17025 (par. 5.9);

2) adding to already high costs associated with accreditation labs have been rather reluctant to participate extensively here without any "rewards", e.g. reductions in surveillance visits by accreditors.

Therefore, ILAC has had to lay down a policy of minimal participation but in some countries, like in the Czech Republic, rather extensive programs have been implemented for a couple of years now with a lot of data accumulated.

It has been demonstrated in the past that the situation in chemical labs (one kind of testing lab) has not been satisfactory (see [5] and Figure 1) – the interpretation of those results is simply that up to 50% of labs would fail and out of those who failed up to 50% are accredited labs, repeatedly in the long-term. This is applicable to PTs where great care is taken, like in the IMEP program in Europe (www.irmm.jrc.be/imep), to properly establish the certified values (with uncertainties) of RMs used for PTs and to shut out any possible leaks of information about the certified values to participants in those PTs (to use here commercially available RMs does not reveal much). However, the situation in metrology in chemistry is complicated by deficiencies in traceability in some areas, especially where it is extremely difficult to separate analytes from the matrix material. What is a new tendency in this would-be "stabilized" situation is that unsatisfactory results are starting to be found in physical metrology as well, especially in areas like temperature with no apparent nominal values, like in case of platinum resistance thermometers in the Czech Republic. It might be a result of that aggressive pressure on cost-cutting and of other factors mentioned above. Moreover, the performance of non-accredited labs is starting to surpass that of accredited ones. If confirmed, it is a new, highly negative trend to raise concern of all those involved in metrology business to shake up the current system of distribution of traceability described above. All the above mentioned problems are getting under the spotlight of metrology community, at least in the Czech Republic. A regulation cannot be employed here, a unifying effort on the part of accreditors to enforce the same rules for everybody is widely believed to be missing (or, at least, the accreditation process is non-transparent enough to judge it from the outside). These rules are of a technical nature, like unified procedures and handling of uncertainties, being now abandoned by accreditors and going back to NMIs where they originally were, NMIs uncertain whether to embrace it (the money will stay with accreditors in any case). The result is confusion and apprehension about where we in metrology community are heading in future.

It has to be pointed out here that, unlike PTs among NMIs for purposes of the CIPM MRA, participants cannot be publicly identified based on a requirement from ISO Guide 43 enforced on PT providers by accreditors. It is easy to imagine the sole positive effects on technical competence of making such results public (the market itself will arrange for self-regulation here) – but it will be a real uphill struggle to achieve any change here, important stakeholders here (low end labs themselves, accreditors) will be against with users and high quality (but more expensive) labs benefiting. In this context it has to be welcome, as a kind of revolution with far reaching consequences, when the mentioned Joint Statement by CIPM and ILAC calls for "work towards transparency and open publication of the results of comparisons between named accredited labs". As shown above, it is arguably the

only way forward to overcome the problems concerning the quality of measurement results in the future in the globalized world.

The less involved the accreditors are in technical matters the more successful they are in marketing accreditation as already mentioned. The fact that accredited labs are allowed to use, concerning traceability, only services of other accredited labs (and NMIs) is logical – but nowadays users of measuring instruments are also starting to require accreditation in increased quantities, the most notable example of this being the automobile industry by its ISO TS 16949:2002 standard written on the platform of ISO 9001:2000. It requires technical competence but in practice it is reduced to accreditation – strictly speaking, calibration has to be made by an accredited lab, not exactly that the calibration itself has to be accredited. Due to mentioned deficiencies in accreditation the other ways to demonstrate technical competence (evidence from PTs, CIPM MRA in case of NMIs – see below) should be encouraged to be used more widely.

Finally, a note to the role of manufacturers of measuring instruments. A discernable change of attitude here can be recently identified in what is made publicly available in terms of proprietary technical information on those instruments necessary for any services on them in use. Originally, information on adjustments, basic maintenance and troubleshooting if needed was readily available, at least to users if not openly, which independent servicing organizations inclusive of calibration labs could use to be able to offer a complete service. However, recently manufacturers in some areas (weighing instruments, CMMs, medical devices) have embarked on a strategy to work through a network of authorized national representatives as their daughter companies releasing this vital technical information only to them and after any sale they would offer to users fixed-term contracts for after-sale servicing of those instruments whereby securing an outstanding position in this business. Ideally, this servicing should include calibration and, if possible, subsequent verification in the regulated area as well. A visible consolidation is going on among manufacturers of measuring instruments, the stronger ones are buying out the smaller ones – afterwards they will start offering complex services around those instruments inclusive calibration by their own regional service centers effectively squeezing any independent labs out of that business. Taken to an extreme, any services on such instruments could eventually be made only by them with no independent bodies around – it would, among others, prevent creation of any single market in these services in free trade areas, a matter rather escaping e.g. any attention on the part of the European Commission.

Similar development is taking place in the European harmonized regulated area (new approach directives) where a number of independent labs are active as notified bodies (1700 bodies altogether, 400 of them in metrology). What is historically going on in the conformity assessment area can be summarized with some simplification in the following 3 stages:

– multiple testing (in each country) exclusively by third party bodies;

– third party testing, only once on the Single Market with some self-declaration options for manufacturers (module A – but exceptionally applied in metrology);

– assessment of the technical documentation (module H).

The whole process is arguably driven by an extensive lobbying on the part of manufacturers at the CEC and is characterized by ever decreasing involvement of independent bodies in these activities. Therefore, the only imaginable next stage is a zero involvement of NBs which some manufacturers might consider the ultimate goal, an ideal situation. Only recently they, at least the high end ones, are coming to the perception that a reasonable regulation generates a level playing field and provides a protection against (mostly unfair) competition of low quality products from low-cost countries.

In summary, it can be argued from the above that independent labs are nowadays under increasing pressure from all the possible angles:

– in accreditation which is in the way of their entering on the market they face a stronger monopoly position that can potentially lead to overcharging and excessive administrative burden, especially for labs being SMEs (a majority of them);

– effective cost-cutting efforts on the part of users of measuring instruments are pushing calibration fees down without any apparent lower limit;

– manufacturers of measuring instruments make a point of not disclosing any necessary technical information for calibration/verification on a larger and larger scale to get an exclusive position in provision of those services;

– all the services (with an interesting exception of accreditation) are being increasingly liberalized in Europe in favour of manufacturers and users (the element of impartiality plays ever decreasing role).

3. The 2005 amendment of ISO/IEC 17025

Some of the above mentioned trends and approaches can be demonstrated on the situation around the amendment to this important standard for laboratories, being a center of conflicting interests, intended to reestablish its alignment with the big revision of ISO 9000 family of standards on QMS (for details see [7]).

4. Proposal of a new long-term strategy

As follows from discussions above, bodies subject to certification/accreditation are now more and more concerned by the following issues:

– real added value of any third party assessment in the present mature situation;

– an excessive dependence of certifiers/accreditors on the financial side of the business sidelining the technical content;

– a hostile environment for small and medium enterprises (SMEs) created by certifiers/accreditors.

Even if calibration (and testing strictly speaking as well) is not a conformity assessment activity (see [8]) accreditation is a legitimate approach for the demonstration of technical competence of labs but it is not the only imaginable approach to be applied here. The other ones can be summarized as follows:

– client audits;

– extensive participation in ILCs or PT schemes;

– self-declaration supported by evidence from PT and ILC programs with openly accessible results (further supported by e.g. client audits and by making openly available a part of the QMS documentation, e.g. on the internet);

– peer assessment on the basis of ISO 17040 (e.g. within national associations of calibration and/or testing labs, especially suitable for mutual recognition arrangements among NMIs – CIPM MRA, OIML MAA);

– certification against ISO/IEC 17025 or cross-referenced ISO 9001:2000 and ISO/IEC 17025 or ISO 9001:2000 + ISO 10012.

At present, accreditation dominates here even if there are legitimate grudges against some aspects of its practical implementation. Big industry is now pressing for self-declaration of their QMS; in the laboratory community a similar idea is being shaped up as well as a consequence of an obvious inability of accreditation to tackle the prevailing deficiencies in technical matters associated with competence of labs no matter what amount of money labs are paying for such services. Self-declaration would be much more acceptable here due to the existence of a supporting, underpinning element, the publicly available and identifiable results from proficiency testing. Accreditation in the area of calibration and testing should therefore withdraw from here to be confined to accreditation of PT providers only. Such a system will be much more efficient in achieving a high level of real technical competence at much lower costs for labs. Such a major change can only be achieved in the long term with a full support of the global laboratory community, now supported by the Joint CIPM – ILAC Statement, and steps would have to be taken to

create a robust and extensive volume of PT activities to be able to effectively support self-declaration claims to be viable and recognizable.

Labs also should take steps in the revision of the new approach aimed at limiting the legislation on accreditation to the regulated area only. The CEC, DG Enterprise should ideally give some perspective to the current notified bodies or to clearly tell them that they should count with their zero involvement here – all the necessary investments on their part in new instrumentation will be simply be wasted.

The existence of the directive on services should be used as a vehicle to influence matters regarding provision of necessary technical information by manufacturers for services on measuring instruments in use. Otherwise, there is a danger that these services will eventually be monopolized by manufacturers of measuring instruments and their authorized representatives with negative consequences for the users and public at large.

On the basis of cooperation of NABs and NMIs a lot of work should be done to make an order in the BMC/CMCs issue, starting with a unified definition of measurement capabilities, and their alignment along the border line to create a fair and technically sound competition among various types of calibration labs in Europe.

As a great deal of lobbying would be necessary to achieve the above outlined goals calibration labs, Europe should consider creating an association of their own (European association of calibration or standard labs?), something similar to NCSLI in the USA. Currently, the only option here is to enter EUROLAB where membership is already dominated by testing labs, especially from the chemical area, so that it is questionable whether this representation would serve calibration labs well. On the other hand, EUROMET as an association of European NMIs is concentrated in its cooperation on a number of matters but surely not on services in metrology – it is obviously, out of a number of reasons, rather unsuitable to represent calibration labs at large.

5. Summary

We are now living in a world where the cheapest products of even a doubtful quality eventually seem to get the upper hand, arguably a result of continuous reductions in various types of safeguards (e.g. various forms of testing) aimed at securing their quality being sacrificed at the altar of lowest possible costs (at any cost) under the buzzwords of globalization, liberalization etc. It puts the quality-related issues where metrology plays a significant role under enormous pressure. More and more information is coming in as evidence of shortcomings in this area – we are even getting an impression that quality in general is poorer now than tens of years ago despite vast financial resources spent on various third party assessments. Globalization is a fact of life, at least at the moment, but approaches should be

explored how to use these resources in a better way to be able to get at the core of quality of a given product. Furthermore, the center of gravity in various assessments and testing should move from objects "as delivered to conformity assessment bodies" to objects "as customarily delivered to customers". Such a step providing for an ideal impartiality would necessarily imply an involvement of public funding which is a problem but things are bound to head in this direction. The area of metrology as a whole is no exception with the luck of possessing a tool able to deal with this problem effectively and the metrology community should do its best to bring it to life, e.g. based on information provided by this paper.

References

[1] M.Kuehne, M.Krystek, A.Odin, "Traceability of measurement results in industrial metrology", *Proceedings of NCSLI Conference*, Washington DC (2005).

[2] *International Vocabulary of Basic and General Terms in Metrology*, second edition, 1993, BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, ISO, Geneva (1993).

[3] ISO/IEC 17025:2005, General requirements for the competence of testing and calibration laboratories, ISO, Geneva (2005).

[4] Ch.D.Ehrlich, S.D.Rasberry, "Metrological Timelines in Traceability", *J. Res. Natl. Inst. Stand. Technol*. 103, 93 (1998).

[5] P.Klenovsky, "The relation of conformity assessment standards to metrology", *Proceedings of 2nd Middle-East Metrology Conference and Exhibition*, Kingdom of Bahrajn (2004).

[6] S.Klenovska, P.Klenovsky, "Interlaboratory comparisons in Europe: the situation and the lessons learnt", *Proceedings of NCSLI Conference*, Washington DC (2005).

[7] P.Klenovsky, "The 2005 amendment of ISO/IEC 17025:1999 and its implications for laboratories", *Proceedings of NCSLI Conference*, Washington DC (2005).

[8] ISO/IEC 17000:2004, Conformity assessment – Vocabulary and general principles, ISO, Geneva (2004).

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The harmonized European gas cubic meter for natural gas as realized by PTB, NMi-VSL and LNE-LADG and its benefit for the user and metrology

D. Dopheide[a], B. Mickan[a], R. Kramer[a], H-J Hotze[a], D. Vieth[a], M. Van Der Beek[b], G. Blom[b], J-P. Vallet[c], O. Gorieu[c]

[a] dietrich.dopheide@ptb.de
PTB-Pigsar, Germany

[b] Mvanderbeek@NMi.nl
NMi-VSL, The Netherlands

[c] cesame@univ-poitiers.fr
LNE-LADG, France

ABSTRACT: This paper describes the background of "The European Harmonized Reference Values for the Cubic Meter of National Gas" realized by three independent national metrology institutes: PTB Germany, NMI VSL, the Netherlands, LNE France.
The outcome of such a process is to define a unique reference value that is more precise and reliable with a long term stability which can be considered now as the worldwide best available realization of the high pressure gas cubic meter since key comparisons carried out have confirmed this reference value.
The perequisites, procedures and results of the harmonization process will be pointed out as well as the mutual benefit for users and metrology.
That is why this paper can be considered as a tool for metrology in general and will be presented partly with video animation showing the potential, and the economic benefits of the harmonization process.

The challenge of gas flow measurement

Since the 1970s an increasing use of Natural Gas as an energy source and in Europe a vast network (gas grid) has been realized to enable an average gas consumption in Europe of more than 400 billion cubic meters per year. In this expanding gas grid more and more points of transfer of ownership are installed, leading ultimately to an increasing demand for reliable and stable reference values for high-pressure gas-flow measurements. The principle of Third Party Access, supported in the future by direct invoicing of energy-shipment, makes it of vital importance that gas-transport organizations have at all times a clear knowledge about the contents of their transport-grid.

Hence, long term stability of reference values is gaining importance. Although small (insignificant) changes in (national) reference values are accepted by metrologists, the impact of variations on e.g. invoicing will probably never be understood nor accepted. The drive for one equivalent reference value in this working field of Natural Gas resulted in extensive cooperation between three NMIs holding test facilities for High Pressure Natural Gas in Europe. The results of these activities have been verified by means of CIPM Key Comparisons conducted under the auspices of the CCM.

Prerequisites for harmonization

In previous papers and mainly during the last FLOMEKO 2003 in Groningen and FLOMEKO 2004 in Guilin, the authors have already described the harmonization procedure between PTB-pigsar and NMi-VSL and LNE-LADG in detail, see e.g. [1], [2]. The participating facilities at NMi-VSL are presented in [3], the German National Standard pigsar has been described in [4] and [5] in detail. The LNE-LADG facilities are presented in [6].

The European Harmonized Reference Level or Gas Cubic Level comprises a weighted average of three different individual national realizations of the gas cubic meter (reference levels). This weighted average is based upon the following metrological prerequisites:

1) PTB, NMi-VSL and LNE operate independently realized Traceability-Chains. At NMi VSL a system based on mass-comparison of gas-flow is in use (basis verification system), whereas the German National facility for high-pressure gas-flow standards, PTB-pigsar has its traceability-chain in operation based mainly on a Piston-Prover as well as on LDA (Volume comparison plus density determination via pressure, temperature) and LNE applies the pVTt-method (mass comparison).

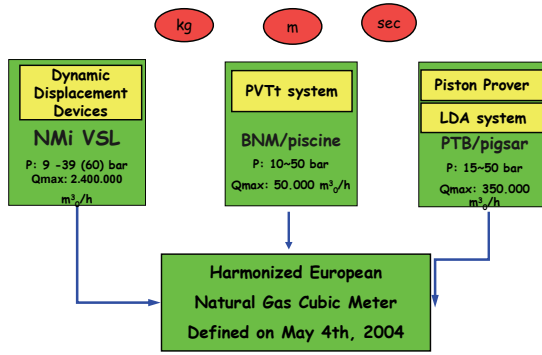


Figure 1. *Visualizes the calibration and measuring range of the three completely independent national primary standards of NMi-VSL, LNE and PTB and their interaction to realize the Harmonized European Natural Gas Cubic Meter*

2) The uncertainty-budget of each of the systems is fully known, understood and mutually accepted.

3) A permissible difference between the three systems smaller than the Root Square Sum of the corresponding uncertainties (2σ) is established.

4) The stability of each chain (sets of reference values) is demonstrated. Stability refers to the reproducibility of the Reference Values over the years.

5) The Degree of Equivalence is established (based on historic performance and on accepted uncertainties).

This procedure has been applied in all overlapping flow rate and pressure ranges of pigsar-PTB, NMi-VSL and LNE. This ancillary condition can be considered as prerequisite.

LNE, PTB and NMi-VSL have applied three to four sets of different turbine meters (two in series) to allow a maximum of overlap. In addition, we have also applied a choked nozzle.

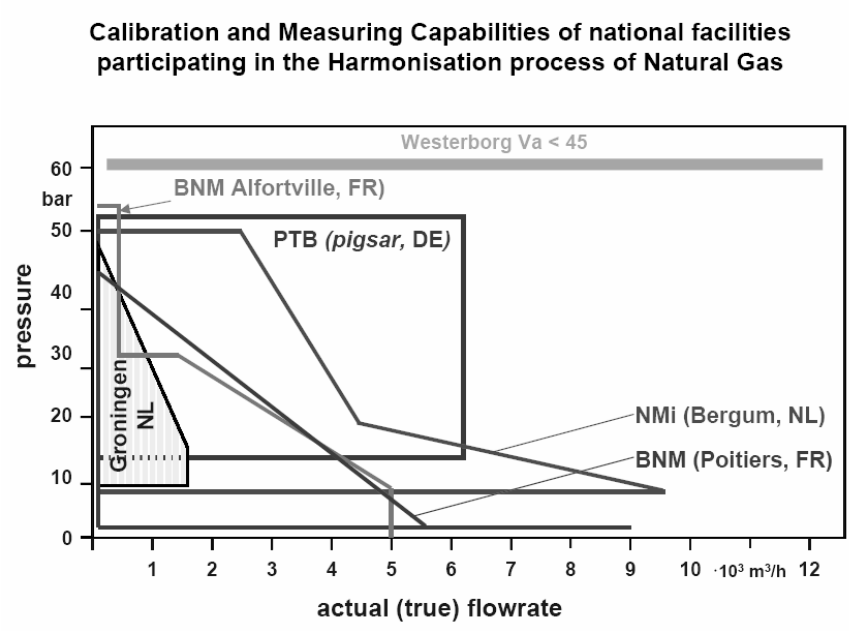


Figure 2. *Calibration and measuring capabilities of PTB-pigsar, NMi-VSL and LNE. Harmonization has been mainly applied in the overlapping ranges*

All partners have agreed to search continuously for improvements of the metrological independent traceability chains to meet future demands for more stable reference values with smaller uncertainties. The main benefit for customers is the same and equivalent calibration of meters at any calibration test rig in Germany, the Netherlands and France. The harmonization as accomplished by PTB, NMi-VSL and LNE is principally open to third parties if all five prerequisites can be met and if it is practically feasible. So far however, there is no other national facility available in the world, which meets all prerequisites.

Harmonization process for reference values

To understand the technique of the weighted average, which has been applied in the harmonization process, let us discuss the method for two partners at first and then it shall be expanded towards all 3 partners using the latest results from 2004.

This method has already been explained in previous papers, e.g. [1], [4], [5] and shall therefore only be summarized.

Based on the fact equivalence and independence of calibration chains, the "true value" f_{Ref} of meter deviation shall be assumed as the weighted average of any pair of results. In Figure 3 an example of one meter calibration pair is given. The meters used in the transfer packages are the Reynolds balanced, therefore the determination of difference $\Delta_{\mathrm{PTB\text{-}Ref}}$ ($\Delta_{\mathrm{NMi\text{-}Ref}}$ resp.) to the common reference level is done with respect to Reynolds number. In practice each pair of measuring points is close together but is not exactly at the same Reynolds number. Thus, polynomial approximation of calibration curve f is used as seen in Figure 3. The weighted average f_{Ref} is now calculated using the polynomials. The differences $\Delta_{\mathrm{PTB\text{-}Ref}}$ and $\Delta_{\mathrm{NMi\text{-}Ref}}$ are determined for each measured point relative to the average polynomial

$$f_{\mathrm{Re}\,f} = W_{NMi} f_{NMi} + W_{PTB} f_{PTB}$$

with:

$$W_{NMi} = \frac{1}{\frac{U_{NMi}^2}{U_{PTB}^2} + 1}$$

and:

$$W_{PTB} = \frac{1}{\frac{U_{PTB}^2}{U_{NMi}^2} + 1}$$

$$\Delta_{PTB-\mathrm{Re}\,f} = f_{PTB} - f_{\mathrm{Re}\,f} \quad \text{and:} \quad \Delta_{NMi-\mathrm{Re}\,f} = f_{NMi} - f_{\mathrm{Re}\,f}$$

f - meter deviation w - weighing factor

Δ - difference U- uncertainty (k=2)

f_{Ref} is the meter deviation of the meter under test based on the harmonized high pressure cubic meter of NMi-VSL and PTB.

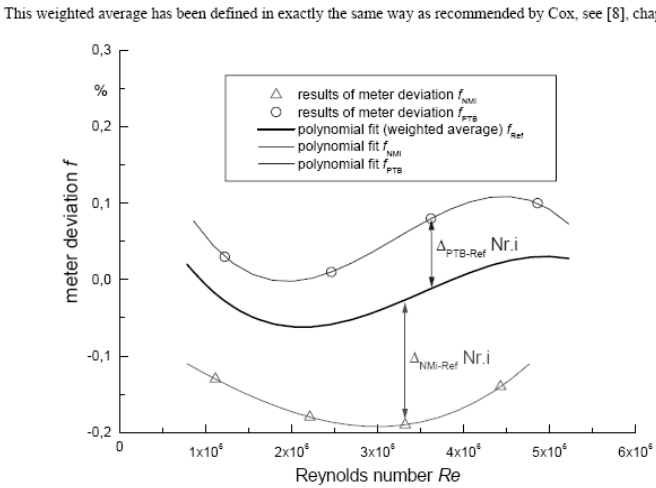


Figure 3. *Results of comparison for one meter at one pressure stage and determination of differences* $\Delta_{PTB\text{-}Ref}$ *and* $\Delta_{NMi\text{-}Ref}$

The outcome is that the participant who offers the smallest uncertainty will pull the reference value towards him heavily.

Finally, all determined differences $\Delta_{PTB\text{-}Ref}$ and $\Delta_{NMi\text{-}Ref}$ for all meters at all pressure stages were put into one graph depending on the Reynolds number (Figure 4) presenting the original data from 1999. The reproducibility (2σ) of calibrations are less than half of the uncertainty budget of each participant. Nearly every result of one participant lies in the uncertainty interval of the other. Although three different meter sizes and two different pressure stages for each size were used, there is no significant discontinuity to be seen. This is an evident demonstration of high quality and reliability of calibration work of both partners, NMi-VSL and pigsar.

The determined difference $\Delta_{PTB\text{-}NMI}$ between NMi-VSL and pigsar increases slightly with Reynolds number. The slope of the results of NMi-VSL is only a mathematical effect of the weighing process because the uncertainty U_{NMi} of NMi's chain increases with pressure stage. The trends for $\Delta_{PTB\text{-}Ref}$ and $\Delta_{NMi\text{-}Ref}$ in Figure 4 can finally be approximated by a linear function depending on the logarithm of the Reynolds number. These linear functions are used as correction functions in order to disseminate a harmonized value of cubic meter high pressure natural gas in both countries.

In Figure 4 the harmonized reference level has been put on the zero line to demonstrate the effects of weighted means.

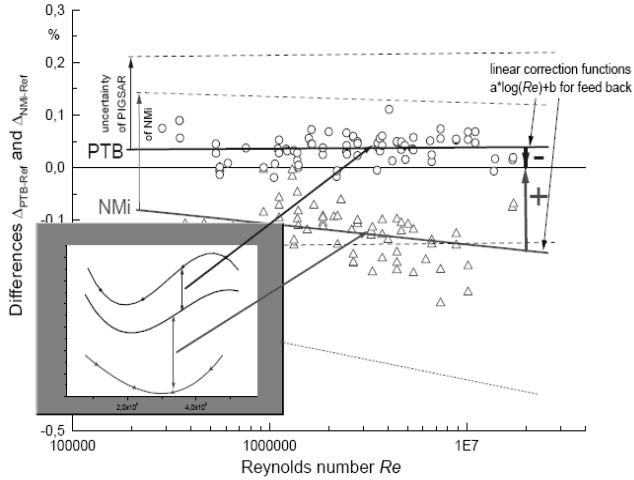


Figure 4. *Summary of all determined differences $\Delta_{PTB\text{-}Ref}$ and $\Delta_{NMi\text{-}Ref}$ for all meters in all pressure stages plotted as function of the Re-number*

The difference between both traceability chains is clearly to be seen but much smaller than the uncertainties. Within the reproducibility of the results there is no significant discontinuity although three different meter sizes and two different pressure stages for each size were used. To implement the feedback of comparison results, linear approximations of differences $\Delta_{PTB\text{-}Ref}$ and $\Delta_{NMi\text{-}Ref}$ were determined.

The uncertainty levels (2σ) shown in the graph are the particular uncertainties of pigsar and NMi-VSL

The following conclusion from Figure 4 can be drawn: the partner with the smaller uncertainty pulls the reference value towards him. In Figure 4 PTB is a little bit closer towards the harmonized reference value.

The cubic meter obtained at pigsar is (was in 1999) a little bit too large and the cubic meter obtained at Bergum is a little bit too small and therefore both sides have to correct their results with a correction factor (wich is actually a function of Re number, pressure and flow rate).

Due to the comparison measurements we have two independent sources of information of the "true value" given by both calibration chains, hence we obtain a lower uncertainty level U_{Ref} of meter deviation f_{Ref} based on harmonization:

$$\mathrm{U_{Ref}} = \sqrt{w^2{}_{NMi}U^2_{NMi} + w^2_{PTB}U^2_{PTB}}$$

with:

$$w_{NMi} = \frac{1}{\frac{U_{NMi}^2}{U_{PTB}^2} + 1}$$

and:

$$w_{PTB} = \frac{1}{\frac{U_{PTB}^2}{U_{NMi}^2} + 1}$$

Adding 1 partner LNE:

$$U_{Ref} = \sqrt{w^2{}_{NMi} U_{NMI}^2 + w_{PTB}^2 U_{PTB}^2 + w_{LNE}^2 U_{LNE}^2}$$

U – uncertainly (k=2) w – weighing factor

U_{Ref} is the uncertainty of the harmonized reference value for high pressure cubic meter of the different partners.

The positive outcome for the customer is, that he obtains always the same calibration in any country at any test facility and he can enjoy the benefit of a very stable and small uncertainty of the harmonized reference value.

The benefit for metrology is the reduced uncertainty of the harmonized reference value.

Since November 1st, NMi-VSL and PTB have disseminated the same "harmonized" high pressure natural gas cubic for all calibrations, which have been performed at their test facilities. They have been joined by LNE-LADG (France) from May 2004.

European gas cubic meter

Harmonization procedure between PTB, NMi-VSL and LNE

As the initiators have been applauded for their work by the users, they completed the harmonization procedure by inviting the French LNE and their "piscine" facility to participate. Again, all 5 or 6 prerequisites have been checked carefully in very long discussions and evaluation procedures. Since May 4th 2004 the French facilities

represented by LNE-LADG have been a partner and full member of the European Harmonization Club.

Some of the harmonization results among all three NMIs shall be discussed here in detail.

Figure 8 presents measured calibration results for a single turbine meter calibrated by the three different national high-press standards in Europe, named here Inst 1 Inst 2 and Inst 3. The claimed uncertainties of all measuring points are indicated. The transfer standard was a dual turbine meter set.

Line 1 in Figure 8 denoted as f_{Ref} is the weighted least square fit of all measured data points at PTB-pigsar, NMi-VSL and LNE for a particular meter. f_{Ref} is the Harmonized European Reference value as realized by this comparison presented in Figure 8.

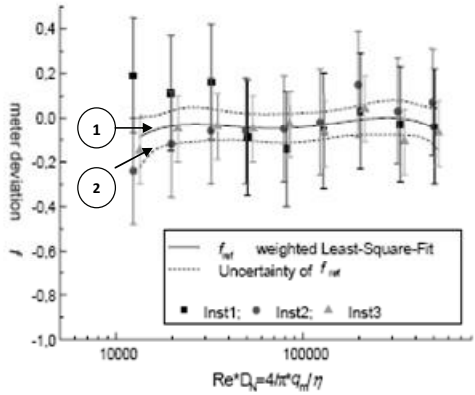


Figure 5. *Practical measured meter deviation f of one transfer meter within a package at one pressure. f_{Ref} is the weighted Least-Square-Fit of all results (measured at 20 bar, package with two turbine meters, meter 1). This is the value which we use as the Harmonized European Reference Level*

The f_{Ref} function has been calculated using the weighted average of all 3 participants PTB-pigsar, NMi-VSL and LNE, line 1. Line 2 is the uncertainty function of the Harmonization Reference Value (Harmonization European Gas Cubic Meter) f_{Ref} and is of course much smaller than the individual uncertainties of the participating institutes PTB, NMi-VSL and LNE.

We recognize a quite good overlap of all institutes with the reference value f_{Ref}

In order to quantify the degree of equivalence between the f_{Ref} function and the participants as well as the degree of equivalence among the participants, we have

also used here the En-Criterion as it is commonly in use in the accreditation area, see e.g [10] and [11] and [8].

From the measured comparison results in Figure 8, we can calculate the following quantities:

$$d_i = f_i - f_{f\,ref}; d_{i,j} = f_i - f_j$$

$$En_i = \frac{|d_i|}{U(d_i)}; U(di) = \sqrt{U_{Mut}^2 - U_{fref}^2}$$

with

$$U_{Mut}^2 = U_{CMC,i}^2 + U_{TM}^2$$

d_i means the bias between f_{Ref} and measured value and $U(d_i)$ is the corresponding uncertainty of this bias between f_{Ref} and the measured value, which will be calculated according to Cox, see [8], Chapter 5, section 4C, equations (3° AND (5).

$U(d_i)$ is the uncertainty associated with the difference d_i.

Uf_{Ref} is the uncertainty associated with the f_{Ref} function, line 2 in Figure 5.

U_{CMCi} is the claimed uncertainty of NMI i.

U_{Mut} is the uncertainty of the meter under test.

U_{TM} is the uncertainty of the transfer meter.

The quantity $En_i = \frac{|d_i|}{U(d_i)}$ will be the degree of equivalence which we suggest to use here to characterize the degree of equivalence of the national facilities.

In this way it will be possible to describe the degree of equivalence of a lab to the f_{Ref} function using a dimensionless number. En should be between 0….1 and may go up to 1.2.

En should be as close as possible to "0". E_n=0 means no deviation betweek the f_{Ref} function and the lab. E_n=1 means, that the error bars do just overlap.

Figure 6 presents the degree of equivalence E_n as determined from all measurement results between the European national standards as presented in Figure 5.

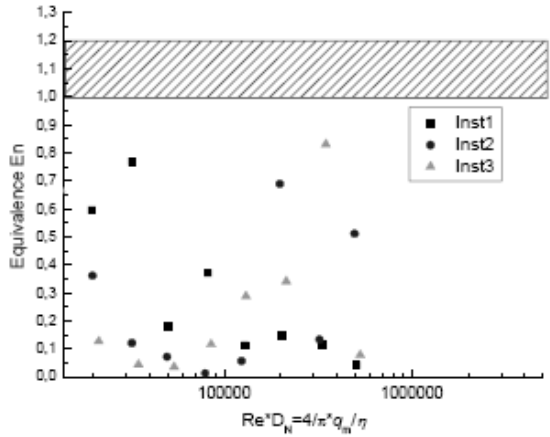


Figure 6. *Degrees of equivalence E_n determined for all measured results in Figure 5*

From Figure 6 we can conclude that in the overlapping range of PTB, NMi-VSL and LNE, all institutes are equivalent (En<1) with respect to the measured data in Figure 8 (single meter, single pressure stage).

In Figure 7 we present the degree of equivalence making use of all measurements at all pressures between the European national facilities to give an impression on the very acceptable agreement between these gas facilities. The outcome is, that the institutes are equivalent to each other in all Re numbers.

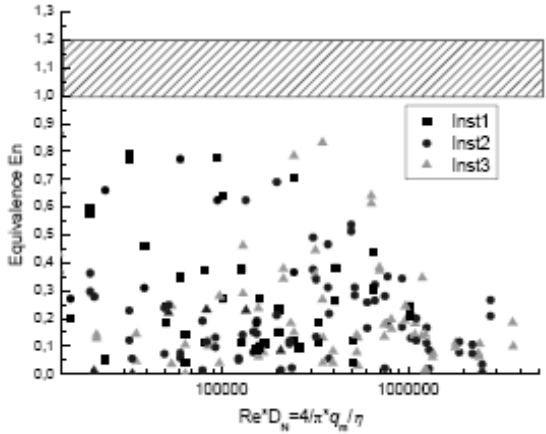


Figure 7. *Degrees of equivalence En determined for all results determined at all high press national facilities in Europe, namely PTB-pigsar, NMI-VSL and LNE (Measured at 20; 40 and 50 bar; packages (DN150, DN250, DN400) with two turbines each, additionally single sonic nozzle (TF200)). The degree of equivalence between all European National gas facilities*

As we find a lot of single values of En for each measured result, it is helpful to define an overall value as a characteristic criterion for each laboratory taking part in the harmonization procedure in order to have a single number per institute.

Starting from the fact that the degree of equivalence is a random variable with a log-normal probability density, it is the simplest approach to use the geometric mean as the characteristic value En_{total}:

$$En_{total} = \left(\Pi_{i=1}^{n} En_i\right)^{\frac{1}{n}} = \exp\left\{\frac{1}{n}\sum_{i=1}^{n}\ln\left(En_i\right)\right\}$$

Institute 1: $En_{total} = 0.24$

Institute 2: $En_{total} = 0.14$

Institute 3: En_{total} 0.13

We finally obtain the following visual presentation for the degree of equivalence between PTB-pigsar, NMi-VSL and BNM using the single En-numbers for each institute.

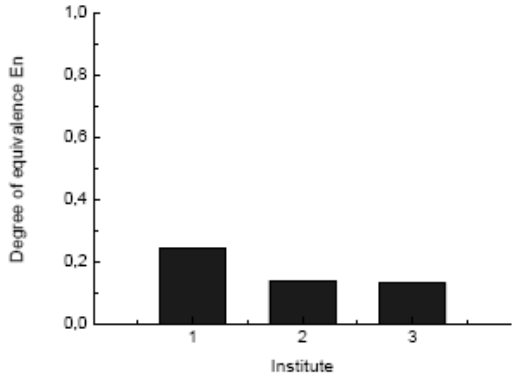


Figure 8. *Characteristic degree of equivalence En for all institutes based on the geometric mean using all results in Figure 7*

The CIPM KCRVs and the european harmonized reference value

The CIPM decided, in accordance with the BIPM Mutual Recognition Arrangement (MRA) [8], to conduct Key Comparisons (KCs) [9] among national primary standards of selected NMIs in the subject field high-pressure gases. This includes natural gas and compressed air and/or nitrogen. The members of the responsible CCM Working Group for Fluid Flow (WGFF) elected PTB and NMI-VSL as the pilot laboratories for this Key Comparison (KC) in order to obtain key comparison reference values KCKU for high pressure gas flow.

The KCs have been done in accordance to the Guidelines for CIMP Key Comparisons (8) and have been performed to fulfill the requirements of the CIPM MRS (7) and the requirements from the CIPM Committee Consultative for Mass and Related Quantities (9).

As obviously the participants of the CIPM Key Comparisons are almost the same as the participants of the procedure for the Harmonized European Natural Gas Cubic Meter, the KCRV of the CIPM-KCs is the same as the harmonized European reference level.

It shall be mentioned here that during the harmonization procedure much more comparison work has been done at all pressures and flow rates to obtain the best available reference values.

In order to obtain the KCRV or the Harmonized Reference Value, the weighted average/mean of the calibration results at all the facilities have been used as described by Cox, see [10]. This is the procedure recommended by an advisory group on statistics as explained in [11]. The pilot lab has decided to follow the

BIPM recommendations and the weighted mean has been taken for the KCRV function. The weighing factors are the claimed and mutually recognized uncertainties (u^{-2}) of the facilities.

Very stable transfer packages using turbine meters have been selected for these KCRVs.

In order to demonstrate the excellent results of the CIPM KCs, some of the comparisons might be shown here in Figures 9 and 10 (natural gas) and Figures 11 and 12 (compressed air).

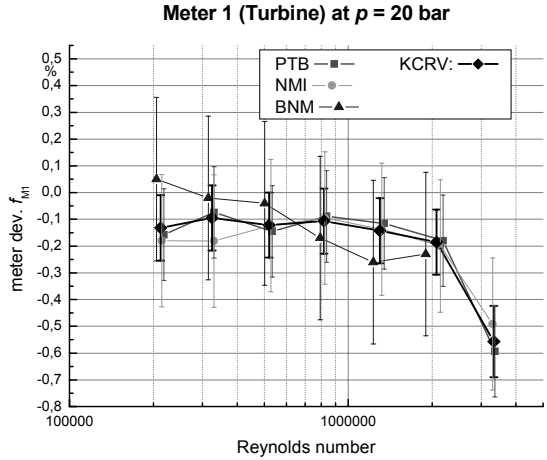


Figure 9.

From this figures the following conclusions can be drawn:

PTB, NMI-VSL and LNE are equivalent to each other within their claimed uncertainties, as all uncertainty bars are overlapping very nicely.

The degree of equivalence of all three NMIs with the KCRV couldn't be much better. The KCRV is considered to be the best available realization of the Natural gas cubic meter (flow) at high pressure.

As mentioned before, this KCRV is identical to the Harmonized European Reference Level or Gas Cubic Meter. This is, because the European NMIs and their calibration facilities do disseminate the KCRV itself since May 4th 2004.

Figure 10 presents similar Key Comparison results using an ultrasonic meter at 47 bar.

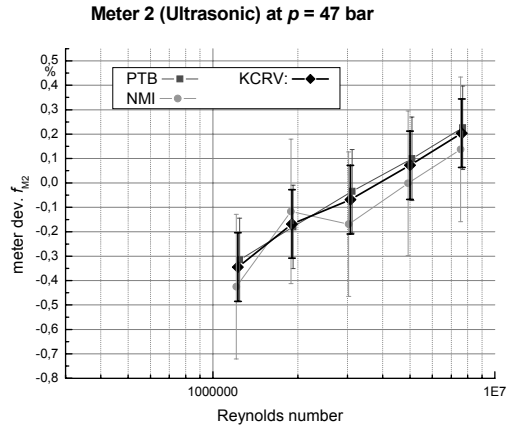


Figure 10. *Calibration of an ultrasonic meter (Instromet G1000) turbine transfer standard at PTB and NMi at 47 bar during the CIPM KCs. The high degree of equivalence is obvious, as all error bars are overlapping very well with all partners and with the KCRV. LNE could not participate here, compare Figure 2*

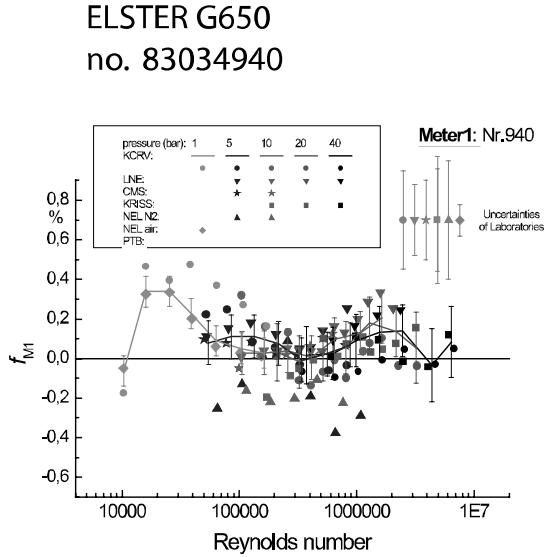


Figure 11. *Calibration results of meter 1 of the transfer package at all participants and all pressures between 1 bar and 40 bar. The claimed uncertainty bars for all labs are indicated in gray. The resulting KCRV at the different pressure stages and its uncertainty is presented throughout the entire flow range*

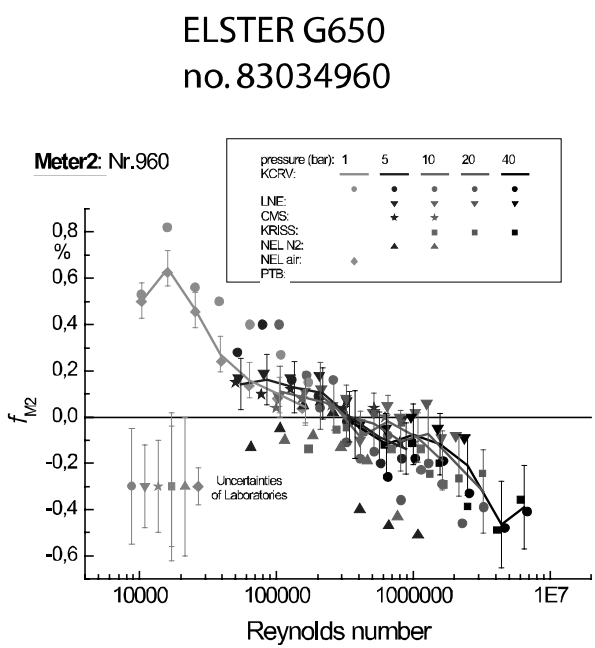


Figure 12. *Calibration results of meter 2 of the transfer package at all participants and all pressures between 1 bar and 40 bar. The claimed uncertainty bars for all labs are indicated in gray. The resulting KCRV at the different pressure stages and its uncertainty is presented throughout the entire flow range*

Final conclusion

This detailed analysis of the harmonization procedure between PTB-pigsar, NMi-VSL and LNE shall demonstrate that a very careful uncertainty analysis has been performed in order to make sure that all the latest BIPM recommendations have been followed truly. The nice outcome for the consumers is that the use always obtains the same calibration in Germany, in the Netherlands and in France at any test facility and he can enjoy the benefit of a very stable and small uncertainty of the harmonized reference value.

All three institutes maintain their individual independent facility and apply permanent improvements in order to provide for a more stable and reliable gas cubic meter.

This Harmonized European Natural Gas Cubic Meter will be disseminated towards all countries and all facilities in Europe (West and East). In the meantime it has already been accepted by the Canadian NMI, the NRC in Canada. In the

meantime nearly all European calibration facilities and authorties have accepted this reference value, which is highly acknowledged by all gas companies.

Finally it shall be pointed out here, that the full report on the CIPM Key Comparisons is available on the Internet at the BIPM website and has been published online at the international journal Metrologia.

References

[1] M.J. van der Beek, I.J. Landheer, B. Mickan, R. Kramer and D. Dopheide: "Unit of volume for natural gases at operational conditions: PTB and NMi-VSL disseminate "Harmonized Reference Values", *FLOMEKO 2003 Groningen, CD-ROM Conference proceedings.*

[2] D. Dopheide, B. Mickan, R. Kramer, M. van der Beek, G. Blom: "The harmonized high-pressure natural gas cubic meter in Europe and its benefit for user and metrology". *Conference Proc. FLOMEKO 2004 Guilin*, China 14-17 September 2004

[3] M.P. van der Beek, I.J. Landheer and H.H. Dijstelbergen: "Developments in the Realization of Traceability for high-pressure Gas-Flow measurements", *International Gas Research Conference (IGRC) 2001*, Amsterdam, the Netherlands.

[4] B. Mickan, R. Kramer, H.- J. Hotze, D. Dopheide: "***Pigsar-*** the extended test facility and new German National Primary Standard for high pressure natural gas", *FLOMEKO 2003 Groningen*, CD-ROM conference proceedings.

[5] I. Krajcin, M. Uhrig, P. Schley, M. Jaeschke, D. Vieth, K. Altfeld: "High-precision measurement and calibration technology as a basis for correct gas billing", World Gas Conference 2006, CD-ROM 2006.

[6] J-P Vallet, C. Windenberger, O. Gorieu: "Traceability chain and uncertainty budget calculation for the french high pressure gas flow measurement laboratory LNE-LADG and the impact of the harmonization process with PTB – pigsar and MNi-VSL on this calculation". *6th ISFFM* – Queretaro Mexico, May 2006.

[7] CIPM MRA 1999: "Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes", *Comité International des Poids et Measures*, October 1999, Paris, France.

[8] BIPM guideline for conducting key comparisons, 1999 (appendix F to MRA [2]).

[9] CCM formalities for KCs, CCM2002-11: Formalities required for the CCM Key Comparisons (2nd Revised Draft), *CIPM Consultative Committee for Mass and Related Quantities*, CCM/2002-11, June, 2001, Submitted by M. Tanaka.

[10] M.G. Cox: "The evaluation of key comparison data", in: Metrologia, 2002, 39, pp. 589-595.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Environment

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Traceability of environmental chemical analyses: can fundamental metrological principles meet routine practice? Example of chemical monitoring under the Water Framework Directive

Ph. Quevauviller

European Commission
Rue de la Loi 200, B-1049 Brussels, Belgium

ABSTRACT: The direct application of the metrological principles including traceability are hardly adapted to a wide range of environmental measurements owing to major differences among the physical and chemical measurement processes which are generally dependent upon the sample composition. Another difficulty lies in the lack of awareness of the analytical community about fundamental metrological principles, in particular when routine analyses are concerned. This presentation looks at the traceability application to a specific case, the chemical monitoring program under the Water Framework Directive (2000/60/EC) and discusses fundamental metrological limitations.

Introduction

The Water Framework Directive (WFD) is certainly the first EU legislative instrument which requires a systematic monitoring of biological, chemical and quantitative parameters in European waters at such a wide geographical scale (covering more than 120 river basins in 27 EU countries) [1]. The principles are fixed in the legislative text and exchanges of information among experts have enabled us to set out a common understanding of monitoring requirements in the form of guidance documents [2-4]. While water monitoring is obviously not a new feature, it should be noted, however, that the WFD monitoring programs are in their infancy in that they had to be designed and reported by the Member States in March 2007.

Monitoring data produced in 2007-2008 under the WFD will form the basis for the design of programs of measures to be included in the first River Basin Management Plan (due to be published in 2009), and thereafter used for evaluating the efficiency of these measures. Monitoring data will hence obviously be used as basis for classifying the water status, and they will also be used to identify possible pollution trends. This is an iterative process in that better monitoring will ensure a better design and follow-up of measures, a better status classification and a timely

identification of trends (calling for reversal measures) which puts a clear accent on the need for constant improvements and regular reviews (foreseen under the WFD) and hence on the need to integrate scientific progress in an efficient way.

Monitoring goes hand in hand with metrology. The "science of measurements" indeed puts a clear accent on issues of data quality and comparability (including traceability aspects) which are covered, albeit in a general way, in technical specifications of the WFD. This paper focuses on chemical monitoring WFD requirements (noting that the directive also requires ecological status monitoring for surface waters and quantitative status monitoring for ground waters) and highlights metrological features that are presently under consideration in European expert groups discussing specific implementation issues.

Main legal requirements

The Water Framework Directive establishes "good status" objectives to be achieved for all waters by the end of 2015. With regard to surface waters, good status criteria are based on biological parameters (ecological status) and chemistry (chemical status linked to compliance to EU Environmental Quality Standards), while for ground waters, good status refers to quantitative levels (balance between recharge and abstraction) and chemistry (linked to compliance to groundwater quality standards established at EU, national, regional or local levels). Monitoring requirements linked to these status objectives are found in Appendix V of the directive.

EU Member States had to design monitoring programs before the end of 2006 and report them to the European Commission in March 2007. Basic requirements are that monitoring data have to provide a reliable assessment of the status of all water bodies or groups of bodies (administrative units defined by Member States). This implies that networks have to consider the representativeness of monitoring points as well as frequency. In addition, monitoring has to be designed in such as way that long-term pollution trends may be detected.

Non-legally binding recommendations

The technical challenges of the WFD have led Member States to request the European Commission to launch a "Common Implementation Strategy" (CIS), which aims to exchange knowledge, best practices, develop guidance documents, etc. in a coordinated fashion. The CIS process is recognized to be a powerful consultative process and an excellent example of governance at EU level [5]. It comprises expert groups discussing various topics, including a Chemical Monitoring Activity (CMA), which is the object of this section.

The Chemical Monitoring Activity is described elsewhere [6]. Along the CIS principles, the expert group (composed of ca. 60 experts from Member States environment agencies or ministries, stakeholder's associations and the scientific community) has discussed various chemical monitoring features concerning surface and ground waters, which resulted in recommendations summarized in two guidance documents [3,4]. Issues such as sampling design and representativeness, sampling and analytical methods, etc. are discussed in fair levels of details in order to provide a common base for Member States when dealing with chemical monitoring programs. These recommendations are directly relevant to metrology and provide background rules that should help harmonized approaches to be developed in Europe. It is important to note is that these guidance documents are not legally-binding, i.e. they provide general rules that may be adapted to local or regional situations without requiring a necessary enforcement. In other words, the enforceable obligations are those found in Article 8 and Appendix V of the WFD, and these are explained with technical details in guidance documents that are themselves not legally binding. Despite this, the guidance documents will have an obvious impact on the way Member States will perform chemical monitoring programs in the future.

Coming back to metrology, the most important aspect was considered to be the demonstration of data comparability, which should be ensured at EU level. In this case only, technical specifications were deemed to be necessary in an enforceable way. This is discussed in the following section.

Further binding rules

As discussed above, the overall water management and decision-making system of the WFD is closely depending upon measurement data. Without clear rules for demonstrating the accuracy (hence the comparability) of data, the system cannot deliver a sound basis for proper decisions. This argument led to open discussions about legally-binding rules concerning minimum performance criteria for analytical monitoring methods. The legal background is paragraph 3 of Article 8 of the WFD, which opens the possibility to develop technical specifications supporting monitoring programs. In this context, a draft Commission Directive "adopting technical specifications for chemical monitoring in accordance with Directive 2000/60/EC of the European Parliament and the Council" has been developed by the CMA expert group. The proposal includes mandatory requirements for Member States' laboratories regarding the validation of methods, minimum performance criteria (linked to target uncertainty and limit of quantification linked to environmental quality standard values), participation in quality assurance programs (including proficiency testing schemes, analyses of reference materials, training) and accreditation to be ensured by the end of 2012. This 10-pages text is still under discussion at the time of publication of the present paper, and it is hoped that it will be adopted by Member States under comitology (i.e. by a regulatory committee

composed of Member State's representatives, including a consultation of the European Parliament) by the end of 2008. With this background, all elements are met to develop a metrological framework that will ensure that chemical monitoring data produced under the WFD are comparable, i.e. traceable to commonly accepted references. The achievement of traceability and related references in the context of the WFD monitoring programs are discussed below.

WFD monitoring and its links with metrology

Introduction

The above sections have highlighted the importance of data comparability and traceability in the context of WFD chemical monitoring. Let us now examine in details what are the different references that need to be considered for the development of a sound metrological system.

Firstly, as a reminder, traceability is defined as "the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties" [6]. Discussions on how these elements apply to chemical measurements have been, and are still, under discussion [7-9]. In this context the basic references are those of the SI (Système International) units, i.e. the kg or mole for chemical measurements. Establishing SI traceability of chemical measurements may, in principle, be achieved in relation to either a reference material or to a reference method [10]. The unbroken chain of comparison implies that no loss of information should occur during the analytical procedure (e.g. incomplete recovery or contamination). Finally, traceability implies, in theory, that the uncertainty of all stated references that contribute to the measurement is duly considered (meaning that the smaller the chain of comparison the better the uncertainty of the final result), which is obviously hardly applicable to complex environmental measurements and to sampling and sample pre-treatment steps [11].

As an additional remark, it should be noted that traceability should not be confused with accuracy which covers the terms trueness (the closeness of agreement of the measured value with the 'true value') and precision (the closeness of agreement between results obtained by applying the same experimental procedure several times under prescribed conditions). In other words, a method which is traceable to a stated reference is not necessarily accurate (i.e. the stated reference does not necessarily correspond to the 'true value'), whereas an accurate method is always traceable to what is considered to be the best approximation of the true value (defined as 'a value which would be obtained by measurement if the quantity could be completely defined if all measurement imperfections could be eliminated').

Chemical measurements in the context of the WFD (including water, sediment and biota) are based on a succession of actions, namely:

- sampling, storage and preservation of representative samples;
- pre-treatment of a sample portion for quantitation;
- calibration;
- final determination; and,
- calculation and presentation of results.

Starting from this, let us examine how traceability may be understood in the context of the WFD chemical monitoring program.

Basic references: the SI units

The unit that underpins chemical measurements is the unit of amount of substance, the mole. However, in practice, as there is no 'mole' standard, the kg is used, i.e. chemical measurements are actually traceable to the mass unit, the kg. In other words, water-related chemical measurements are based on the determination of amount of substance per mass of matrix. For solid matrices (sediment, suspended matters, biota), these are units corresponding to ultratrace (ng/kg) and trace (µg/kg) concentrations for many organic micro-pollutants and trace elements, and mg/kg for major elements. For water, results should also in principle be reported in mass/kg of water but the practice is usually that they are reported as mass/volume, e.g. ng/l, µg/l or mg/l, which is already diverging from basic metrological principles.

Standardized methods in the context of the WFD

Routine monitoring measurements often rely on standard methods adopted by official national or international standardization organizations (e.g. ISO, CEN). Written standards aim to establish minimum quality requirements and to improve the comparability of analytical results. Standardized methods are usually developed to be used on a voluntary basis, but they may also be linked to regulations. This is the case of series of standards that appear in paragraph 1.3.6 of Appendix V of the WFD, which reads "methods for the monitoring of the parameters shall conform to the international standards listed below [NB: six standards are presently listed by the Appendix is opened to any relevant standards] or such other national or international standards which will ensure the provision of data of an equivalent scientific quality and comparability".

As underlined above, the on-going discussions on minimum performance criteria for analytical methods used in chemical monitoring (proposal for a Commission Directive) are embedded into paragraph 3 of Article 8 of the WFD, stipulating that "technical specifications and Standardized methods of analysis and monitoring of

water status shall be laid down in accordance with the procedure laid down in Article 21" (NB: this latter article concerns "comitology", i.e. the possible adoption of legally-binding regulations by Member States with consultation of the European Parliament). In this draft decision, the use of alternative methods providing data of equivalent or better scientific quality and comparability than Standardized methods is highlighted, providing that these methods are properly validated. This actually enables all kind of methods to be used for chemical monitoring of "total" amounts of chemical substances. For operationally-defined parameters, however, the use of Standardized methods is made mandatory.

For operationally-defined parameters (e.g. extractable forms of elements using a specific extraction method), results are obviously linked to the methods used, which thus requires that these methods are Standardized if comparability of data has to be ensured. This justifies standard methods for these parameters to be made mandatory. In this context, they will represent a key element in the traceability chain considering that analytical results are directly linked to the analytical protocol. In other words, the traceability chain is broken if the protocol is not strictly followed [12].

In the context of the WFD chemical monitoring program, recommendations to use standard methods will thus certainly focus on analytical steps that are based on technical operations that may differ from one country to another. Examples are sampling, sample pre-treatment (e.g. filtration for water, sieving for sediment), measurements of 'extractable' forms of pollutants, etc. Efforts are, therefore, on-going to identify existing standard methods that could readily be recommended in the light of WFD chemical monitoring, as well as standards for which research would be needed prior to adoption by the European Standardisation Organization (CEN). The principle will be to develop documented guidelines about the methods, their limits of applicability, and other relevant aspects (e.g. sampling representativeness, frequency and techniques).

At the present stage, a range of standardized methods is available as references for specific parameters or monitoring steps. However, there are still many cases for which there are no such standards, hence the reference has to be found elsewhere else. This is discussed in the next section.

Role of reference methods

Complex environmental monitoring measurements are generally based on successive analytical steps, such as extraction, derivatization, separation and detection. This succession multiplies the risk that the traceability chain is broken because of a lack of appropriate references (e.g. pure calibrants, reference materials). Reference methods usually refer to methods with high metrological qualities (they may actually be standardized methods), examples of which are primary methods that are directly traceable to SI units without the need to use an external calibration.

If we consider that reference methods correspond to methods exempt from systematic errors (this is the case of primary methods) and only affected by few random errors, we have to admit that these mainly exist for inorganic determinations at the present stage and solely for analyzing samples in the laboratory (i.e. there are no "reference methods" possible for sampling). In the case of determinations of organic or organometallic compounds, the need to include pre-treatment steps, e.g. extraction or derivatization, will lead to breaking of the traceability chain and stated references will rely on approximations (recovery estimates). The better these approximations, the closer the traceability of the measurement of the substance in the sample to the true value. In the case of substances requiring an extraction step, primary methods hardly exist since there are no means at present to give proof that extraction or chemical reactions (e.g. derivatization) have yielded a 100% recovery.

Examples of primary methods used in determining multi-isotope trace elements are isotope dilution-based techniques (e.g. isotope dilution mass spectrometry), enabling traceability of the results to SI units. For organometallic compounds, the use of these techniques will thus guarantee traceability to SI units for the compounds in the extracts.

For methods that are based on internal or external calibrations, the link will rely on the availability of calibrants of high purity and verified stroichiometry, which represents the last part of the traceability chain (i.e. calibration of the detector signal). A reference method should provide guarantees that all steps prior to determination are recorded and documented in such a way that the result of the final determination is linked to an unbroken chain of comparisons to appropriate standards. In other words, establishing traceability for reference methods implies that several "primary" chemical reference materials in the form of ultrapure substances are interlinked by well-known, quantitative, high-precision, high-accuracy chemical reactions [13]. This is actually not achieved for a vast majority of chemical monitoring measurements as there are still many "weak links" in the traceability chain (e.g. extraction recovery estimates, derivatization yield verifications using "secondary standards").

As a conclusion, the denomination of a method as a "reference method" in the context of wide-scale monitoring program as the WFD one will need to be considered very carefully, clearly highlighting weaknesses in the traceability chain. As stressed above, methods with analytical steps requiring recovery estimates need to be validated using independent methods (based on different principles) in order to provide data of good scientific comparability. It should be stressed again, however, that comparability is not synonymous with accuracy. Hence, few of the methods may be considered as reference methods unless they are documented with a full description of all the analytical operations and the limits of applicability of the methods. This is the case, for example, for "official methods" required in the food sector.

Regarding WFD chemical monitoring, there are no technical prescriptions about analytical methods at this stage. The principles of validation are now being discussed for agreement at EU level, but discussions do not go as far as describing methods with specific technical requirements. We have seen that this may be envisaged in the context of standardisation, but we also have to realise that WFD monitoring relies on the Member State's practices and it is not aimed at establishing an unnecessarily too rigid system, even if it is recognized that an operational coordination and common understanding are desirable.

Role of reference materials

The role and use of reference materials are in principle well known, in particular Certified Reference Materials (CRMs) used as calibration materials or matrix materials representing – as far as possible – "real matrices" used for the verification of the measurement process, or (not certified) laboratory reference materials (LRMs also known as quality control (QC) materials) used, for example, for interlaboratory studies or internal quality control (control charts). Examples of reference materials relevant to WFD monitoring (water, sediment and biota) are described in the literature [14-15].

Control charts used for monitoring the reproducibility of methods (repeated analyzes of one or several reference materials) may be considered as long-term references for analytical measurements since they allow us to monitor analytical variations according to an anchorage point, i.e. the reference material(s). This concerns reproducibility checking but not necessarily trueness which evaluation relies on relevant CRM analysis.

Regarding CRMs, certifying organizations attempt, wherever possible, to estimate true values of parameters in representative matrix materials, which is achieved mainly in employing a variety of methods with different measurement principles in the material certification study. A good agreement among the various methods enables us to assume (but not firmly demonstrate) that no systematic error was left undetected, and that the certified values are the closest estimate to the true value. Wherever possible, this approach should include primary methods (see above). In many instances, the use of various (independent) methods enables us to obtain consensus values that are accepted as true values reflecting the analytical state-of-the-art and hence ensuring data comparability [11].

In the case of water-related measurements, the wide variety of matrices and substances encountered, in principle implies that various types of matrix materials (e.g. waters, sediments and biota) should be available. However, the best match between a CRM and a natural sample can never be achieved and compromises are in most instances necessary. As discussed previously [11,12] reference materials represent "physical" references to which measurements can be linked, but which call for a cautious evaluation resulting from matrix differences. It was stressed that a

correct result obtained with a matrix CRM does not give a full assurance that "correct results" will be achieved when analyzing unknown samples because of differences in matrix composition [9].

The traceability of water, sediment and biota measurements on the basis of matrix CRMs (representing complex chemical systems) to SI units is largely debated. This results from the lack of certainty that the certified values correspond to true values, in particular with regard to complex matrix reference materials. This does not diminish the value of CRMs since the "consensus values" play the role of reference in achieving traceability for a given water-related measurement (linking results to a wide analytical community and thus ensuring data comparability) but do not necessarily allow trueness to be checked.

In addition, there are still gaps in CRM availability or representativeness, or situations in which they cannot be prepared owing to instability. In these cases, other approaches have to be followed to achieve traceability, e.g. interlaboratory studies (see paragraph 8.6). Where there is good correspondence between the matrix of samples and the matrix of CRMs, this reference is certainly the most appropriate one for checking the accuracy of analytical methods and comparing the performance of one method with other methods (and other laboratories). Similar comments with respect to representativeness also apply to not certified reference materials used for internal QC purposes.

In the light of the WFD monitoring requirements, we may highlight that RMs are key QC tools. They are not specifically referred to in the WFD core text, but are clearly part of the related Commission Decision on minimum performance criteria (see the above section on "further binding rules").

Proficiency testing schemes

Proficiency testing schemes enable laboratories to establish "external" references for evaluating the performance of their methods. In this framework, one of several reference materials are distributed by a central organization to participating laboratories for the determination of given substances. Comparing laboratory results of different methods allows us to detect possible sources of errors linked to a specific procedure or the ways a method is applied in a given laboratory. When the testing focuses on a single method, this enables performance criteria (e.g. limits of detection or quantification, precision etc.) to be checked. The references establishing the traceability link are again based on reference materials which have to fulfil the usual requirements. However, by contrast with reference materials used for internal QC, proficiency testing schemes may involve samples with a limited shelf life, which may be distributed to laboratories for analysis of particular parameters that could not be evaluated using stabilized RMs. Examples of water samples containing unstable compounds with a short-term preservation capacity.

Proficiency testing schemes may also be based on reference sites e.g. to evaluate sampling procedures or a bulk common sample to be analyzed by participating laboratories at a given time (e.g. a tank full of water for analyzing microbiological parameters).

As discussed above for RMs, the measurement values obtained in the framework of interlaboratory studies (using different techniques) may be considered as an anchorage point, representing the analytical state of the art: this offers laboratories a means to achieve comparability (i.e. traceability) of their results to a recognized reference, which in this case is a consensus value (generally the mean of laboratory means). Let us stress again that this reference does not enable traceability to the true value of the substance in the medium to be achieved, but it represents a very useful method for achieving comparability of environmental measurements using an external QC scheme [12].

Similarly to RMs, proficiency testing schemes are not specifically referred to in the WFD core text, but are clearly part of the related Commission Decision on minimum performance criteria (see above paragraph on "further binding rules").

Laboratory accreditation

Accreditation of laboratories following rules established by the ISO 17025 contributes to the overall framework contributing to the achievement of traceability. It is, however, an organizational framework which as such does not guarantee traceability can be achieved. It instead requests laboratories to proceed with internal and external QC of analytical methods as described in the preceding sections.

Conclusions, perspectives

This paper highlights needs various features related to metrology in the context of a large-scale chemical monitoring program such as the one required by the WFD. It thus mixes policy considerations with technical issues which highlight the complexity of the overall approach that will need to be developed to demonstrate the traceability of monitored chemical data at EU level. In comparison to physical metrology, the challenges are tremendous as they involve many different chemical specificities (e.g. operational steps such as sampling, sample pre-treatment, different matrices, thousands of substances, need for reference materials for validating methods, etc.) which physical measurements do not have to tackle. A "simple" translation of physical metrological principles to chemical metrology is, therefore, not possible. This has been extensively discussed by chemical experts within the past 10 years [7-9], and theoretical discussions have not yet led to practical solutions that might be implemented at on the scale of an EU-wide monitoring program. These on-going discussions have the merit that the principles are now well settled and that we more or less know which direction should be taken to improve the metrological chemical framework. Making it operational will, however, require a

huge coordination of efforts and it would not be realistic to think that what has taken one century for physical metrology will be achieved in a few years for chemical metrology. The WFD presents the advantage of offering a very wide testing framework, and the metrological community should take this opportunity to examine how theory may match the practice. This is being studied in the context of an EU funded project (EAQC-WISE stemming for "European Analytical Quality Control in support of Water Information System for Europe). The actual improvements will come out progressively with discussions on the soundness of chemical data produced by Member States in the context of the preparation of river basin management plans. There are, therefore, milestones that could and should be used to not only improve the metrological system but also to enforce technical rules upon which the actual implementation of an EU-wide system will have all the chances to be successful. This perspective is to be conceived over the next decade, taking the opportunity to use reviewing milestones (2015 and every six years thereafter) of the WFD to constantly improve the monitoring metrological basis, not solely on chemical monitoring but also ecological status monitoring.

As a final consideration, we should keep in mind that achieving traceability of water, sediment and biota chemical monitoring measurements in the context of the WFD will have direct implications on the way programs of measures will be designed and made operational to achieve 'good status' objectives by 2015. This puts a strong emphasis on metrology as possible erroneous data might have tremendous (social and economic) consequences.

References

[1] Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, Official Journal of the European Communities L 327, 22.12.2000, p.1.

[2] Guidance Document N°7 on Monitoring under the Water Framework Directive, European Commission, Brussels, 2003.

[3] Guidance Document N°15 on Groundwater Monitoring, European Commission, Brussels, 2006.

[4] Guidance Document on Surface Water Monitoring, in press.

[5] Common Implementation Strategy for the Water Framework Directive, European Communities, ISBN 92-894-2040-5, 2003.

[6] ISO, International Vocabulary of Basic and General Terms in Metrology, 2nd Edition, International Standardisation Organization, Geneva, Switzerland, 1993.

[7] M. Válcarcel and A. Ríos, Anal. Chem., vol. 65, 78A, 1999.

[8] M.C. Walsh, Trends Anal. Chem., vol. 18, 616, 1999.

[9] Ph. Quevauviller, J. Environ. Monitor., vol. 2, 292, 2000.

[10] Ph. Quevauviller and O.F.X. Donard, Trends Anal. Chem., 20, 600, 2001.

[11] Ph. Quevauviller, in: Z. Mester, B. Sturgeon (Eds.), Sample Preparation for Trace Analysis, Elsevier, Amsterdam, 2004.

[12] Ph. Quevauviller, Trends Anal. Chem., vol. 23(3), 171-177, 2004.

[13] P. de Bièvre, in: H. Günzler (Ed), Accreditation and Quality Assurance in Analytical Chemistry, Springer, Berlin, 1996.

[14] P.h. Quevauviller, Trends Anal. Chem., vol. 13(9), 404-409, 1994.

[15] Ph. Quevauviller and E.A. Maier, Interlaboratory Studies and Certified Reference Materials for Environmental Analysis, Elsevier, Amsterdam, ISBN: 0-444-82389-1, pp. 558, 1999.

[16] Ph. Quevauviller, Quality Assurance for Water Analysis, John Wiley & Sons, Chichester, ISBN: 0-471-89962-3, pp. 252, 2002.

Disclaimer

The views expressed in this paper are purely those of the author and may not in any circumstances be regarded as stating an official position of the European Commission.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Main standard for refrigerant liquid leak throughputs

I. Morgado[(1)], J.C. Legras[(1)] and D. Clodic[(2)]

[(1)] Laboratoire National de Métrologie et d'Essais
1 rue Gaston Boissier – F 75724 Paris Cedex 15
[(2)] Centre d'Energétique et des Procédés, France
Ecole des Mines de Paris, 60, bld Saint-Michel – F 75272 Paris Cedex 06, France

ABSTRACT: Leak detection is nowadays widely used in various fields such as the automotive and refrigeration industries. The leak tightness of installations loaded with refrigerants must be checked periodically by qualified refrigerant detectors using refrigerant leaks. However the latter are not yet traceable. Therefore, a project involving the Laboratoire national de métrologie et d'essais (LNE), the Centre d'énergétique et procédés (CEP) and the Agence de l'environnement et de la maîtrise de l'énergie (ADEME) was initiated to develop a national standard for calibrating refrigerant leaks. The following document presents the method, the achieved installation and the first results.

Context

At the international level

The Protocol of Montreal relating to substances depleting the ozone layer has led to the ban of CFC use (in 1995 in Europe) and to the gradual ending of HCFC fluid use, resulting in the use of HFC hydrofluorocarbon fluids as well as other molecules such as CO_2 which do not deplete the ozone layer.

However, HFC gases with a high potential for global warming are part of the Kyoto protocol negotiated by signatory country members in 1998. This protocol relating to the greenhouse gas effect and signed by more than 100 countries demands the reduction of gas emission.

At the European Union level

As for the 2002/358/CE decision of the April 25 2002 convention concerning the approval, in the name of the European Community, of the Kyoto Protocol, the community and its member states are required to reduce their greenhouse gas emissions by 8%, and in particular a large part of fluoride gas such as R-134a between 2008 and 2012.

European Regulation no. 842/2006 with the objective of reducing the greenhouse gas effect intends to achieve a policy of control in terms of confinement, use, recycling, labeling and elimination of products and equipment containing these gases. This means a leak tightness quality control of any equipment containing more than 3 kg of fluid by certified personnel, and the control frequency depends on the fluid load.

Leak tightness control method

In France, the decree from January 12 2000 specifies the leak tightness quality control method for refrigerated and climatic equipment. This control must be carried out by a 5 g/yr hand leak detector or an atmospheric controller with a sensitivity of 10 μmol/mol.

The qualification of leak detectors is defined by the French E 35-422 standard written in 1999, which to this day is used as a basis for the European standard EN 14624 published in July 2005. This standard supports the use of refrigerant fluid calibrated leaks – R-134a – with an mass flow rate of 1 to 50 g/yr [3]. The standard specifies that the leaks to be used must be capillary leaks "which must be calibrated in secondary standards". In addition, it specifies that these secondary standards must be calibrated by a primary standard. This primary standard dedicated to direct calibration of refrigerant leaks does not exist to this day.

In order for a calibration chain of these calibrated leaks to occur, the LNE and CEP developed a national reference from an infrared detection method created by the CEP in the context of ADEME financing.

Objectives

The project is part of an LNE research and development program. It has been submitted to the ANRT and is today part of a thesis from both ENSMP and LNE.

As stipulated in the regulations, calibration of refrigerant leak outflows is intended to ensure traceability of leaks used for qualification of hand leak detectors ensuring periodic mandatory leak tightness quality over refrigerating and climatic installations. The project is intended to implement a primary reference based on a technique of detection by accumulation.

The measuring principle was developed by the CEP. The method consists of measuring the concentration variation of the gas transmitted by a leak over time accumulated in a closed volume. During the trial, a continuous circulation between the accumulation enclosure and the cell of a gas analyzer made up of a photo-acoustic infrared spectrometer enables the concentration elevation measurement in R-134a or in CO_2.

The project also plans for the implementation of a test bench for leak detector and atmosphere controller qualification in cooperation with CEP. From these methods, a rigorous analysis of the operation of refrigerant leak detectors, and phenomena occurring during liquid transfer from the leak to the detector's sensing element, will make it possible to establish the precautions to be taken during leak tightness tests.

The different components of the reference were selected from performance tests. From these preliminary studies, the reference was made possible at CEP and delivered to LNE in February 2007. The reference qualification is currently being developed. In particular, calibration methods of the photo-acoustic spectrometer and accumulation volume must be finalized. On these bases, measurement uncertainty components of calibrated lead throughputs will be defined and validated by the comparison of these methods from the LNE experiment for helium reference leaks. The complete documentation of the metrologic file will be created in order to receive a COFRAC accreditation toward the end of 2007.

Measure methods of refrigerant leak throughputs

Principle

The method consists of measuring the variation over time t of concentration C of gas emitted by the leak to calibrate in a closed accumulation volume V previously filled with dry clean air – made up of 20% oxygen and 80% nitrogen, with approximately 99.99% purity. Volume is at pressure P close to atmospheric pressure and at a temperature close to 20°C. The mass flow rate is then deducted from the law of perfect gases:

$$Q_m = \frac{M.V}{R} \cdot \frac{\partial\left(\frac{C.P}{T}\right)}{\partial t} \qquad (1)$$

with R the constant of perfect gases, M the volume mass of the gas and T the thermodynamic temperature of the gas [1].

Each minute, a sample containing the accumulated gas – R-134a or CO_2 – in the volume is collected by the spectrometer and analyzed. Once the concentration measure is achieved, the analyzer rejects the sample in the accumulation volume. The concentration value obtained is then corrected by simultaneous temperature and pressure measures. Time variation of gas concentration studied in the closed volume is established by the least squares method. A good knowledge of the accumulation volume then makes it possible to calculate mass flow rate.

Advantage of the method

The first advantage with this method is the use of refrigerant gas as a tracer gas. As it turns out, throughput standards for helium leaks are not adapted to refrigerant leaks. Throughput measurement methods for helium leaks, which still exist today, are rightly considered as the measurement methods adapted to the lowest throughputs between 10^{-14} and 10^{-9} mol/s. For this technical field, methods were implemented in several countries, for example a pressure variation method to LNE [13] or a flow variation method to NIST [12]. Throughput leaks between 10^{-14} and 10^{-9} mol/s correspond to typical molecular flows of vacuum systems with molecular mass as physical influencing parameter, whereas refrigerant leak throughputs detectable by manual detectors or atmosphere controllers are leak throughputs with a flow that is characterized by a viscous state. In addition, refrigerant leak throughputs are approximately between 10^{-9} and 10^{-8} mol/s, corresponding to mass flow rates of approximately 1 to 10 g/yr. Typically, the upstream and downstream pressure gap of the leak is approximately 100 kPa, with a pressure close to the atmospheric pressure downstream. The flow state is then laminar. The significant influencing physical parameter is viscosity. Because of liquid thermophysical properties, HFC type refrigerant liquids, such as R-134a, are excellent tracer gases. In fact, detection sensitivity is all the greater as gas viscosity is low. For example, viscosity of R-134a is 30% lower than that of helium ([2], [4]).

Using refrigerant liquids as tracer gas is also interesting because of the absorption properties of these liquids in infrared: infrared spectrometry constitutes a path offering many advantages. Contrary to mass spectrometry which measures throughput, infrared spectrometry measures concentration. Consequently, infrared spectrometry measures gas accumulation in a closed circuit. In addition, measuring a variation of accumulated gas concentration instead of an absolute concentration decreases the sources of uncertainty – such as for example zero shifting of the gas spectrometer – and decreases uncertainties by extending the measurement period.

Preliminary studies

The method to measure refrigerant leak mass flow rate is mainly based on variation measurement of the accumulated concentration in the closed benchmark volume – called accumulation volume – and knowledge of the accumulation volume. In fact, the pressure value can be reasonably known with an uncertainty of 10 Pa, or impact on pressure measure – close to the atmospheric pressure – by 0.010%. Similarly a temperature can be known within 0.10°C, with a relative uncertainty of 0.034% in a climate controlled room at 20°C. Because of this, the development of the benchmark was particularly based on the following two key points:

- the choice of infrared spectrometer, and

- the design and development of the volume standard in order to measure the accumulation volume by the so-called expansion volumes method [9].

Choice of the photo-acoustic spectrometer

The measurement method for leak throughputs consists of measuring the concentration variation over time. Parameters such as repeatability and linearity of spectrometers are the most relevant. A comparative study considering these parameters made it possible to highlight the performance of photo-acoustic analyzers.

Photo-acoustic spectrometer description

The detection method is based on a phenomenon called photo-acoustic effect. The photo-acoustic effect is the transmission of a vibration by a gas sample in a closed space; the transmission is triggered by the absorption of radiation by a modulated light source. Gas molecules absorb infrared light and unexcite by releasing kinetic energy ([6], [7]). In a closed cavity, this is translated by an increase in pressure. If the infrared source is pulsed, then an acoustic wave in the cavity is created. The acoustic pressure thus created in then measured by microphones.

During measurement, a sample is inserted into a small sealed compartment, the photo-acoustic cell (Figure 1). It is then radiated by light pulsed through this compartment. This light is transmitted by an infrared source. In order to optimize radiation power, an ellipsoidal mirror is used to focus all the radiation into the transparent window of the photo-acoustic cell. Finally radiation is modulated by a 20 Hz beam chopper and filtered at the characteristic wavelength of the measured gas: R-134a or CO_2. When the light is absorbed by the gaseous sample, the gas will heat and dilate, leading to an increase in pressure. Since radiated light is transmitted in the form of frequencies, the heat is also transmitted in the form of frequencies (photo-acoustic effect). The sample then transmits pressure waves measured by one or more microphones, if their frequency band is between 20 Hz and 20 kHz, which is the case. The idea then is to process the signal to learn the concentration of the target gas, after appropriate calibration [8].

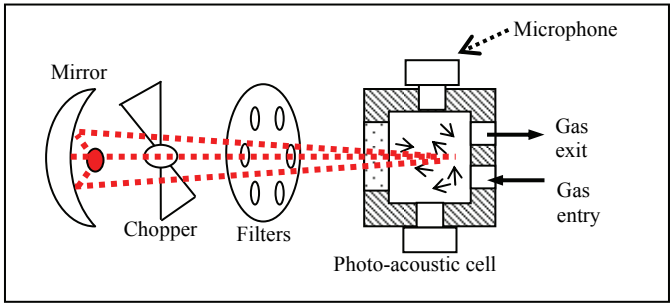


Figure 1. *Photo-acoustic spectrometer diagram*

Detection modeling

A mechanical, energetic and thermal study of the system makes it possible to establish a relation that is directly proportional between the acoustic pressure and the concentration of the sample:

$$p_{rms} = 2\sqrt{2} \times 0.85 \times \frac{(\gamma - 1)}{\omega} \times .|S|.(\alpha.\sin^2\theta.c).\int F(\upsilon - \upsilon_0).k(\upsilon).Tf.L_\upsilon.d\upsilon$$

The integrant $\int F(\upsilon - \upsilon_0).k(\upsilon).Tf.L_\upsilon.d\upsilon$ is calculated by numerical methods, from absorption spectrums. S describes the system behavior in low frequency and θ is the angle of the light cone.

In this way, it turns out that the photo-acoustic detection is a theoretically direct and linear method. It is all the more linear as electrostatic microphones are used. Electrostatic microphones are made up of a thin metallic diaphragm suspended over a rigid back plate. This results in a flexible condenser. When sound waves stimulate the diaphragm, the distance between this and the rigid back plate varies. Consequently, the capacitance varies. This capacitance variation in turn produces a variation of the tension. The associated circuit converts these tension modifications into a signal which is sent to a preamplifier. Microphones are used in low frequency. They are very linear, very stable and not very susceptible to humidity. They are also very sensitive: the electronic background noise of the device measured by a voltmeter is 0.125 mV, which corresponds to an acoustic pressure of approximately 2.5×10^{-6} Pa, whereas the pressures to measure are approximately 10^{-4}-10^{-3} Pa.

Determination of performance

In order to determine the spectrometer performances, a mixer makes it possible to maintain a known gas concentration by an upstream dilution of the gas studied in clean air. For a given standard concentration, several measures are performed. The ambient temperature in the benchmark varies from 23 to 47°C. Since the measuring method of leak thoughputs consists of measuring a concentration variation over time, the study has mainly focused on characteristics such as repeatability and linearity of analyzers.

Results (Figure 2) show that repeatability of the photo-acoustic analyzer is compliant with the level announced by the photo-acoustic spectrometer provider. The repeatability associated with 25 measures is lower than the device resolution, specifically 10 nmol/mol. In addition, results show that the photo-acoustic spectrometer linearity gap is 0.011% of the measurement range which interests us, specifically from 0 to 100 μmol/mol. Waste is lower than the analyzer resolution. Gas analyzer performances are such that very low changes in concentration – between 0 and 1.2 μmol/mol – are perfectly significant (see Figure 3).

However, noted results only show results over a small range, or from 0 to 1.2 μmol/mol. Measures for a concentration of 40 μmol/mol tend to indicate that results on the 0 to 1.2 μmol/mol range can apply for a range between 0 to 40 μmol/mol. The measurement range was voluntarily limited to a low and restricted range, since the goal of this study was to determine the most powerful spectrometer in a lowest possible range of concentrations in order to determine device sensitivities from 0.010 to 0.5 μmol/mol. Optimizing the quality of leak throughput measures and measure time comes down to choosing the lowest sensitivity and most linear spectrometer. Results obtained have established that the photo-acoustic infrared detection was the most adapted to the measurement method of refrigerant leak throughputs.

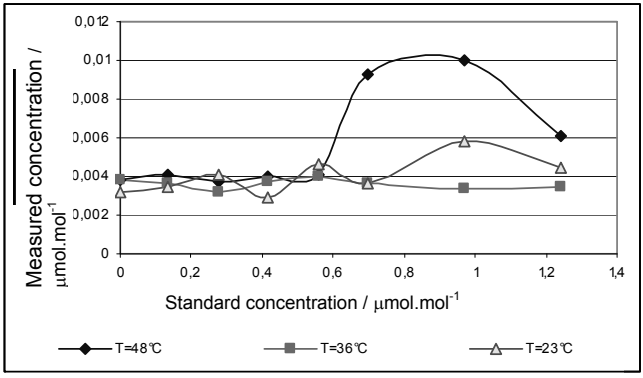


Figure 2. *Determination of repeatability of a photo-acoustic spectrometer measuring a sample of 0 to 1.20 μmol/mol from R-134a (25 measures)*

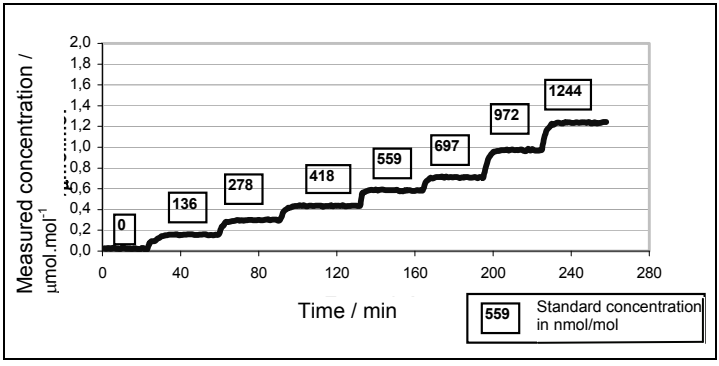


Figure 3. *Measures by a photo-acoustic spectrometer of a sample of 0 to 1.20 μmol/mol for R-134a*

Development of standard volume

Expansion volume method

In order to calibrate refrigerant leaks by the infrared accumulation and detection method, it is necessary to measure the concentration variation as well as the accumulation volume of the installation. The accumulation volume includes the chamber volume and the internal analyzer volume (see Figure 4). Determining this volume with the expansion volume method is considered: an adiabatic expansion of the gas contained in one of the volumes is carried out between two volumes initially isolated by a valve. In fact, the expansion volumes method consists of measuring pressure and temperature of the suppressed accumulation volume V (P_0;T_0) and pressure and temperature of the vacuum standard volume (P_{res};T_{res}), then to measuring the balance pressure and temperature (P_e;T_e) of both volumes after valve opening. By ignoring the thermal effects and residual leaks, conservation of the gaseous mass and the law of perfect gases enables the following relation:

$$R = \frac{V}{V_e} = \frac{P_e - P_{res}}{P_0 - P_e} \tag{2}$$

If one of the volumes is known, the value of the second volume can be deduced. A standard volume was thus designed and developed by LNE.

Description of standard volume

In order to calibrate the standard volume by dimensional measures, its geometry must be very simple. It has been chosen as tubular, for a simplified volume calculation and calibration. Its optimal capacity mainly depends on the dimensions of pressures involved and the accumulation volume.

Since the gas analyzer must be included in the accumulation volume, pressure in the accumulation volume must be included in the acceptable range by the instrument so as not to be degraded. The pressure must therefore theoretically be between 90 kPa and 110 kPa. In this way, by setting the initial and final pressures in the accumulation volume, a simulation of uncertainty components of pressures over the volume ratio makes it possible to establish from relation (2) that the optimal theoretical ratio of both volumes is 4.3 for an initial pressure in the standard volume of 5.3 kPa.

Therefore, the optimal standard volume is four times lower than the accumulation volume. Knowing that the response time is in inverse proportion to the accumulation volume, limiting the volume becomes interesting. The calibration benchmark of leak throughputs is intended for calibrating leaks from 1 g/yr to 50 g/yr. Considering a leak of 1 g/yr, a little more than 2 minutes would be needed to reach a concentration of 1 µmol/mol for a volume of 1 L, approximately

4 minutes for a volume of 2 L, and around 11 minutes for a volume of 5L. Considering a leak of 50 g/yr, approximately 6 minutes would be needed to reach a concentration of 150 μmol/mol for a volume of 1 L, approximately 13 minutes for a volume of 2 L. But the photo-acoustic analyzer is linear for the range of measure 0.015 to 150 μmol/mol and carries out a measure every minute on average. Because of this, a volume with a nominal value of 2 L was preferred over a volume of 1 L, since for a leak of 50 g/yr it makes it possible to carry out at least a dozen measures in the linearity scale of the gas analyzer, contrary to a cell of 1 L. Consequently it is preferable that the standard volume has a capacity of approximately 0.5 L.

Description of the reference

The reference (Figure 4) includes:

- a photo-acoustic spectrometer,
- an accumulation enclosure of a nominal 2 L capacity,
- a clean air supply system,
- a vacuum pump.

The benchmark can be divided into 4 circuits.

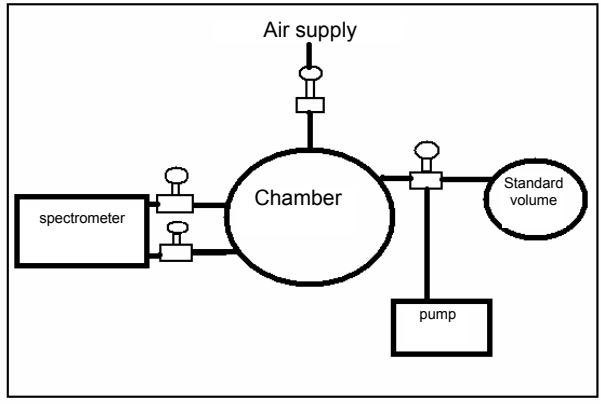


Figure 4. *Pneumatic diagram of the reference*

Description of circuits

Measure circuit

The measure circuit represents all benchmark measure elements connected to the measure enclosure, particularly the circuit connected to the photo-acoustic

spectrometer and isolated from the chamber by two valves. This circuit represents the accumulation volume.

Clean air supply circuit

The clean air supply circuit is isolated from the enclosure by a valve. This circuit's main role is to ensure that liquid circuit pressure is similar to atmospheric pressure, while ensuring a clean atmosphere mainly made up of clean air, particularly void of H_20 and CO_2.

Vacuum circuit

The vacuum circuit of the accumulation volume (or only of the measure enclosure) and/or of the standard volume includes a dry vacuum pump and is delimited by a valve isolating the vacuum pump from the standard volume and the measure enclosure.

This circuit has two objectives:

- vacuum the standard volume during accumulation volume calibration with the expansion volumes method,
- drain impurities from the reference's pneumatic circuits.

Calibration circuit of the accumulation volume

The calibration circuit of the accumulation volume calibrates the accumulation volume with the expansion volumes method. It connects both volumes, isolated from each other by a series of valves.

Description of a leak throughput measure

Before any measure, the leak to measure is connected to the accumulation volume. The leak then flows into the volume. This volume is then rinsed: the vacuum pump is activated and evacuates the accumulation volume. Then, the air feed circuit regulates the pressure in the enclosure to a pressure close to atmospheric pressure. The accumulation volume is then filled with dry clean air. After a leak stabilization time, the measures are launched.

Time, temperature, pressure and concentration are logged at the analyzer measure frequency, on average every minute. This data makes it possible to calculate (1) the mass flow rate of the connected leak.

Qualification of the reference

Calibration method of the photo-acoustic spectrometer

Two methods are considered to calibrate the photo-acoustic spectrometer. The first one considers concentration as measurand. This measurand is a quantity similar to the one measured by the spectrometer. The calibration is based on the method used to evaluate spectrometer performances on the market. Several standard concentrations of the gas involved are inserted in the accumulation volume and are measured by the spectrometer. A polynomial fitting of the measured concentration according to the standard concentration can then be established over a range of 0 to 100 μmol/mol.

The second method consists of comparing the mass flow rate of a leak calibrated by another method with a throughput that is measured by the reference. Calibrating would then be indirect. The implementation of this type of calibration will mainly be based on the calibration methods of helium capillary leaks and on the adaptation possibility of these methods to refrigerant capillary leaks. A Euromet project is currently carried out between INRIM (Istituto Nazionale di Ricerca Metrologica) and the LNE in order to establish a conversion method of a leak throughput measure of a gas A –such as helium – into a throughput of this same throughput emitting another gas B such as a refrigerant gas.

This conversion method requires knowledge of the flow regimen. In the case of a molecular system, the conversion is clearly established. But that is not the case with the intermediate system which is not well known [11]. The Knudsen formula describes an influence from the solar mass as well as from viscosity in an intermediate system. Few studies have been performed in this field and most are based on the Knudsen relation [10]:

$$C = \frac{\pi}{128} \cdot \frac{D^4}{\eta L} \overline{P} + \frac{1}{6} \cdot \sqrt{\frac{2\pi \cdot kT}{N_a \cdot M}} \cdot \frac{D^3}{L} \cdot \left[\frac{1 + \sqrt{\frac{N_a \cdot M}{kT}} \cdot \frac{D}{\eta} \cdot \overline{P}}{1 + 1{,}24 \cdot \sqrt{\frac{N_a \cdot M}{kT}} \cdot \frac{D}{\eta} \cdot \overline{P}} \right]$$

Leak behavior in the intermediate flow regime is still not known well enough. The first objective of the project is to exchange information concerning equipment developed for this purpose and measures of available leak throughputs, and to verify the coherence of a few measures between both laboratories. A program will then be established to optimize the efforts.

Calibration of standard volume

Calibration consisted of the determination of necessary dimensions to reach a volume, diameter and height, and to estimate uncertainty components, particularly

uncertainties linked to the standard, to flatness and to cylindricality. It was calibrated in an air conditioned room at (20°C ± 0.5°C) by a three-dimensional measuring machine connected to national standards.

Total volume is broken down into several elements (Figure 5):

- main cylinder, Table 1 presents the uncertainty budget,
- capacity of the fittings between the standard volume and a manometer,
- added capacity of the fittings between the standard volume and the volume to measure.

Calibration consists of the determination of geometric parameters for the calculation of volume. Concerning the cylinder diameter, its value is taken at several heights in order to reflect volume cylindricality. The operator carries out a first location and determines a flat reference virtual surface modeling one side of the cylinder. The diameter range at every point of the cylinder is then determined in relation to the reference diameter after manipulation of the geometric element. Concerning the cylinder height, two measures are made: one by touching the lower face of the cylinder, upper face taken off; the other by touching the upper face of the cylinder, the lower face taken off.

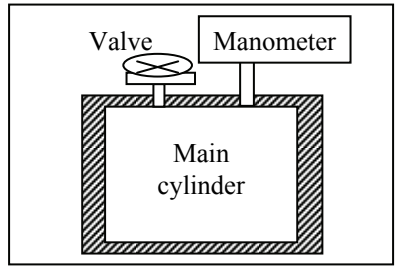


Figure 5. *Diagram of standard volume*

For example, Table 1 presents the uncertainty budget of the main cylinder. Dimensional calibration estimates the standard volume at 509.56 cm^3 with a relative uncertainty of 9.1 x 10^{-4} (k=2).

Main cylinder						
Components	**Unit**	**Uncertainty**	**Distribution law coefficient**	**Standard uncertainty**	**Sensitivity coefficients**	**Standard uncertainty [mm^3]**
Uncertainty of the diameter	*mm*	0.0034	2	0.0017	12320 mm²	21
Deviation of the diameter from the mean value	*mm*	0.011	3	0.0037	12320 mm²	45
Uncertainty of the measurement of the length of the cylinder	*mm*	0.010	2	0.0050	5346 mm²	27
Uncertainty in the determination of the flatness of the flanges	*mm*	0.0015	2	0.0008	5346 mm²	4
Flatness of the upper flange	*mm*	0.0070	√3	0.0040	5346 mm²	22
Flatness of the lower flange	*mm*	0.047	√3	0.027	5346 mm²	150
Deviation in the definition of the reference plane	*mm*	0.024	√3	0.013	5346 mm²	74
Thermal expansion	*K*	1 K	√3	0.58 K	24.4 $mm^3.K^{-1}$	14

Table 1. *Uncertainty budget of the "main cylinder" of the standard volume*

Accumulation volume calibration: first results

Once the standard volume is qualified, the calibrating method for the accumulation volume must be developed. Two aspects of the expansion volumes method are not present in the equation (2): the imperfect tightness of the reference questions the conservation aspect of the matter and thermal effects.

Before the expansion, compression is carried out in the accumulation volume so that the pressure is slightly higher than the atmospheric pressure. The standard volume is depressed. A stabilization delay of both pressures is taken into consideration. Once both pressures are stabilized, they are measured and their values are extrapolated at the opening of the valve.

After expansion, the temperature in the accumulation volume drops and the temperature in the standard volume increases. To limit the uncertainty of the standard volume, the temperature inside the standard volume is not measured. Only the wall temperature of the standard volume is checked. Gain in temperature is therefore insignificant. A temperature correction is not possible as long as the

thermal balance of gas in both volumes is not reached. Therefore, pressure and temperature measurements can only be taken into account after stabilization of the pressure and temperature inside both volumes. An example is provided in Figure 6.

However, the longer the stabilization, the more mass conservation can be questioned. It is therefore necessary to model variation of pressure P and temperature T in order to establish the value of the P/T ratio at valve opening. A thermal study and tests are currently being performed to validate this method and establish uncertainty components. The first results seem to estimate the volume at 2.16 L with a reproducibility of approximately 0.10%.

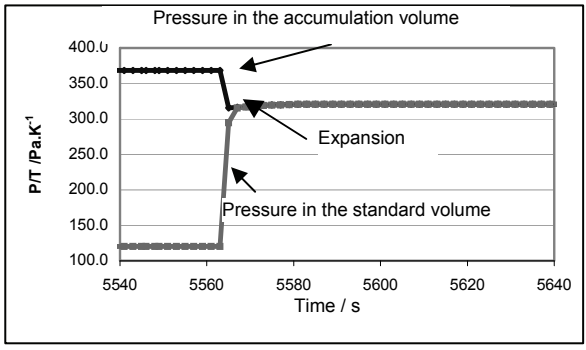


Figure 6. *Measurement results for the 10:1 ratio of the 100 Ω and 10 Ω standard resistors. Vertical lines denote accuracy specification limits of the bridge*

Conclusion and perspectives

On the bases presented in this document, measurement uncertainty components of calibrated leak mass flow rates will be defined and validated by the comparison with methods from the LNE experiment on helium throughput calibration. Early in 2008, the LNE will be able to provide reference leak calibration services for users or original equipment manufacturers in order to respond to regulation text requirements.

As described in these texts, calibration of refrigerant leak mass flow rates is intended to ensure traceability of leaks used for the qualification of hand leak detectors, ensuring mandatory periodic leak tightness controls over refrigerant and climatic installations. The project also plans the implementation of a qualification benchmark for leak detectors and ambiance controllers. Using these means, a rigorous analysis of the operation of refrigerant leak detectors, as well as phenomena occurring during mass transfer from the leak to the detector sensing element, will make it possible to establish the precautions to take during leak tightness tests.

Studies carried out on leak behavior, in particular according to the nature of the gas, and on the operation of leak detectors, should appear in 2008 in propositions in the prescriptive field, to confirm or amend existing texts, or to propose others.

References

[1] N. Torbey, D. Clodic, Méthode d'étalonnage sur un étalon primaire des fuites calibrées permettant la vérification des performances des détecteurs de fuites des fluides frigorigènes, ADEME (01 74 084/11243), December 2002.

[2] D. Clodic, N. Torbey, F. Fayolle, "Qualification des détecteurs de fuite – Etalonnage des fuites calibrées", in *Colloque à effet de Serre III*, 2002.

[3] D. Clodic, "Measurement and Control of Refrigerant Leaks", in *AICARR Conference*, 2000.

[4] D. Clodic*, Zero Leaks,* ASHRAE, 1998, ch.6, pp89-94.

[5] F. André, "Spectroscopie photo-acoustique", in *Techniques de l'Ingénieur, Mesures et Contrôles treatise*.

[6] S. Schäfer, A. Miklós, P. Hess, "Photo-acoustic Spectroscopy, Theory", in *Encyclopedia of Spectroscopy & Spectrometry*, Lindon J, Tranter G, Holmes J, Eds, 1999, pp. 1815-1822.

[7] A. Rosencwaig, *Photo-acoustics and Photo-acoustic Spectroscopy*, John Wiley & Sons, 1980, ch3-4, pp.15-41.

[8] J. Christensen, The Brüel & Kjaer Photo-acoustic Transducer system and its Physical Properties, Denmark: *The Brüel & Kjaer Technical Review*, 1990, n°1, pp. 4-19.

[9] J.D. Wright, A.N. Johnson, M.R. Moldover, "Design and Uncertainty Analysis for a PVTt Gas Flow Standard", *Journal of Research of the National Institute of Standards and Technology*, Vol. 108, n°1, pp. 21-47, Jan/Feb 2003.

[10] J.Delafosse, G. Mongodin, *Les Calculs de la Technique du Vide*, Société Française des Ingénieurs et Techniciens du Vide, 1961, ch.3, pp 29-35.

[11] M. Bergoglio, G. Brondino, A. Calcatelli, G. Raiteri, G. Rumiano, "Mathematical model applied to the experimental calibration results of a capillary standard leak", in *Flow Measurement and Instrumentation*, Vol. 17, pp. 129–138, 2006.

[12] C. D. Ehrlich, S. A. Tison, "NIST Leak Calibration Service", in *NIST Special Publication*, n°250-38.

[13] J.C. Legras, J. LE Guinio, "L'étalonnage des fuites de référence au BNM/LNE", in the *8ème Congrès International de Métrologie*, 1997

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Quality aspects of determining key rock parameters for the design and performance assessment of a repository for spent nuclear fuel

B Adl-Zarrabi[1], R Christiansson[2], R Emardson[1], L Jacobsson[1], L R Pendrill[1], M Sandström[1], B Schouenborg[1]

[1]SP Technical Research Institute of Sweden, Box 857, SE-50115 Borås, Sweden
[2]SKB Swedish Nuclear Fuel and Waste Management Co., Box 5864, SE-102 40 Stockholm, Sweden
Leslie.pendrill@sp.se

ABSTRACT: In planning for the appropriate disposal of radioactive waste, it is necessary to understand and predict the long-term changes that take place in a final waste repository and how these changes can influence the repository's ability to maintain adequate isolation of the spent nuclear fuel. The value of performing detailed measurement system analyses and interlaboratory comparisons as an important means of providing confidence in test results is demonstrated for two key mechanical and thermal bedrock parameters. Systematic measurement quality assurance makes it possible to identify the degree the variations in measurement data depend on actual natural variability in the rock mass as opposed to uncertainties from the laboratory testing.

Introduction

The mechanical (compressive and shear strength, elasticity modulus, etc.) and thermal (expansion coefficients, thermal conductivity, etc.) properties of the bedrock are crucial parameters for the construction and long-term safety of underground repositories for spent nuclear fuel [1]. These mechanical properties are, together with rock stress measurements, initially used for a safe construction phase. In the early post-closure phase (up to 1,000 years), the thermal properties become equally important and are, together with the mechanical properties, important when assessing the coupled thermo-mechanical processes that are expected to affect the repository.

The practical use of all test results is dependent on a correct treatment of measurement quality, particularly metrological traceability and measurement uncertainty. Estimations of test uncertainty will include a statistical analysis of repeated measurements of rock properties and a total measurement system analysis, covering the instrument used, the operator, chosen test method, environmental

conditions and even the test objects themselves. In estimating these uncertainties, the test engineer's experience is used together with regular evaluation of measurement method accuracy using inter-laboratory experiments; see ISO 5725 [2]. We also have to take into account core sampling on site and questions of statistical representativeness. A second aspect of quality-assured measurements covers the actual mechanical strains and stresses and thermal gradients to which the rock is submitted underground on site.

The final step is to combine the measurement data on rock properties and compare the values of the different parameters with tolerance values for these, such as specified based on detailed, three-dimensional simulations of the site. Important steps include the correct treatment of measurement uncertainty when making decision of conformity assessment as well as realistic estimates of the risks and costs of incorrect decisions.

The paper will present examples of measurement quality assurance, system analysis and conformity assessment applied to rock mechanics and nuclear waste disposal.

Reliability of data used in analysis

Uncertainty in problem assessment can arise from many factors [3]. These could comprise both the conceptual uncertainty in understanding the problem, including the impact on analysis results of various simplifications made when the problem description is established, as well as the uncertainty in the data used. Geoscientific analyses additionally focus on all of the general geostatistical issues.

Uncertainty in data depends primarily on three factors;

– how typical are the data for the actual geological formation? This deals with issues such as geological homogenity/heterogenity and spatial variability, for example trends with depth;

– the strategy for collecting representative samples, and the samples needed for the problem analysis;

– the quality of the data collected.

Two examples are discussed that are relevant for the design of a nuclear facility in a geological environment, illustrating how the assessment of data uncertainty can contribute to the improvement of confidence in the prediction outcome.

Example no. 1, influence of UCS on the analysis of stress-induced spalling

The problem of stress-induced spalling on an underground opening has been assessed by many authors. Martin [1999] proposed an empirical relation based on the maximum tangential stress at the opening σ_T and the uniaxial compressive stress *UCS* at which spalling would occur:

$$\sigma_T/UCS \geq 0.55 \pm 0.20 \qquad (1)$$

The uncertainty in σ_T depends on how the geometry of the opening influences stress concentrations at the contour of the opening, but the largest uncertainty is related to estimation of the in-situ state of stress. This need for high quality stress estimation has been highlighted by, for example, Christiansson & Hudson [2003]. Generally speaking, the problem of an adequate determination of the state of stress at a large depth is greater than the determination of the *UCS*, but the discussion in this paper is limited to the latter.

Example no. 2, Influence of thermal conductivity on the canister spacing in a KBS-3 repositor

Heat generation from the spent fuel causes a temperature gradient from the deposited canister. The transfer of heat from the canister depends on the thermal properties of the buffer and the rock, as well as the dimensions of the deposition hole and the buffer. For the Swedish reference case with a heat generation of 1 700 W/canister and assumptions on the parameters such as density and water content that determine the thermal properties of the buffer, Hökmark and Fält [2003] calculated how the canister spacing depends on the thermal conductivity of the rock, highlighting the impact of uncertainty in the thermal conductivity on the estimation on how large the repository should be.

Measurement quality assurance

Measurement quality and reliability in the competence of laboratories making the tests can be assessed and assured with both internal and external measures.

- Internal means of quality assurance are e.g. use of reference material, control charts, regular internal audits etc.

- External means are on the one hand the surveillance audits e.g. within the legal authorisation/notification or accreditation procedure and on the other hand participation in inter laboratory comparisons of results of the same test item.

A key observation about measurement quality assurance is that the **test object** occupies a special place in the measurement system since it is both the entity whose intrinsic characteristics are to be determined as the prime aim of the test, but is at the same time an integral part of the measurement system. It will be important to distinguish between situations where:

– on the one hand the aim of the test is in fact to evaluate how, for instance, the object is affected by its environment or how much the object varies, and

– on the other hand, where uncontrolled environmental effects on the test sample lead to measurement uncertainty.

The test object is crystalline rock, which causes additional concerns [4] since geological heterogenities cause natural variations in the mechanical and thermal properties. In addition, the mechanical properties may be affected by the stress paths the object has been subject to when it has been cored out from depth in the bedrock. The stresses depend both on the induced stresses and heat generation during core drilling, as well as the stress relaxation caused by the release of the core from its in-situ state of stress. The overall strategy in sample selection for testing in the SKB site investigation programme is based on:

– consultation with site geologists on the representativeness of potential test objects;

– collection of samples in bathes at a depth with a number of specimens as close to each other as practically achievable.

In addition, the results have to be critically reviewed. In particular, test results that are anomalous compared to the majority of results from a batch have to be scrutinized. Local heterogenities may have large influences on test results because of the small volume of rock involved in the testing methods.

The steps in evaluating measurement quality will be exemplified for two examples of measurement of critical properties of bedrock samples taken from proposed SKB site investigations for deep geological disposal of radioactive waste [5]. The particular rock determinations exemplified here are the uniaxial compressive strength test and thermal conductivity determined according to the TPS-method.

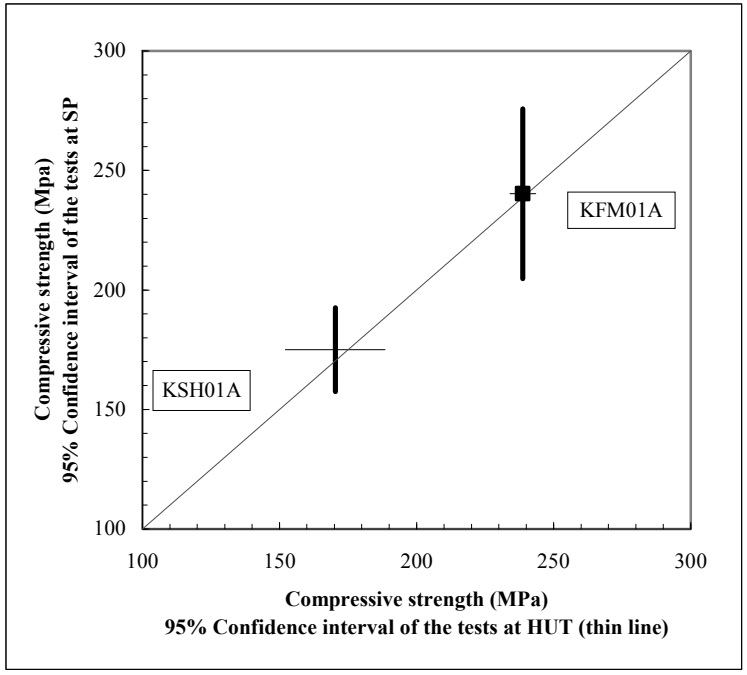


Figure 1. *Combined results of the uniaxial compressive strength tests*

Actual measurements I: Uniaxial compression test

The uniaxial compression test consists of the loading of cylindrical specimens in the axial direction up to and beyond failure (post-failure) in the actual tests [6]. The stress, axial and radial strains are recorded during the test and the elasticity parameters, Young's modulus and Poisson ratio as well as the uniaxial compressive strength are deduced from the measured sets of data.

Inter-laboratory experiment [7]

Determinations of uniaxial compressive strength of the bedrock samples were performed at two testing laboratories: the Swedish National Testing Institute (SP) and the Helsinki University of Technology (HUT). Both laboratories tested the samples following the same standard. The main differences between the laboratories were the use of different test machines and different types of equipment for measuring the axial and radial deformations.

Detailed information about the samples and tests is given in the P-reports for example [8].

Figure 1 shows the mean values of the measurements at each laboratory plotted against each other. The 95% confidence intervals are shown in the figure as vertical and horizontal lines for SP and HUT respectively.

The diagonal line illustrates where the cross should be if the results were identical. From the figure, we can see that the confidence intervals include the diagonal "zero-line". Hence, with 95% confidence, we can conclude that the combined systematic difference between the batches at the two laboratories and the systematic differences in methods and equipment is smaller than the scatter between the samples at each laboratory.

Actual measurements II: Thermal conductivity test

The determination of the thermal properties (thermal conductivity, thermal diffusivity and volumetric heat capacity) is based on a direct measurement method, the so-called "Transient Plane Source Method" [9]

Interlaboratory experiment

The samples were first tested at SP and then at the Hot Disk AB laboratory. Figure 2 shows the mean results for thermal conductivity for the two lots of samples, with three different temperatures for each lot. The mean values of the measurements at each laboratory are plotted against each other. The 95% confidence intervals are shown in the figure as vertical and horizontal lines for HD and SP respectively. The diagonal line illustrates where the cross should be if the results were identical. From the figure, we can see that the confidence intervals in about half of the cases include the diagonal "zero-line". Hence, systematic difference between the batches at the two laboratories and the systematic differences in methods and equipment are more significant than the scatter between the samples at each laboratory in contrast to the measurements of the *UCS*.

However, care should be exercised in making due allowance for other sources of measurement error: both laboratories have used nominally the same TPS method. Hence, additional systematic components to the measurement error may exist without being apparent in the interlaboratory comparison.

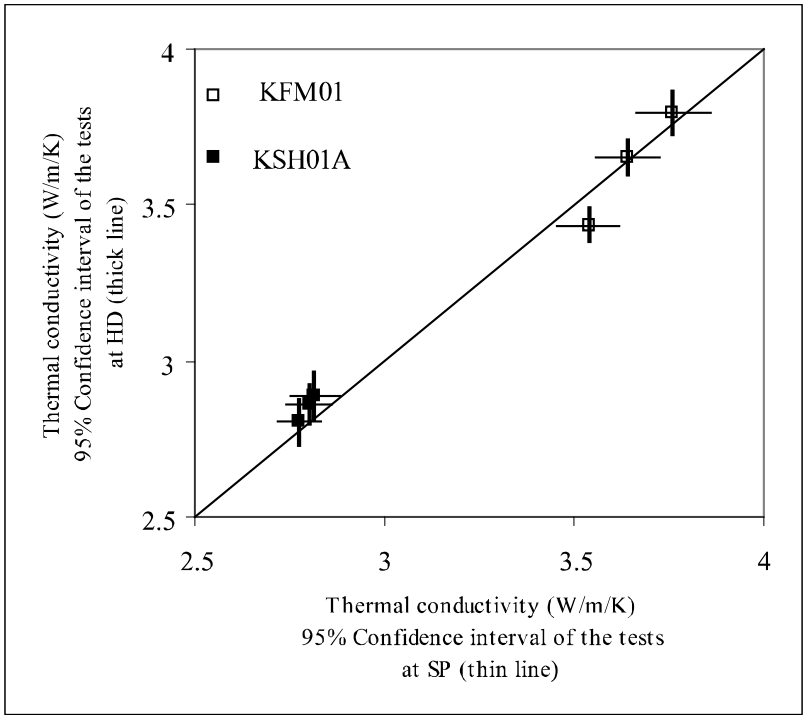


Figure 2. *Diagram showing the measurements of thermal conductivity*

Discussion

The scatter in a set of laboratory data is dependent on many factors but can be simplified to the following relation:

$$\sigma_y^2 = \sigma_m^2 + \sigma_w^2 \tag{1}$$

where σ_y^2 is the total scatter in a set of data, σ_m^2 is the natural variation amongst the tested specimens [10] and σ_w^2 is the total [11] variation due to the testing procedures, i.e., the measurement noise.

Estimation of natural variability and uncertainties in testing

Based on the results of the interlaboratory tests, we can deduce information on the measurement noise in the two test cases UCS and TPS and for the different laboratories involved. The natural variation among the samples can be written as a function of the variances of the sum and difference of the measured signals

$$\sigma_m^2 = \frac{1}{4}\left(\sigma_{sum}^2 + \sigma_{diff}^2\right) \tag{2}$$

and the measurement noise can then be found for the two laboratories as

$$\sigma_w^2 = \sigma_y^2 - \sigma_m^2 \tag{3}$$

For the KSH site the measurement noise in the UCS measurements is typically 11-12 MPa. For the Forsmark site, it is difficult to draw conclusions about the measurement noise due to the large discrepancy in the results.

In the TPS measurements, the measurement noise is approximately 0 and 0.02 W/m/K for the SP and HD laboratories respectively at the KSH site and 0.08 W/m/K and 0 for the SP and HD laboratories respectively at the Forsmark site. Based on these results, it is reasonable to assume that the measurement noise in the TPS measurements is approximately 0.05 W/m/K.

The above analysis and conclusions about the measurement uncertainties at the different laboratories would not have been possible to perform without the interlaboratory comparisons. Additional labs in such comparisons would presumably help in the determination of measurement uncertainties especially for the USC measurements where we at present have difficulties in separating the effects due to the fact that the samples at the labs are different.

The two Swedish study sites enclose some 5 – 8 km^2 in the focused area for a tentative repository. The investigated depth is down to 1,000 m with a target depth in the interval 400 – 700 m. In addition, there are also investigations carried out outside the focus area to study the boundary conditions at each of the sites. It is obvious that there is the need for a strategy for investigating and sampling that has to focus on geological characterizations. These works divide the rock mass into domains that are estimated to be homogenous. The sampling for characterization is focused on collecting batches of data in clusters at different locations and depths. The main purposes are to investigate how the natural variation of design parameters may vary in the scale of a tunnel (approximately sampling within a 2 – 5 m borehole length) and how consistent this natural variation is over the site, and with depth.

The amount of testing is a trade-off between the general geological understanding of the homogenity/heterogenity of a site, and cost. It is not realistic to establish a specific level of confidence in data to be achieved. Instead, using step-wise analyses of data, feed-back can be given to the data collection activities. Many of the issues that are studied, such as for example the risk of a deposition hole spalling, have limited value where stability problems may be expected. It is relatively straight-forward to identify if a parameter value range for a geological domain may be of concern or not. So if for example the insitu stresses are estimated to be large, it may be more important to also define the most likely range of UCS for the actual rock mass. In the Forsmark example the estimated state of stress has been estimated to be relatively high [SKB 2005 = SDM v1.2]. Spalling has been a concern that has been subject to special studies [Martin, 2005]. Even though the uncertainty of estimating the state of stress has the largest impact on the estimated risk of spalling, even a reduction of the uncertainty span of the UCS mean value in accordance with Table 1 is important in improving confidence in the prediction outcome.

Table 1 summarizes the uncertainties in the testing procedures and the natural variability of the samples. In the table, we present a standard deviation in all data for the gneiss granite in Forsmark according to SKB 2006 [SDM 2.1]. The standard deviation of the measurement noise is taken from (3) and discussed above. Assuming that the all measurements in the larger study are subject to measurement noise of the same size as reported in this paper, we can estimate the natural variability of this data set from (1).

	Variability *UCS* (MPa)	Variability *TPS* (W/m·K)
Larger Population	29	0.17
Measurement noise from (3)	12	0.05
Natural variability from (1)	26	0.16
Natural variation as percentage of mean value	9%	4%

Table 1. *Uncertainties in testing procedures and natural variability of rock samples*

In the present case, the strategy for sampling and laboratory testing seems to be adequate for the characterization of the mechanical and thermal properties of the bedrock. It remains to be evaluated if the actual strategy would also be sufficient in a more heterogenous rock mass. It must also be considered that the sampling strategy has to be defined from the purpose with the project and the need of confidence in predictions.

Conclusions

The paper has illustrated the value of performing detailed **measurement system analyses** and interlaboratory **comparisons** as an important means of providing confidence in test results for two key mechanical and thermal bedrock parameters. A well-functioning system of interlaboratory comparisons in all fields and at different levels may support the wider acceptance of test results. Furthermore, frequent participation in interlaboratory comparisons may result in a collective learning effect, leading hopefully to the results of the different laboratories converging successively and uncertainties becoming smaller. It is important that in all the different fields regular interlaboratory comparisons are made available in view of the overall improvement of measurement this brings.

It has been emphasized that it is advantageous to give a clear separation between **intrinsic object characteristics** – such as for instance actual heterogenity – and an apparent heterogenity arising from investigation with a particular measurement system. This separation is necessary both in assessing uncertainties as well in deducing the result of the test. The test objects in the present investigations were cores obtained from the drilling of bore hole. Therefore, the properties of the test objects vary, depending on variation of the petrographic composition, structure etc.

Account has also to be taken of **core sampling on site** and questions of statistical representativeness arise. The testing of geological materials such as crystalline rock requires a systematic strategy for sample selection to reduce the effect of natural heterogeneities. In addition, the stress paths the specimen may have experienced from its release from the bedrock may also influence the results:

- The strategy for sampling and testing of the mechanical and thermal properties in the Swedish ongoing site investigations for a nuclear disposal facility has been focused on batches of tests within short distances along the boreholes. This has revealed the extent of the natural variability in the rock mass on a scale that is relevant for the planned tunnels (up to approximately 5 m).

- All data from several boreholes has been used to assess the natural variation in the rock mass.

- Systematic assurance of measurement quality has made it possible to sort out to what extent the scattering in data depends on natural variability in the rock mass, as opposed to the sum of all uncertainties associated with the laboratory testing.

- It is crucial that the test methods are properly defined; systematic quality procedures are followed and that the operators and the laboratories are experienced.

– In the examples given in this paper, it has been shown how the systematic assessment of measurement uncertainties helps to determine the natural variability in rock properties.

References

[1] B. Adl-Zarrabi, R Christiansson, R Emardson, L Jacobsson, L R Pendrill, M Sandström, B. Schouenborg 2007, "Quality aspects of determining key rock parameters for the design and performance assessment of a repository for spent nuclear fuel", forthcoming.

[2] ISO 5725 1994 "Accuracy (trueness and precision) of measurement methods and results".

[3] J Helton *et al.* 2006 "Survey of sampling-based methods for uncertainty and sensitivity analysis", *Reliability Engineering and System Safety* **91**, 1175 – 1209.

[4] SKB site description models (platsbeskrivande modeller) Forsmark and Laxemar, SKB 2005 and 2006.

[5] SKB, 2001, Site investigations. Investigation methods and general performance procedure (Platsundersökningar. Undersökningsmetoder och generellt genomförandeprogram), SKB R-01-10, Svensk Kärnbränslehantering AB.

[6] ISRM 1999 "Draft ISRM suggested method for the complete stress-strain curve for intact rock in uniaxial compression", *Int J Rock Mech Min Sci* **36**(3), 279 – 289.

[7] Matz Sandström, SP Swedish National Testing and Research Institute, Forsmark and Oskarshamn site investigation boreholes KFM01A and KSH01A: Inter-laboratory comparison of rock mechanics testing results, SKB P-report, P-05-239, December 2005, ISSN 1651-4416.

[8] Jacobsson, Lars, SP Swedish National Testing and Research Institute, Forsmark site investigation, Borehole KFM01A, Uniaxial compression test of intact rock, SKB P-report, P-04-223, December 2004, ISSN 1651-4416.

[9] Adl-Zarrabi, Bijan, SP Swedish National Testing and Research Institute, Forsmark site investigation, Drill hole KFM01A, Thermal properties: heat conductivity, and heat capacity determined using the TPS method and Mineralogical composition by modal analysis, December 2004, P-report, P-04-159, ISSN 1651-4416.

[10] Kurfürst *et al.* 2004 "Estimation of measurement uncertainty by the budget approach for heavy metal content in soils under different land use", *Accred. Qual. Assur.* **9**, 64 – 75.

[11] Ramsey M H 2004 "When is sampling part of the measurement process?", *Accred. Qual. Assur.* **9**, 727 – 8.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Environment protection measurements by the technical inspection of vehicles

Edi Kulderknup, Jürgen Riim

Institute of Mechatronics, Tallinn University of Technology
Ehitajate tee 5 Tallinn Estonia
edikuld@staff.ttu.ee

ABSTRACT: Environment protection measurements are being increasingly requested due to environment pollution problems. Among the most significant sources of air pollution are emissions' from road vehicles. The main reason for the pollution is exhaust gas components which have a more harmful effect – CO, CO2, O2 and HC from ignition-type engines and NOx and smoke from compression engines.
Nowadays it is common to use the periodical inspection of road vehicles with the aim of obtaining knowledge of the technical level of each item and assuring safety and environmental protection. The above gases are on the list of inspected parameters.
To give more confidence to the measurement results, they contain uncertainty estimation. To correctly control the theoretically estimated uncertainty, a suitable way is to carry out comparisons of vehicle exhaust gas measurements.
In this paper we give a model of uncertainty estimation of exhaust gas measurements and results and conclusions of practical measurement comparisons carried out by the technical inspection bodies in Estonia.

Introduction

Nowadays environment protection is more requested. Among the most significant sources of air pollution are road vehicles. The main reason for the pollution is the exhaust gases emitted by the combustion engines of vehicles. Exhaust gases include components which have a harmful effect as greenhouse gases and initiators of the smog. The above gases generally include carbon monoxide CO, carbon dioxide CO_2, oxygen O_2 and hydrocarbons HC for ignition engines and NO_x and smoke (fine particles) for compression engines. Measurement of the gas components and its concentration can be conducted rather easily using a gas analyzer. The measurement process is visually uncomplicated but to ensure a proper result a wide scale of influence factors will be taken into account.

During the measurement of exhaust gas concentration some influence factors can give exaggerated uncertainty if not limited correctly. As influence factors, the state of the measuring equipment sensors, environment temperature, humidity and cleanness and vehicles hose faultlessness have special importance.

The quantity and/or concentration of exhaust gases are regulated by legal norms and they shall be controlled during the periodical inspection of road vehicles as a

rule. Correct conformity estimation to the vehicle technical state shall include uncertainty of the exhaust gas measurement.

For this study work were used GUM principles for estimation of the uncertainty components. Component estimation was carried out using concrete measuring equipment and methods and condition data.

To give more real bases and confidence to the estimated uncertainty components, practical control activities shall be carried out. This is suitable to perform the measurement comparisons.

For this study, estimation of the correctness of theoretical values was controlled through the practical comparison measurements. Practical comparison measurements of exhaust gas concentrations were carried out over the time period from May up to September 2005. Three vehicles with ignition engines were used and over 40 Estonian vehicle inspection stations were involved. To assure equal conditions were used meant comparison object stability and limiting the extreme measurement influence factors.

During comparison road vehicles were also involved with other parameters which had a link with safety. The list of items to be inspected and/or tested was as prescribed in the European Union directive 96/96/EC. Results of measurement of the other safety parameters were presented in [1].

This was the main origin for this study and the task was to estimate the influence factors and its uncertainty components of the exhaust gas quantity measurements and to control these results using from the practical comparison. This study gives models for the uncertainty component estimation.

As a result of this study, obtained guidance allowing the improvement of the estimation of measurement capability of the exhaust gases components was given and used for accreditation of measurement activities of the road vehicle inspection stations. Results are and would be used in practice in Estonian road vehicles inspection stations (about 60 stations).

Need for environmental protection measurements of road vehicles

In Estonia about 0.4 million road vehicles are registered and those have an average age 8 – 10 years. This means that exhaust gases from vehicles make up an important part of the air pollution. Some studies give the estimate that the vehicles part is ca 40% of all the air pollution. To limit overvalue of exhaust gas concentration for vehicles norms are determined by legal acts and those are controlled during vehicle technical inspections.

Technical inspection age of the vehicle. During shall be carried out over a prescribed time period which is one or two years depending on the vehicle inspection mandatory inspections are performed, including any associated testing and measurements, of road vehicles according to prescribed requirements and determine the conformity of the inspected vehicles to the requirements on the basis of a professional judgment. Such a declaration shall not limit vehicle owners' legal rights to use the vehicle, but on the other hand increase safety for the other traffic.

During the vehicle inspection mainly 5 – 6 measurements are used as shown in Table 1. The parameter which shows the vehicles conditions but also assure a level of pollution that is as low possible, is the concentration of CO, CO_2, O_2 and HC for the ignition engines and NO_x and smoke particles quantity for compression engines.

Parameter	Measurement range
Ignition engines exhaust gases	CO content (0.03÷10) vol% CO_2 content (0.5 ÷ 18) vol% CH content (5÷9999) vol ppm O_2 content (0.1 ÷ 22) vol% λ-number (0.8 ÷ 1.2)
Compression engines exhaust gases	(0.5 ÷5) m^{-1}
Braking force and vehicles mass	0.1 kN ÷ 40 kN 100 kg ÷ 15 t
Geometrical measures	0.5 mm ÷ 30 m
Noise when stationary	(30÷130) dB
Windscreen transparency	(5 ÷ 60)% (60 ÷ 100)%

Table 1. *Measurement of vehicles emission and safety parameters with measurement range*

Technical inspections are carried out by inspection bodies, of which there are 65 - 67 in Estonia. They are all private companies. To give more confidence, inspections and measurements in Estonia are settled on the system of estimation of measurement competence and the measurement comparisons are regularly carried out.

Exhaust gas concentration measurement

Main components. Exhaust gas components whose content is limited are oxygen O_2, carbon oxide CO, carbon dioxide CO_2, hydrocarbon HC and NO_x for ignition engines and smoke particles and NO_x for compression engines [2].

Oxygen concentration in the air is about 21 vol% and it is needed in the burning process for the engine. Theoretically, oxygen content should be near 0% in exhaust gases for correctly working engines. Practically, exhaust gases includes (1…2) vol% of O_2. O_2 content is not a source of air pollution, but it shows how good the engine's working state is.

Carbon monooxide CO is a strongly poisonous gas without color or odor. About 0.3 vol% over 30 minutes is a deadly quantity for humans. CO joins with the hemoglobine and this cause serious damage to human health. In air pollution CO is most extensive and involved in up to 50 mass% of the total air pollution. CO is produced when the engine does not work in the correct way and the burning of fuel is not completed. Taking into account technical aspects, it is just as bad to have a CO content that is too low, as it is to have a CO content that is too high. A reason for the high CO content can be high fuel consumption and for the low content, defective functioning of the ignition systems.

Carbon dioxide CO_2 is produced when engine working conditions are normal. For the effective working a concentration of 13 up to 15 vol% is needed. CO_2 participates in photosynthesis and is not poisonous and in this way can appear to be useful. **** harmful influences can be found when CO_2 concentration increases in the air. This is because carbon dioxide is a greenhouse gas. Practically, it is possible to achieve low CO_2 quantity when vehicles use less fuel.

Hydrocarbons HC are the main component in petrol and theoretically should burn completely. Practically, burning, HC does not destroy it and it stays in the exhaust gases. Various types of HC exist and they are dangerous especially for human eyes and lungs. It is the bases of smog and dangerous for green forests. New car engines have a HC fraction up to 20 – 30 ppm and older engines have a limit up to 1000 ppm.

NO_x is a poisonous gas and the basis for smog development. It is raised by high temperature and pressure. This means when engines work correctly the NO_x content is much higher.

Lambda number shall be 1±0.003 and shows the combined influence of the above components.

Principles of measurements. The above gases can be measured directly using an exhaust-gas analyzer which works mainly using the non-dispersive infrared method. Vehicle exhaust gas analyzers are produced widely and have the measurement parameters and accuracy parameters shown in Table 2. Accuracy parameters are the maximum permissible errors for the measuring instrument under rated operating conditions according to directive 2004/22/EC.

Parameter	Class I	Class II
CO fraction	$\pm 0.06\%^{v}/_{v}$ ±5%	$\pm 0.2\%^{v}/_{v}$ ±10%
CO_2 fraction	$\pm 0.5\%^{v}/_{v}$ ±5%	$\pm 1\%^{v}/_{v}$ ±10%
HC fraction	$\pm 12\ ppm^{v}/_{v}$ ±5%	$\pm 30\ ppm^{v}/_{v}$ ±10%
O_2 fraction	$\pm 0.1\%^{v}/_{v}$ ±5%	$\pm 0.2\%^{v}/_{v}$ ±10%
λ	±0.3%	±0.3%

Table 2. *Exhaust-gases analyzers accuracy data*

Absolute values are expressed in $\%^{v}/_{v}$ or $ppm^{v}/_{v}$ and percentage values are a percentage of the true value.

The practical measurement process of exhaust gas values is visually easy to carry out, but in real conditions rather complicated and has various influence factors which shall be tackled during the measurement and the uncertainty estimation. Some final points are given as follows.

The first group of points is linked to the vehicle's engine condition before and during the time of measurement. Those are measurement process characteristics like the engine's temperature and speed and the engine's working conditions roughness directly before measurements.

The second group of moments is related to the gas analyzer and its appliance conditions and environment – hose and filter clearness and ventilation capability.

This last group is tackled with the measurement personnel's competence – how strictly and smoothly they carry out the measurement action.

Uncertainty of measurement

General principle. For this study GUM principles were used for estimation of the uncertainty components. Component estimation was carried out using concrete measuring equipment data, method particularities and specific condition influences.

Essential for the vehicles exhaust gas parameter measurement is the fact that a considerable number of the influence factors exist. Some of them are difficult to estimate in practice, but can give an exaggerated uncertainty if not correctly limited. The number of influence factors can reach up to 5 – 7.

The main technical components involved in exhaust gas parameter measurements are the measurement method, measuring instruments, the item e.g. vehicle part correctness, external environment conditions and also the people carrying out testing. The above are components for the measurement process of uncertainty estimation.

An exact list of influence factors as components for the uncertainty is as follows:

– idle speed and working stability of engine;

– vehicle exhaust gas system's working conditions;

– gas inlet hose correct length and not broken;

– gas analyzer drift;

– gas inlet hose cleanness from the previous measurements;

– measuring instrument calibration uncertainty;

– nearby environment separation from exhaust gases;

– temperature and general cleanness of the working environment;

– testing personnel - competence to carry out procedures in a similar way.

Theoretical values. Allowed values for exhaust gas concentrations are given by legal requirements. Also exist measuring instruments accuracy data and generally approved measurement methods.

Theoretically the measurement uncertainty should be up to 3 to 5 times more than the measuring device uncertainty. Taking into account that the exhaust gas measurement procedure is complicated and using I class measuring instruments we can assume the expanded uncertainty on the probability level 95% for the measurements of CO, CO_2, HC and O_2 fraction up to 25% (percentages as a relative value) e.g. 5 × 5% from Table 2.

To estimate the relation between measuring instrument and measurement process uncertainty more correctly we will find the sensitivity coefficient of influence factors. At an initial stage we can give only a rough estimation of them and they were further controlled as a summarized value through measurement comparisons.

Proficiency testing as a tool to confirm uncertainty component values

The estimation of correctness of the theoretical results can be controlled through practical measurements as comparisons using the road vehicle inspection station measurements. To assure the comparison results for proficiency scheme ISO Guide 43 principles will be used.

Interlaboratory comparisons are conducted for various purposes and may be used by participating laboratories and other parties:

– to determine the performance of laboratories for specifics measurements and to monitor laboratories' continuing performances;

– to identify problems in laboratories and give a better quality of measurements;

– to monitor established methods;

– to provide additional confidence to laboratory clients and also to the authorities;

– identified interlaboratory differences and for the determination of systematic deviations;

– as evidence for best measurements capability estimation;

– to help development and validation of in-house methods or methods based on the normatives general requirements;

– as preventive action, an analysis of interlaboratory comparison results carried out by laboratory themselves taking account specific requirements related to measurements.

Participation in proficiency testing schemes provides laboratories with an objective means of assessing and demonstrating the reliability of the results they produce.

Proficiency testing performance mathematical model on exhaust gas measurement area. On the exhaust gas measurement area it is practically impossible to find standard values as they do not exist in moving standard engines with an exhaust system. Thus laboratory Z-values can be used calculated by equation:

$$Z = \frac{D_{\mathrm{i}} - D_{\mathrm{a}}}{s}, \tag{1}$$

where D_{a} is the participant results' mean value, D_{i} is the participant measurement value and s is the standard deviation.

The participant result by ISO Guide 43:1 is good if the Z-value is less than 1 and satisfactory if the Z-value is less than 2.

The laboratory deviation from the mean value expressed in a percentage is also important, which can be calculated by equation:

$$D_{ri} = \frac{D_i - D_a}{100\, D_a} \ \%. \tag{2}$$

Practical comparison. Practical measurements for the comparisons were carried out over the time period from April up to September 2005. Three road vehicles with ignition engines were used and Estonian inspection stations for road vehicles were involved. Some stations had few gas analyzers. For the assurance of equal conditions means giving stability of comparison object were used and to limit the extreme measurement influence factors, but some influence factors can't be equally minimized as the time period was long.

Problems which shall be taken into account and solved so that influence was on minimal level can be described as follows:

- comparison object stability assurance; the comparison period was long, up to 5 months. Comparison vehicles were also used in practice and they traveled distances up to 4,000 km. Influence was controlled using analysis results from the figures and the drift of results cannot be viewed,

- there is no possibility of obtaining a reference value for measured parameters; estimation only on a statistical basis;

- measurements were carried out by station personnel, in various ways;

- various environment influence, temperature variation from 5°C up to 20°C and ventilation of rooms was different;

- measuring instrument maintenance, can be stay influences from the previous measurement;

- the presences of outlier results; how should they defined and removed for performance estimation.

Used conditions gives a more or less real measurement situation and gives the possibility of having an average value which is on the level of expanded uncertainty for Estonian measurements in this area.

Comparison results

Exhaust gas concentration measurement results are given in Table 3 (measurements by minimal idle) and Table 4 (measurements by fast idle). As an example only one vehicle's data is shown. Using comparison measurement data the main general characteristics for uncertainty estimation are calculated.

Measured parameters						
		CO	CO_2	O_2	CH	λ
°C	min^{-1}	vol %	vol %	vol %	ppm	
92	690	0.015	15.91	0.03	41	0.999
72		0.01	14.9	0.08	60	error
81	690	0.00	14.7	0.55	8	1.025
90	690	0.003	15.17	0.03	6	1.001
86	1040	0.004	15.23	0.69	2	1.032
86	700	0.005	14.98	0.15	25	1.005
81	700	0.007	15.29	error	21	1.067
80	930	0.01	16.8	0.02	15	1.000
82	700	0.009	15.58	0.17	10	1.007
66	720	0.01	14.7	0.24	1	1.010
93	700	error	15.36	0.08	68	0.999
81	690	0.00	14.7	0.13	10	1.005
Mean		0.007	15.277	0.197	22	1.014
Median		0.007	15.20	0.13	13	1.005
Min		0.003	14.70	0.02	1	0.999
Max		0.015	16.80	0.69	68	1.067
Rectangular distribution, standard deviation, s						
		0.003	0.5833	0.186	19	0.019
Normal distribution, standard deviation, s						
		0.005	0.6077	0.222	22	0.021
Realistic deviation ±2s						
		0.010	1.100	0.40	40	0.040

Table 3. *Exhaust gas parameter measurement initial results and calculated values for vehicle No 1; minimal idle*

The characteristics found are: general mean $\overline{X}$, median *Me*, minimal and maximal values in the set (except outliers), standard deviation s_{ra} if expected would be rectangular distribution, standard deviation s_n if expected would be normal distribution and estimated expected deviation (realistic through practice value) Δ_{rp} on probability level P = 95% taking into account both standard deviations. For Δ_{rp}

the expert estimation was used finding a value taking into account standard deviations and min and max values related to the mean value.

Measured values						
C	p/min	CO	CO2	O2	CH	λ
		vol %	vol %	vol %	ppm	vol %
93	2550	0.005	15.90	0.03	13	1.000
82	2300	0.02	14.9	0.07	14	1.002
83	2500	0.00	14.6	0.53	8	1.025
92	2540	0.006	15.18	0.03	7	1.001
86	2790	0.004	15.23	0.70	1	1.032
93	2570	0.009	15.05	0.05	28	1.000
96	2480	0.004	15.39	1.41	14	1.062
94	2340	0.016	16.72	0.01	17	0.999
99	2390	0.01	15.52	0.10	6	1.004
69	2560	0.02	14.7	0.10	1	1.010
94	2500	0.047	15.48	0.00	72	0.995
83	2550	0.01	14.7	0.07	10	1.002
Mean		0.0126	15.28	0.258	15.9	1.011
Median		0.010	15.21	0.070	12	1.002
Min		0.00	14.60	0.00	1.00	1.00
Max		0.047	16.72	1.41	72	1.062
Rectancular distribution, standard deviation, s						
		0.0131	0.589	0.392	19.7	0.018611
Normal distribution, standard deviation,s						
		0.0126	0.596	0.426	19.1	0.019522
Realistic deviation, ±2s						
		0.025	1.10	0.80	40	0.039

Table 4. *Exhaust gas parameters measurement initial results and calculated values for vehicle No 1, fast idle*

Based on the above results average characterizing parameters taking into account all three vehicles are calculated. These summarized results are given in Table 5.

Vehicle	Idle	Realistic deviation				
		CO	CO_2	O_2	CH	λ
		$\%^{v}/_{v}$	$\%^{v}/_{v}$	$\%^{v}/_{v}$	$ppm^{v}/_{v}$	
1	Min	0.037	0.650	0.19	15	0.008
1	Fast	0.050	0.550	0.12	13	0.006
2	Min	0.015	2.200	0.10	13	0.004
2	Fast	0.010	2.100	0.07	12	0.002
3	Min	0.010	1.100	0.40	40	0.040
3	Fast	0.025	1.100	0.80	40	0.039
Realistic deviation, ±2Δ for 3 vehicles						
		0.030	1.500	0.20	20	0.010

Table 5. *Exhaust gases concentration measurement comparison summary results as realistic deviation for 3 vehicles*

In the end part of Table 5, probable realistic values assured using the comparison expanded uncertainty of exhaust gas concentration measurements are shown.

Table 6 shows the theoretical and realistic deviations values. The theoretical value is found by multiplying gas analyzer accuracy parameter by 5.

As a conclusion we can show that the realistic measurement deviation is 2 – 3 bigger than class I measuring instrument accuracy parameter.

	CO	CO_2	O_2	CH	λ
	$\%^{v}/_{v}$	$\%^{v}/_{v}$	$\%^{v}/_{v}$	$pm^{v}/_{v}$	
Realistic in practice, $2\Delta_{rp}$	0.030	1.500	0.20	20	0.010
Theoretical	0.300	2.500	0.50	60	0.015

Table 6. *Exhaust gases concentration measurement theoretical and assured through comparison realistic deviations*

In the four figures are shown measurement data from Table 4. As an example only one vehicle's data is shown but the other three vehicle's results look similar.

Figure 1 shows CO, Figure 2 shows O_2, Figure 3 shows CO_2 and Figure 4 HC concentration measurement results. The Horizontal axis shows body code and also time.

No systematic influence factor was found for the above case; results have a random up and down movement.

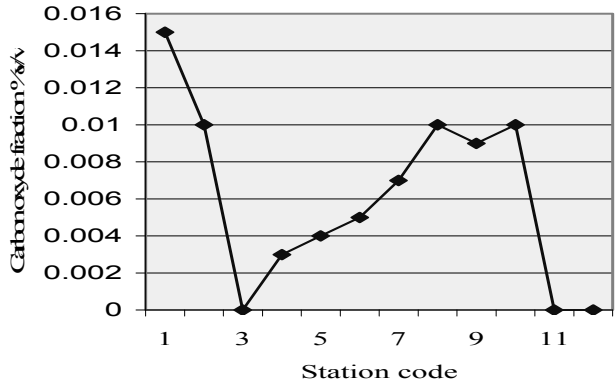


Figure 1. *Carbon monoxide concentration measurement results for vehicle No 1*

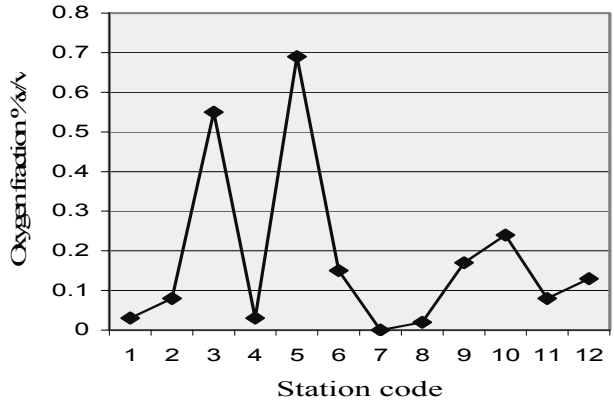


Figure 2. *Oxygen concentration measurement results for vehicle No 1*

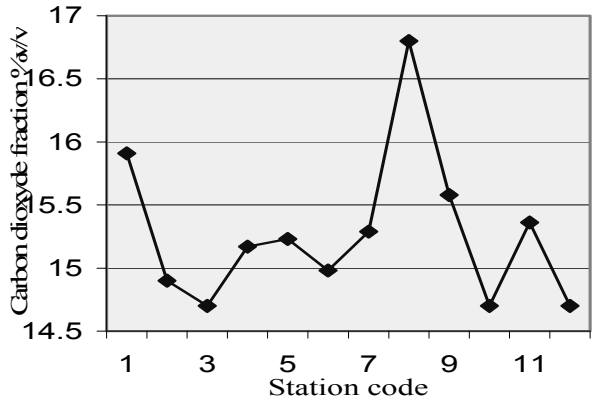


Figure 3. *Carbon dioxide concentration measurement results for vehicle No 1*

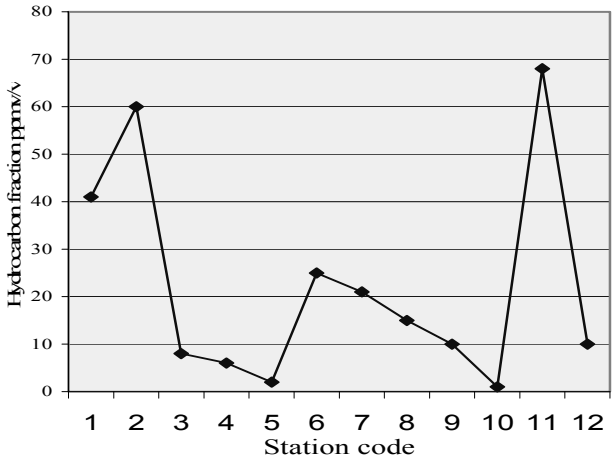


Figure 4. *Hydrocarbon concentration measurement results for vehicle No 1*

Conclusion

Vehicle exhaust gas concentration measurement has some specific moments and those shall be carefully included in the uncertainty estimation. This study gives uncertainty values and confirms uncertainty estimation results through practical measurement comparison. The study shows that expected theoretical uncertainty is a good take 3 times bigger than the measuring device uncertainty.

Results of this study were used by Estonian verification laboratories for improving its procedures and quality of work.

References

[1] E.Kulderknup, J.Riim, V.Krutob. Uncertainty of braking parameters measurements by vehicles technical inspection. IMEKO XVIII World Congress. Proceedings. Rio de Janeiro: 2006.

[2] Charles White. Avtomobilnõje dvigateli. Sistemõ upravlenia i vprõska topliva. Rukovodstvo (in Russian). Alfamer. St. Petersburg, 2001

This research was carried out with help of Estonian Ministry of Education Science Grant No 142506s03 and Estonian Science Foundation Grant No 6172

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Dimensional Metrology and Uncertainty

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Behavior of touch trigger dynamic probes: correction and calibration methods

Thierry Coorevits, François Hennebelle, Benjamin Charpentier

Laboratoire de Métrologie et Mathématiques Appliquées (L2MA)
Ecole Nationale Supérieure d'Arts et Métiers – C.E.R de Lille
8, boulevard Louis XIV – 59046 Lille Cedex, France

ABSTRACT: We propose a model of three-dimensional behavior of the dynamic probes. This model uses the hypothesis that for at least one contact point, the effort becomes zero when the probe records the measure. An experimental verification shows the quality of the model. We then propose a new model which combines an experimental 2D measure with a very simple theoretical equation. This model constitutes a method of correction and allows the creation of a new method of calibration.

Introduction

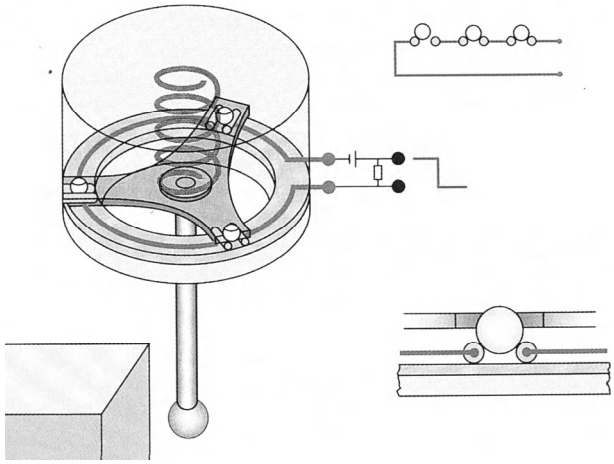


Figure 1. *Principle of the touch trigger probe*

A three-dimensional measuring machine (MMT) positions a probe in space in relation to the part to be measured. We will only focus on the mechanical technologies consisting of placing a ball, generally a red ruby, on the part.

There are two basic measurement strategies, the dynamic measurement also called on the fly measurement and static measurement.

On the fly measurement consists of approaching the ball at a constant speed, or zero acceleration, and detecting the contact point between the part and the ball in the form of a reading cue of measurement rules of MMT axes movement. Measuring with zero acceleration cancels out inertia efforts.

Static measurement consists of placing the ball on the part and measuring while the machine is stopped or moving very slowly. These measurements are reserved for another probe technology that will not be covered here.

Regardless of the technology, the ball describes a surface that is parallel to the part at a distance equal to the ball radius. Correct knowledge of this radius is therefore key in this technology.

To achieve on the fly measurements, the basic technology is described in Figure 1. It is a probe based on a symmetry 3 link, called Boyce. This type of technology is the largest in the industry. This technology evolves by integrating sensitive elements (piezoceramics, semi-conductor gauges etc.).

The goal of this chapter is to describe the behavior of probes based on a Boyce link.

Operation principle

The Boyce link consists of 6 contact points regularly distributed in a circle in 3 groups of 2 points. A traditional development consists of 6 balls on which we place a three branch star made up of three cylinders at 120 degrees from one another.

An electric circuit is established between the balls and cylinders (Figure 1). When the ruby probe ball is not placed on the part, the star is in isostasy, all 6 contacts are closed. Circuit resistance is then of a few ohms. Imagine the ruby ball getting closer to the part. The measurement speed suggested by Renishaw for these type TP2 or TP6 probes is of 488mm/mn or approximately 8 micrometers per millisecond. From the moment where the contact starts, the star is in hyperstatic mode. Certain contact efforts between balls and cylinders increase and others decrease. Insofar as contact points are electrically connected in series, we can wonder what the resistance of the evolution is.

In fact, contact point efforts evolve in a linear way according to the contact effort between the ball and part but contact resistances do not evolve linearly and the resistance increases globally. At the end of the cycle, when MMT stops, the probe is completely open and resistance is infinite. The probe works as a switch.

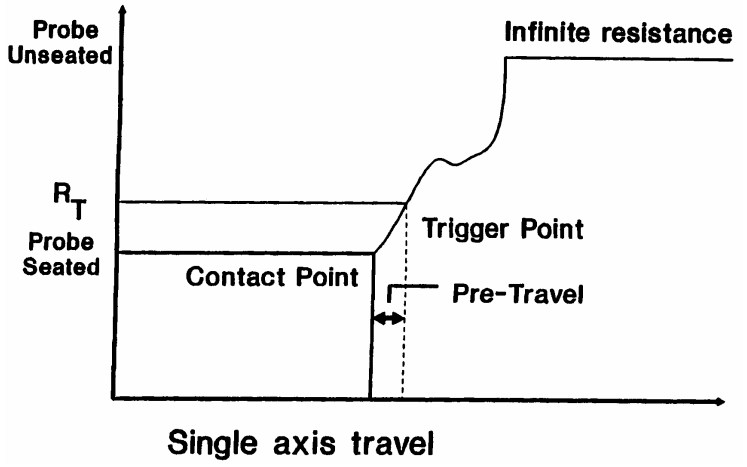


Figure 2. *Evolution of the resistance (doc Renishaw)*

When the resistance is infinite, at least one of the contact points moves and the position of the star is no longer controlled. The idea is to follow the evolution of resistance and to decide on a threshold value. In this way, when the probe detects contact, no contact is interrupted, the star position remains under control. Consequently, we cannot establish a purely mechanical rigorous model necessarily based only on the idea that a contact is interrupted.

The distance travelled between the contact and the trigger is called pre-travel (Figure 2). A classic pre-travel value is a few micrometers, or at 8 micrometers per millisecond of measurement speed, phase time described here is of a fraction of a millisecond.

Pre-travel is linked to all probe distortions during triggering, the main ones are stylet bending, star and contact distortions at the ball level, especially at the 6 contact points. Models generally integrate stylet bending. Wozniak and Al [1] consider a small star movement during contact but without clearly connecting it to Hertz type distortions or to star distortions. And yet, these distortions are significant because the star is small and far from the measurement point, there is therefore a slight effect of amplification. They also integrate contact distortion between the ruby ball and the part but the contribution is very small.

In a practical viewpoint, the fitting quality between the star and stylet (Figure 3) is very important because that is where the bending moment is at its most significant. The smallest machining drip or the smallest impurity has large consequences.

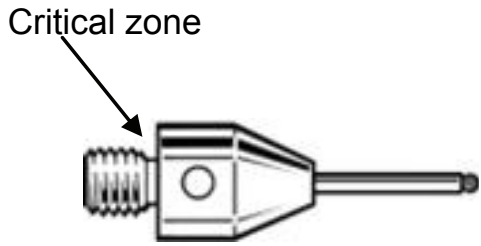


Figure 3. *Quality of fitting*

Three-dimensional theoretical model

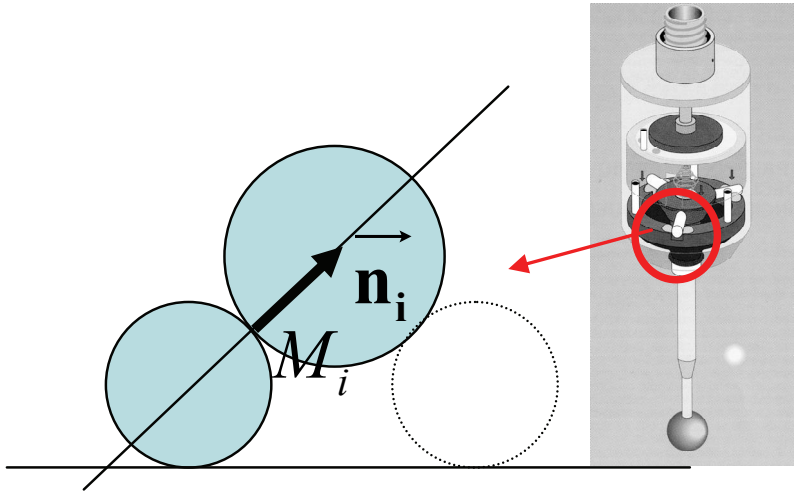


Figure 4. *Probe modeling*

We isolate the star with three branches placed on 6 contact points characterized by M_i and the average $\vec{n_i}$ such that $\left\|\vec{n_i}\right\| = 1$ for i={1.6}. We suppose the absence of contact rubbing, therefore the effects are noted as $\vec{F_i} = F_i.\vec{n_i}$. The contact point of the spring is noted as O and the corresponding effort is noted as $\vec{T}$. The contact point between the ruby ball and the part is noted as C and the contact effort is noted

as $\vec{P}$. We decide the writing of the balance of isolated solid at point 0, we then have:

$$\begin{cases} \sum_{i=1,6} F_i.\vec{n_i} + \vec{T} + \vec{P} = \vec{0} \\ \sum_{i=1,6} \overrightarrow{OM_i} \wedge F_i.\vec{n_i} + \overrightarrow{OC} \wedge \vec{P} = \vec{0} \end{cases}$$

Projections provide 6 equations. We have 7 unknowns, contact efforts F_i and effort P if we set the contact direction. The count corresponds to a hyperstatic system. To solve the system, we make the hypothesis that one of the contact points detaches with the limit described. The problem is that physically there is a configuration change during detachment but, mathematically, the equations correspond to a bilateral contact and are linear. The idea is to calculate the 6 efforts F_i for $P=0$ on one hand and for a sufficiently large P value to ensure detachment, for example, $P=10N$.

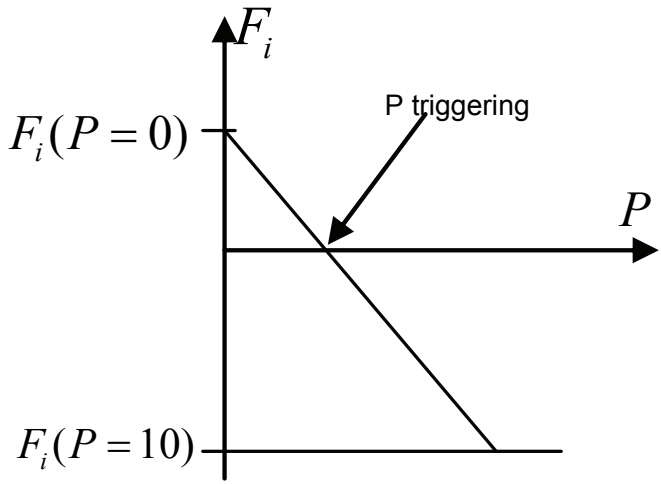


Figure 5. *Detection of the least loaded support*

Since for *P*=0 all supports are loaded in the same way, we can test the support corresponding to the most negative "contact effort" for $P=10N$. The support detaches the first one (Figure 5). A simple rule of thirds provides the corresponding *P* effort.

If we suppose that the stylet is a fixed beam that is l long, with a constant quadratic moment I and with Young's modulus E then we have the pre-travel evaluated by the following formula:

$$\mathrm{Pr} = \frac{P.l^3}{3EI}$$

In this formula, if the stylet has a diameter d, we have $I = \frac{\pi.d^4}{64}$. Pre-travel is very sensitive to the choice of l and d. For example, choosing d=1 mm instead of d=1.5 mm comes down to multiplying pre-travel by 1.5^4 or approximately 5. When the operator chooses a stylet, there is a major influence on measurement uncertainty. Figure 6 is a result of calculation in the equatorial plane of the ball, i.e. for a slicing effort in relation to the stylet. The 6 points at the center of the figure correspond to the position of contact points. The other points describe the effort necessary for the star tilt. The spring adjustment effort is 1N. The stylet is 30 mm long.

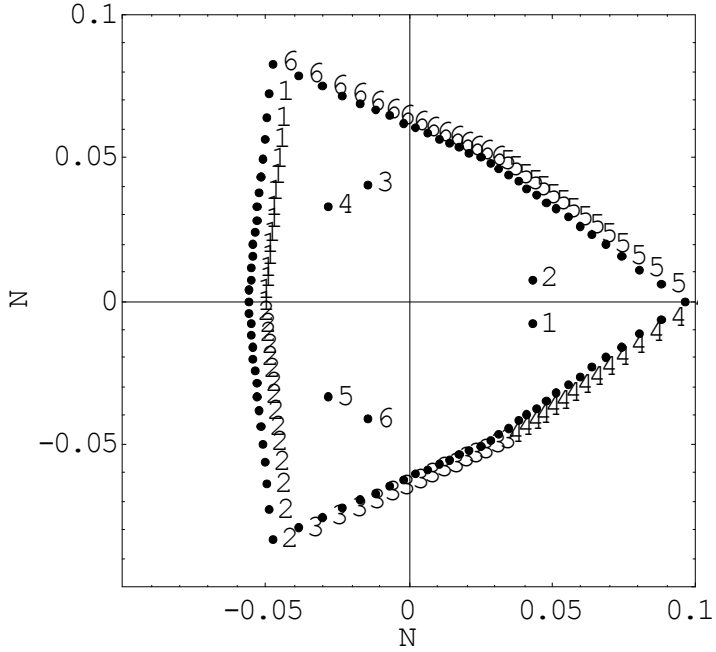


Figure 6. *Detachment effort for a TP2 Renishaw*

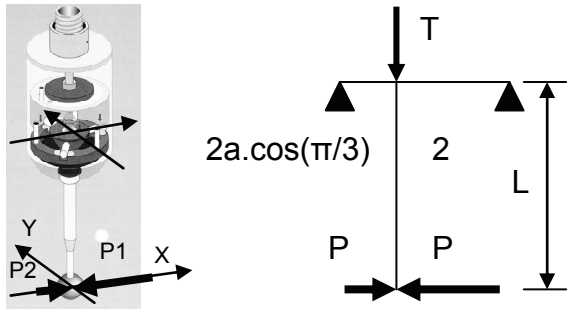


Figure 7. *Maximum and minimum triggering efforts*

Figure 7 corresponds to detachment efforts in the direction of the abscissa in Figure 6. For effort P1, the approximate balance equation is written as: $2a.T = L.P1$. For P2 effort, the equation is written as: $a.T = L.P2$. The ratio between maximum and minimum effort provided by the simulation program is approximately 1.7.

Experimentally, Figure 8 is a typical result with a ration between the longest and shortest arrows of 1.8. For this measurement, gaps were re-centered to least squares, the symmetry of the measured points gives the same weight to each point.

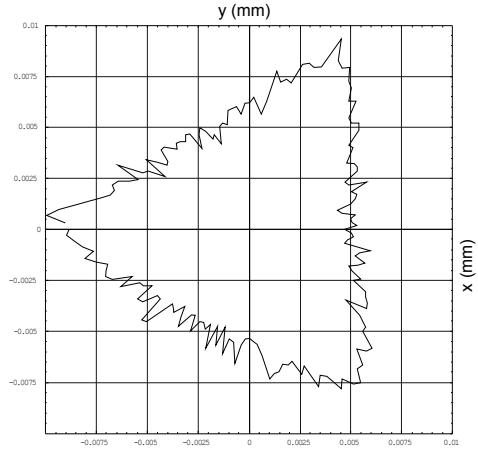


Figure 8. *Measurement in an equatorial plane*

Three-dimensional mixed model

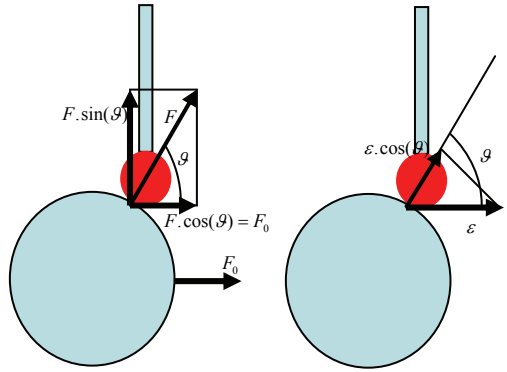


Figure 9. *Analysis of the 3D measurement*

It is easy to measure stylet bendings in the ball's equatorial plane. Going into 3D requires high point density. The idea is to replace this measurement by a simple analysis of the probe behavior.

Figure 9 illustrates on one hand that the effort following the average (we ignore rubbing) can be broken down into a slicing effort and a compression effort. The compression effort has an insignificant influence on the stylet distortion. The slicing effort tilting the star is the same as in the equatorial plane, therefore bending is the same. On the other hand, the stylet arrow is not seen in real size, we obtain $\varepsilon.\cos(\vartheta)$. At the pole, the stylet bending is therefore zero.

To experiment with this very simple model, we use the previous acquisition and we find a plane corresponding to a maximum arrow on one side and a minimum arrow on the other side (Figure 10) by analyzing the gradient in Figure 8.

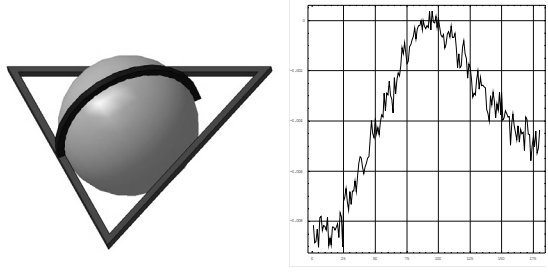


Figure 10. *Measurement in a preferred star plane*

We obtain the result in Figure 10 after result updating. In fact, the arrow is almost zero when the stylet is solicited in pure compression and we know that in the plane studied, the arrow ratio is 1.8.

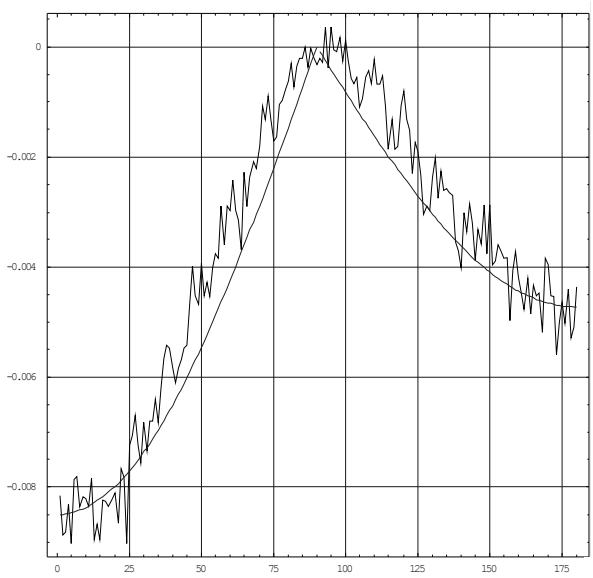


Figure 11. *Cosine model*

Figure 11 illustrates the cosine model which provides an excellent approximation of the probe behavior.

We can improve the model by noting that component $F.\sin(\vartheta)$ tends to decrease the pre-constraint created by the spring. The minimum tilt effort equals: $a.(T - F.\sin(\vartheta)) = L.F.\cos(\vartheta)$ or $a.T = (a.\sin(\vartheta) + L.\cos(\vartheta))F$ hence $L.F\cos(\vartheta) = \dfrac{L\cos(\vartheta)}{a.\sin(\vartheta) + L\cos(\vartheta)} a.T$. The term $\dfrac{L\cos(\vartheta)}{a.\sin(\vartheta) + L\cos(\vartheta)}$ appears as a multiplying coefficient of the arrow. Figure 12 is an adjustment with this new model.

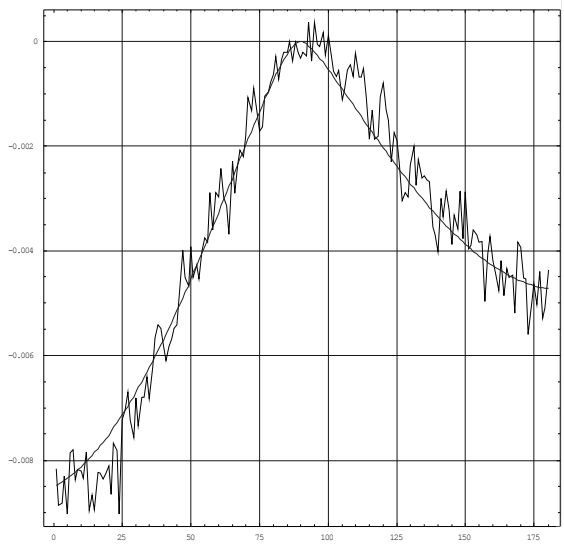


Figure 12. *Adjustment of 3D behavior*

This model is better than the previous one since it accounts for the fact that at the pole, the star is solicited in pure compression. The cosine model simply corresponds to the consideration that distance a is insignificant compared to L.

Conclusion

The theoretical three-dimensional model well accounts for the behavior of touch trigger probes but the complexity of stylet geometry and the real electric behavior makes its use tricky. A 2D measurement in an equatorial plane is simple and with the proposed mixed model going to 3D is simpler and the quality remains excellent.

This model constitutes an efficient method for the correction of stylet bending.

In the absence of the model, we can propose a calibration method of the probe responding to two criteria. On one hand, we must balance uncertainties of the determination of the different parameters describing the measurement ball position and its radius and on the other hand, we must minimize the gap based in the maximum average linked to the probe behavior.

References

[1] M. Dobosz, A. Wozniak. CMM touch trigger probes testing using a reference axis. *Precision Engineering* 29 (2005) 281–289.

[2] A. Wozniak, M. Dobosz. Influence of measured objects parameters on CMM touch trigger probe accuracy of probing. *Precision Engineering* 29 (2005) 290–297.

[3] A. Wozniak, M. Dobosz. Metrological feasibilities of CMM touch trigger probes. *Part I: 3D theoretical model of probe pretravel. Measurement* 34 (2003) 273–286.

[4] M. Dobosz, A. Wozniak. Metrological feasibilities of CMM touch trigger probes. *Part II: Experimental verifcation of the 3D theoretical model of probe pretravel.* Measurement 34 (2003) 287–299.

[5] Y. Shen, S. Moon. Investigation of point-to-point performance test of touch trigger probes on coordinate-measuring machines. *Robotics and Computer Integrated Manufacturing* 17 (2001) 247-254.

[6] W. Estler and Al. Practical aspects of touch-trigger probe error compensation. *Precision Engineering* 21:1-17, 1997.

[7] W. Estler and Al. Error compensation for CMM touch-trigger probes. *Precision Engineering* 19:85-07, 1996.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Angular measurements at the primary length laboratory in Portugal

F. Saraiva, S. Gentil

Laboratório Central de Metrologia, Instituto Português da Qualidade
Rua António Gião 2, 2829 – 513 Caparica, Portugal;
e-mail: fsaraiva@mail.ipq.pt; sgentil@mail.ipq.pt

ABSTRACT: This paper describes the work performed by the Primary Length Laboratory (LCO) of the Central Metrology Laboratory (LCM) in one of the dimensional metrology fields, the angular field.
At Portuguese Institute for Quality (IPQ), the possibility to perform the practical realization of the definition of the radian is based on the circle subdivision principle.
In the laboratory this aim is obtained using several angular artefacts (index tables, polygons, autocollimator) and following international standards and specific proceedings.
With this work, LCO can offer the reference values and dissemination of the unit to the Accredited Calibration Laboratories, regional directions, the Portuguese Army and industry, in order to assure the traceability in the calibration of their working standards and to the customers.
This paper also describes and presents the results regarding the calibration of an index table, using the cross calibration method, with two different procedures, one using a second index table and the other using a 12-sided polygon.

Introduction

The Primary Length Laboratory (LCO) of the Central Metrology Laboratory (LCM) is responsible for the development of the length and plane angles national metrology standards. In order to implement the Angular Metrology field, LCO has use since 1998, an Autocollimator (optical system used to measure small angular differences), an Index table (allows the generation of angular positions), polygons and angular gauges.

With these systems the LCO execute calibrations of several angular standards, such as angle gauges, polygons, index tables, pentaprism, optical square and angular optics for interferometric systems.

In angle metrology, the LCO gives answers to sectors like the Accredited Laboratories, Portuguese Army, regional directions and industry, guaranteeing the traceability of the angular measurements in Portugal.

SI unit

The derived unit of the International System of Units (SI) of the quantity Plane Angle is the radian (symbol: rad).

The radian is defined as:

The plane angle situated between two radius that, in a circumference of a circle, intersects an arc length equal to the radius of the mentioned circle (Resolution 12 of the 11th General Conference on Weights and Measures (CGPM), 1960).

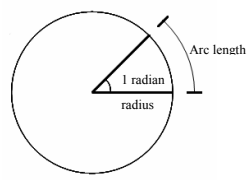


Figure 1. *Angle measuring 1 radian*

The ratio of a circle's circumference to its diameter is π.

Since π is an irrational number, for practical angular measurements in engineering the units used belong, almost exclusively, to the sexagesimal system. This system dates back to the Babylonian civilization and is composed by the grade (º), minute (') and second ('').

1 degree (º) = 1/360 of a complete circle

1 minute (') = 1/60 of a degree

1 second ('') = 1/60 of a minute

Since the complete circle corresponds to 2π rad:

360º = 2π rad

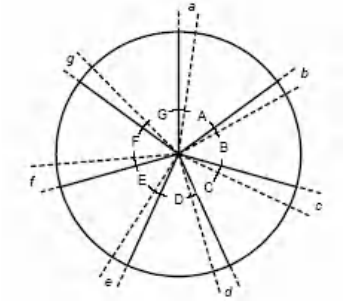
1º = (π/180) rad

1' = (1/60)º = (π/10 800) rad

1'' = (1/60)' = (π/648 000) rad

Realization of the Radian Definition

The realization of the radian definition is made in terms of the circle subdivision. As the sum of all intermediate angles of a circle is 360º and the sum of all errors of the same angles is zero, it is possible to produce matrices and equations to determine the error of each intermediate angle of the circle.



A+B+C+D+E+F+G=360º;
a+b+c+d+e+f+g =0º

Figure 2. *Circle-closing principle [1]*

This method is based on the circle-closing principle. The Circle-closing principle is one of a number of self-proving comparator techniques employing multiple measurements together with suitable rearrangements of the components of a measurement system [2]. This is the method used at the Length Laboratory to calibrate polygons and index tables using as reference an autocollimator.

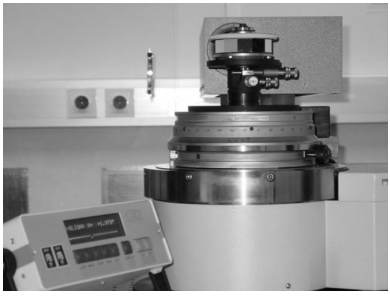


Figure 3. *Set up used in the cross calibration against a polygon*

Measurement system and standards of angle at LCO

To perform the practical measurements of angular standards, the LCO has the following systems.

Autocollimator

The Autocollimator is a system used to detect and measure small angular deviations of a reflecting surface that is positioned in front of the lens of the autocollimator. To minimize the wavefront errors, it is important that its optical axis crosses the center of the reflecting face of the angular standard, and that the distance between the reflecting surface and the lens of the autocollimator should be the lowest possible.

The LCO have an autocollimator, Moller Wedel with two measurement axes, x, y, perpendicular to each other, with a range of 310'' and a resolution of 0.005''.

Index tables

The index table allows the generation of angular positions, dependent of their resolution.

The LCO have two index tables. The Moore index table is a mechanical system with 1440 radial serrations on each of its upper and lower elements, with a resolution of 15'.

Recently the IPQ purchased an automated Index table (Tekniker), with a resolution of 0.01'', whose system is based on a rotary encoder of high accuracy, installed on a table supported and oriented through a pneumatic support device, assuring a right rotation without friction. The rotation of the table is operated by an engine of continuous current, which transmits the movement through a mechanical friction device. The final movement, of high accuracy, will be performed using a piezoelectric actuator and a highly accurate deflection mechanism.

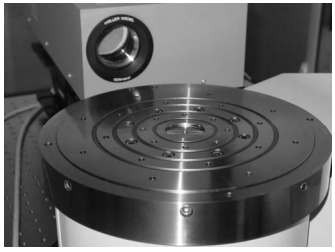


Figure 4. *Autocollimator and index table*

Polygons

Polygons are prisms with a polygonal regular section, through which angular values are materialized in a whole number of degrees, according to the number of faces. The polygons are constituted by reflecting faces whose number usually varies between 3 and 72. They are used together with optical autocollimators as angle transfer standards to calibrate some angle measuring devices.

LCO has two polygons, one with 9 faces and an other with 12 faces.

Angle gauges

A set of angle gauge blocks is another type of angular standard that enables the realization of angular units. This set is composed of gauges with different nominal angles, that when wrung together, in an additive or subtractive way allows the reproduction of different angular values.

It is important to point out that measurement angle standards do not require a primary standard, since any angle can be achieved by dividing a circle to corresponding parts.

Calibration of Angle Measurement Systems

In order to assure the traceability chain and the credibility of the results, the angular instruments/standards have to be regularly calibrated. With regards to this, LCO provides the following calibration services.

Angular gauges, pentaprisms and optical squares

The angle given by the two measuring faces of the angular gauges is evaluated using an autocollimator, AK, an index table and a levelling table. This last device is essential for minimizing the pyramidal error. The pyramidal error of an angle standard is caused by non-alignment between the measuring faces in the y-axis, due either to imperfections in the manufacture of the standard or to misalignment in the set used in the calibration [3]. To calibrate pentaprisms and optical squares, as well as the above-mentioned systems, it is necessary to use a high quality mirror.

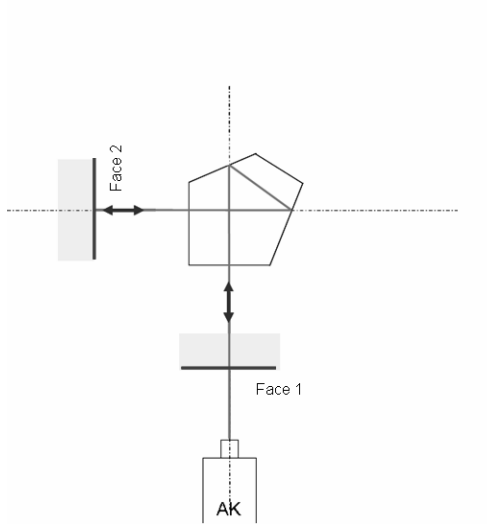


Figure 5. *Pentaprism calibration*

Polygons and index tables

For this type of calibration, the LCO uses the cross calibration method against a high quality accurate polygon.

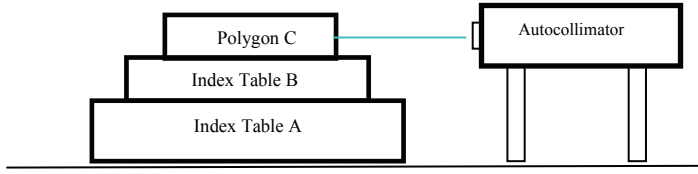


Figure 6. *Cross calibration method against a polygon*

In this process, the two coaxially aligned index tables A and B are compared in all possible relative positions. Simultaneously, this set up enables us to carry out the relative positioning of the polygon (also coaxially aligned with the A and B).

The calibration of the two index tables is performed in steps equal to the number of faces of the polygon used.

With this method “n” measurement series, a total n^2 values are obtained.

A1+B1+C1	A6+B2+C1	A5+B3+C1	A4+B4+C1	A3+B5+C1	A2+B6+C1
A2+B1+C2	A1+B2+C2	A6+B3+C2	A5+B4+C2	A4+B5+C2	A3+B6+C2
A3+B1+C3	A2+B2+C3	A1+B3+C3	A6+B4+C3	A5+B5+C3	A4+B6+C3
A4+B1+C4	A3+B2+C4	A2+B3+C4	A1+B4+C4	A6+B5+C4	A5+B6+C4
A5+B1+C5	A4+B2+C5	A3+B3+C5	A2+B4+C5	A1+B5+C5	A6+B6+C5
A6+B1+C6	A5+B2+C6	A4+B3+C6	A3+B4+C6	A2+B5+C6	A1+B6+C6

Table 1. *Example of a matrix for n = 6*

The values obtained are put in the matrix, as shown above, and by summing rows, columns and diagonals, with the condition:

$$\sum A_i = \sum B_i = \sum C_i = 0,$$

the directional deviations for A, B and C angular standards are obtained.

It is necessary to pay special attention to the convention of signs. The signal of the final results is dependent on factors such as the direction of reading of the autocollimator, the sense of rotation of index tables and the way indexed tables and polygon are numbered. The effect of all these factors should be checked in every calibration [4].

Nowadays the traceability is achieved through the calibration of the autocollimator at another primary laboratory.

Calibration of an index table

At LCO, the index table can be calibrated using two different methods, one is the above referred, “Cross calibration against a high quality precision polygon” and the other is the “Cross Calibration against another index table”:

Cross calibration against another index table

To implement this method it is necessary to have two index tables and a high quality mirror. The index table to be calibrated, index table B, is placed above index table A, being necessary to guarantee the coaxiality of the two index tables. The mirror is positioned above index table B and in front of the lens of the autocollimator. Index tables A and B are compared in all possible relative positions, positioning the mirror in every calibration step of index table B.

The mirror is placed in a levelling support to minimize the pyramidal error.

In order to implement this calibration service and to establish the proceeding to use, LCO analyzed the results obtained by the two ways of performing the calibration.

The results were the following.

Angular step (º)	Cross calibration against a polygon (″)	Cross calibration against another index table (″)
0	-0.02	0.04
30	0.23	0.34
60	0.43	0.56
90	0.25	0.26
120	-0.07	-0.08
150	-0.31	-0.29
180	-0.02	-0.08
210	0.21	0.16
240	0.27	0.17
270	-0.10	-0.11
300	-0.47	-0.52
330	-0.40	-0.44

Table 2. *Results of the two methods*

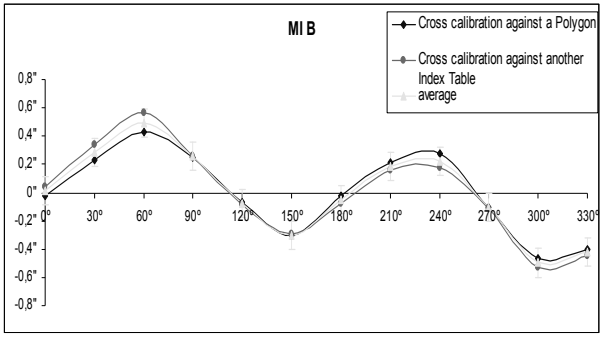


Figure 7. *Graphic of the results of the two methods, for the index table B (MI B)*

The figure shows that in the two processes similar results were achieved.

The "cross calibration against another index table" is a longer process and the expanded uncertainty was slightly higher than that achieved with the "cross calibration against a polygon" method.

Conclusion

The results allow the selection of the "cross calibration against a polygon" as the reference method to implement at LCO and to use in the process of calibration of index tables.

With the purchase of the new index table, it was also possible to reduce the uncertainty of measurement, being more significant in case of polygons and index tables.

The future participation in international comparisons allows for the validation of the calibrations methods.

The next step is to prepare theoretical and practical works for the implementation of the autocollimators calibration at LCO.

Acknowledgments

The authors would like to thank the contributions of the individuals who have contributed to this work. In particular they acknowledge Rui Saraiva and Rui Gomes and colleagues (Laboratório Central de Metrologia) for their contributions for stimulating the work in this field.

References

[1] Yandayan T., Akgoz S.A., Haitjema. "A new calibration method for polygons with a pitch angle which does not match with the subdivision of the used indexing table", *Proceedings of the 3rd Int. Conference*, euspen, pp 481-484, 2002.

[2] Estler W.T, "Uncertainty Analysis for Angle Calibrations Using Circle Closure" *J. Res. Natl. Inst. Stand.Technol*. Vol. 103, Nr 2; pp 141-151, 1998

[3] Kruger O.A,"Methods for determining the effect of flatness deviations, eccentricity and pyramidal errors on angle measurements" *Metrologia*, Vol.37, pp 101-105, 2000

[4] Hewit P.L., "Modern Techniques in Metrology", *World Scientific*,1984

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Uncertainty of plane angle unit reproduction by means of a ring laser and holographic optical encoder

M.N. Bournachev[a], Y.V. Filatov[a], D.P. Loukianov[a], P.A. Pavlov[a], E.P. Krivtsov[b], A.E. Sinelnikov[b]

[a] St.Petersburg Electrotechnical University
Prof. Popov str. 5, 197376, St.Petersburg, Russia

b D.I.Mendeleev Institute for Metrology,
Moskovski pr. 19, 198005, St.Petersburg, Russia

ABSTRACT: Development of the measurement standard of a plane angle for reproduction of the size of plane angle unit in dynamic mode is described in the report. The development is based on the dynamic goniometer with the ring laser (RL) as an angle sensor of super high resolution. The important feature of development is integration of angle sensors operating with totally different physical principles. In this purpose the holographic incremental optical encoder (OE) with 324,000 lines is included in the measurement system. The data processing in the system is carried out with the use of a cross-calibration technique providing the determination of the systematic error of the RL as well as the systematic error of the OE. Experimental research showed that the random and residual systematic errors of the angle unit reproduction are on the level of 0.01 arc-sec.

Introduction

Laser methods and means occupy one of the leading places in the development of modern measuring and control technologies of civil and military applications. The unique properties of modern lasers have stipulated their broad application in national standards of leading industrially developed countries. During recent years, a ring laser (RL) has become one of the most accurate instruments in angle measuring and particularly in dynamic goniometers [1].

The purpose of the RL implementation in the ring laser goniometer (RLG) is the same as a circular scale's use in the conventional goniometer. Owing to the features of the RL the RLG can be used for circular measurements in static and dynamic modes of operation with high speed and accuracy and the ability to automate the whole process of measurement. For example it is possible with the RLG to carry out the calibration of a circular optical encoder with a number of lines of about several thousands during a few minutes of time. Systematic error of the RLG is defined in the main by an external magnetic field influence on the RL (usually it is the magnetic field of the Earth).

Experimental research has shown that this error is on the first harmonic of the rotation frequency with an amplitude of about 0.03…0.05 arc-sec (for the RL with magnetic screen). On the other hand, this error can be taken into account and compensated algorithmically. As the main metrological characteristic, the RL offers the advantage of a nearly perfect circular division due to the laser wavelength as its base. It therefore offers the potential as an angle standard similar to the linear laser as a length standard.

Concept of reproduction of plane angle unit

Previous results [2] show that the dynamic RL goniometer should be considered as a high precision measurement system and it can be used as the basis of development of an angle unit reproduction system. The features of the RL are super high resolution and equability of the circular scale. At the same time the RL's scale factor and respectively the angular value of its division can vary slightly over time which induces the necessity of the RL's current calibration on the 2π radian angle or on some other known angles. These particularities of the RL are good reasons for development of the angle standard on the base of combination of the RL with an optical encoder using other physical principles. Optical encoders do not have such good equability of the circular scale but they are very stable in time.

Besides these advantages the usage of two different devices in the measurement system allows us to carry out the procedure of cross-calibration that provides determination of systematic errors of both devices. The cross-calibration procedure [3] consists of step-by-step calibration of one device with the use of another device. Usually after each step the stator of one of the devices is being turned in relation to the housing of the system at the angle $360°/n$, where n is the number of cross-calibration steps. The result is the n x m data array that allows to obtain, after its processing, the systematic errors of both devices.

On the basis of the reasons stated above the schematic of the angle unit reproduction can be considered with the use of the following angle measurement devices:

1. The He-Ne ring laser of monoblock design with 0.4 m resonator perimeter and corresponding scale factor about one million or 1.3 arc-sec per pulse.
2. Optical encoder with the holographic principle of circular scale formation (HOE) and with 324,000 lines per revolution. Resolution is about 4 arc-sec without interpolation [4].
3. Optical polygon (OP) together with the optical null-indicator (NI) intended for transfer of reproduced angle unit to other reference systems or for comparisons with other angle standards.

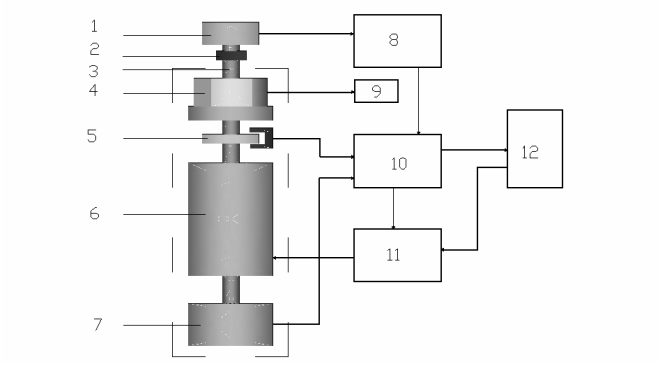


Figure 1. *1 – calibrated OE, 2 – coupling, 3 – spindle, 4 – optical polygon, 5 – holographic OE, 6 – drive, 7 – RL, 8 – OE electronics, 9 – NI, 10 – electronic unit, 11 – drive control unit, 12 – PC*

The possible scheme of the system for angle unit reproduction is shown in Figure1. On the spindle of the air-bearing the RL, the rotor of the holographic OE (HOE) and the OP are mounted. On the housing of the system, the stator of the HOE and the NI are fixed. The drive rotates the spindle with approximately constant rotation rate. The output signals of the RL, the HOE and the NI are fed to the electronic unit for their pre-processing and then to the computer (PC). On the top of the system the OE is mounted which can be calibrated by means of the whole system. This OE is connected with the spindle by means of coupling.

Experimental results

The system shown in Figure 1 was researched for determination of the random and systematic components of error. Figure 2 shows results of the HOE's calibration conducted by means of the RL. In this procedure the output signal of the HOE was divided by a digital divider with division factor equal 900. After dividing the number of output pulses of the HOE per revolution reduced to 360, i.e. the nominal angle value between pulses after the divider was equal to 1 degree. The results presented in Figure 2 were obtained for different positions of the RL in relation to spindle in correspondence to the procedure of cross-calibration. The angles of the RL's turns were equal to 60 degrees. As it mentioned above the processing of obtained data allows us to obtain the systematic error of the RL as well as the HOE.

The systematic error of the HOE can be determined by means of simple averaging of the curves presented in Figure 2. The results of such averaging are shown in Figure 3.

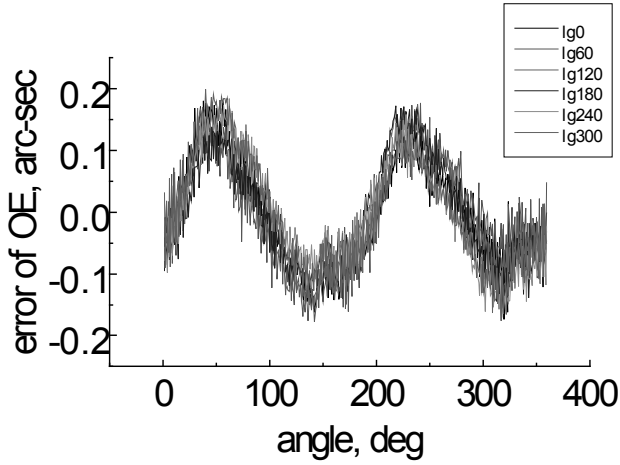


Figure 2.

It can be seen from the figure that the systematic error of the HOE is mostly on the second harmonic of the rotation frequency with the amplitude of about 0.15 arc-sec. The spectrum of the systematic error of the HOE is shown in Figure 4. It can be seen that except for the second harmonic there is also a rather strong first harmonic.

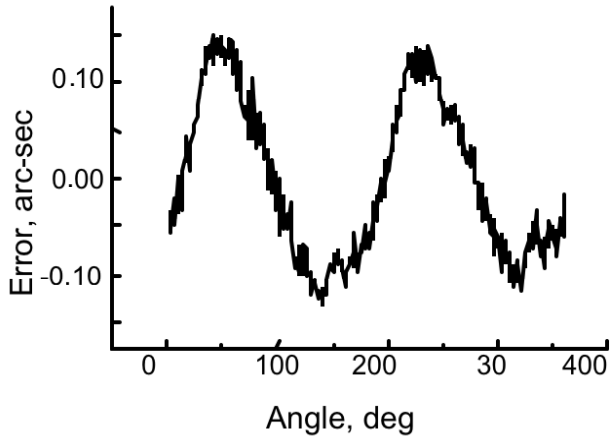


Figure 3.

The systematic error of the RL was determined by means of subtracting the curve presented in Figure3 from the curves shown in Figure2. Each of the six curves obtained after subtracting presents systematic error of the RL.

However, each of these curves has its own phase defined by position of the RL in relation to spindle during each measurement. Averaging over all six curves after phase shift yields the systematic error of the RL.

Averaging was conducted after shift of each curve in phase (second curve on 60 deg., third curve on 120 deg. and so on). Results of this processing are shown in Figure 5.

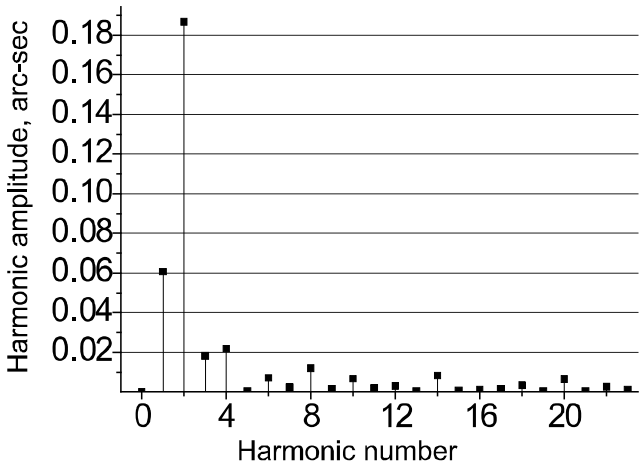


Figure 4.

From the curve in this figure it can be seen that the systematic component of the RL's error is the first harmonic of the rotation frequency with the amplitude of about 0.035 arc-sec.

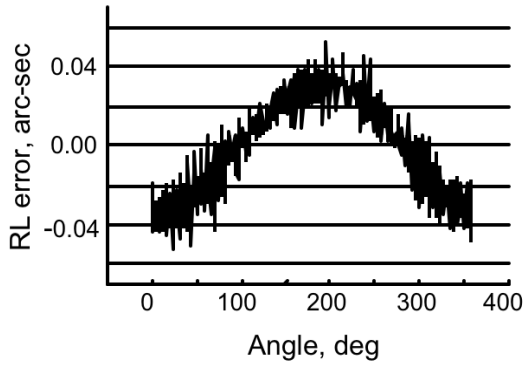


Figure 5.

The random error is at the level of 0.01 arc-sec. Fitting the obtained curve with the corresponding harmonic function allows us to take it into account and subtract algorithmically from the results of the measurements. The difference between fitting the harmonic function and curve in Figure 5 represents the residual systematic error of the RL.

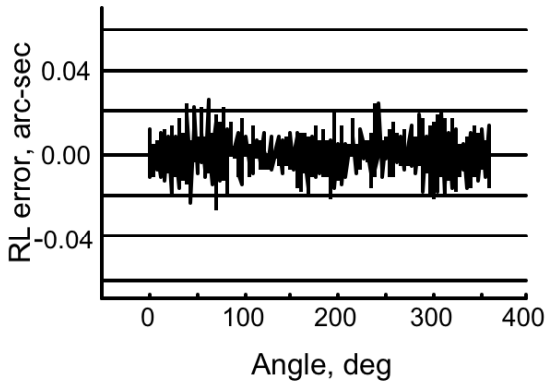


Figure 6.

This difference is shown in Figure 6. From this figure it can be seen that the residual systematic error of the RL is in the limits of 0.01 arc-sec.

Obtained results show that the error of reproduction of the angle unit by means of the described system is in the limits of 0.01 arc-sec. However, to be sure in that figure it is necessary to repeat the measurements over quite a long time (about several months) and to find that systematic component of the error (curve in Figure 5) does not change in time.

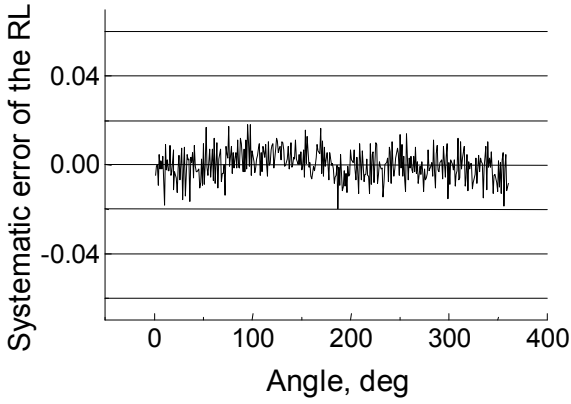


Figure 7.

Another way is the reduction of the external magnetic field, which is the main source of the systematic error of the RL. To reduce the external magnetic field we moved the experimental set-up to the room with additional shield.

The results of the measurements carried out in this room showed reduction of systematic error of the RL to value about 0.01 arc-sec. These results are presented in Figure 7.

Calibration of optical encoders

The developed system was modified for calibration of optical encoders. The skeleton diagram of the RLG intended for optical encoder (OE) calibration is shown in Figure 8. The optical polygon (OP) and the ring laser are mounted on the moving part of the RLG - rotary platform. The rotor of the OE (which should be calibrated) is connected with rotary platform by a coupling unit (CU). On the base of the RLG interference null indicator (NI) and stator of the OE are mounted. The output signals of the RL, the OE, and the NI come to interface placed inside the personal computer (PC).

The motion control system of the RLG rotary platform ensures a condition of a stabilized angular velocity in a range 0.05-10 rev/s with a possibility of a varying the direction of rotation.

There is the following principle of operation of the RLG. At rotation of a rotary platform the photoreceiver of the RL reads out an interference pattern of counter-propagating waves generated in the RL cavity. A harmonic output signal of the RL has the angular discreteness of 1.3 arc-sec per period.

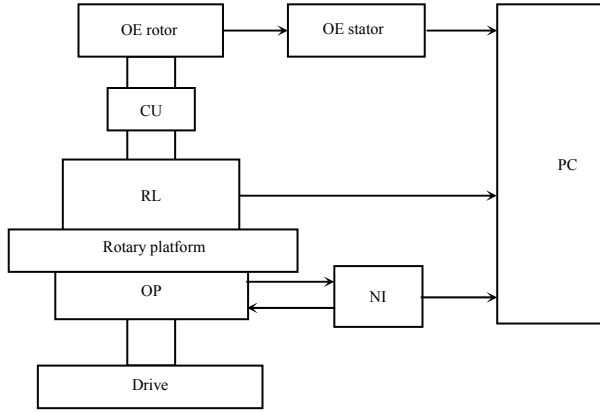


Figure 8.

To increase an angular resolution of the system it is fed to the broadband frequency multiplier with multiplication factor of 10. After this, the TTL-form output signal is fed to an information input of a pulse counter, which is controlled by one of signals of the OE, formed by a system of the analysis of OE codes. The analysis of signals, selection of a signal of the necessary code and signal of a zero code and data processing are carried out by the personal computer (PC). If there is no need to calibrate all lines of the OE, the frequency of the OE output signal is divided with the factor allowing the wishful discreteness of calibrated angular intervals. The overview of the system is shown in Figure 9.

The developed RLG was used for calibration of incremental optical encoders and absolute optical encoders with information capacity of 12…18 bits. With the use of the RLG it is possible to check the following precision and information parameters of OE:

- Systematic errors of OE;
- Information capacity of the angle channel (total number of lines);
- Correct sequence of code changes;
- Statistical characteristics of angle measurements errors;
- Dynamic errors of OE.

Using the RLG, we calibrated various OE manufactured by Heidenhain Co. (RON255, RON905, etc.), holographic encoders developed in the St Petersburg Institute of Nuclear Physics, prototypes of absolute optical encoder produced by *Avangard* Co. and others.

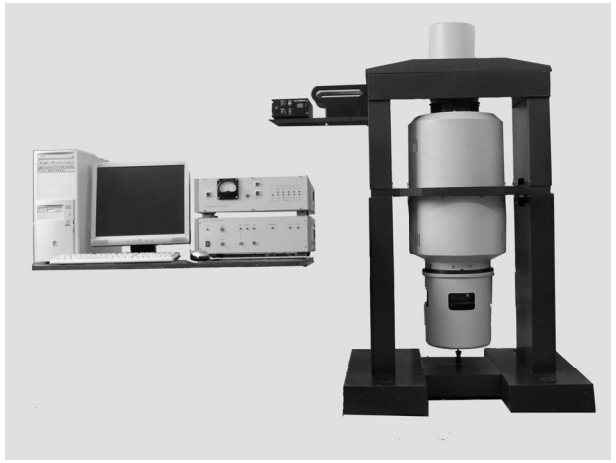


Figure 9.

As an example of the RLG implementation for optical encoder calibration we can use the results obtained during investigation of the system. These measurements were aimed at obtaining the estimation of repeatability and accuracy of the RLG in this mode of operation. The commercial device LIR-3170A of SKB IT company was examined for this purpose. This device is an incremental angle encoder of medium accuracy with 18000 signal periods in 360 deg. The accumulated error for this encoder is specified within 2.5 arc-sec. The LIR-3170A was mounted on the RLG in coaxial direct coupling with the shaft of the goniometer. The integral coupling device of the LIR-3170A allowed the influence of shaft wobbling and the non-coincidence of RLG and OE axes to be partly eliminated. Our electronic measuring system included a frequency divider for the encoder output signal, intended to reduce the number of angular intervals to be calibrated.

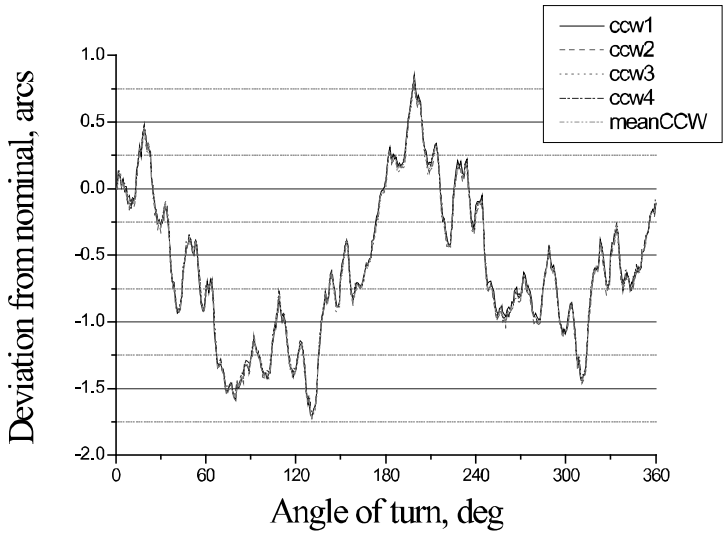


Figure 10.

A holding capacity of our instrumentation allowed us to test only 1/100 part of full set of OE lines but as we can see from Figure 10, this dose is quite sufficient for the validated assertion of this angle scale quality. In this set of testing, the rotation rate was 250 deg/s.

In Figure 10 the outcome data of calibration are shown. As we can see from the figure, in our setup a systematic error of the sensor (deviation from a nominal value) ranges within the limits of ±1.6 arc-sec. The repeatability of the measurement results is in the limits of ±0.05 arc-sec.

Conclusion

The use of the system with the RL and the holographic OE allows reproduction of the angle unit with systematic and random errors about 0.01 arc-sec. One of the limiting factors that enable us to come to errors at the level of several thousandths of arc-sec is the random error of the RL. It is possible to reduce this error using the RL with larger dimensions. The usage of the RL with a perimeter of about 1.5 m (side about 40 cm) will allow us to obtain the random error about 0.001…0.003 arc-sec, to determine the systematic error with the same accuracy and to subtract it.

References

[1] “Requirements and recent developments in high precision angle metrology", Proceedings of 186 PTB-Seminar, 5 November 2003, Braunschweig, Germany.

[2] Filatov Yu.V., Loukianov D.P., Probst R., 1997, *Dynamic angle measurement by means of a ring laser*, *Metrologia*, 1997, Vol. 34, pp. 343–351.

[3] Sim P.J., in “Modern techniques in metrology” (edited by P.L.Hewitt), Singapore, World Scientific, 1984, pp.102–121.

[4] Gordeev S.V., Turukhano B.G. *Opt Laser Technol*, 1996, Vol. 28, pp.255-261.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Advanced 2D scanning: the solution for the calibration of thread ring and thread plug gauges

R. Galestien

IAC Geometrische Ingenieurs B.V.
www.iac-instruments.com
Jules Verneweg 13-17 – 7821AD Emmen, The Netherlands
Tel.: +31 591 644103 – fax: +31 591 648064 – reginald.galestien@iac-instruments.com

ABSTRACT: The calibration of thread gauges by means of advanced two dimensional scanning methods is meeting in a practical way the latest requirements of guidelines such as VDI 2618 Parts 4.8 and 4.9
Calibration techniques as performed on one axis measuring machines in combination with balls or wires used to give unreliable results.
Just measuring the distance over wires or between balls is not enough for a good determination of the pitch diameter of a thread gauge if the other thread parameters are assumed to have nominal values.
Well known calculation methods by Berndt for wires or balls, recommended by EA as the most precise formulae for the calculation of pitch diameter, are suffering accuracy if the actual values of lead, pitch and flank angles are different from the values used in the calculation.
Thus it is clear that pitch and flank angles should also be measured. This is exactly what is in the calibration scope of the newest guidelines.
With conventional equipment the calibration of thread gauges is becoming very elaborate and time consuming. However, after fulfilling all these requirements the influence of straightness deviations of the flanks due to wear or poor manufacturing is not covered in the pitch diameter calculations.
The conclusion is that the determination of the pitch diameter with wires and balls cannot give optimal accurate results and is not economical anymore.
For this reason it is preferred to measure the pitch diameter or simple pitch diameter with a method that takes care of all mentioned criteria.
This method was initially developed in 1992 by IAC Geometrische Ingenieurs for the automatic post process inspection of gas generators for airbags.
One of the important requirements for acceptance by TÜV was the in line measurement of all the parameters for each thread including the straightness of the flanks of Buttress threads with a capacity of one airbag part per 30 seconds.
IAC solved this problem successfully with the development of tactile 2D thread scanners.
The same technology is now applied in the MasterScanner for the calibration of thread gauges. The complete two dimensional intersection of the surface of the gauge with a mathematical plane through the reference axis is composed by the sequential scanning of two opposite thread contours by a probe with two styli. Direct after finalizing the scanning process the MasterScanner calculates and presents the parameters: effective diameter, simple pitch diameter, virtual diameter, major diameter, minor diameter, pitch, flank angles, straightness deficiencies of each flank, taper, etc.
Special calibration methods are developed to guarantee the traceability of the results.

Introduction

Definition of simple pitch diameter:

The thread groove diameter of an imaginary cylinder the surface of which passes through the thread profiles at such points as to make the width of the thread groove equal to one-half of the basic pitch [4,5].

Definition of pitch diameter:

The diameter of an imaginary cylinder, which intersects the surface of the thread profile in such a manner as to make the width of the thread groove equal to one-half of the *basic* (nominal) pitch [6].

The effective pitch diameter and the simple pitch diameter of thread gauges are measured traditionally with the distance m over wires and between balls (see Figure 1).

However, the effective diameter is not only based on the distance m as can easily be seen in the formulae in Figure 2. In the latest guideline VDI 2618 (2006) parts 4.8 and 4.9, it is now compulsory to also measure pitch and angles for the first inspection of thread gauges [2,3].

The most advanced formulae for calculation of the pitch diameter d2 or D2 on the basis of the measured value m are developed by Professor Berndt.

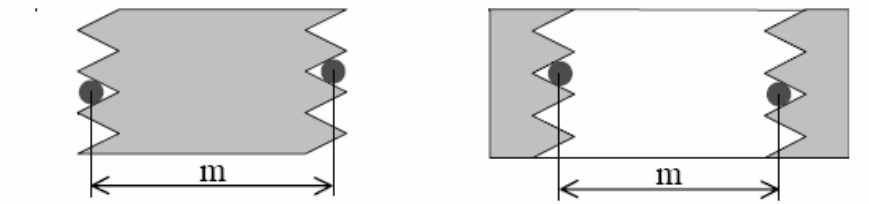


Figure 1. *External thread with wires and internal thread with balls*

It is recommended by EA to make use of these formulae [1].

$$d_2, D_2 = m \cdot \cos\theta \; \mathrm{md_D} \cdot \frac{\cos\frac{\beta-\gamma}{2}}{\sin\frac{\beta+\gamma}{2}} \cdot \sqrt{1 - \frac{m^2 \cdot \sin^2\theta}{{d_D}^2 \cdot \cos^2\left(\frac{\beta-\gamma}{2}\right)}} \pm \left(\frac{l}{n} - \frac{2 \cdot l \cdot \theta}{\pi}\right) \cdot \frac{\cos\beta \cdot \cos\gamma}{\sin(\beta+\gamma)}$$

$$\arcsin(\theta_k) = \frac{d_D \cdot \ell}{\pi \cdot m^2} \cdot \frac{\cos\beta \cdot \cos\gamma \cdot \cos\frac{\beta-\gamma}{2}}{\cos\frac{\beta+\gamma}{2}} \cdot \frac{\sqrt{1 - \frac{m^2 \cdot \sin^2\theta_{k-1}}{{d_D}^2 \cdot \cos^2\left(\frac{\beta-\gamma}{2}\right)}}}{\cos\theta_{k-1} \mp \sin\left(\frac{\beta+\gamma}{2}\right) \cdot \cos\left(\frac{\beta-\gamma}{2}\right) \cdot \frac{d_D}{m} \cdot \sqrt{1 - \frac{m^2 \cdot \sin^2\theta_{k-1}}{{d_D}^2 \cdot \cos^2\left(\frac{\beta-\gamma}{2}\right)}}}$$

Figure 2. *The Berndt formulae*

The problems with the 2-Ball and 3-Wire measurement

However, just a value of m is not good enough.

Even the well known calculation methods by Berndt for wires or balls, are suffering accuracy if the actual values of lead, pitch and flank angles are different from the values used in the calculation. Thus, it is clear that for accurate results pitch and flank angles should be also measured.

Furthermore, an assumption by Berndt is that the actual profiles have perfect straight flanks.

In most cases this not due to wear or poor production quality.

In the case of straightness deficiencies of the flanks the formulae by Berndt are not valid anymore.

The conclusion is that the measurement of the pitch diameter with wires and balls cannot give optimal accurate results and is not economic anymore. The actual shape including the form deficiencies of the thread profiles is playing an important role in both mentioned definitions of pitch diameter and simple pitch diameter and should also be taken into account.

The solution: 2D scanning

The 2D scanning method was initially developed in 1992 by IAC Geometrische Ingenieurs for the automatic post process inspection of gas generators for airbags.

One of the important requirements for acceptance by TÜV was the in line measurement of all the parameters for each thread including the straightness of the flanks of Buttress threads with a capacity of one airbag part per 30 seconds.

IAC solved this problem successfully with the development of advanced tactile 2D scanning techniques.

Further development of these techniques made possible the application for the calibration of thread gauges. The developed instrument obtained the name: IAC MasterScanner (see Figure 3).

IAC has been manufacturing MasterScanners since 1995.

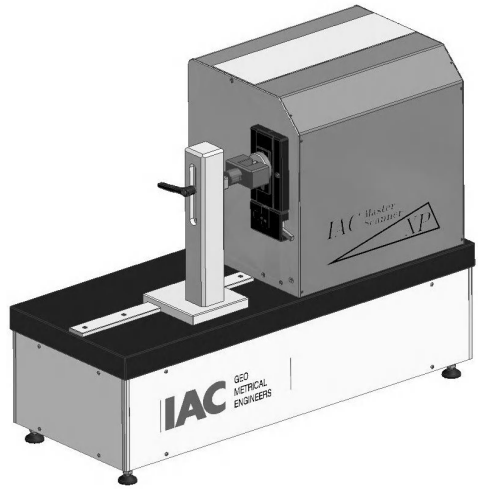


Figure 3. *IAC MasterScanner*

The challenge of 2D scanning

The challenge of 2D scanning is that after solving some problems inherent to tactile scanning it is possible to obtain in one fast automatic cycle with very good measurement uncertainties:

effective diameter, simple pitch diameter, virtual diameter, major diameter, minor diameter, pitch, flank angles, straightness deficiencies of each flank, taper, etc.

Also the visual presentation of the scanned thread profiles is very helpful for analyzing all this information.

Correction of the tips

For this method it is very important to obtain on the flanks the actual border between steel and air. The actual shape of the tips of the scan probe must be available for correction of the profile scans. By means of the scanning of certified contours of the Master Gauge (see Figure 4) and advanced fitting techniques the influence of the scanning styli including wear of the tips is mathematically compensated for.

Figure 4. *Master Gauge for Probe Correction*

For the each direction of the left and right flank the actual functional stylus radius is estimated, including the influence of wear and deflection, etc. (see Figure 5).

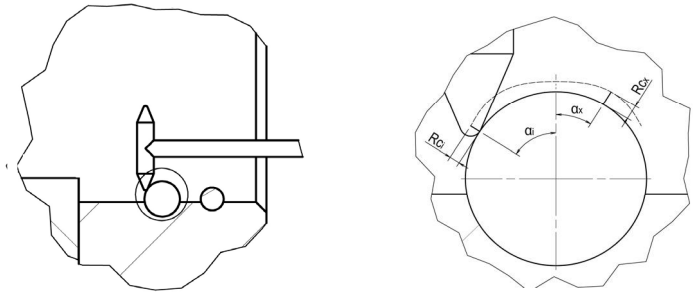


Figure 5. *Mapping of the stylus on the Master Gauge*

Fixation and alignment

The gauges are supported by self-centering quick exchangeable click-on supports. The function of the click-on supports is to position each gauge at a reproducible out of center position relative to the field of scanning. This makes it possible to use the so called intermediate calibration (see Figures 6 and 7).

Intermediate calibration

Intermediate Calibration of each support eliminates the need for manual centering without loss of accuracy and in the scanning of a certified setting ring or setting plug.

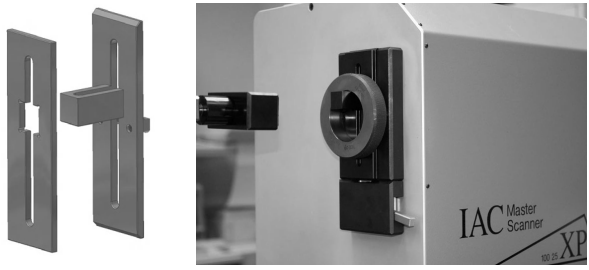


Figure 6. *Support for ring gauges*

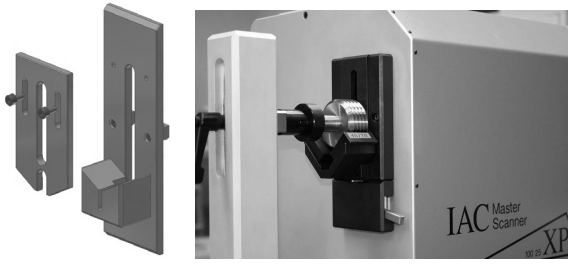


Figure 7. *Support for plug gauges*

With this procedure it is estimated how much smaller the measured diameter of the gauges to be calibrated is due to the out of center position of this particular support on. The accuracy is regained by adding up this compensation value to the measured diameters (see Figure 8).

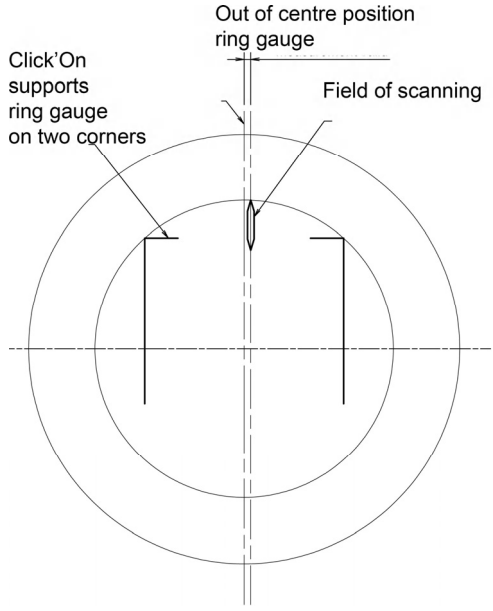


Figure 8. *Intermediate calibration of support with calibrated setting ring*

Parallel alignment of the reference axis of the gauge to the axis of the MasterScanner is accomplished by a coordinate transformation.

From the upper and lower contour the reference axis of the gauge can be derived.

Sources of measurement uncertainty

In this example the sources of measurement uncertainty of the effective diameter of a thread ring gauge are listed.

The calibration of the thread ring gauge can be divided into of two processes:

1) intermediate calibration of setting ring;
2) scanning of gauge.

In the case of a 30 mm setting ring for the intermediate calibration and the calibration of a M30-6g thread ring gauge the combination of process 1 and 2 is estimated:

Uncertainty effective diameter: 1.9 µm

This estimation of uncertainty is based on a standard uncertainty multiplied by a coverage factor k = 2, and on the working conditions of a MasterScanner XP10025 in our Laboratory.

For a MasterScanner XP 10025 we specify: 2.5 µm.

Process 1: intermediate calibration of setting ring

1.1 To be measured object Reference Ring Steel
Uncertainty Certficate of the ring
change diameter due to material drift mm
1.2 To be measured object in the measurement environment
Deviation ambient temperature from 20°C
Diameter variation due to Standard deviation thermal exp. Coef. ppm/°C
Diameter variation due to temperature variation on ambient temp. °C
1.3 Transfer of information
Standard Deviation Master Scanner
Parallel deficiency of the probe styli correction
1.4 Processing of information
Internal rounding

Process 2: calibration of effective diameter

2.1 To be measured object Pitch Diameter to be calibrated Ring Gauge
Deficiency of the contact surfaces mm
2.2 To be measured object in the measurement environment
Deviation ambient temperature °C
Diameter variation mm due to Standard deviation thermal exp. Coef. ppm/°C
Difference Nominal Diameters
Diameter variation due to temperature variation of Ring °C
2.3 Transfer of information
Standard Deviation Master Scanner
Excentricity error E

Heidenhain Glass Scales displacement mm

Variation Heidenhain Scales due to Standard deviation thermal exp. Coef. ppm/°C

Variation Heidenhain Glass Scales due to temperature variation °C

Linear deviation Heidenhain Glass Scales

Upper and lower contour Flank positions of intersection actual profile with pitch lines:
axial X Meas.Uncertainty

Residual stylus radiuscorrection uncertainty/due to Master Gauge

2.4 Processing of information

Internal rounding

Available diameter ranges

At this moment the following ranges are covered by the IAC 2D-Scanners:

2D-Scanner	Internal Diameter (mm)		External Diameter (mm)	
	smallest	largest	smallest	largest
MasterScanner XP	2.5	100	0.8	90
MasterScanner XL	40	330	40	330

Conclusion

The comparison of 2D scanning versus Wires and 2 Ball methods in this presentation makes it clear that the 2D scanning method has many advantages over the conventional methods.

2D scanning estimates the effective pitch diameter and the simple effective diameter more in accordance with the definitions from the thread standards.

The requirements of new guidelines such as VDI 2618 parts 4.8 and 4.9 can in fact only feasibly be fulfilled by application of 2D scanning.

References

[1] EA-10/10: EA Guidelines on the Determination of Pitch Diameter of Parallel Thread Gauges by Mechanical Probing

[2] VDI/VDE/DGQ 2618 part 4.8 (March 2006) Test instructions for setting plugs and plug gauges for cylindrical threads

[3] VDI/VDE/DGQ 2618 part 4.9 (April 2006)
Test instructions for setting rings and ring gauges for cylindrical threads

[4] DIN 2244

[5] ANSI/ASME B1.7M

[6] ISO 1502, 1996-12 Metric Threads

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Geometry and volume measurement of worn cutting tools with an optical surface metrology device

R. Danzl[1], F. Helmli[1], P. Rolland[2] and S. Scherer[1]

[1]Alicona Imaging GmbH, Teslastrasse 8, A8042 Grambach, Austria
[2]Alprimage, 11 rue de Savoie, 91940 Les Ulis, France
Reinhard.Danzl@alicona.com

ABSTRACT: We present a method to measure geometry and wear of cutting tools using an optical surface metrology device based on the focus variation technology. In order to measure the wear, the cutting tools are measured once before and once after usage. Afterwards, the two obtained 3D models are automatically registered to each other which allows us to calculate a difference height map from which the volume of the worn material is extracted. In addition to volume measurements, various other analysis methods such as form measurements of the cutting tool angle or its radius are presented.

1. Introduction

The quality of cutting tools is highly dependent on their geometry and surface properties. Durability, feed rates and overall machining properties are influenced by the cutting geometry such as cutting angle and radius. Especially important for the understanding of the cutting process is the measurement of the amount of wear that occurs during use of the tools in the machining process [8]. This knowledge helps to improve the durability of cutting tools and leads to an improvement of machining speed.

Traditionally, quality assurance methods for cutting tools have been restricted to 2D image analysis methods such as profile projectors or to tactile methods. 2D optical methods are limited in use since they are not able to detect various features such as concavities. Tactile measurement devices on the other hand do not allow the measurement of specific radii or angles due to the mechanical contact of the stylus to the surface. Typically, cutting edge radii vary between 2 µm and 100 µm.

The full 3D wear measurement is a complex task that requires several key aspects. First, a whole area with dense measurement points has to be measured which takes long time when done by contact stylus instruments. Second, the measurement device has to be able to measure even steep surface flanks which is typically problematic for many optical metrology devices such as white light interferometry [10]. Third, a method has to be provided that allows to register the worn cutting tool to the original model.

The optical 3D measurement device InfiniteFocus is a highly valuable quality assurance tool which perfectly meets all necessary requirements for this rather complex measurement task. Its outstanding measurement capabilities and automatic registration processes allow highly accurate and easy to perform wear analysis. In contrast to previous attempts for optical wear measurement [9] that use a fusion of structured light and white light interferometry InfiniteFocus is able to measure both, large scanning areas at high vertical resolutions, at the same time.

In section 2 we describe the metrology device InfiniteFocus and its basic technical principle. Afterwards the method for registration of surfaces is presented which is the core method for wear analysis. After providing results for form analysis we provide results for wear analyis.

2. 3D Measurement with InfiniteFocus

Optical surface measurement techniques are well established in industry but they still hardly provide robust measurements of surfaces that meet the demands for speed, accuracy, repeatability and inline-capability. However, “Focus-Variation” [6] allows advanced optical 3D surface metrology enabling the robust and repeatable measurement of components with steep flanks and varying reflective properties with a vertical resolution up to 10 nm. This new optical technology is the core of the optical measurement system InfiniteFocus (Figure 1) and delivers dense measurements over large areas with a density of 2 Mio – 25 Mio measurement points.

The different components of the system and the basic principle are presented in Figure 2. The main component of this optical metrology instrument is a precision optic consisting of various lens systems. It can be equipped with different objectives allowing measurements with different resolutions. With the aid of a beam splitting mirror, light emerging from a white light source is inserted into the optical path of the system and focused onto the specimen via the objective. Depending on the topography of the specimen, the light is reflected in several directions as soon as it hits the specimen. All rays emerging from the specimen and hitting the objective lens are bundled in the optics and gathered by a light sensitive sensor. The focus variation technology exploits the small depth of field of the optics due to which only small regions of the object are sharply imaged. To perform a complete detection of the surface with full depth of field, the precision optic is moved vertically along the optical axis while continuously capturing data from the surface. This means that each region of the object is sharply focused. All sensor parameters are optimized at each vertical position according to the reflective properties of the surface. After the scanning process the acquired data from the surface are computed to then deliver 3D information as well as an image with full depth of field. This is achieved by analyzing the variation of focus along the vertical axis. Due to the large amount of data mechanical restrictions can be eliminated allowing measurement results with a

high resolution. Once all height measurements are determined, an image with full depth of field is calculated.

A key characteristic of the system is that it does not only deliver topographical information but also an optical color image of the surface which is perfectly registered to the height data. The technology on which the system is based has recently been included in ISO standards classifying different methods for surface texture extraction [6].

In contrast to many other optical devices, InfiniteFocus is capable of measuring surfaces with steep flanks of 80° and more [4]. This feature is especially important for wear measurement applications, where deep scars or e.g. the wear of cutting tools shall be analyzed.

The vertical resolution can be as low as 10 nm making the instrument ideal for wear measurements in the sub-micron range. It can be used in both the laboratory and near production environment. Automation of functions and analysis can also be added to make the instrument useable for the majority of surface metrology and inspection requirements. Recent studies have shown very accurate results for roughness [3] and form measurements [5].

Figure 1. *The InfiniteFocus system*

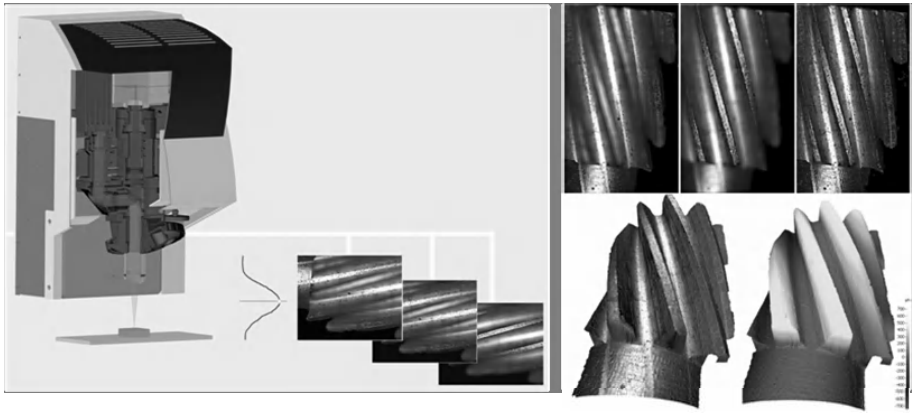


Figure 2. *Schematic view of the technical principle "Focus-Variation". The specimen is scanned vertically along the optical axis which means the sharpness is calculated for each position on the specimen. This results in a 3D object with perfectly registered color information and full depth of field*

In Figure 2 the measurement principle is demonstrated with a typical measuring example of a micro-gear wheel. Figure 2 illustrates the technical principle via the 3D measurement of a micro gear wheel. The images right (left and middle in the raw of three, above) show the raw data from the color sensor obtained at two different positions. Due to the small depth of field each image contains only small sharp imaged regions. By analyzing the variation of focus while vertical scanning the system computes an overall sharp image (see right image in the raw of three and 3D model below)

3. Registration

In order to measure the wear of a cutting tool, InfiniteFocus performs a 3D measurement with true color information before and after the usage of the cutting tool. In order to measure its wear the difference has to be calculated which requires the perfect alignment of the two models. Whereas it is almost impossible to realize this alignment manually we present a fast and robust way to perform this registration automatically.

The registration of surfaces is a long-researched problem in the field of computer vision. The task is to find an Euclidean transformation that transforms one 3D model to another 3D model so that the distance between the two models becomes a minimum. The transformation consists of three rotation and three translation parameters, thus having six degrees of freedom.

Almost all fine registration algorithms are based on the Iterative Closest Point Algorithm (ICP) developed by Besl and McKay [1]. The ICP algorithm iteratively performs the following steps:

- for each point of the first model, search for the closest point in the second model;

- build an error function as the sum of differences between the point pairs;

- calculate the Euclidean transformation that minimizes the error function.

Since this approach was very successful, many different variants of the original algorithm have followed, mainly trying to improve it in terms of speed and robustness [2]. Based on this framework we have developed a variant of this algorithm that is specially designed to handle worn regions.

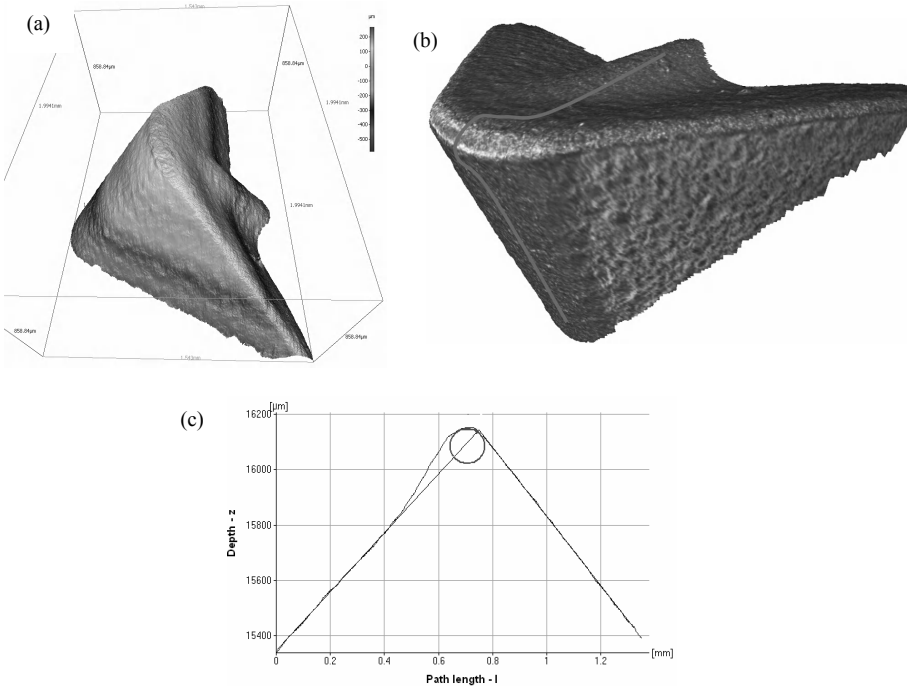


Figure 3. *Form measurement and extracted profile path of a cutting tool with InfiniteFocus (a) and (b) surface measurement via fitted circle at the top of the surface profile (c)*

4. Form measurement

InfiniteFocus provides a series of modules for the geometry measurement of cutting tools. Among these are methods for extraction and analysis of surface profiles, ISO conforming roughness measurements, 2D image analysis, 3D volume calculations or the latest ISO conforming surface texture analysis methods [7].

Key parameters for the characterisation of cutting tool quality are the angle and the radius of the cutting tool tip. Both parameters can be easily obtained by InfiniteFocus as demonstrated in Figure 3. Figure 3 shows a 3D model in pseudo color where each color represents a specific height. InfiniteFocus additionally delivers the perfectly registered color information of the object what helps to identify special regions on the tool, such as the worn tip that is much brighter than the rest of the tool. Additionally a profile path is drawn, along which a height profile has been extracted. From this profile the angle of the tool and the radius of the tip have been extracted using robust methods as angle = 81° and radius = 66 µm (Figure 3c).

5. Wear analysis

The measurement of the total wear volume is presented in Figure 4 using a cutting tool that has been measured with InfiniteFocus once before (4a) and once after usage (4b). The true color image that is provided by InfiniteFocus in addition to the 3D information easily allows us to identify the worn regions. In Figure 4c, both 3D models (light: before usage, dark: after usage – usually shown in two different colors) are overlaid before registration. In order to calculate the difference between the two 3D models they are automatically registered as proposed in the section on registration. The registration result in Figure 4d shows the good registration and allows us to calculate the difference between the measurements as provided in Figure 4e. The differences have been color-coded where the worn parts (here: light) show a different color from the unworn parts (here: dark). From this difference model various parameters can be calculated such as the total wear volume, the maximal worn height, the total worn area or various other parameters as described in [8]. The good measurement and registration quality can also be judged from two surface profiles that have been extracted from the registered original surface and the worn surface as provided in Figure 4f.

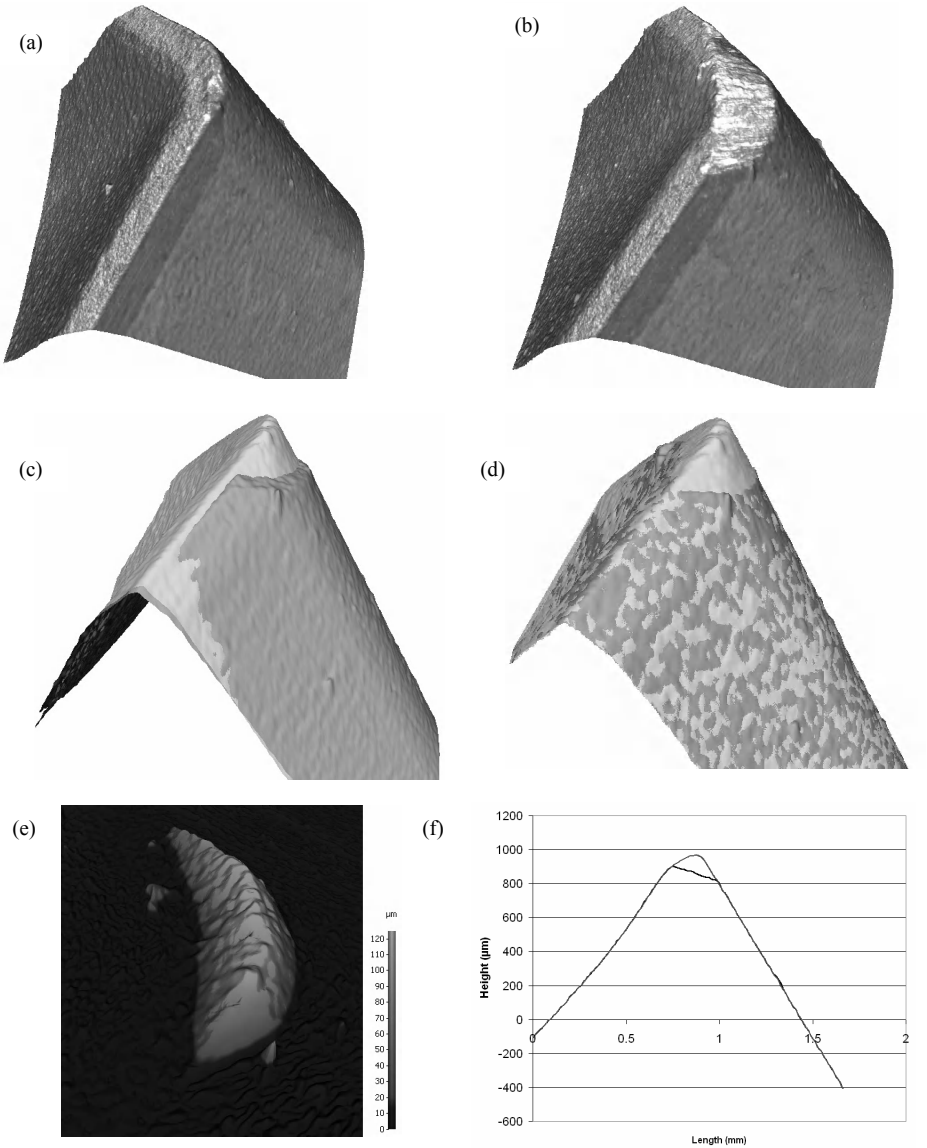


Figure 4. *(a) Cutting tool before usage; (b) cutting tool after usage; (c) overlay of the two cutting tools before they have been automatically registered; light: cutting tool before usage, dark: cutting tool after usage (d) overlay of automatically registered cutting tools; (e) difference between the two 3D models for automatic volume measurement (dark "ground": unchanged, bright "peak": worn parts); (f) two profiles are overlaid and perfectly match each other, expect at the top to visualize worn material*

6. Conclusions

We have presented a method to measure the wear of cutting tools using the optical surface metrology device InfiniteFocus which is based on focus variation technology. Before and after the cutting process, a 3D model is obtained using this metrology device. In order to measure the wear of the tool the difference between the two 3D models is calculated using an automatic registration procedure which performs the alignment and calculates a difference model. Afterwards, InfiniteFocus is used to automatically extract various parameters such as the total volume of the worn parts, the area and the maximum worn height. In addition, geometry features such as the tool angle or tool tip radius can easily be extracted.

In contrast to many other optical devices, InfiniteFocus is able to measure surfaces with steep flanks and does not only measure the surface topography but also calculates a perfectly registered true sharp color image of the surface. This makes it especially useful for the proposed task since the tip of the worn cutting tools contains surfaces with steep flanks. Additionally the worn parts usually have a different color from the original tool which makes the reconstructed sharp color image a valuable source of information for the cutting process.

As far as quality assurance of high-performance cutting tool inserts is concerned, optical measurement of surfaces can provide valuable surface analysis since it removes inaccuracies caused by tactile inspection methods. The optical measurement device InfiniteFocus enables such highly accurate measurements of wear parameters and provides useful insights into the cutting process.

References

[1] P. J. Besl, N.D. McKay, “A method for registration of 3d shapes”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), pp. 239-256, 1992.

[2] S. Rusinkiewicz, M. Levoy, “Efficient variants of the ICP algorithm”, in *Proc. 3rd International Conference on 3D Digital Imaging and Modeling,* pp. 145-152, 2001.

[3] R.Danzl, F.Helmli and S. Scherer, “Comparison of Roughness Measurements Between a Contact Stylus Instrument and an Optical Measurement Device Based on a Color Focus Sensor”, *Nanotech*, Boston, (2006) Vol. III, pp. 284-287.

[4] R. Danzl, F.Helmli. “Three-dimensional Reconstruction of Surfaces with Steep Slopes using an Optical Measurement System based on a Color Focus Sensor”, *6th International Euspen Conference*, (2006), Vol. I, pp. 516-519.

[5] M. Neugebauer and U. Neuschaefer-Rube. “A new Micro Artefact for Testing of Optical and Tactile sensors”, *5th Euspen International Conference and 7th Annual General Meeting of the European Society for Precision Engineering and Nanotechnology*, (2005) 201-204, ISBN 92-990035-0-5.

[6] ISO 25178-6 (draft), Geometrical product specifications (GPS) -- Surface texture: Areal -- Part 6: Classification of methods for measuring surface texture.

[7] ISO 25178-2 (draft), Geometrical product specifications (GPS) -- Surface texture: Areal -- Part 2: Terms, definitions and surface texture parameters.

[8] V.P. Astakhov, “The assessment of cutting tool wear”, *Int. Journal of Machine Tools & Manufacture*, 44 (2004), pp. 637-647.

[9] A. Weckenmann and K. Nalbantic, “Precision Measurement of Cutting Tools with two Matched Optical 3D-Sensors”, in *2003 CIRP Annals*, Vol. 52/1, pp. 24-30, August 2003.

[10] H.-G. Rhee, T. V. Vorburger, J. W. Lee and J. Fu, “Discrepancies between roughness measurements obtained with phase-shifting and white-light interferometry”, *Applied Optics*, Vol. 44, Issue 28, pp. 5919-5927.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Innovation and Knowledge Transfer

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The iMERA/EUROMET joint research project for new determinations of the Boltzmann constant

J. Fischer[a], B. Fellmuth[a], Ch. Gaiser[a], N. Haft[a], W. Buck[a], L. Pitre[b], C. Guianvarc'h[b], F. Sparasci[b], D. Truong[b], Y. Hermier[b], Ch. Chardonnet[c], Ch. Bordé[c], R. M. Gavioso[d], G. Benedetto[d], P. A. Giuliano Albo[d], A. Merlone[d], R. Spagnolo[d], L. Gianfrani[e], G. Casa[e], A. Castrillo[e], P. Laporta[f], G. Galzerano[f], G. Machin[g], M. De Podesta[g], G. Sutton[g], J. Ireland[g], E. Usadi[g], N. Fox[g]

[a] Physikalisch-Technische Bundesanstalt (PTB), Abbestrasse 2-12, 10587 Berlin, Germany

[b] Institut National de Métrologie (LNE-INM/CNAM), 61 rue du Landy, 93210 La Plaine-Saint-Denis, France

[c] Université Paris Nord, 99, ave. J.-B. Clément, 93430 Villetaneuse, France

[d] Istituto Nazionale di Ricerca Metrologica (INRiM), Strada delle Cacce 73, 10135 Turin, Italy

[e] Seconda Università di Napoli, Via Vivaldi 43, 81100 Caserta, Italy

[f] Politecnico di Milano and IFN-CNR, Piazza Leonardo da Vinci 32, 20133 Milan, Italy

[g] National Physical Laboratory (NPL), Hampton Road, Teddington, TW11 0LW, United Kingdom

ABSTRACT: To support new determinations of the Boltzmann constant, which has been asked for by the International Committee for Weights and Measures (CIPM) concerning preparative steps towards new definitions of the kilogram, the ampere, the kelvin and the mole, an iMERA/EUROMET joint research project is coordinating the European activities in this field in France (LNE-INM/CNAM, University of Paris North), Italy (INRiM, Universities of Naples and Milan), the United Kingdom (NPL) and Germany (PTB). In this major European research project the Boltzmann constant will be determined with various methods. The aims and the progress to date of the project are reviewed in this paper.

Introduction

The unit of temperature *T*, the kelvin is currently defined by the temperature of the triple point of water. Thus, the kelvin is linked to a material property. Instead, it would be advantageous to proceed in the same way as with other units: to relate the unit to a fundamental constant and fix its value. Using this no temperature value and no measurement method would be favored. For the kelvin, the corresponding constant is the Boltzmann constant *k,* because temperature always appears as thermal energy *kT* in fundamental laws of physics. For fixing the value, the current value of *k* needs to be confirmed by several independent measurement methods.

To support new determinations of the Boltzmann constant, the Consultative Committee for Thermometry (CCT) recommended "that national laboratories initiate and continue experiments to determine values of thermodynamic temperature and the Boltzmann constant" [1], which is also asked for by the recent recommendation of the CIPM concerning preparative steps towards new definitions of the kilogram, the ampere, the kelvin and the mole [2]. As one result of a workshop held at PTB in 2005 [3], the iMERA/EUROMET 885 joint research project is now coordinating the European activities to determine the Boltzmann constant in France (LNE-INM/CNAM, University Paris North), Italy (INRiM, Universities of Naples and Milan), the United Kingdom (NPL) and Germany (PTB).

In this major European research project, nearly all partners involved use different methods, such as acoustic or dielectric constant gas thermometry, or spectroscopic (Doppler broadening) and radiometric methods. All partners will determine the Boltzmann constant using their own equipment aiming at a relative standard uncertainty close to 1 ppm. These new values will be compared with the present CODATA value of *k* [4] determined by acoustic gas thermometry of NIST [5].

Acoustic method

At equilibrium, the thermodynamic state of a dilute real gas may be described by the virial equation of state:

$$p = RT\rho(1 + B(T)\rho + C(T)\rho^2 + \ldots), \qquad (1)$$

where p is the pressure, ρ the molar density, T the thermodynamic temperature and $B(T)$, $C(T)$, etc. are the second, third, etc. density virial coefficients. Given its basic thermodynamic definition $u^2 = (\partial p/\partial \rho)_S$, the adiabatic speed of sound in a real gas may be expressed by a virial expansion similar to equation (1):

$$u^2(p,T) = A_0(T) + A_1(T)p + A_2(T)p^2 + \ldots, \qquad (2)$$

where the $A_n(T)$ are called acoustic virial coefficients. The zero pressure limit of equation (2) is:

$$u_0^2(T) = A_0(T) = \gamma^0 RT/M = \gamma^0 kT/m, \tag{3}$$

where γ^0 is the ideal gas heat capacity ratio (5/3 for a mono-atomic gas), that links the speed of sound and the thermodynamic properties of a gas at a macroscopic (M is the molar mass, R the molar gas constant) or microscopic level (m is the molecular mass). The expression in equation (3) shows that primary thermometry can be performed by acoustic methods. The experimental determination of u_0 at the triple point of water with T_w = 273.16 K leads to a value of the gas constant R and, through the accepted value of the Avogadro constant N_A to the Boltzmann constant $k = R/N_A$. The currently recommended values for R and k [4] are based on measurements of the speed of sound in argon [5, 6]. The method can be employed to perform thermometry to determine the difference (T-T_{90}) between the thermodynamic temperature and international temperatures T_{90} defined by ITS-90 [7]. The ratio of speed of sound measurements at two different empirical temperatures yields the ratio of the corresponding thermodynamic temperatures, without the uncertainties related to the quantities R, γ^0 and M. With this method, remarkably accurate and mutually consistent determinations of (T-T_{90}), which are shown in Figure 1, have been obtained [8-12].

Spherical acoustic resonators are generally used for the measurement of the speed of sound in dilute gases. This is because of the high accuracy achievable by the method and that many of the major perturbations, which influence the determination of the resonance frequencies, are well understood. Recently, cavities having a quasi-spherical geometry have been employed. These permit operation as combined acoustic and electromagnetic resonators and their quasi-spherical form lifts the degeneracy of the microwave normal modes. In primary acoustic thermometry experiments, microwave measurements allow accurate assessment of the thermal expansion of the cavity. The use of these devices for the determination of R or k looks promising, as the need for an independent measurement of the cavity dimensions appears not to be required.

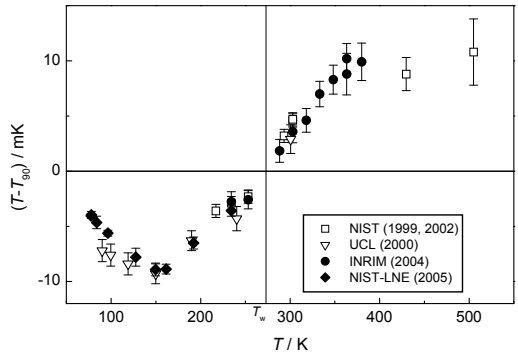


Figure 1. *Comparison of recent primary acoustic thermometry data [8-12]*

In fact, exploiting the exact definition of the speed of light in vacuum c_0, we may write down the following simplified equation for a mono-atomic gas:

$$R = kN_A = \frac{3}{5} c_0^{\,2} \frac{M}{T_{\mathrm{w}}} \lim_{p \to 0} \left(\frac{f'_{\mathrm{a}}(p)}{\langle f'_{\mathrm{m}}(p) \rangle} \right)^2 , \tag{4}$$

where f'_x are the experimental acoustic and microwave resonance frequencies corrected for the main known perturbative effects and normalized by the corresponding eigenvalues. Thus, the metrological problem of determining the speed of sound is shifted from an independent measurement of time and space to measuring the adimensional ratio of acoustic and microwave frequencies as a function of pressure. The ultimate accuracy achievable will depend on the capability of modeling appropriate corrections for the acoustic and microwave eigenfrequencies to account for the imperfections affecting a real cavity. State-of-the-art dimensional measurements of the cavity form will be required in the initial stages of the project to validate the microwave approach.

Research projects which aim at a determination of R and k with the acoustic method are currently being developed at LNE-INM/CNAM, INRIM and NPL. All these research groups intend to use the experimental method and procedures described above with minor differences in their respective approach. In particular, LNE is currently testing the acoustic and microwave performance of copper ellipsoidal cavities and working to develop an improved mathematical model of the major acoustic perturbations; INRIM is striving to obtain an accurate dimensional characterization of a misaligned spherical stainless steel resonator and analyzing the results of speed of sound measurements in helium in comparison to the predictions of recent ab-initio calculations; NPL is completing its first experimental apparatus

and is planning for independent /alternative validations of the microwave determination of the resonator volume by pycnometry and accurate coordinate-measuring machine methods.

Dielectric-constant gas-thermometer method

The basic idea of dielectric-constant gas thermometry (DCGT) is to replace the density in the state equation of a gas by the dielectric constant ε and to measure it by incorporating a capacitor in the gas bulb. The dielectric constant of an ideal gas is given by the relation $\varepsilon = \varepsilon_0 + \alpha_0 N/V$, where ε_0 is the exactly known electric constant, α_0 is the static electric dipole polarizability of the atoms and N/V is the number density, i.e. the state equation of an ideal gas can be written in the form $p = kT(\varepsilon - \varepsilon_0)/\alpha_0$. Absolute DCGT requires knowledge of the static electric dipole polarizability α_0 with the necessary accuracy. Nowadays this condition is fulfilled for ^{4}He, which has become a model substance for evaluating the accuracy of *ab initio* calculations of thermophysical properties. Recent progress has decreased the uncertainty of the *ab initio* value of α_0 well below one part in 10^6 [13]. The molar polarizability A_ε is defined as $A_\varepsilon = N_A \alpha_0/(3\varepsilon_0)$, thus the Boltzmann constant is related to α_0, A_ε and R by

$$k = \frac{R}{A_\varepsilon}\frac{\alpha_0}{3\varepsilon_0} \quad . \tag{5}$$

The measurement of the ratio A_ε/R of two macroscopic quantities therefore allows the determination of k.

For a real gas, the interaction between the particles has to be considered by combining the virial expansions of the state equation and the Clausius-Mossotti equation. When neglecting higher-order terms and the third dielectric virial coefficient, this yields

$$p = \frac{\chi}{\frac{3A_\varepsilon}{RT} + \kappa_{\text{eff}}}\left[1 + \frac{B^*(T)}{3A_\varepsilon}\chi + \frac{C(T)}{(3A_\varepsilon)^2}\chi^2 + \ldots\right] \quad , \tag{6}$$

where $\chi = \varepsilon/\varepsilon_0 - 1$ is the dielectric susceptibility, $B^*(T) = B(T) - b(T)$, $b(T)$ is the second dielectric virial coefficient, and κ_{eff} is the effective compressibility of a suitable capacitor used to measure the susceptibility χ. For determining $3A_\varepsilon/RT$, isotherms have to be measured, i.e. the relative change in capacitance $(C(p) - C(0))/C(0) = \chi + (\varepsilon/\varepsilon_0)\kappa_{\text{eff}} p$ of the gas-filled capacitor is determined as a function of the pressure p of the gas. The capacitance $C(p)$ of the capacitor is measured with the space between its electrodes filled with the gas at various

pressures and with the space evacuated so that $p = 0$ Pa. A polynomial fit to the resulting p versus $(C(p) - C(0))/C(0)$ data points, together with the knowledge of the pressure dependence of the dimensions of the capacitor (effective compressibility κ_{eff}), yields $3A_\varepsilon/RT$.

New DCGT measurements were performed at PTB in the temperature range from 2.4 K to 26 K in order to establish a temperature scale with reduced uncertainty compared with the first DCGT scale of PTB described in [14]. Progress has been achieved concerning the measurement of capacitance changes, temperature and pressure that resulted in a reduction of the overall uncertainty by about a factor of two. For the capacitance measurements, a more symmetric setup has been realised. Further improvements concern the parasitic impedances of the capacitance bridge, its resolution and the data acquisition. The reduction of the uncertainty of temperature and pressure measurement is based on better thermal conditions in the new cryostat and a calibration of the measuring devices that is fully traceable to the national standards.

At more than 20 temperatures, DCGT isotherms were measured and evaluated both performing single- and multi-isotherm fits. For the data evaluation by multi-isotherm fits, appropriate expressions for the temperature dependencies $B(T)$ and $C(T)$, respectively, of the virial coefficients of the measuring gas ^{4}He were selected, see the detailed discussion in [15]. A theoretically based expansion for $C(T)$ yielded the most stable results. Evaluating a large data set of 172 triplets of pressure, temperature and dielectric constant, a standard uncertainty of the established new DCGT scale ranging from 0.15 mK at 4 K to 0.4 mK at 26 K has been achieved. The differences between the obtained $B(T)$ results and new *ab initio* values are well within the uncertainty estimates, which confirms the quality of both the DCGT data and its evaluation. The comparison of the results with literature data corroborates the thermodynamic accuracy of the scale, see Figure 2. It also corroborates the value for the triple-point temperature of hydrogen of 13.80365(50) K at a deuterium content of 35 µmol/mol that is based on the first DCGT data set [14]. This value yields a triple-point temperature of 13.8039 K for the recently prescribed SLAP (Standard Light Antarctic Precipitation) deuterium concentration of about 89 µmol/mol. All these results support the potential of the DCGT method for determining the Boltzmann constant at the triple-point of water with the claimed uncertainty [16] as a basis for the new definition of the base unit kelvin.

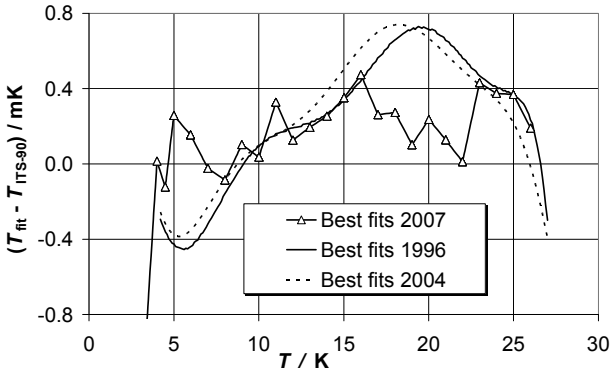


Figure 2. *Deviation of T_{fit} corresponding to different DCGT multi-isotherm fits from the ITS-90: Open triangles: Fit yielding the new DCGT scale of PTB that was performed using carefully selected expressions for the temperature dependencies B(T) and C(T), see [15]. Solid line: The average of the best fits published in [14]. Dotted line: The average of best fits obtained in 2004 by re-evaluating the data given in [14] using a more appropriate expression B(T), see [15]*

Doppler-broadening method

Recently, a method based on laser spectroscopy was proposed to perform a direct determination of the Boltzmann constant [17]. The principle consists of recording the linear absorption in gas phase and measuring the Doppler width of an atomic or molecular line in a cell at the thermodynamic equilibrium. In the Doppler limit, the line shape is a Gaussian (for an optically thin medium) and kT is given by:

$$kT = \frac{mc_0^2}{2}\left(\frac{\Delta_D}{\nu}\right)^2 \tag{7}$$

Δ_D is the e-fold half-width, ν is the frequency of the molecular line and m is the molecular mass. Δ_D and ν are determined experimentally. The probed atoms or molecules belong to a single quantum level of a well-defined isotopic species, which avoids uncertainties coming from macroscopic quantities as it is the case for acoustic gas thermometry. Strictly speaking, we are sensitive to the temperature of one translational degree of freedom of a subset of molecules. Since the temperature is measured on the walls of the cell containing the gas, the determination of k using different rovibrational lines can lead to a verification of the equipartition of energy principle. This experimental situation benefits from a very straightforward analysis, since it turns out that the only systematic effect to be taken into account is the pressure broadening.

Ion mass ratios can now be measured in Penning traps with 10^{-9}-10^{-10} accuracies and binding energies for a molecule can be easily calculated to keep the accuracy of the molecular mass expressed in atomic units at the same level. However, because of the definition of the kilogram, its absolute value implies the Avogadro constant, N_A, as in the previous experiments. Since atom interferometry yields a direct determination of the quantity h/mc_0^2, the current experiment yields directly the ratio, k/h where h is Plancks constant. There are two different projects under development based on these ideas.

The first one in Villetaneuse (University of Paris-North), demonstrated already the feasibility of the principle [18]. The experiment consists simply of recording an absorption spectrum. The selected line was the ν_2 as Q(6,3) rovibrational line of the ammonia molecule $^{14}NH_3$ at the frequency $\nu = 28{,}953{,}694$ MHz. The gas pressure varies from 0.1 to 10 Pa, which is 4 orders of magnitude smaller than the pressure of the acoustic gas thermometry or permittivity experiments. The 30 cm-long absorption cell is placed in a large thermostat filled with an ice-water mixture, fixing the temperature at 273.15 K. The laser source is based on a CO_2 laser stabilized on an absorption line of OsO_4. The laser frequency can be controlled at the level of 1 Hz which contributes to a negligible 10^{-8} to the relative uncertainty on k. The art of this experiment is the capability to record a signal which reflects perfectly the Doppler profile and to suppress any parasitic optical signals. In this first trial we deduced the pure Doppler width from 2000 single spectra by extrapolation of the line width to zero pressure (Figure 3) and obtained a first value for k by this method of $k = 1.380\ 65\ (26) \times 10^{-23}$ JK^{-1} (1.9×10^{-4}) in full agreement with the accepted value. Now, we are developing a second generation of experiment with several significant improvements: an absorption length 10 times larger to reduce the pressure by one order of magnitude, a frequency scan ten times faster, a new modulation/detection scheme which should improve the signal-to-noise ratio by one order of magnitude, which is currently the main limiting factor. And finally, a temperature control better than 1 mK relative to the TPW. With this second generation of experiment we can hope to gain the two orders of magnitude in the accuracy to compare with the CODATA value of k.

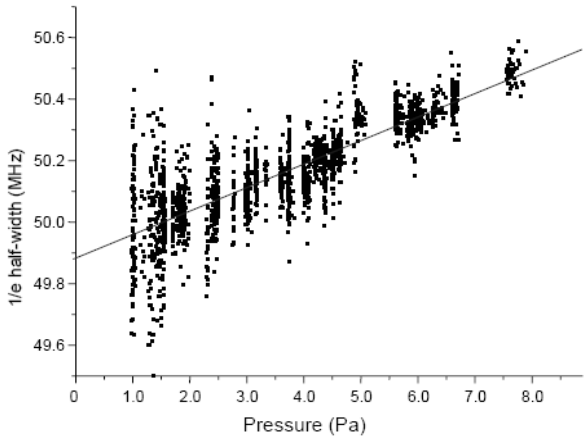


Figure 3. *Extrapolation of the measured line width to zero pressure*

In the second project, which is being performed at the Second University of Naples, in cooperation with the Polytechnic of Milan, high-resolution diode laser spectroscopy in the near-infrared has been implemented for highly accurate determinations of the Doppler line width of a rovibrational transition in a gaseous sample of carbon dioxide, at thermodynamic equilibrium. The selected line is the R(12) rovibrational component of the $\nu_1+2\nu_2^0+\nu_3$ combination band, around a wavelength of 2 μm. The Italian experiment, which is based on an intensity-stabilized absorption spectrometer, basically consists of the precise detection of the absorption line profile, with the lowest levels of distortion on both the frequency and intensity scales. Observed in correspondence of a pressure range for which Dicke-narrowing and speed-dependent effects can be neglected, this line shape is given by a Voigt convolution that accounts for both Doppler and collision broadening mechanisms. Laser-gas interaction takes place in a 12 cm-long cell consisting of a cylindrical cavity inside an aluminium block, whose temperature is stabilised by means of Peltier elements. Precision platinum resistance thermometers (Pt100), calibrated with respect to a standard thermometer, provide information on the temperature of the cell's body while a proportional integral derivative (PID) controller is used to keep the temperature uniform along the cell and constant within 0.1 K. This system also allows us to change the gas temperature between 270.0 K and 330.0 K.

Once the Doppler width (Δ_D) is accurately extrapolated from the absorption line shape, a spectroscopic determination of the gas temperature can be carried out. The relative uncertainty, accounting for both statistical and systematic deviations, is presently at a level of 8×10^{-4}. Several improvements are under development in order to increase the accuracy of the spectrometer. Simultaneously, also in cooperation with the Italian Institute of Metrology (INRIM), a new system is being constructed, based on a pair of phase-locked extended-cavity diode lasers. The new

spectrometer would allow for an improvement of 3 orders of magnitude in the capability of controlling and measuring laser frequency variations.

Radiometric method

Using the Absolute Radiation Detector (ARD) NPL aims to measure k using a radiometric measurement. The ARD experiment builds on the radiometric determination of the Stefan-Boltzmann constant by Quinn and Martin in the 1980s at NPL [19].

The Boltzmann constant is related to the Stefan-Boltzmann constant, σ_{SB} as

$$\sigma_{SB} = (2\pi^5 k^4) / (15 h^3 c_0^2). \quad (8)$$

Since the total power, $M(T)$ radiated by a blackbody is given exactly by the Stefan Boltzmann equation:

$$M(T) = \sigma_{SB}\, g\, T^4, \quad (9)$$

where T is the temperature of the blackbody and g is a geometric factor, we can obtain a measurement of k, using equations (8) and (9), by measuring the total power output of a blackbody at a known temperature.

An absolute measurement of optical power can be made using a cryogenic radiometer; an electrical substitution device which compares the temperature rise of a cold cavity for optical and electrical forms of heating and hence provides a measurement of optical power traceable to the electrical Watt.

As shown in Figure 4, the ARD experimental system involves a cryogenic radiometer operating at 5 Kelvin, a light trap which defines the geometric factor g by supporting two apertures a known distance apart, and a blackbody radiator operating at the triple point of water [20]. The ARD experiment has been developed over a number of years and has benefited from two major improvements over the Quinn Martin radiometer. Firstly, a much larger, wider light trap has been designed using Monte Carlo modeling to optimize the baffle structure. This ensures that radiation not in the direct solid angle of view is absorbed and eliminates previous problems with scatter from the baffles. The surfaces are coated with Nextel black which has well known optical properties. The trap is constructed of diamond turned copper and the diameter and spacing of the apertures is measured mechanically which enable us to determine g to within 5 ppm.

Secondly, the blackbody has been redesigned to overcome diffraction problems. It consists of a black grooved base and black cylindrical sidewalls. The base is maintained at the triple point of water to within 2 mK using an external temperature controlled oil bath. Calculations indicate that the temperature gradient of the side-

walls, in this conventional design, will be small enough that it will not make a significant contribution to the uncertainty in σ_{SB}.

The new ARD has been fully assembled and at the time of writing we anticipate that data will be obtained within the next few months. The expected uncertainty in k is < 10 ppm [20].

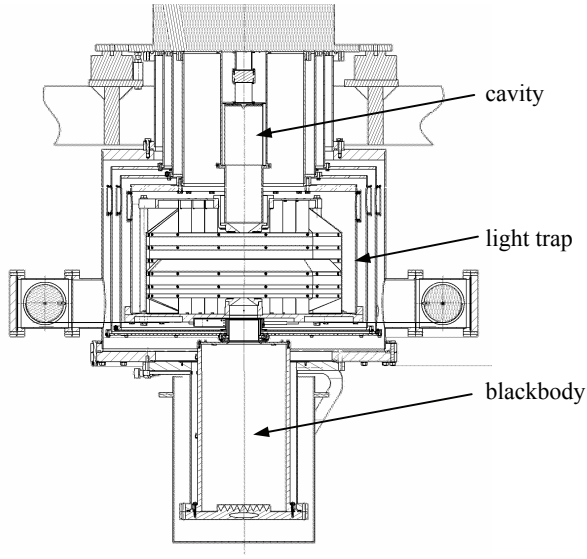


Figure 4. *Schematic of absolute radiation detector*

Conclusions

It is assumed that the experiments currently underway to measure the Boltzmann constant k will achieve by the end of 2010 consistent results [16], so that the CODATA task group can recommend in its 2010 constants adjustment a new value for k with a relative standard uncertainty about a factor of two smaller than the current u_r of approximately 2×10^{-6}. Then, in 2011 the General Conference of Weights and Measures (CGPM) could adopt the new definitions of the kilogram, the ampere, the kelvin, and the mole.

The new definition of the kelvin would be in line with modern science where nature is characterized by statistical thermodynamics, which implies the equivalence of energy and temperature as expressed by the Maxwell-Boltzmann equation $E = kT$. In principle, temperature could be derived from the measurement of energy. In practice however, we have no simple and universal instrument for measuring energy

and it appears in different forms, e.g. temperature. The fundamental constant *k* converts the value of this measurable quantity into energy units.

For the unit kelvin, the new definition based on a fixed value of *k* would generalise the definition, making it independent of any material substance, technique of realization, and temperature or temperature range. In particular, the new definition would improve temperature measurement at temperatures far away from the triple point of water. For example, in the high temperature range, the radiometry community could apply absolute radiation thermometers without the need to refer to the triple point of water. It would also encourage the use of direct realizations of thermodynamic temperatures in parallel with the realization described in the International Temperature Scale. In the long term it will enable gradual improvements to the temperature scale in respect of lower uncertainties and extended temperature ranges, without the high transitional costs and inconvenience that has been incurred with previous changes in temperature scales.

References

[1] Recommendation T 2 (2005) to the CIPM: "New determinations of thermodynamic temperature and the Boltzmann constant", *Working Documents of the 23rd Meeting of the Consultative Committee for Thermometry* (BIPM, Document CCT/05-31, 2005).

[2] Recommendation 1 (CI-2005): "Preparative steps towards new definitions of the kilogram, the ampere, the kelvin and the mole in terms of fundamental constants", (CIPM, Sèvres, 2005)

[3] B. Fellmuth, J. Fischer, C. Gaiser, and W. Buck, "Workshop on Methods for New Determinations of the Boltzmann Constant", *Working Documents of the 23rd Meeting of the Consultative Committee for Thermometry* (BIPM, Document CCT/05-02, 2005).

[4] P. J. Mohr and B. N. Taylor, "CODATA Recommended Values of the Fundamental Physical Constants: 2002", *Rev. Mod. Phys.*, Vol. 77, pp. 1-108, January 2005.

[5] M. R. Moldover, J. P. M. Trusler, T. J. Edwards, J. B. Mehl and R. S. Davis, "Measurement of the Universal Gas Constant *R* using a Spherical Acoustic Resonator", *Phys. Rev. Lett.*, Vol 60, pp.249-252, January 1988.

[6] A. R. Colclough, T. J. Quinn and R. D. T. Chandler, "An Acoustic Redetermination of the Gas Constant" *Proc. R. Soc. Lond., A*, Vol. 368, pp. 125-139, September 1979.

[7] H. Preston-Thomas, "The International Temperature Scale of 1990 (ITS-90)", *Metrologia*, Vol. 27**,** pp. 3-10 and 107, January and April 1990.

[8] M. R. Moldover, S. J. Boyes, C. W. Meyer and A. R. H. Goodwin, "Thermodynamic Temperatures of the Triple Points of Mercury and Gallium and in the Interval 217 K to 303 K", *J. Res. Natl. Inst. Stand. Technol.*, Vol. 104, pp. 11-46, January 1999.

[9] G. F. Strouse, D. R. Defibaugh, M. R. Moldover and D. C. Ripple, "Progress in primary acoustic thermometry at NIST: 273–505 K" in *2003 Temperature: Its Measurement and Control in Science and Industry, 8th Int. Temperature Symp.,*Vol. 7 (Edited by D. C. Ripple), Chicago, (*AIP Conf. Proc*. Vol. 684), pp. 31-36, 2002.

[10] M. B. Ewing, J. P. M. Trusler, "Primary acoustic thermometry between T = 90 K and T = 300 K", *J. Chem. Thermodynam.*, Vol. 32, pp. 1229-1255, September 2000.

[11] G. Benedetto, R. M. Gavioso, R. Spagnolo, P. Marcarino and A. Merlone, "Acoustic measurements of the thermodynamic temperature between the triple point of mercury and 380K", *Metrologia*, Vol. 41, pp. 74-98, January 2004.

[12] L. Pitre, M. R. Moldover and W. L. Tew, "Acoustic thermometry: new results from 273K to 77 K and progress towards 4 K", *Metrologia*, Vol. 43, pp. 142-162, January 2006.

[13] G. Łach, B. Jeziorski and K. Szalewicz, "Radiative Corrections to the Polarizability of Helium", *Phys. Rev. Lett*., Vol. 92, pp. 233001-1-4, June 2004.

[14] H. Luther, K. Grohmann and B. Fellmuth, "Determination of thermodynamic temperature and ^{4}He virial coefficients between 4,2 K and 27,0 K by dielectric constant gas thermometry", *Metrologia*, Vol. 33, pp. 341-352, August 1996.

[15] C. Gaiser, B. Fellmuth and N. Haft, "Primary Gas Thermometry in the Range from 2.4 K to 26 K at PTB", *Int. J. Thermophys.*, Vol. 29, pp. 18-30, January 2008.

[16] B. Fellmuth, Ch. Gaiser, and J. Fischer, "Determination of the Boltzmann constant - status and prospects", *Meas. Sci. Technol.* Vol. 17, pp. R145-R159, October 2006.

[17] Ch. J. "Bordé, Base units of the SI, fundamental constants and modern quantum physics", *Phil. Trans. R. Soc. A*, Vol. 363, pp. 2177-2201, September 2005.

[18] C. Daussy, M. Guinet, A. Amy-Klein, K. Djerroud, Y. Hermier, S. Briaudeau, Ch.J. Bordé, and C. Chardonnet, "Direct determination of the Boltzmann constant by an optical method", arXiv/quant-ph/0701176 and *Phys. Rev. Lett.,* Vol. 98, 250801, June 2007.

[19] T. J. Quinn and J. E. Martin, "A radiometric determination of the Stefan-Boltzmann constant and thermodynamic temperatures between -40 °C and +100 °C", *Phil. Trans. R. Soc., Lond.*, Vol. 316, pp. 85-189, November 1985.

[20] J. E. Martin and P. R. Haycocks, "Design considerations for the construction of an absolute radiation detector at the NPL", *Metrologia*, Vol. 35, pp. 229-233, July 1998.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

The "Measurement for Innovators" program: stimulating innovation in the UK through improved understanding and appreciation of measurement issues

G.E. Tellett, C. Mackechnie, V. Ralph

National Physical Laboratory, Teddington, Middlesex, TW11 0LW, UK

ABSTRACT: The UK Government have funded a program of work, "Measurement for Innovators", developed and delivered by the UK NMI laboratories, which forms an effective bridge between the underpinning research undertaken in the National Measurement System (NMS) and the end-user measurement needs that are required to stimulate innovation in the UK.
This paper outlines the criteria established for the program and includes some success stories. Results of an independent study on the impact and effectiveness of the program are included and the benefits to the NMIs, as well as the value to industry will be discussed.

Introduction

Measurement is of critical importance in encouraging and stimulating innovation. Therefore, it is a priority to provide organisations, particularly those of small to medium size, with a route to accessing the knowledge, people and facilities of the National Measurement Institutes. When engaging with such organisations, time and cost is often of paramount importance to them, so any intervention needs to be unbureaucratic and timely.

All NMIs operate within a framework of limited resources. Therefore, it is also important to put in place processes that ensure the NMI only engages in supporting innovations where the NMI can add most value and where the NMI is not competing with other suppliers or potential suppliers in the market place i.e. where the NMI has a unique contribution. Measurement support of this nature has an impact which is both broader and deeper than the companies directly supported with a particular innovation, it influences the unit measurement programs and alters the perspective of many of the scientists involved in delivering those programs.

Measurement for Innovators

The Measurement for Innovators program commenced in Autumn 2004 with funding of £3.3 million over two years. It has been led by the National Physical

Laboratory but delivered by a consortia of UK National Measurement Institutes including NPL, LGC, NEL and NWML. The program is made up of three different knowledge transfer based activities:

- consultancies to SMEs,
- secondments, both into and out of the NMIs to industry and other relevant organisations,
- co-funded collaborative projects between the industrial communities and the NMIs (Joint Industry Projects).

The program aims to enable companies to work with the NMI to provide measurement solutions by assisting:

- the development of innovative new products and processes.
- the improvement of existing products and process,
- the provision of solutions to generic measurement problems,
- the development of new test methods or models,
- the provision of benchmarks for different test procedures.

The specific targets that were agreed for the program were as follows:

- to execute a program of 15 Joint Industry Projects, 20 Secondments and 250 Consultancies,
- to involve at least 30 companies in collaborative projects co-funding to a level of 50%,
- to ensure wider dissemination of research outputs and case studies to at least 1,000 organisations.
- to change the measurement practice in 250 SMEs leading to improved processes and new products.

This was a new progamme and the targets were based on scant experience and in reality the outturn in Table 1 has been achieved.

Activity	**Target**	**Outturn**
Joint Industry Projects	15	34
Companies involved in collaborative projects	30	122
Secondments	33	87
Covnsultancies	250	157

Table 1. *Targets for the measurement for Innovators program*

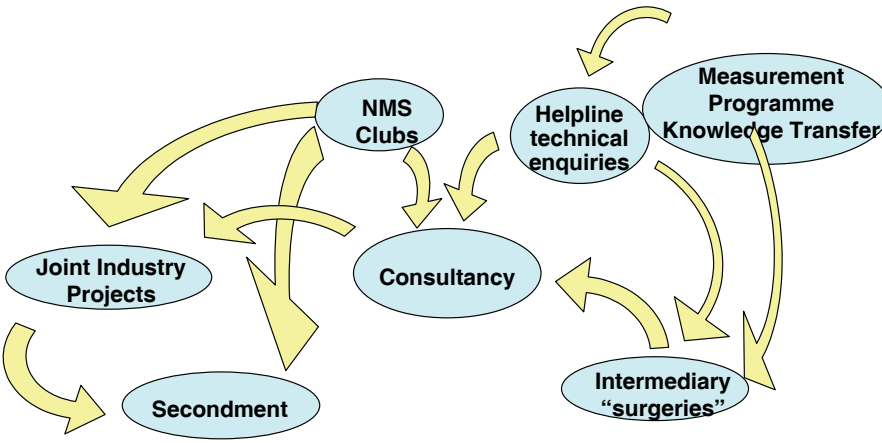


Figure 1. *Interrelationship of support products*

The program has, in general, been more successful than anticipated with greater impact being achieved. For all interventions, demand has far outstripped supply of resources and expectations have had to be closely managed.

Integrated support

In order to extract maximum value and to ensure clarity of offering to the end-user audience it has been important to link the different types of support available from the UK NMS in a holistic and coherent manner. These interventions have been delivered in conjunction with the NMS technical programs and by the staff deployed on them. This program complements their portfolio of knowledge transfer projects and services to their own and new audiences.

It is difficult to access and engage new companies in the activities of the UK NMS partly because of the sheer quantity of businesses and their diversity and partly due to resource constraints. However, the program has now formed a focus for the NMI Laboratories to engage with other Government business support organisations. These intermediaries have both taken the NMS offering to new audiences and provided the first stage filtering to ensure that the most relevant enquiries are taken forward. The relationship therefore works both ways with the NMI Laboratories providing dedicated resources to deploy on supporting their businesses and the intermediary providing a steady stream of new contacts.

Secondments

Secondments are designed to transfer tacit knowledge by exchange of people.

The National Measurement Secondment Scheme is designed to encourage closer collaboration between the UK's National Metrology Institutes and other organisations by means of the following types of secondment:

- to and from UK industry, universities, trade associations and other bodies,
- to and from UK Government Departments,
- to and from Regional Organisations and other business support organisations.

The objectives have been to:

- improve the performance of the UK NMS by exchanging knowledge and skills between scientists and specialists, using a two-way secondment scheme,
- to help companies innovate and develop new products and services,
- explore new technical areas which may be strategically important to the UK Technology Strategy for the NMS programs.

Case Study: Big steps for small sensors

An SME, Applied Nanodetectors, has devised a novel sensor array that has the combined advantage of greater sensitivity with no false positive signals in a device that is physically smaller than those currently available. This has great potential as a reliable sensor for security systems. However, in order to take forward some of their ideas they required access to specialist measurement facilities in order to better understand the measurement issues.

During a secondment to NPL, Victor Higgs (MD) gained first hand experience of how metrology of nanomaterials could be used to accelerate the development of his products and intellectual property. An important feature of the MFI scheme is that it fully supports the generation of new IP for the company and does not seek to stake a claim in the arising IP.

The encouraging results, and Victor's improved understanding of the measurement issues resulting from access to NPLs facilities and expertise, has greatly strengthened his application for MOD funding to further develop the gas sensors, with NPL as the collaborating metrology partner.

Applications are reviewed and considered against the following criteria in order for the approvers to assess the suitability of the application:

- strategic importance to UK NMS,
- relation to Core Programme,
- strategic importance of partner organisation,
- ability to access capabilities,
- secondment candidate suitability,
- value for money,
- impact potential.

Secondments between 1 week and 9 months were allowable under the scheme.

Consultancies

Consultancies are designed to allow NMS laboratories to offer UK SMEs free innovation support. It is possible to offer between 2 hours and 4 days advice and help a company with a specific product, process or service. The project must be capable of delivering a case study at the end and must not compete with commercial offerings already available in the market place. It must also involve specific physical intervention such as inspection of parts or site visits.

Case study: GPS device interference

GSD Navigation is a small UK company marketing commercial GPS navigation systems. In order to supply the police service with GPS navigations systems they are advised to seek approval from the National Policing Improvement Agency (NPIA) who test to ensure that radiated emissions from the device do not exceed a certain limit. The company carried out some preliminary tests to investigate the interference signals but found that they needed additional support and use of specialist facilities to take the investigation further.

The SME were able to benefit from a four-day consultancy with NPL to look into the problem. To begin with ElectroMagnetic Compatibility (EMC) tests were carried out in NPL's fully anechoic room to try to determine the source of interference and to identify which component of the circuit needed shielding. Then various solutions were tried to reduce the emission signal to an acceptable level, using metallized polymer adhesive sheets and RF gaskets to shield the units and cables, with the assistance of Shielding Solutions Ltd.

An improved performance was ultimately achieved. GSD Navigation were very pleased with the results and believe that this will enable them to supply the police with effectively shielded sat-nav devices.

Joint Industry Projects

A Joint Industry Project (JIP) is an industrially co-funded project of short duration, which focuses on end user needs. It allows organisations to obtain leverage on researching specific measurement problems and to access the expertise and equipment of the NMS laboratories. It provides a vehicle for companies to collaborate with like-minded companies on common ground and allows them to play a key role in steering and executing the research.

The spectrum of project types that can be supported by the JIP approach is broad but the following are typical:

- desk-based studies and scientific research and development,
- feasibility studies,
- new test methods, models or databases,
- new standards and reference materials,
- comparisons of measurement test procedures,
- best practice and benchmarking.

An important feature of the project is that the knowledge generated must be exploited beyond the project team, although a short delay (say, 12 months) in disseminating results beyond the immediate partners is acceptable. Projects typically involve different parts of the supply chain. Table 2 below shows some of the participants. The application process is simple with only two pages to be completed. This short proposal is judged by a management board against specific criteria such as:

- impact potential,
- partnership fit,
- promotes innovation,
- fit with technology strategy,
- clearly defined exploitation route,
- multi-disciplinary/cross program.

Category of Partner	Typical Company
Large Company	BP, BOC,NCR, GlaxoSmithKline, Dyson, Renishaw, Taylor Hobson
SME	Gooch and Housego, IndexSAR, EMEC, Precision Acoustics, Bentham
Agency/public Sector	Civil Aviation Authority, Maritime and Coastguard Agency, Trinity house, Health and Safety Executive, Scottish Enterprise

Table 2. *Examples of JIP participants*

Key characteristics of JIPs are:

- short-term (no longer than 12 months duration),
- involvement of industrial partners (no minimum),
- industrial co-funding (minimum of 50% total, no minimum for cash contribution, but will rank high),
- exploitable outputs (clear route for dissemination/uptake of outputs).

Case study: international comparison of biometric techniques

Biometrics are considered essential to improve security in many aspects of life from identifying citizens to protecting new born babies in maternity wards. Current technologies have widely varying error rates, and there are no formal schemes for certifying biometric system performance. Without an independent and scientifically credible evaluation, it is very difficult to assess the relative merits of different biometric systems. This makes the adoption of new and innovative biometric systems particularly difficult, as there is no history of previous use.

A Joint Industry Project with NPL and 12 international manufacturers, integraters and users of biometric systems was set up to evaluate the performance of innovative biometric identification technologies such as skin texture, vein pattern and 3-D face recognition, together with more established techniques such as iris recognition and 2-D face imaging. The comparison demonstrated that biometric performance has improved significantly over the last six years and will enable manufacturers to improve their products further. A database of biometric data has been produced that can be used for development of methods for combination biometrics where different test are used together to improve recognition, and for future evaluations. The project also made progress in developing standards for biometric system performance testing and reporting.

Key findings

Towards the end of this program an independent assessment was commissioned [1] to examine the nature and scale of the impact and to learn key lessons.

- **Projects were proposed by scientists and industry –** just over half of the projects were instigated by the NMI scientist. We would expect this balance to be more end-user lead as the scheme becomes more established.
- **Measurement for Innovators is reaching first time users of the NMIs –** one third of Measurement for Innovators participants had not worked with an NMI prior to using the program. Consultancies, in particular proved a good introductory project for organisations working with an NMI for the first time as they are very low risk to the organisation.
- **The program allows organisations to take business risks –** Measurement for Innovators appears to be helping organisations further riskier or more speculative problems. Over two thirds of participants agreed that the support of the NMI helped them to reduce the risk associated with the development.
- **Project administration is simple –** a majority of participants agreed that the level of paperwork was appropriate. Secondments and consultancies involved a 1-2 page application form and approval was often given on the same day. JIPs were more complex and therefore required more details. Decisions were taken by a management board that met every 2-3 months.
- **Outcomes are achieved where businesses' expectations are managed –** multi-partner collaborative projects are difficult to manage effectively
- **Measurement for Innovators encourages innovation –** around a quarter of participants say they have developed a new product or process and another quarter say that they have proved a new product or process. A fifth of participants have already been able to market the new or improved product or process that was developed or proved in the Measurement for Innovators project. Half of these organisations have accessed new markets with these innovations, the other half have marketed to their existing customers.
- **Measurement for Innovators provides business benefits –** just over half of participant organisations have achieved business benefits. Participants in the program have seen a total annual sales increase of £5.3 million and a total annual profit increase of £5 million. In addition, 60% of participants reported that the program had saved them time using the skills and resources of the NMI and 40% of participants had made cost savings using the skills and resources of the NMIs.

- **Measurement for Innovators facilitates knowledge transfer between industry and NMIs –** the largest benefits for NMI scientists were knowledge

transfer benefits (cited by nearly all scientists). In addition, 51% of participants reported an increase in skills in a measurement area.

- **Measurement for Innovators establishes relationships between NMIs and industry –** Measurement for Innovators strengthens the relationship between the NMIs and industry on three levels:
  - raising awareness – 70% of participants report raised awareness of the NMIs and what they can offer industry,
  - increasing understanding – increasing scientists' understanding of measurement in a business context and increasing participants' understanding of how they can work with the NMIs and the benefits of doing so (reported by 70%),
  - increased interaction – half of scientists say that the time they spend working with industry has increased since the introduction of the Measurement for Innovators program and 69% of participants have continued their relationship with the NMI subsequent to the support they have received under the Measurement for Innovators program.

Outcomes and impact

End user organisations have achieved a number of different outcomes from the programs as shown in Figure 2 below.

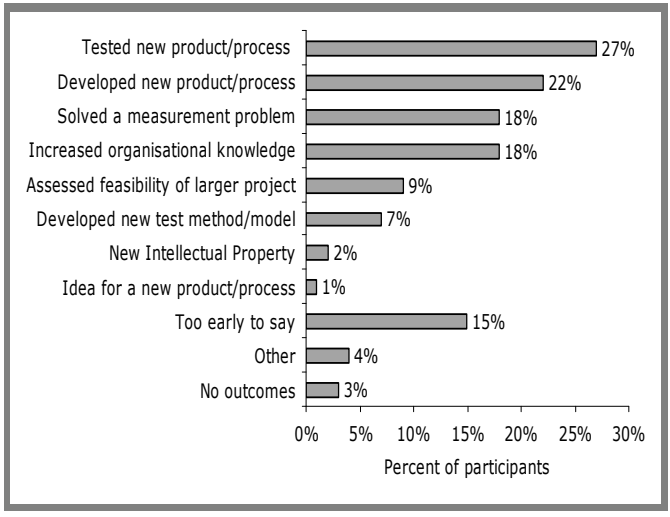


Figure 2. *Impacts recorded by participants in MfI*

Aside from these specific technical outcomes, the association with the NMI is reported to:

- raise the credibility of the innovation and thereby increase the reliability of the product from their customers' perspective;
- speed up the time to market, as the NMI has the expertise to develop the product, process or test model quicker than the organisation on its own.

Even though it is at a very early stage, some companies could already identify bottom line benefits, as shown in Table 3.

NMI support	Total annual profit increase (£000)
Essential	3,784
Helpful	1,188

Table 3. *Profit increases identified*

In addition, two thirds of participants reported that the program had saved them time. These participants could have done some or all of the work themselves, but using the skills and resources of the NMI, the outcome had been achieved in a much shorter timescale. These participants were then asked to estimate how much time the program had saved them and two thirds of these could provide an estimate of time saved, saving a total of 3,600 days:

- participants of secondments saved an average of seven person months;
- participants of JIPs saved an average of five person months;
- participants of consultancies saved an average of six person weeks.

Over one third of the participants had made cost savings using the skills and resources of the NMIs e.g. using equipment owned by the NMI rather than purchasing the equipment. The total amount of money saved through using the skills and resources of the NMIs through Measurement for Innovators is £850,000 an average of £15,000 for those who have made a cost saving.

In total, the value of the program to participants is around £11 million (this includes the value of sales and profit margin increases resulting from the outcomes of the program and cost savings to participants).

The largest personal benefits were improved knowledge for both the individuals from the participating organisations **and the NMI scientists.** The interaction with the participant's organisation as part of the program had helped the scientists refine their work or increase their technical understanding through:

- better understanding the measurement problems faced by industry;
- increase their knowledge in a measurement area that they were unfamiliar with before;
- find out more about the business context of a measurement issue they were researching.

This contextual information will continue to deliver benefits for the organisation.

Summary

The Measurement for Innovators program has been successful in supporting innovation, drawing new companies into a relationship with the NMIs and increasing the understanding of the business context for scientists. Demand has far outstripped resources dedicated to the pilot program and a much larger (£14.5 million) program has just commenced.

More information on the program, including many other examples of success stories can be found at the Measurement for Innovators website www.npl.co.uk/measurement_for_innovators/.

References

[1] C Michaelis and M McGuire, “Measurement for Innovators Impact Assessment”, an independent assessment carried out by Databuild Ltd.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Uncertainty

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Variance model components for the analysis of repeated measurements

E. Filipe

Instituto Português da Qualidade,
Rua António Gião 2, 2829-513 Caparica, Portugal
efilipe@mail.ipq.pt

ABSTRACT: Metrology laboratories, especially those concerned with Scientific Metrology, perform the calibration of the reference and working standards for the dissemination of the related unit. This calibration work is designed in order to control or evaluate the different sources of variability. The experimental design is a statistical tool concerned with the planning of the experiments in order to obtain the maximum amount of information from the available resources and may be applied to metrology, especially for the analysis of a large amount of repeated measurements. An application example of a calibration-nested structure with the corresponding estimation of the variance components is described.
KEYWORDS: experimental design, uncertainty, nested.

Introduction

Experimental design is a statistical tool concerned with the planning of experiments in order to obtain the maximum amount of information from the available resources [1]. This tool is used generally for the improvement and optimization of processes, where the experimenter, controlling the changes in the inputs and observing the corresponding changes in the outputs, is able to make the inference by rejecting the *zero hypothesis* (H_0) of the outputs statistically different for a significance level α, also known as the "producer's" risk.

In addition, it can be used to test the homogenity of a sample(s) for the same significance level, to identify the results that can be considered as "outliers", or to evaluate the variance components between the "controllable" factors.

This tool may be applied to metrology especially for the analysis of a large amount of repeated measurements [2] in short-term repeatability, day-to-day and long-term reproducibility, and include this "time-dependent variability source" information at the uncertainty calculation.

These uncertainty component are evaluated by the statistical analysis of the results obtained from the experiments using the Analysis of Variance (ANOVA) for designs consisting of nested or hierarchical (ISO 3534-3, §2.6) [3] sequences of measurements. The Analysis of Variance as defined by the standard ISO 3534-3,

§3.4, "is a technique that subdivides the total variation of a response variable into meaningful components, associated with specific sources of variation".

An application example in a calibration-nested structure with the corresponding estimation of the variance components is described.

Purpose

Metrology laboratories, especially those concerned with Scientific Metrology, perform the calibration of the reference and working standards for the dissemination of the related unit.

This calibration work is designed in order to control or evaluate the different sources of variability. These laboratories usually have several standards and these are compared in a regular basis in order to have a record of its regular behavior. When the laboratory performs the calibration of an external standard the acquired knowledge is included in this new calibration through the written procedures and experience allowing the validation of the obtained measurements.

The nested or hierarchical design. General model

The nested design is defined [3, 4] as "the experimental design in which each level of a given factor appears in only a single level of any other factor". The objective of this model is to deduce the values of the variance components that cannot be measured directly. The factors are hierarchized like a "tree" (see Figure 1) and any path from the "trunk" to the "extreme branches" will find the same number of nodes.

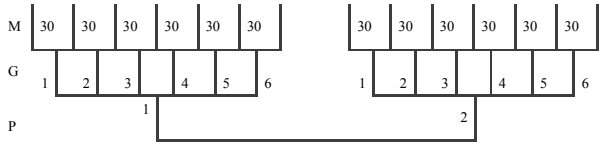


Figure 1. *A [2 × 6 × 30] nested or hierarchical design*

In the calibration experiment to be described, the calibration of a 25 ohm standard platinum resistance thermometer (SPRT) at the aluminum point, the factors are the repeated measurements obtained at regular intervals and the plateau/run measurements. These factors are considered as *random samples* of the *population* from which we are interested to draw conclusions.

In this design, each factor is analyzed with the one-way analysis of variance model (ANOVA), nested in the subsequent factor.

Model for one factor

Considering firstly only one factor with a different levels taken randomly from a large population, [1, 5] any observation made at the i^{th} factor level with j observations, will be denoted by y_{ij}.

The mathematical model that describes the set of data is:

$$y_{ij} = M_i + \varepsilon_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (i = 1, 2, ... a \quad and \quad j = 1, 2, ... n) \tag{1}$$

where M_i is the expected (random) value of the group of observations i, μ the overall mean, τ_i the parameter associated with the i^{th} factor level designated by i^{th} factor effect and ε_{ij} the random error component. This model with random factors is called the random effects or components-of-variance model.

For hypothesis testing, the errors and the factor effects are assumed to be normally and independently distributed, respectively with mean zero and variance σ^2 or $\varepsilon_{ij} \sim N(0, \sigma^2)$ and with mean zero and variance σ_τ^2 or $\tau_i \sim N(0, \sigma_\tau^2)$. The variance of any observation is composed by the sum of the variance components, according to:

$$\sigma_y^2 = \sigma_\tau^2 + \sigma^2 \tag{2}$$

The test is unilateral and the hypotheses are:

$$\begin{vmatrix} H_0 : \sigma_\tau^2 = 0 \\ H_1 : \sigma_\tau^2 > 0 \end{vmatrix} \tag{3}$$

i.e., if the zero hypothesis is true, all factors effects are "equal" and each observation is made up of the overall mean plus the random error $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The total sum of squares, which is a measure of total variability in the data, may be expressed by:

$$\begin{aligned} &\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}\left[(\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)\right]^2 = \\ &n\sum_{i=1}^{a}(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_i - \bar{y})(y_{ij} - \bar{y}_i) \end{aligned} \tag{4}$$

As the cross-product is zero [6], the total variability of data (SS_T) can be separated into: the sum of squares of differences between factor-levels averages and the grand average (SS_{Factor}), a measure of the differences between factor-level, and the sum of squares of the differences of observations within a factor-level from the factor-levels average (SS_E), due to the random error. Dividing each sum of squares by the respective degrees of freedom, we obtain the corresponding mean squares (MS):

$$MS_{Factor} = \frac{n}{a-1}\sum_{i=1}^{a}(\bar{y}_i - \bar{y})^2$$
$$MS_{Error} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2}{a(n-1)} = \sigma^2 \tag{5}$$

The mean square between factor-levels (*MSFactor*) [7] is an unbiased estimate of the variance σ^2, if H_0 is true, or a overestimate of σ^2 (see equation 7), if it is false. The mean square within factor (error) (MS_{Error}) is always the unbiased estimate of the variance σ^2.

In order to test the hypotheses, we use the statistic:

$$F_0 = \frac{MS_{Factor}}{MS_{Error}} \sim F_{\alpha,\, a-1,\, a(n-1)} \tag{6}$$

where F[1], the *Fisher-Snedecor* sampling distribution with a and $a \times (n\text{-}1)$ degrees of freedom.

If $F_0 > F_{\alpha,\, a\text{-}1,\, a\,(n\text{-}1)}$, we reject the zero hypothesis and conclude that the variance σ_τ^2 is significantly different from zero, for a significance level α.

The expected value of the MS_{Factor} is [8]:

$$E(MS_{Factor}) = E\left[\frac{n}{a-1}\sum_{i=1}^{a}(\bar{y}_i - \bar{y})^2\right] = \sigma^2 + n\sigma_\tau^{\,2} \tag{7}$$

The variance component of the factor is then obtained by:

$$\sigma_\tau^2 = \frac{E(MS_{Factor}) - \sigma^2}{n} \tag{8}$$

1 F distribution – sampling distribution. If χ_u^2 and χ_v^2 are two independent chi-square random variables with u and v degrees of freedom, then its *ratio* $F_{u,v}$ is distributed as F with u numerator and v denominator degrees of freedom.

Model for two factors

Considering now a two "stage" nested design as the structure of the example to be described, the mathematical model is:

$$y_{pgm} = \mu + \Pi_p + \Gamma_g + \varepsilon_{pgm} \tag{9}$$

where y_{pgm} is the $(pgm)^{\text{th}}$ observation, μ the overall mean, Π_p the P^{th} random level effect, Γ_g the G^{th} random level effect and ε_{pgm} the random error component.

The errors and the level effects are assumed to be normally and independently distributed, respectively with mean zero and variance σ^2 or $\varepsilon_{pgm} \sim N(0,\ \sigma^2)$ and with mean zero and variances $\sigma_P{}^2$ and $\sigma_G{}^2$. The variance of any observation is composed by the sum of the variance components and the total number of measurements, *N,* is obtained by the product of the dimensions of the factors ($N = P \times G \times M$).

The total variability of the data [4, 9] can be expressed by:

$$\begin{aligned} &\sum_p \sum_g \sum_m (y_{pdtm} - \bar{y})^2 = \sum_p GM(\bar{y}_p - \bar{y})^2 + \\ &+ \sum_p \sum_g M(\bar{y}_{pg} - \bar{y}_p)^2 + \sum_p \sum_d \sum_t (\bar{y}_{pgm} - \bar{y}_{pg})^2 \\ &or \\ &SS_T = SS_P + SS_{G|P} + SS_E \end{aligned} \tag{10}$$

This total variability of the data is the sum of squares of factor P (SS_P), the P - factor effect, plus the sum of squares of factor G for the same P ($SS_{G|P}$) and SS_E, the residual variation. Dividing by the respective degrees of freedom, $(P - 1)$, $P \times (G - 1)$ and $P \times G \times (M - 1)$ we obtain the mean squares of the nested factors, which are estimates of σ^2, if there were no variations due to the factors. The estimates of the components of the variance are obtained by equating the mean squares to their expected values and solving the resulting equations:

$$\begin{aligned} E(MS_P) &= E\left[\frac{SS_P}{P-1}\right] = \sigma^2 + M\sigma_G{}^2 + GM\sigma_P{}^2 \\ E(MS_{G|P}) &= E\left[\frac{SS_{G|P}}{P(G-1)}\right] = \sigma^2 + M\sigma_G{}^2 \\ E(MS_E) &= E\left[\frac{SS_E}{PG(M-1)}\right] = \sigma^2 \end{aligned} \tag{11}$$

Calibration of a SPRT at the Aluminium Freezing Point

Here we present the results/evaluation of two plateaus used for the calibration of a 25 ohm SPRT at the aluminum freezing point (660.323ºC), performed on two different days. The measurements were taken in sequences of 30 values with a 5 minute interval between each sequence and 6 groups of data were obtained, during approximately 1 hour. At each plateau/run the measurements started one hour after the stabilization of the freeze. These freeze plateaus usually last more than 10 hours.

The 360 values grouped by the 6 groups at each plateau were analyzed in order to study the stability of each plateau and include this information in the evaluation, by a type A method, of the uncertainty components.

From the measurements were obtained the values W_{Al} (ratio between the measured resistance at the aluminum point and the measured resistance at the triple point of water, $t = 0.01$ºC), corrected by the hydrostatic pressure and extrapolated to a current of 0 mA.

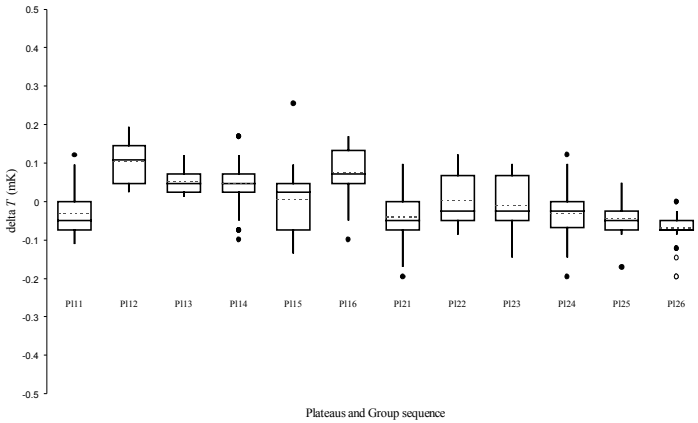


Figure 2. *Boxplot diagrams of the groups of measurements from the first and second plateaus. Dot line – average of each group measurements*

The Boxplot diagrams of the groups are shown (see Figure 2) enabling a "view" of the variability of the measurements and its distribution symmetry. Indeed in a single chart, they display the first quartile, median, third quartile, the interquartile range, minimum and maximum values, that are not considered as outliers (whiskers), and also plot the "suspected" outliers as considered by John Tukey, the values outside of the ±1.5 × interquartile range (the difference between the upper and lower quartile), are marked. The full line corresponds to the median and the dotted line to the average of the measurements.

We verify that all the values are included in an interval of ±0.3 mK which at first glance is very acceptable for a SPRT calibration at the aluminum point.

Analyzing this data with the analysis of variance (ANOVA) as described in [10] we see (Table 1) that there is an effect between the plateaus and between the groups of measurements.

Equating the mean squares to their expected values, we can calculate the variance components and include them in the budget of the components of uncertainty drawn at Table 2.

Source of variation	Sum of squares	Degrees of freedom	Mean square	F_0	Critical values $F_{v1, v2}$
Plateaus	0.5037	1	0.5037	11.4433	4.9646
Groups	0.4402	10	0.0440	10.6483	1.8579
Measurements	1.4386	348	0.0041		
Total	2.3825	359			

Table 1. *Analysis of variance table of the complete set of measurements;* $\alpha = 5\%$

Expected value of mean square	Components of Type A uncertainty	Variance $(mK)^2$	Standard - deviation (mK)
$\sigma^2+30\sigma_I^2+180\sigma_P^2$	Plateaus	0.003	0.05
$\sigma^2+30\sigma_I^2$	Groups	0.001	0.04
σ^2	Measurements	0.004	0.06
	Total	0.008	0.09

Table 2. *Uncertainty budget for the components of uncertainty evaluated by a Type A method of the mean* $W_{Al} = 3.37579875$

For the purpose of the calibration of the SPRT these values are acceptable as they are consistent with the experimental equipment, the aluminum cell and the stability of the SPRTs at this temperature.

Residual Analysis

As mentioned (3.1 and 3.2) that, in this model, the errors and the level effects are assumed to be normally and independently distributed. This assumption was checked drawing a normality plot (Henry line) of the obtained residuals (see Figure 3).

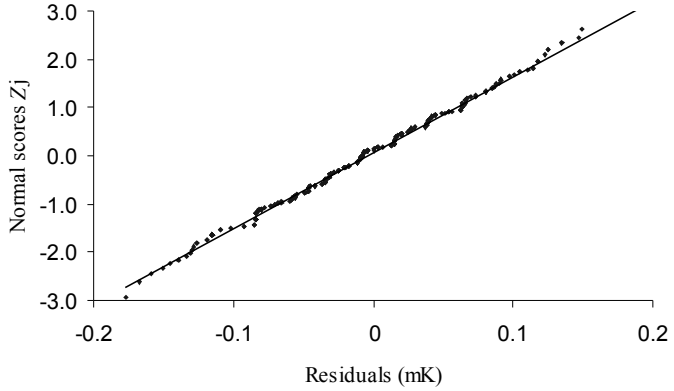


Figure 3. *Normal probability plot of the residuals of the experiment*

Remarks

At the aluminum point the SPRTs suffer stability problems that can be due to an insufficient previous annealing or to some "quenching" when withdrawing the SPRTs from the hot furnace. Repeating the plateaus allows us to verify the stability of the measuring instrument.

The plateau values are continuously taken where previously analyzed in terms of its normality. The size of the samples and the time interval were procedure options. This stability studies are important for the validation of the obtained values.

This model for "Type A" uncertainty evaluation that takes into account these time-dependent sources of variability is foreseen at the GUM [1]. The obtained value using this nested design u_A= 0.09 mK (Table 1) is considerably larger than that obtained by calculating the standard deviation of the mean of the 360 measurements u_A= 0.0043 mK. This last approach is generally used and evidently underestimates this component of uncertainty.

Conclusion

The nested-hierarchical design was described as a tool to identify and evaluate components of uncertainty arising from random effects.

An application of the design has been drafted to illustrate the variance components analysis in a SPRT calibration at the aluminum point in a two stage nested model. This same model can be applied to other number of factors, easily treated in an *Excel* spreadsheet.

References

[1] G. A. Milliken, D. E. Johnson, *Analysis of Messy Data*. Vol. I: *Designed Experiments*. 1st ed., London, Chapman & Hall, 1997.

[2] BIPM *et al*., *Guide to the Expression of Uncertainty in Measurement (GUM)*, 2nd ed., International Organization for Standardization, Geneva, 1995, pp. 11, 83-87.

[3] ISO 3534-3, Statistics – *Vocabulary and Symbols – Part 3: Design of Experiments*, 2nd ed., Geneva, International Organization for Standardization, 1999, pp. 31 (2.6) and 40-42 (3.4).

[4] ISO TS 21749 Measurement uncertainty for metrological applications — Simple replication and nested experiments Geneva, International Organization for Standardization, 2005

[5] G.E.P. Box, W.G. Hunter, J.S. Hunter, *Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building*, 1st ed., New York, John Wiley & Sons, 1978, pp. 571-582

[6] R. C. Guimarães, J. S. Cabral, *Estatística*, 1st ed., Amadora: McGraw Hill de Portugal, 1999, pp. 444-480

[7] B. Murteira, *Probabilidades e Estatística*. Vol. II, 2nd ed., Amadora, McGraw Hill de Portugal, 1990, pp. 361-364.

[8] D. Montgomery, *Introduction to Statistical Quality Control*, 3rd ed., New York, John Wiley & Sons, 1996, pp. 496-499.

[9] J. Poirier, "Analyse de la Variance et de la Régression. Plans d'Expérience", *Techniques de l'Ingénieur,* **R1**, 1993, pp. R260-1 to R260-23.

[10] E. Filipe, "Evaluation of Standard Uncertainties in nested structures", *Advanced Mathematical & Computational Tools in Metrology VII*, Series on Advances in Mathematics for Applied Sciences – Vol. 72, World Scientific, 2006, pp. 151-160.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Optimized measurement uncertainty and decision-making

L.R. Pendrill[1]

[1]SP Technical Research Institute of Sweden, Measurement Technology, Box 857, SE-50115 Borås, Sweden
Leslie.pendrill@sp.se

ABSTRACT: This work re-examines, in the new light provided by a decision-theory approach, traditional rules in conformity assessment of setting limits on measurement uncertainty and risks – including maximum permissible uncertainty – specified as a fixed fraction (typically 1/3) of the region of permissible values – the "shared risk" principle, as well as operating characteristics in traditional statistical sampling planning. One of the first applications of an optimized uncertainty methodology in testing and sampling by attribute is reported. This goes beyond traditional statistical sampling plans in that it assigns costs to both testing and consequences of incorrect decision-making. Examples are taken from the measurement of utilities, the environment and pre-packaged goods covered by legal metrology – where costs are in most cases reasonably well defined for both consumer, producer and authorities – as well as other important conformity assessment areas.
KEYWORDS: Measurement – Uncertainty – Sampling – Decision theory – Conformity assessment – Risks – Costs

Introduction

There is increasing interest in various sectors of society in conformity assessment of product in general, and efforts are underway to further develop a uniform approach to conformity assessment – for instance, in the European Union [EU Commission 2006]. This is important when regulating safety and other essential collective protection requirements for public interest issues, such as health, consumer or environmental protection etc. with regard to potential risks which freely-marketed products could present while at the same time enhancing innovation, growth and welfare.

Product measurement and testing provide valuable, often quantitative, evidence on which conformity decisions can be based. Many product tests are however made in practical situations where time and resources are limited. A balance has to be struck between expenditure in product measuring and testing and the potential costs associated with various risks – to both producer and consumer – associated with incorrect decisions arising from limited measurement accuracy and test uncertainties [Williams and Hawkins 1993, Thompson and Fearn 1996, AFNOR 2004, Pendrill 2006, 2007, Pendrill and Källgren 2006].

While much effort has been expended in the metrology community over the last decade or so in giving guidance about and harmonizing *how* to evaluate measurement uncertainty [JCGM GUM 1995], arguably less attention has been paid for the *use* of uncertainty in conformity assessment and many questions remain as yet unanswered, such as a clear and objective rule for decision-making, which is arguably of prime interest to stakeholders in general [Pendrill and Källgren 2006].

The current presentation covers the treatment of measurement errors and test uncertainties in the context of conformity assessment of product, in particular, formulating a general model in statistical and economic terms and highlighting an increasing interest and enhanced insight into decision-making gained when extending classical, purely statistical treatment of consumer and producer risks [ILAC 1996], towards a more "stakeholder" motivated (and ultimately optimized) approach in terms of effective costs associated with manufacture, testing and incorrect decision-making.

The application of a unified approach is a step towards establishing clearer procedures for setting and specifying tolerances and associated uncertainties, and in facilitating acceptance of conformity by both customer and supplier. In the context of conformity assessment, there is a need to identify, in addition to essential requirements for each specific sector, even important "horizontal" technical issues which, together with legislative issues, span a range of sectors.

Costs, economic and statistical risks in conformity assessment

Arguably of most concern (for both consumer and producer) are the product value and the estimation of various potential losses associated with, not only regular manufacture, but also with incorrect decision-making in conformity assessment.

A general formulation of the overall profit [Williams and Hawkins 1993, Pendrill 2007] is in terms of a sum of the various incomes and losses shown in equation (1):

$$\begin{aligned} Profit \quad &= \sum_{\hat{x} \le a} [\left(k_{\mathrm{P},S}(\hat{x}) - k_{\mathrm{P},II}(\hat{x}) \cdot R(\hat{x})\right) \cdot \Pr_{entity}(\hat{x})] \\ &- \sum_{\hat{x}} [\left(k_{\mathrm{P}}(\hat{x}) + k_{\mathrm{M}}(\hat{x})\right) \cdot \Pr_{entity}(\hat{x})] - \sum_{\hat{x} > a} [\left(k_{\mathrm{P},I}(\hat{x}) + k_{\mathrm{M},I}(\hat{x})\right) \cdot R^{*}(\hat{x})] \cdot \Pr_{entity}(\hat{x}) \end{aligned} \quad (1)$$

which is, term-by-term, the sum of:

a) income from sales of passed, conforming product: $\sum_{\hat{x} \le a} \left(k_{\mathrm{P},S}(\hat{x})\right) \cdot \Pr_{entity}(\hat{x})$

b) loss associated with consumer risk (passed, non-conforming product): $\sum_{\hat{x} \le a} \left(k_{\mathrm{P},II}(\hat{x}) \cdot R(\hat{x})\right) \cdot \Pr_{entity}(\hat{x})$

where the classical "Consumer" (type II or "false accept" decision error [Montgomery 1996]) in purely statistical terms where non-conforming products are incorrectly accepted is expressed as a global consumer's statistical risk $R = \sum_{\hat{x} \le a} R(\hat{x}) \cdot \Pr_{entity}(\hat{x})$ [Rossi and Crenna 2006]

where $R(\hat{x}) = \Phi_{test}(x > a|\hat{x})$ is the specific consumer statistical risk, associated with a test uncertainty distribution, in terms of the (cumulative) probability Φ_{test} of an entity subject to conformity assessment lying outside specification, when the mean value $\hat{x}$ for an individual product is inside a specification limit, *a*;

a) cost of (all) manufacturing product (exclusive test): $\sum_{\hat{x}} \left(k_{\mathrm{P}}(\hat{x})\right) \cdot \Pr_{entity}(\hat{x})$

b) cost of testing (all) product: $\sum_{\hat{x}} \left(k_{\mathrm{M}}(\hat{x})\right) \cdot \Pr_{entity}(\hat{x})$

c) loss associated with re-manufacturing with producer risk (failed, conforming product): $\sum_{\hat{x}>a} [\left(k_{\mathrm{P},I}(\hat{x})\right) \cdot R^*(\hat{x})] \cdot \Pr_{entity}(\hat{x})$

d) loss associated with re-testing with producer risk (failed, conforming product): $\sum_{\hat{x}>a} [\left(k_{\mathrm{M},I}(\hat{x})\right) \cdot R^*(\hat{x})] \cdot \Pr_{entity}(\hat{x})$

The sum is over all of the products (and becomes an integral for continuous varying product) and the classical "producer" risk (type I or "false reject" decision error [Montgomery 1996]) where conforming products are incorrectly rejected is expressed as a global producer's statistical risk $R^* = \sum_{\hat{x}>a} R^*(\hat{x}) \cdot \Pr_{entity}(\hat{x})$ [Rossi and Crenna 2006].

Here $R^*(\hat{x}) = \Phi_{test}(x \le a|\hat{x})$ is the specific producer statistical risk, associated with a test uncertainty distribution, in terms of the (cumulative) probability Φ_{test} of an entity subject to conformity assessment lying within specification, when the mean value $\hat{x}$ for an individual product is outside a specification limit, *a*.

Consequence, testing and sampling costs

Entity		*Costs*
Product	$k_{P,S}$	Product sale price
Producer cost	k_P	Cost of manufacturing product (excluding testing)
	$k_{P,I}$	Cost of re-manufacturing product (excluding testing)
Consumer costs	$k_{P,II}$	Cost of incorrectly accepted non-conforming product
Measurement costs	k_M	Cost of testing product (including calibration, certification, etc)
	$k_{M,I}$	Cost of re-testing product

Table 1. *Examples of different cost factors*

Product and consequence costs

Estimates of product value and how it varies with product error (deviation from nominal) enter into a number of terms in equation (1), for example $k_{\mathrm{P},S}(\hat{x})$.

Depending on how product quality is valued, there will be different models of how product value varies with product error [Pendrill 2007], ranging from constant value in some cases to non-linearly varying costs, such as in the Taguchi [1993] loss function model of customer (dis-)satisfaction.

For products valued by quantity, costs will in many cases vary linearly with entity value. Examples are common, for instance in the metering of utilities, energy, fuels [Pendrill and Källgren 2006a] and environmental emissions [Pendrill and Källgren 2006b], where product price is quoted per quantity (€/kg, for example). In the simplest case, a consequence cost will be proportional to the amount of product sold times the unit price per quantity.

The conformity assessment of pre-packaged goods also belongs to this category where consumer costs vary linearly with product error. Pre-packaged goods cover a wide range of commodities and levels of trade are rapidly expanding, in part due to changing consumer tastes and advances in production, transport and other supply-chain technologies, as noted by Birch [2007].

Testing costs

Another consideration is how to model the higher cost of increased efforts made to reduce measurement uncertainty. There are of course a number of conceivable

models of how test costs could vary with measurement uncertainty [Fearn *et al.* 2002]. In this work, the test cost is assumed to depend inversely on the squared (standard) measurement uncertainty $u^2_{measure}$, i.e.,

$$k_{\mathrm{M},I}(x) = \frac{D}{u^2_{measure}} \tag{2}$$

where D is the test cost at nominal test (standard) uncertainty $u_{measure}$. Such a model was suggested by Thompson and Fearn [1996], based mainly on the argument that N repeated measurements would reduce the standard deviation in the measurement result by $\sqrt{N}$ while costing (at most) N times as much as each individual measurement. In this paper, this model of measurement costs is not only used where the statistical distribution associated with measurement and sampling uncertainties are known (such a type A evaluation) but is also extended to cover more generally even other components of uncertainty (including the expanded uncertainty in the overall final measurement result) where the underlying statistical distribution is often not known (type B evaluation).

Where re-testing becomes necessary, as in producer risk, it is assumed here that re-test costs will be approximately the same cost as ordinary testing of product, although we could imagine scenarios where re-test costs might be different.

Optimized uncertainty methodology

An optimized uncertainty methodology, in which the analysis costs are balanced against the costs associated with the consequences of incorrect decision-making was introduced by Thompson, Fearn and co-workers (Thompson and Fearn [1996], Ramsey *et al.* [2001]) who sought a criterion for "fitness for purpose" in analytical measurement.

Generally, test costs increase when attempting to reduce consumer consequence costs – and *vice versa* – and a trade off can therefore be made between some of the cost terms in the general model shown in equation (1). In this case, balancing test costs against consequence costs leads to a U-shaped cost curve according to test uncertainty and the possibility of identifying an optimal test uncertainty where costs are minimized, in particular, trading the cost of testing (all) product against the loss associated with consumer risk (passed, non-conforming product).

$$E_{np,linear} = \left(k_{\mathrm{M}}(x) + k_{\mathrm{P},II}(x) \cdot R(\hat{x})\right) = \frac{D_{np}}{u^2_{measure}} + C_{np} \cdot \left[\int_{USL_x}^{\infty} x \cdot \frac{1}{\sqrt{2\pi} \cdot u_{measure}} \cdot e^{-\frac{(x - x_{measure})^2}{2 \cdot u^2_{measure}}} \cdot dx \right] \tag{3}$$

Consumer risk by variable

As an application of an optimized uncertainty methodology in which test costs are balanced against consequence costs, we take the case of household electricity meters, where the costs of instrument error are estimated as the consequence cost, C_{np}, of non-conforming instruments which requires an estimate of the annual cost based in the present case on the taxed energy consumption.

The overall cost, E_{np}, of consumer risk for a single specimen takes the form shown in equation (3) for cases where consumer costs vary linearly with entity error, exemplified here for the case of consumer risk with respect to an upper specification limit, *USL*, and where the distribution of entity value under test is dominated by measurement uncertainty, $u_{measure}$ [Pendrill 2006].

Evaluation of expression (3) for the overall cost, E_{np}, of consumer risk leads to plots of cost versus test uncertainty. Opportunities to optimize test uncertainties at a level corresponding to a minimum in overall costs are evident from Figure 1 shown for the case of electricity meters [Pendrill 2007]. At an instrument error $y_{measure} = USL + u_{measure}$, (where $USL = + 2\%$) if the standard measurement uncertainty can be reduced[1] below $MPU/2 = 0.36$ %, this will only lead to increasing test costs. On the other hand, relaxing measurement uncertainties towards higher levels will allow overall costs to approach a minimum of about 10 € at about 0.5% test uncertainty, before costs rise again at higher uncertainty levels as consequence costs become significant. The corresponding losses at other levels of instrument error are, as expected, less for test results more clearly within the zone of permissible values (integrating costs over the known distribution of instrument errors (discussed in [Pendrill 2006]) leads to the "dashed-dot" curve in Figure 1).

This example gives, as we have earlier pointed out [Pendrill 2006], added significance to the traditional rule in conformity assessment of limiting measurement uncertainty to some fraction of the region of permissible values. From Figure 1, it is evident that, although the economic losses only vary slowly with uncertainty, if we were to allow the test uncertainty to increase beyond the target, or maximum permissible uncertainty, *MPU* (=1/3·*MPE*), then additional costs would be incurred compared with the optimum. In this context, it will be a matter of the two parties – producer and consumer – in the conformity assessment, coming to a mutually acceptable agreement as to how risks are to be shared.

For both consumer and producer, risks in economic terms are arguably more tangible than more traditional percentage risk estimates.

1 There are of course practical limits to how far test uncertainties can be reduced.

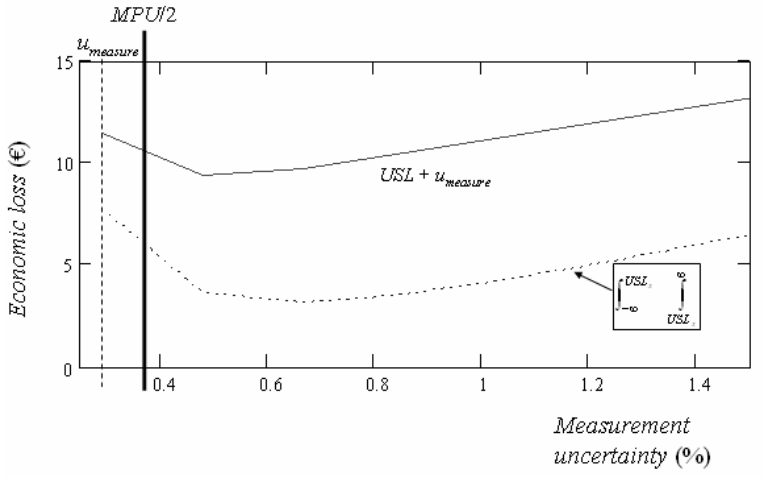


Figure 1. *Consumer risk costs per meter (€/a) as a function of standard measurement uncertainty for electricity meters at one level (upper curve) and a distribution of instrument error at a nominal load of 0.5I_{max} [Pendrill 2007]*

Estimates of consumer risk could also include the product warranty costs; failure analysis costs; potential loss of future sales; environmentally induced waste caused by defective products, etc. In those cases, it may be appropriate to use other models of how consequences vary with product error [Pendrill 2007].

Attribute consumer risks

Analogously, an optimized sampling uncertainty can be derived from the sum of sampling costs and consequence costs according to the expression:

$$E_{USL,attr} = \frac{D_{np}}{u^2_{sample}} + C_{USL} \cdot \Phi_{binomial}(d, n, SL_{\hat{p}}) \tag{4}$$

modeled on a statistical uncertainty in the number of non-conforming units, where the probability of selecting d conforming and $(n - d)$ non-conforming entities when a sample of size n follows a binomial distribution.

Figure 2 illustrates how overall costs vary with the number of instruments sampled (assuming an infinite population) and how an optimized sampling uncertainty [Pendrill 2006] can be identified in relation to traditional sampling planning limits *AQL* and *LQL* [Montgomery 1996] for the example of pre-packaged coffee powder.

This new optimized sampling uncertainty methodology, extends traditional attribute sampling plans to include economic assessments of the costs of measuring, testing and sampling together with the costs of incorrect decision-making.

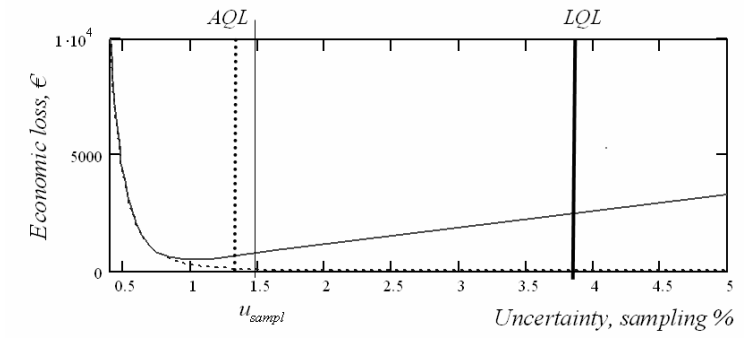


Figure 2. *Costs according to attribute sampling uncertainty for pre-packaged coffee powder*

Conclusions

Wherever there are uncertainties, there are risks of incorrect decision-making in conformity assessment associated both with sampling and measurement. This paper reviews how an earlier criterion for "fitness for purpose" in analytical measurement, as recently introduced by Thompson and co-workers, can be extended to more general measurements in conformity assessment. This approach weighs economic factors, such as the costs of analysis and associated with the consequences of incorrect decision-making, into more traditional percentage analyses of consumer and producer risk. This is considered to be an important step towards establishing clearer procedures for setting and specifying tolerances and associated uncertainties, and in facilitating acceptance of conformity in general by both customer and supplier.

Acknowledgements

Thank you for discussions about the performance of the testing of different measurement instruments with Håkan Källgren and other colleagues at SP.

This work has been partially financed by the European Thematic Network "Advanced Mathematical and Computational Tools in Metrology", NICe project 04039 and by grant 38:10 National Metrology of the Swedish Ministry of Industry, Employment and Communication.

References

1. EU Commission 2006 N560-2 EN 2006-0906 Annex to "A HORIZONTAL LEGISLATIVE APPROACH TO THE HARMONISATION OF LEGISLATION ON INDUSTRIAL PRODUCTS", European Commission, http://ec.europa.eu/enterprise/newapproach/review_en.htm.

2. R H Williams and C F Hawkins 1993 "The Economics of Guardband Placement", *Proceedings, 24th IEEE International Test Conference*, Oct. 17-21, 1993, Baltimore (MD).

3. M Thompson and T Fearn 1996, *Analyst*, **121**, 275 – 8

4. AFNOR "Metrology and statistical applications - Use of uncertainty in measurement: presentation of some examples and common practice", FD x07-022 (2004).

5. L R Pendrill 2006 "Optimized measurement uncertainty and decision-making when sampling by variables or by attribute", *Measurement*, **39**, 829-840, Advanced Mathematical Tools for Measurement in Metrology and Testing, http://dx.doi.org/10.1016/j.measurement.2006.04.014.

6. L R Pendrill 2007 "Optimized Measurement Uncertainty and Decision-Making in Conformity Assessment", *MEASURE* NCSLi, vol. 2, no. 2, pp. 76-86.

7. L R Pendrill and H Källgren 2006a "Decision-making with uncertainty in attribute sampling" *Advanced Mathematical and Computational Tools in Metrology VII* in Series on Advances in Mathematics for Applied Sciences Vol 72 pp. 212 - 220, World Scientific, Singapore, ISBN 981-256-674-0.

8. L R Pendrill and H Källgren 2006b "Exhaust gas analysers and optimized sampling, uncertainties and costs", *Accred. Qual. Assur.* DOI: 10.1007/s00769-006-0163-3 http://dx.doi.org/10.1007/s00769-006-0163-3.

9. JCGM GUM 1995, *Guide to the expression of uncertainty in measurement* edition, 1993, corrected and reprinted 1995, International Organisation for Standardisation (Geneva, Switzerland). ISBN 92-67-10188-9.

10. ILAC guide no. 8 1996 "Guidelines on Assessment and Reporting of Compliance with Specification".

11. G B Rossi and F Crenna 2006 "A probabilistic approach to measurement-based decisions", *Measurement* **39**, 101 – 19.

12. G Taguchi 1993 "*Taguchi on Robust Technology*", New York, ASME Press.

13. D C Montgomery 1996 "Introduction to statistical quality control", J Wiley & Sons, ISBN 0-471-30353-4.

14. Birch J 2007 "The economic importance of legal metrology in pre-packaging", *OIML Bulletin* **XLVIII**, 19 - 24.

15. T Fearn, S Fisher, M Thompson and S Ellison "A decision-theory approach to fitness for purpose in analytical measurement", *Analyst* **127**, 818 – 24 (2002).

16. M H Ramsey, J Lyn and R Wood 2001 *Analyst* **126**, 1777 – 83.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Giving early ideas to tests: uncertainty project implementation in the field of railway testing

V. Regragui

Agence d'essai ferroviaire (AEF)
21 avenue du Président Allende, 94407 VITRY-SUR-SEINE, Cedex, France

ABSTRACT: The Agence d'Essai Ferroviaire provides test engineering services – planning and implementation of experiments and tests – for rolling stock.
300 tests are carried out and more than 70 have COFRAC testing accreditations.
In order to satisfy the requirements of ISO 17025 about uncertainty in measurement, an "uncertainty" project of great width has been implemented.
To date, 140 people are trained, 65% of methods are validated and 13% are about to be validated.
The persons in charge of testing are autonomous and use works on uncertainty as tools for controlling and eventually improving their methods.

Presentation of Agence d'Essai Ferroviaire (AEF)

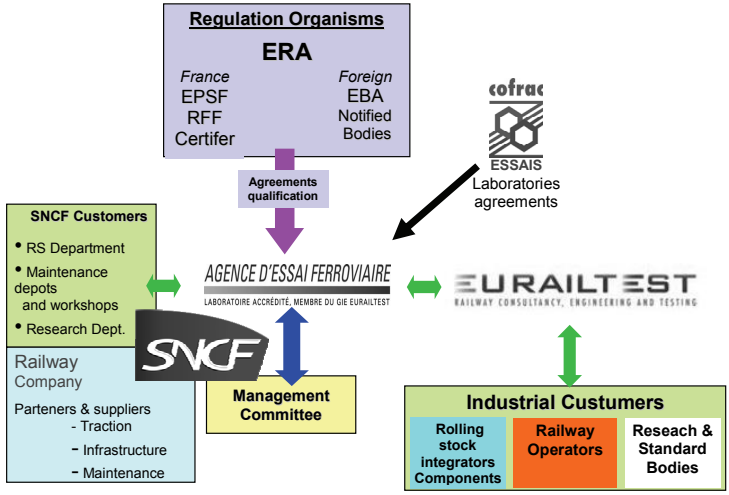
AEF mission

The main goals of Agence d'Essai Ferroviaire are: design, organization, preparation, development and operation of tests and experiments relating to rolling stock, their components and their interactions with the railway industry. For this purpose, scientific, technical and organizational skills as well as infrastructures and technical means necessary to these missions are assembled at Vitry-sur-Seine:

- **company laboratories** for providing to SNCF entities testing and expertise services in research, development or transformation projects for rolling stock as well as maintenance support;

- **railway test operator in the French network**, by ensuring organization and security of test circulations;

- **test and expertise service provider in the competitive railway testing market**, by providing its resources and skills to industrial clients and players in the railway industry (infrastructure operators, certification organizations etc.) with which it would contract work through **GIE EurailTest** ensuring the commercial interface.



Scope of tests

The AEF provides its customers with **test engineering** services where a large part of the added value is design activities based on the knowledge of measurement in the railway industry. These services are based on skills and technical resources developed in the following technical fields.

The different AEF activities lead its personnel to implement its technical skills over the SNCF infrastructures, as well as over foreign infrastructures and **industrial sites of its clients and partners.**

AEF is made up of 250 people broken down as follows: 45% in management, 35% knowledge agents and 25% operations (technician, operator, assistant, administrative).

Measurement uncertainties

Problem

In 2002, a major problem emerged which involved all AEF test scopes: a new requirement is added to the NF ISO 17025 standard, a requirement that needs to be satisfied in order to maintain or obtain the COFRAC accreditation for each test.

In fact, test laboratories must estimate measurement uncertainties relating to each test that is subject to accreditation. What does this requirement imply?

It involves providing a numerical uncertainty value associated with test results based on a recognized methodology.

The AEF controls over 300 different tests, over 70 of them are accredited. The NF ISO 17025 requirement is therefore a major project for AEF.

In order to understand the project better, its justification had to be defined.

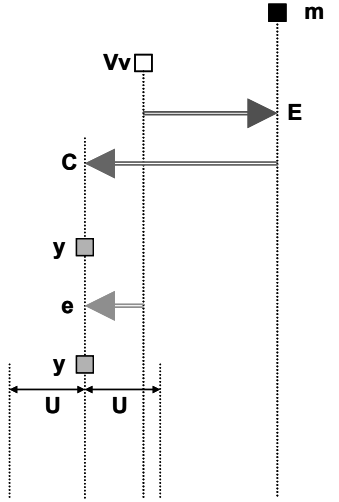
Why calculate measurement uncertainties of our tests?

The response is obvious: so we can provide our client with the dispersion associated with the measurement result offered, to guarantee test control and ensure the declaration of a product's compliance.

The following diagram summarizes the importance of the notion of uncertainty. The object of the test is to measure an unknown value (real value). In order to do this, a test method is created and implemented by human and material means.

Test results are not perfect; they are inevitably subject to dispersion. If the test is done 300 times in repeatability conditions, we will never have 300 identical results. Errors appear in the test process, some are systematic and can be corrected, others are random and cannot be eliminated.

For this reason, the assurance for providing the client with a result equal to real value does not exist, only a measurement result associated with an uncertainty can be provided and guarantee an interval in which the real value is located.



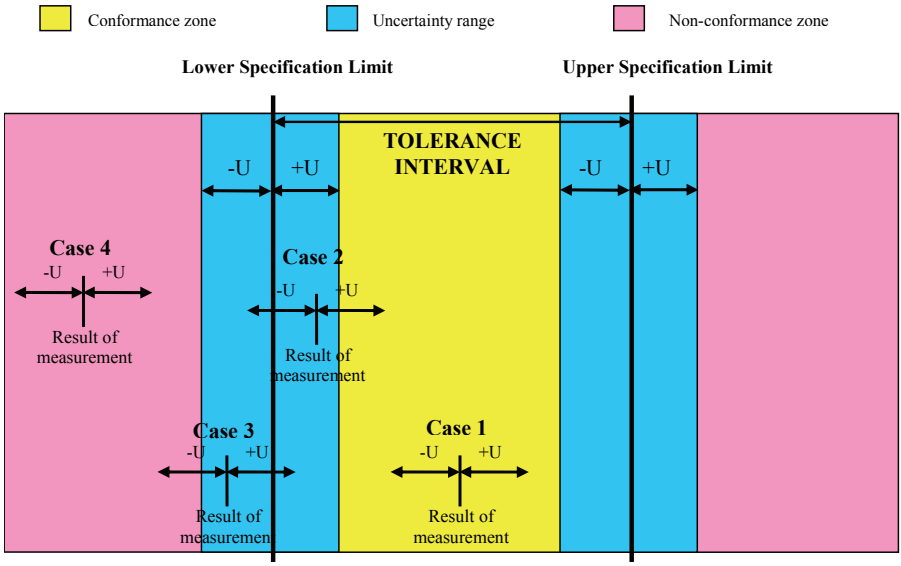
m: raw result
Vv: real value
E: measurement error
C: correction of known errors
y: corrected result
e: remaining unknown error
U: measurement uncertainty

The uncertainty associated with a measurement result then makes it possible to offer quantitative information on the quality of this measurement result provided to our client.

From then on, dispersion of test laboratory results can be compared against each other.

The consequence of this new requirement in the railway testing world is significant: from now on, tests can be compared through the measurement uncertainty value, price ranges can rely on sold uncertainties.

On the other hand, compliance declarations must consider the existence of a zone of doubt, or zone of uncertainties, around defined specification limits. The following diagram summarizes this zone of doubt, the 2nd and 3rd cases emphasize the importance of correctly estimating the measurement uncertainty, in order to decrease the risk of accepting a product which may not be compliant, or to reject one when it may be compliant.



Risks facing this problem

The risks taken if AEF does not satisfy the NF ISO 17025 requirement are as follows:

- lose COFRAC accreditations and consequently markets,
- provide incomplete result to the client,
- declare a product compliance when a non-compliance risk exists.

Constraints

This report and its requirements raise several questions:

- who will control the methodology?
- who will carry out these calculations?
- how to convince AEF agents of the usefulness of this requirement?

Knowing that constraints are as follows:

- workload increases,

- consequently, personnel do not have enough time,
- 70 COFRAC tests are accredited and include several measurement results, so as many uncertainty methodologies have to be carried out on each test result provided to the client,
- the methods are complex, several are specific to the railway world, maybe even inherent to the AEF.

Since the risks and constraints are substantial, AEF administration has therefore chosen to control this requirement in the form of a project managed at AEF level.

Organization of the uncertainties project

The objective set in mid 2002 was to estimate measurement uncertainties of all COFRAC tests in order to build an optimal organization for quickly reaching this objective, choosing a method and defining a policy.

Pilot

The pilot must have a general understanding of AEF and must be able to operate on this project independently. He or she guarantees the uncertainty methodology regardless of the test Aactivities. It is the reference in this field.

The pilot is part of the Progrès et Conformité department of AEF (quality service). This mission was entrusted to the Metrology Department head.

Method

The chosen reference standard is known internationally, the NF ENV 13005 standard, as the guide for the expression of measurement uncertainty (GUM).

We should note that very few of our COFRAC tests (less than 5 out of 70) can be submitted to inter-laboratory tests because our test methods are very specific. The uncertainty methodology could therefore not rely on ISO 5725.

Corporate policy

Because of the substantial amount of work on uncertainties to carry out, it was not possible to entrust all this work only to the Metrology Department head. On the other hand, the methodology can only be in perfect appropriateness with the test studied if the person or persons performing the study control the test method.

It was decided that uncertainty studies would be carried out by each person responsible for test methods supported by their team. The head of the Metrology Department plays the role of reference-consultant, controlling the method and validating all AEF studies.

Since the GUM is not so easy to analyze for people not trained for statistics or mathematics, it was decided to implement training with the objective of providing a practical and educational tool relating to measurement uncertainties, as well as to ensure the correct appropriation of the method by all.

Methodology for the estimation of measurement uncertainties

The methodology chosen was developed by LNE: the four step methodology based on GUM.

It is broken down as follows:

Step 1: Analysis of the test process

- definition of the measurand,
- scales of influence,
- mathematical model.

$$Y = f(x_1, x_2 \ldots x_n)$$

Step 2: Quantification of scales of influence

$$u(x_1) \quad u(x_2) \quad \ldots \quad u(x_n)$$

Step 3: Calculation of compound uncertainty

$$u_c^2(y) = \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]^2 \times u^2(x_i)$$

Step 4: Calculation of expanded uncertainty

$$U = k \times u_c$$

The most important step is step 1 because it is the analysis part of the test process:

- how many measurands are associated with the studied test,

- how to define with precision this measurand to avoid continuing the process in an incorrect measurand. Because in all cases, the methodology can be applied and a numerical value can be obtained, the major problem remains to apply this methodology to the test process that is really used.
- list scales of influence in detail, so as not to under estimate the measurement uncertainty and thus provide an incorrect indication to the client,
- finally, write a mathematical model including the definition of the measurand as well as the listing of scales of influence.

  A mathematical notion is not necessary to complete this vital step.

Step 2 is used to affect a numeric value to each scale of influence listed previously. To evaluate these typical uncertainties, two methods can be applied according to the available source of information (type A or B). Three laws are used at the AEF for the type B method: normal law, uniform law and law derived from arcsinus. They can be used in most quantifications. In some cases, other laws may be more appropriate but the methodology chosen is maximization.

In fact, the objective is to estimate the measurement uncertainty as quickly as possible (sometimes studies take two years because the tests are complex and the agents obviously are not working full time on them). Because of this, in order not to waste time uselessly on trying to quantify a scale as precisely as possible, hard-line estimations are often performed, particularly in the use of distribution laws as well as in the choice of original data.

Subsequently, if the uncertainty value obtained is too high compared to the one desired or expected, the calculation will be restudied only using predominant values. This saves time and avoids studies that are too detailed on a non-affecting parameter. The core business of AEF is to carry out tests and not to calculate measurement uncertainties.

In the case of a hard-line approach, the uncertainty value provided to the client will not be incorrect. The dispersion associated with the result will be higher than reality, but it always guarantees that the real value is located in the calculated interval. An opposite approach (minimalist) will result in providing an interval that is too low, not necessarily including the real value, and consequently providing a false result to the client.

Step 3 makes it possible to combine the different typical uncertainties into a single one: the compound uncertainty. For this, the law of uncertainty composition is applied. The hypothesis chosen regardless of the test is the consideration of scales of influence independent from one another. In this way, the formula applied is simplified (it is used in the diagram above).

Finally, the compound uncertainty is expanded with a k expansion factor, equal to two, which corresponds to a level of confidence of 95%.

Methodology implementation

Training

The methodology must be understood by all method heads as well as by the measurers, i.e. it is addressed to agents with a bac+2 to bac+5 level or higher.

A partnership with LNE was put in place: an on-site training session involving measurement uncertainties adapted to AEF requirements was created.

This training was broken down into three parts: four step theory on the methodology, statistical demonstrations and practical case study on an AEF test. It lasted three days. The targeted audience were people responsible for activity and method. Four sessions were taught in the first semester of 2003.

Since then, training has been adapted internally. It is done by the head of the Metrology Department over two days and involves the theoretical methodology (one half day) and a concrete case study on an AEF test (one and a half days). The statistical demonstrations part did not seem useful for the application of the methodology. The target audience corresponds to the new people responsible for method, test engineers and measurers.

In addition, internal training on metrology awareness addresses the notion of uncertainties in a general way in order to sensitize agents working in the measurement field of this subject.

To this day, 25 training sessions have been given and 140 people have been trained on uncertainties.

Study procedure

Generally, the head of method works alone or with measurers if there are any or with a workgroup, depending on the complexity of the test method. He or she carries out step 1 of the test process analysis. This step 1 will be submitted to the head of the Metrology Department, to re-center the study in case of incorrect direction.

Steps 2, 3 and 4 are also carried out by the workgroup in relation to the internal Metrology Service to obtain information on measurement equipment, and with other players in activity to estimate particular scales or carry out repeatability or reproducibility tests (to quantify the operator effect in particular).

Note: the methodology is sometimes given to interns once we are confident that step 1 is carried out in close cooperation with the person responsible for the method.

Tool: Excel spreadsheets

Excel spreadsheets are used to make reading the different calculations easier and to render a more dynamic view of the methodology.

A tab corresponds to a stratum of the mathematical model.

Each tab uses the scales of influence listed, quantification and sensitivity factors. The remaining calculation is then done automatically: we obtain not only the compound and expanded uncertainty, but also and especially influence percentages associated with each scale.

This is how predominant scales of influence will be emphasized. In this way, the method described in detail by the mathematical model can be easily understood: the person responsible for the method knows which parameter to trigger to decrease the uncertainty and which parameter is insignificant.

Products of the methodology output

Two output products are required:

- the data sheet relating to the written and validated study,
- Excel spreadsheets completed and validated based on the current AEF model.

The person responsible for the Metrology Department is responsible for validating the studies.

The data sheet ensures the study's permanence.

All justifications and hypotheses relating to the four steps are included. The objective is to quickly update the uncertainty calculation without having to go through all the steps in the study later on.

The Excel worksheets remain the only numerically updated support. In this way, if a sensor has a different calibration uncertainty from one year to the next, the calculation will have to be updated. The Excel file will then be modified numerically. The sheet will not be updated because the uncertainty estimation method associated with this test has not changed.

Advantages

There are several advantages to this methodology.

The first one is a detailed view of the test method. It is dissected and analyzed to highlight the predominant scales of influence. The Excel sheets then help us target these values and simulate improvement in order to forecast the impact of the modifications considered.

We can also observe the relevance of MTEs (maximum tolerated errors) applied during measurement equipment verification for example: we can simulate an increase or decrease of MTEs and observe the impact of these adjustments on the test method. We can do the same with calibration uncertainties. The definition of connection requirements and critical products is then simplified and controllable.

Similarly, the influence of the operator, roundings made in software used, or the assembly method can be highlighted.

The methodology has convinced AEF agents because of dynamic spreadsheets and the display of percentages linked to scales of influence.

At first, the people were reluctant to get involved in this uncertainty project. The different participants in the methodology found a source of motivation with:

- uncertainty training provided with the study of a real AEF case,
- the assurance of technical support,
- the will to reach these percentages of influence,
- the ability to simulate considered improvements before their implementation.

Several sectors that completed the uncertainty estimations relating to COFRAC tests were involved logically in the non-COFRAC tests.

The wish was not only to satisfy a requirement of COFRAC, but even more so to use the uncertainties to control the methods and their evolutions.

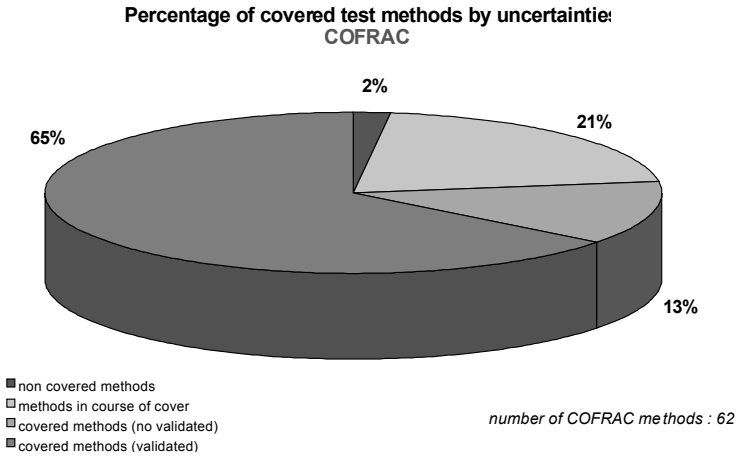
Indicators

Indicators were created in order to follow up on the state of progress of uncertainty studies and the tests covered by uncertainties.

We should note that a test can have several test results, which leads to several uncertainty studies. A test is only considered covered by uncertainties when all its

results are associated with a calculated measurement uncertainty and that the quality sheet and Excel spreadsheets are validated.

In March of 2007, the state of indicators is as follows:



Complex test methods are still not covered, they will be by the end of 2007. They are difficult methods to model and require a lot of time to be correctly analyzed.

Conclusion

The uncertainty project in the field of railway tests is a large scale project with the primary goal of estimating measurement uncertainty of COFRAC tests. The direction and implementation of this project have led to an implication from all the measurement players in this methodology. We are then able to obtain a numeric value of the uncertainty as well as to control the scales of influence and the ability to improve the test method.
The uncertainty project is destined to continue for the purpose of retaining skills in uncertainties in order to:

- keep calculations updated,
- carry out the methodology for non-COFRAC tests,
- carry out the methodology for new tests for new accreditations.

It is now considered as its own process.

The undeniable contribution of this methodology is to:

- control the different parameters of the test method,
- possibly improve the method,
- adapt maximum tolerated errors of verifications or calibration uncertainties of measurement equipment, and then better define our requirements concerning our metrologic connections and better control associated costs.

In the future, the direction will be to:

- declare a compliance by considering the risks taken (based on the associated uncertainty),
- possibly adapt the current specifications and understand those to come,
- sensitize clients to the notion of uncertainties.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Interlaboratory comparisons in Europe: which way to choose?

Simona Klenovska, Pavel Klenovsky

Czech Metrology Institute
Okruzni 31
638 00 Brno
Czech Republic
e-mail: sklenovska@cmi.cz, pklenovsky@cmi.cz

ABSTRACT: Proficiency testing/interlaboratory comparisons (PTs/ILCs) represent very effective method of the laboratory technical competence assessing. This powerful quality assurance tool enables individual labs to compare their performance with that of similar labs, to take any remedial actions needed and can be used for the purposes of a continual improvement of their quality systems as required by new ISO 17025:2005 standard. Recently, this activity has spread up very rapidly.
The infrastructure of PTs in Europe consists now of basically 3 elements, all of which are aimed at providing support to the accreditation system:

- IMEP program of EU JRC IRMM;
- EA-organized PTs and
- various national schemes.

In the area of physical metrology (the world of calibration laboratories) there is a regional level of ILCs operated by accreditation bodies which is complemented by providers on a national level. An example of a national system is presented: the Czech Metrology Institute (CMI) has been a provider of ILCs in the Czech Republic since 1995 with nearly 90 comparisons completed since that time – during the last two years even with international participation.
The following text presents operation of such national system and its results not only on the national level but also expanding to the international level. On a practical example of an internationally organized recent ILC the importance of ILCs to underpin mutual recognition agreements both in accreditation and metrology is being demonstrated.

1. Introduction

The arrival of accreditation has brought forward the issue of laboratory proficiency testing as a highly technical means of systematic supervision over their technical competence. This powerful quality assurance tool enables individual labs to compare their performance with that of similar labs, to take any remedial actions needed and can be used for the purposes of a continual improvement of their quality systems as required by ISO 9000:2000 series of standards. Recently, this activity has spread very rapidly up to a point when some metrologists believe that proficiency testing would be able to effectively replace accreditation as ultimate evidence of technical competence.

It is important to note that ILCs can in principle also be designed for purposes other than proficiency testing, e.g.:

- validation of methods – to determine key performance characteristics as defined in ISO 5725 B1 [2];
- characterization of reference materials – it is the only method to assign a certified value and to estimate its uncertainty for matrix-type reference materials;
- self-assessment of a laboratory's performance in a test;
- to assign a value and estimated uncertainty of a primary measurement standard;
- to establish so-called horizontal traceability of a measurement result when necessary.

Firstly, some clarification in terminology should be made. The definitions of the underlying terms are as follows (ISO/IEC Guide 43:1997 [1]):

(Laboratory) proficiency testing (PT): determination of laboratory testing performance by means of interlaboratory comparisons.

Interlaboratory comparisons (ILCs): organization, performance and evaluation of the test on the same or similar test items by two or more laboratories in accordance with pre-determined conditions.

2. Current situation in Europe

Traditionally, ILCs as such have been more frequently used in the family of chemical laboratories, the most populous type of testing labs. It is basically given by prevailing problems to establish robust traceability chains up to primary standards; in the case of matrix-type reference materials (RMs) it is mostly the only method to assign metrological characteristics to them. Many models of PT schemes with a lot of providers involved are now in use throughout Europe. An example of a supranational model to address this problem is a European regional program called IMEP operated by one of the EC research institutes, Institute for Reference Materials and Measurement (IRMM) Geel, Belgium, in the area of chemical labs, in support of the European legislation where various chemical analyses play a vital role (present in some 30% of EU legislative acts). Labs are invited to participate free of charge, certified RMs (specially prepared for these PTs so that their reference values can not be found on the internet) as test samples and the evaluation being provided by IRMM. A considerable effort is invested in the establishment of the reference values: these are the weighted averages of measurement results from a number of leading labs in Europe (IRMM, LGC, BAM, NMi etc.) – this arrangement enables us to apply an innovative approach in the chemical area to the evaluation of

measurements based on the normalized error (to replace an inferior scheme, so called z-score). The program has been consistently in operation since the early 1990, the number of participants always exceeds 50 from all parts of Europe. The results are available at www.jrc.irmm.be/imep. On national levels there is also a number of providers in each country working on a sectorial basis (medicine, quality of water or air, etc.). The analysis of results of past PT schemes clearly shows that the performance of chemical laboratories is rather poor, even among the accredited ones – failure rates of 50% are not uncommon. Even if accreditation bodies associated with the European Accreditation (EA) are supposed to take these results up with the labs and to enforce necessary corrective actions the unsatisfactory situation has prevailed since the 1990s. To the outside world, it gives the impression that this powerful technical tool does not play a key role in the whole system, the results being downplayed by all the stakeholders including the accreditors themselves and regulatory bodies. The impact of PT schemes is somehow suppressed by formal accreditation – instead, various forms of validation of methods are mainly subject to discussions during accreditation but the procedure is more or less of an administrative nature. Accreditation bodies seem to be preoccupied by efforts to strengthen their current position or to bring ever new areas under accreditation instead of working on underlying technical fundamentals. The quality of reference values established by leading institutes (IRMM, LGC, BAM etc.) is a considerably strong point of this top down approach, making this type of PT absolutely essential to the whole system. On the other hand, there are also severe limitations consisting of a low repetition frequencies and a low number of participants (relatively, from the European perspective).

In the area of physical metrology (the world of calibration laboratories) the situation is somewhat different. Again, there has been a regional level of ILCs operated by accreditation bodies which is complemented by providers on a national level. The former is the original form started in the early 1990, partially as a tool to supervise the competence of national accreditation bodies (NABs): accreditation bodies earmarked funds to support these comparisons which were organized in coordination within the EA – each NAB was assumed to organize one European-wide ILC every year under an approved plan. Major labs of a regional character were mostly the participants, DKD of Germany being one of the most active accreditation bodies here. These ILCs were of the piloted type with reference values provided mostly by various national metrology institutes (NMI). Although on a technically sound basis, these ILCs suffered from the fact that it would take an excessive period of time to finalize them (e.g. 7 - 10 years!). This made any corrective actions almost irrelevant and added to the relatively high associated costs of ILCs (each over 30,000 EUR). It was clear that without a more businesslike approach based on charging fees to the laboratories for participation, this type of ILC was bound to die out. The free-of-charge participation might have been a strong point here (especially for labs) but weak points (a limited participation, extremely long periods to finalize them and a quickly decreasing commitment to them on the part of NABs) rather prevailed at the end.

As the EA-organized ILCs faded away those operated nationally were coming more into the foreground. Here national metrology institutes (NMIs), traditionally more involved in physical metrology than in anything else, were the most natural choice, especially smaller, more service-oriented NMIs would consider this activity as a part of their mission. Such national ILCs can obviously be much less costly than those organized regionally with a higher penetration. As an example, the Czech Metrology Institute (CMI) has been a provider of ILCs since 1995 in the Czech Republic. The costs are carefully calculated and subsequently shared by all the participants – in this way an average fee constitutes ¼ of the fee charged by accreditors for a surveillance visit. In this way, an ILC in every major field of measurement can be run every year with a high number of participants. The long-term experience from such ILCs clearly shows that there are persisting deficiencies in laying down accreditation scopes of labs especially in terms of BMCs (ranges, uncertainties, non-existence of service categories), the reasons may vary from incompetence up to trying to obtain a competitive advantage over other similar labs. However, to deal with this disorder was exactly one of the main reasons to set up accreditation and it raises a natural question of what is the added value of accreditation after many years being around. Another conclusion to be drawn from running ILCs in the long term is a continual evolution of ILCs to become a major tool to demonstrate technical competence: the growing number of customers for calibration accepts the evidence from ILCs as a full replacement of accreditation – such evidence is simply tale-telling. It has to be pointed out here that an accreditation system as such is quite non-transparent to the outside world: due to the system of self-assessment among NABs no outside party has, even in cases of reasonable doubt, any insight into the underlying facts and evidence collected during the accreditation process (not even names of experts, results of ILCs, etc.). This is becoming a major drawback when it comes to disputes or when customers need a clear picture about the real performance of a lab. PTs on national levels have a potential to be the core of the whole system with a potential to expand abroad to cover a whole region, strong points being their high penetration, low costs and short periods of finalization. The operation is facilitated by the business-like approach to the operation of the system based on fees. The only weak point of this bottom up system is the establishment of reliable reference values in the chemical area where approaches based on some public support have to be sought if the system should ideally be based on the normalized error (e.g. through public bodies like IRMM, LGC, BAM in Europe, NIST in the USA, universities, etc.).

At the beginning of 2003, allegedly in an effort to concentrate on accreditation, the EA approached EUROMET, the association of European NMIs, with the proposal to take over technical activities associated with accreditation: provision of proficiency testing (ILCs) and preparation of calibration procedures. It came as a surprise to the metrology community: here the idea prevails that just the opposite should take place. Nevertheless, the idea was debated in the Executive Committee with controversial views: mainly larger NMIs considered it a distraction from their primary occupation of research. A joint EA-EUROMET task force was therefore

established to study the matter. Such ILCs could be considered a logical extension of key comparisons among NMIs themselves in implementation of CIPM MRA down to calibration labs on a national level, in such a way making a metrologically integrated system out of it. Eventually, the EA proposal was in principle accepted in the 1st quarter of 2004 by EUROMET while the EA Laboratory Committee started, all of a sudden, dragging its feet on this issue and decided to embark on an obviously lengthy process of analyzing the situation in individual countries before taking any action on their policy. Instead, a new concept is now being developed by the EA LC whereby the EA would launch a special type of regional PT/ILC, organized by tendered providers on behalf of the EA LC, aimed solely at supporting the regional MLA. The concept is still a bit controversial: participating accredited labs should be chosen by NABs (what are the criteria?) even if the participation should be based on a fee recovered from those labs (in countries where such a system is normally used the majority of them?).

The matters discussed above can be exemplified by a recent ILC in the field of pressure organized by CMI (see figures). This ILC, originally conceived as a national one, has attracted participants from other EU and non-EU countries – a relatively modest fee of 350 EUR was charged to participants. Eventually labs from 8 countries participated making it a regional ILC in the EA region of interest. Despite relatively complicated logistics, the ILC was finalized within 12 months with a rate of non-compliance of 10%. The ILC is directly linked to a preceding key comparison in absolute pressure under CIPM MRA through the CMI primary pressure lab which provided the reference values for this ILC. In this way a link along the vertical line of traceability was established putting both metrology and accreditation systems under scrutiny whereby the consistency of both has been verified, at the same time identifying targeted areas for improvements. This highlights the benefits of regional ILCs organized by NMIs being signatories of CIPM MRA rendering the above-mentioned special EA regional ILCs unnecessary. Another current feature at the interface of both MRAs in metrology and accreditation can be demonstrated on this ILC: BMCs of leading private accredited labs would surpass registered CMCs of NMIs, even in cases of NMIs to which they are directly traceable – a matter outside the scope of this paper.

During the debate EUROMET decided to start a discussion on the sensitive issue of making the results of PT/ILCs public with an identification of the laboratories as it is the case of key comparisons under the mutual recognition arrangement drawn by CIPM in the Metre Convention (CIPM MRA, see www.bipm.org). Such a move would bring the importance of PT to technical competence to a climax point, even overshadowing the importance of accreditation itself: it would overcome a non-transparency of accreditation for the outside world in this respect and it would expose labs to public scrutiny by their customers interested in their real technical competence. This would be extremely beneficial for customers of calibration/testing services at large and for those labs considering their technical competence the highest priority while being very much opposed by “low cost” labs. In general, there

are growing doubts about the quality of products while their providers are squeezed by a globalized competition to slash costs but no readily available tool is, after a wave of liberalization decreased the value of various certifications, at the disposal of their customers to make their own judgment: in the case of calibration/testing, such a tool is available but it is not used to its full potential – participants cannot be publicly identified based on a requirement from ISO Guide 43. It would be a real uphill struggle to achieve any change here, important stakeholders (labs themselves, accreditors) will be against with users and high quality (but expensive) labs benefiting. In this context it has to be welcome, as a kind of revolution with far reaching consequences, when the Joint Statement by CIPM and ILAC made at the end of 2005 calls for "work towards transparency and open publication of the results of comparisons between named accredited labs." As shown above, it is arguably the only way forward to overcome the problems concerning the quality of measurement results in future in the globalized world. It might be a matter of discussions in the on-going work on a new ISO 17043 standard on PT.

The above mentioned issue is associated with a policy on participation of accredited labs in PT. In Europe, this is laid down in the document EA-2/10 [3] even if some interpretation controversies among NABs still prevail. The recommended minimum amount of appropriate PT activities per laboratory is fixed as follows:

– one activity prior to granting accreditation;
– one activity relating to each major sub-discipline of a laboratory's scope of accreditation within the period between 2 subsequent assessments (3 to 5 years).

The policy has been established after a heated debate with accreditors, on one side, trying to lay down a firm framework for the participation and laboratories represented by EUROLAB and EURACHEM, on the other side, stressing cost-effectiveness: the latter argued that the accreditation costs by themselves are quite a financial burden for them and they are not ready, without any concessions on the part of accreditors, to follow any policy based on regular participation in PT with short periods. This topic is still a matter of lively debate at various levels, e.g. within EURACHEM itself, in the joint working group of EA, EUROLAB and EURACHEM on PTs, etc. It is a natural requirement on the part of laboratories to obtain some sort of "reward" for their successful participation in PT exercises, e.g. a reduced frequency of surveillance visits but accreditors are not ready, obviously out of purely financial reasons, to accept this idea so far. Given the principal technical importance of PT it raises concern: accreditors often sell accreditation as a public-type of activity and when it comes to reasonable decisions on such important matters to improve the system they block them basically on financial grounds. Some related documents are available on the EURACHEM website at www.eurachem.ul.pt.

We would logically expect, given the importance of PTs, that good performance shall be considered (rewarded) when fixing other surveillance measures in the accreditation process as it is stressed even in the document EA-2/10 [3]. The policy

of surveillance and reassessment is fixed by another document, EA-3/09 [4]. The interval between surveillance visits should not exceed 18 months with a recommended length of 12 months. In fixing this interval the accreditation body (only) **may** take into account the performance of the accredited body under consideration in previous visits and activities. However, in section 5.2 it is, quite surprisingly, stated: "Proficiency testing can almost never replace surveillance visits because it can cover only a part of the scope for which the laboratory is accredited, and it cannot reflect the overall performance of the laboratory and its quality system." It is an important point to be discussed. Accreditors have always argued that the respective standard ISO/IEC 17025 [5] should be only about technical competence of labs, other quality issues associated with calibration and testing processes (as stipulated e.g. by ISO 9000:2000 series of standards) should not be taken on board. Therefore, the entire (quality) management system here is fully concentrated on the correct performance of delivering the end product of each lab, i.e. provision of correct results of calibration or testing summarized in the corresponding certificate. PT is aimed at testing this performance in its entirety making the above mentioned sentence in section 5.2 of EA-3/09 absolutely false and unacceptable. It implies a ridiculous idea for any technician that a lab with a good (quality) management system as established during accreditation but showing badly in PT exercises is undoubtedly technically competent which, on the other hand, could not be said about a lab having a history of successful results in PT without a formal accreditation. As regards the first part of the sentence, it may currently partially be true – but it is only the result of such discouraging statements and practices on the part of accreditors that PT is not developed to its full potential which would bring enough robust data to make this statement irrelevant. The other way round: when successful participation in PT exercises can lead to reductions in surveillance visits the reduced costs would enable laboratories to participate in a higher number of ILCs, making the whole concept internally consistent as it could be demonstrated on the example of the Czech Republic. The only reason for this unfortunate development holding PT in an undeveloped state can arguably be self-interest of accreditors: PT schemes are normally operated and charged by providers outside NABs and after being fully developed they can potentially tap into the revenues of accreditors as a result of reductions in surveillance activities. Currently in Europe, there is no major evidence that participation and performance of accredited labs in PT are taken into account in any way in planning surveillance activities. This is especially unpleasant in geographically large economies (e.g. USA) where on-spot surveillance visits can be quite costly compared with participation in PT exercises.

Eventually, based on contemporary situation, a discussion of the approach implemented in section 5.9 of ISO/IEC 17025 [5] towards assuring quality of measurement results should be initiated: there is a number of options here how to monitor the validity of measurements undertaken but these options are by far not identical as to their importance to this goal, some of them being of doubtful benefit at all. Such open ends seriously undercut the use of PT in practice and can easily be

misused to the advantage of some labs, giving an opening to a non-transparent and unfair treatment. For every measurement, a way to establish traceability, even if not rigorously, has to exist nowadays so that a corresponding ILC can arguably be designed for PT purposes. This would enable us to make the policy in 5.9 [5] more tight, basically relying on PT. The system will thus be made more streamlined and transparent, establishing a level playing field for laboratories.

3. Proposal of a new system

Based on the analysis above the following integrated system of PT provisions in Europe being applicable globally is proposed along the following lines:

1. At the next revision of ISO/IEC 17025 [5] PT should be made the principal way of assuring quality of test and calibration results (if not the only one).
2. The system will be based on PT provision on a national level, expandable to a regional level if there is an interest on the part of labs and any regional comparisons organized by accreditors (EA) could be terminated after the system is up and running.
3. As a principle, local NMIs will preferably be the providers of PT for calibration labs. In case of a low number of participants in a country or of no resources or interest in the local NMI, the NAB may contact and assign the job to another NMI, e.g. from a neighboring country.
4. The competence of NMIs is derived form their participation in CIPM MRA. Other providers should be accredited in the case when they are not able to demonstrate their competence differently.
5. The justifiable share of the costs associated with participation in PT shall be borne by participants themselves, i.e. laboratories.
6. When the underlying ILC is of the piloted type, the reference values should be provided preferably by a signatory of the CIPM MRA or by an accredited lab. In such a case it will not be necessary to organize any regional European loops because the providers of the reference values will derive their calibration measurement capabilities from the CIPM MRA with an underlying network of key and supplementary comparisons. In this way the ILCs will be indirectly linked to these higher-level comparisons under the CIPM MRA.
7. Associated technical issues that may arise during PT (ILCs) could be attached to the corresponding EUROMET technical committees where a harmonized solution to those problems applicable throughout Europe could be sought.
8. The accreditation system should ideally become more transparent, at least the results of PTs should be publicly available where the participants should be identifiable.

9. A clear-cut binding policy on participation of accredited labs in PT should be issued and implemented on the regional level but ideally on a global level (within ILAC), e.g. on the basis of the document EA-2/10 [3]. In a broader perspective this policy could be based on the requirement that every lab has to cover its entire accreditation scope by participation in PT/ILCs in the 5 year period between successive accreditations.
10. The accreditation system should be amended in a way to reduce the frequency of surveillance visits in labs complying successfully with the policy discussed in point 9. In a broader perspective, surveillance visits between accreditations (5 year period) could be totally eliminated when there is no new measurement procedure added or a lab scores badly in a PT/ILC in the meantime.

The implementation of the outlined scenario will have the following benefits:

- the whole system will be more streamlined and cost effective;
- the intensity of assessments (including surveillance visits) in labs could be reduced;
- at the same time the level of technical competence of labs will be improved;
- technical competence of PT/ILC provision will be enhanced;
- customers of laboratories will have a quantitative tool at their disposal to assess the real performance of these external providers of calibration and testing services;
- data would be systematically gathered to verify the validity of MLAs in accreditation.

4. Summary

With the example of Europe it has been demonstrated that the current system of PT provision is more or less still in an infant stage and there are a lot of loopholes in the system. The question is what is the situation in other regions of the world, but at least a couple of lessons can be taken from that. The aim of the paper is to generate a broader discussion at least among laboratories themselves to arrive, hopefully, at a common position on the issue which could subsequently be used for negotiation and action at various levels, especially with accreditors and towards any new revisions of ISO/IEC 17025. Laboratories would clearly benefit from the proposed new system in a number of ways starting with a reduction of costs associated with maintaining their technical competence up to a more robust level playing field.

5. References

[1] ISO/IEC Guide 43:1997, Part 1: "Development and Operation of Proficiency Testing Schemes", 2nd Edition, 1997.

[2] ISO 5725, B1, 1-6, Accuracy (Trueness and Precision) of Measurement Methods and Results, February 1996

[3] EA-2/10, EA Policy for Participation in National and International Proficiency Testing Activities, EA, August 2001.

[4] EA-3/09, Surveillance and Reassessment of Accredited Organizations, December 1999.

[5] ISO/IEC 17025:1999, General Requirements for the Competence of Testing and Calibration Laboratories.

[6] ILAC G13:2000, Guidelines for Requirements for Competence of Providers of Proficiency Testing Schemes, ILAC, Sydney, January 2000.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Common approach to proficiency testing and interlaboratory comparisons

Oswin Kerkhof, Michel van Son, Adriaan M.H. van der Veen

Nederlands Meetinstituut
Thijsseweg 11, 2629 JA Delft, the Netherlands

ABSTRACT: Proficiency Testing (PT) and Interlaboratory Comparisons (ILCs) are similar activities in the sense that they provide a tool for laboratory performance assessment. The ISO/IEC 17025:2005 standard explicitly mentions these tools as a method for quality monitoring. The performance assessment programs are usually operated using an SI–traceable reference value for benchmarking. The uncertainty evaluation of the reference value appreciates as appropriate effects of non–homogenity and instability. In particular the regular participants in these programs benefit from the demonstrated consistency between reference values in subsequent rounds in that they can monitor long–term effects in their own measurement data.

Introduction

Since the early 1990s NMi VSL, the Dutch National Metrology Institute has organized proficiency testing (PT) and interlaboratory comparisons (ILCs). The main aim of these proficiency assessment programs (PAPs) is to offer a platform for laboratories to assess their performance. Participation in this kind of activities has meanwhile become part of quality standards such as ISO/IEC 17025 [1]. Now it has become an obligatory part of a laboratory's measures to obtain or maintain accreditation [2,3], provided of course that PAPs are offered for a particular calibration or testing service. Accreditation bodies in Europe have agreed to specifically look at laboratory participation and performance in these programs.

In ISO Guide 43-1 [4], *proficiency testing* is defined as an interlaboratory comparison aiming at rating the performance of the participating laboratories. In practice, in the calibration areas this activity is better known as "interlaboratory comparison", whereas in the testing area the term "proficiency testing" is well established. The key distinction of PT in relation to various kinds of interlaboratory comparisons is that in PT the laboratory results are *rated*, that is, criteria are set and applied to determine whether the performance is acceptable or not. Various schemes for performance rating exist, but the most commonly used ones are the Z– and E_n – scores [4,5]. Other types of interlaboratory comparisons include, e.g., the collaborative study (formerly also known as *certification study*) for characterizing reference materials [6,7].

The vast majority of the PAPs are run with reference values. A common, basic model has been developed for developing protocols for these programs [8]. The basic model is very flexible and allows different ways for reference values, batch inhomogenity and stability issues to be addressed. In this paper, it is shown how from this common model for interlaboratory comparisons, accurate models are developed for a particular program, and what factors come into play.

The approaches initially chosen for the testing and calibration area have been further harmonized as a consequence of the introduction of the concept of uncertainty of measurement in the testing area [9]. Until 2000, testing laboratories rarely provided uncertainty information with their data. Since the publication of ISO/IEC 17025 [1], reporting uncertainties is becoming common practice, so that also in the testing area the uncertainty information can be taken into consideration [10]. In essence, it even does not matter whether a consensus or reference value is used for benchmarking the laboratories, although it has some implications for the uncertainty evaluation of the departures from the assigned value, and hence, for the performance rating of the laboratories [11].

Reference versus consensus values

From a quality assurance point of view, the preferred method for establishing an assigned value is to determine the reference value for the artefact(s) in a PAP. The use of reference values has great advantages over the use of consensus values [10,12], although it is not always possible to come up with an accurate reference value. The latter particularly holds for method-defined parameters in the testing area. An example is the determination of mineral oil content in soil, where the measurement result not only depends on the definition of the measurand, but also on the exact measurement procedure applied. In cases where the participating laboratories have their measurements well under control, the use of consensus values for evaluation is a valid method. Even if measurements are under control, it is often very difficult to demonstrate the metrological traceability of a consensus value. In order to do so, it would require assessing the metrological traceability of the laboratories' results, which is often impossible in the context of this type of comparison.

The main disadvantage of consensus values is they can be unstable in time, especially with a varying group of participants in each round of the program. Furthermore, in the case of (relatively) poor datasets of some possible outliers, the consensus value can simply be inaccurate. As an example of such a dataset, the results of a PT for the determination of the amount fraction carbonyl sulphide in methane [13] are shown in Figure 1.

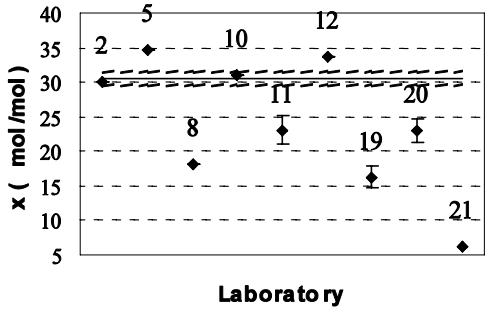


Figure 1. *Results of the determination of the amount fraction carbonyl sulphide in methane [13]*

Figure 1 shows that most results are significantly different from the reference value (indicated with the dotted lines (U, k = 2)). If in this PT the consensus value had been used for performance rating, those participants that have reported satisfactory results (labs 2 and 10), would have obtained an unsatisfactory score. In the actual PT, however, the results of these participants were the only ones qualified as satisfactory. This example underpins the use of reference values, where the qualification of a participant's result is not influenced by the performance of the other participants, in contrast to the use of consensus values.

An example of a situation where the measurements are under control is shown in Figure 2. In this figure the laboratory results are shown for methane in round 16 of PT scheme natural gas [14]. As can be seen in the figure, all participants have reported results that are not significantly different from the reference value. Consequently, the consensus value is not significantly different from the reference value either and the evaluation using the consensus value will not lead to a significantly different qualification of the laboratory results.

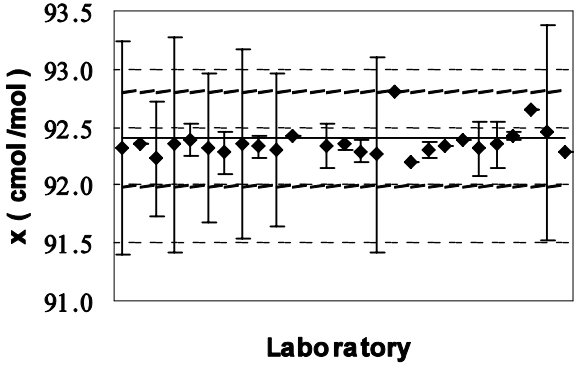


Figure 2. *Results of the determination of the amount fraction methane in natural gas [14]*

An important property of the reference values provided in the schemes discussed is that they are made metrologically traceable to SI (International System of Units). Calibrations or measurements of artefacts are performed against NMi VSL's own primary measurements standards. These measurements standards are subject to a maintenance program and the performance of calibrations and measurements is assessed in comparison with other National Metrology Institutes (NMIs). Interlaboratory comparisons between NMIs are called key comparisons, and the results of a key-comparison for the determination of the amount fraction of propane in natural gas are shown in Figure 3 [15].

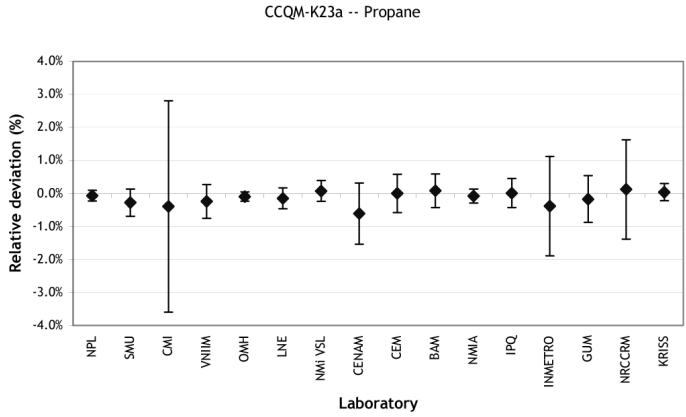


Figure 3. *Results of propane for CCQM key-comparison K23a "Natural gas Type I" [15]*

As can be seen Figure 3, the results of the metrology institutes, differing less than 1% relative, are very close. This means that, if a laboratory participating in the PT Scheme Natural Gas has a result for propane that is close to the reference value, the result would also be rated satisfactory when comparing to those of, e.g., NPL (UK) or KRISS (Korea).

Usually the reference value is established before the participants carry out their measurements. In order to be valid, the reference value should be stable during the lifetime of the comparison [4,5]. Any kind of in-stability should be accounted for, either in terms of uncertainty [8] or as a drift correction (with of course a contribution to the uncertainty budget of the reference value). The full model reads as [8]

$$x_{ref,i} = x_{ref} + \delta x_{bb,i} + \delta x_{stab,i} \tag{1}$$

where xref denotes the reference value, δxbb a correction due to batch inhomogenity and δxstab one due to instability. The index denotes the artefact sent to laboratory i. With regard to the stability correction, δx_{stab} can either be

1. zero, with a certain band-width in which the artefact(s) have been demonstrated to be stable (this band-width goes into the uncertainty budget of the assigned value);

2. a function of time (with in case of a travelling artefact), with associated uncertainty.

In the PAPs in the chemical area, usually the first approach is chosen. In the various fields of physical calibration, a drift function may be established, often by means of 1 measurement at the beginning, 1 at the end, and sometimes several extra in between. The travelling plan is designed accordingly. In case instability is observed, the variation/drift in time is mathematically modeled and for each participant an individual reference value is calculated (see equation (1)).

Figure 4 gives an example of how the reference value is used. In this case a relative humidity data logger was used as a transfer standard in an interlaboratory comparison. From the measurements of the reference laboratory at the beginning and at the end of the loop it was evident that some drift had occurred. From the investigation of the participant's results it was inferred that the drift was probably caused by a specific event halfway the loop. The report of one of the laboratories corroborated this view on the data. Consequently, the drift was modelled with a step function and reference values for each participant were determined. The reference laboratory's results were leading and the participant's results were used as support for establishing the drift function.

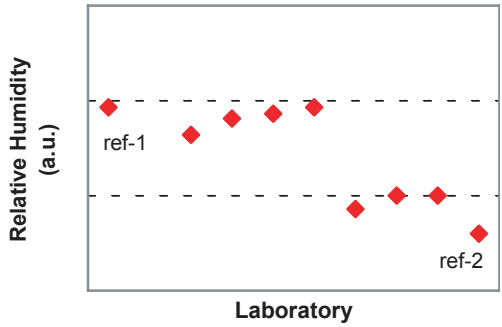


Figure 4. *Drift in a transfer instrument that apparently was caused by a single event in time*

Figure 5 shows an example where the transfer standard (an air temperature datalogger) exhibits a drift. From this data set it is not possible to determine the temporal drift behaviour during the course of the loop. The reference value was determined by taking the average of the two values assigned to the artefact at the beginning and end of the loop. The difference between the reference values was used for calculating an additional uncertainty component (due to instability) with a rectangular probability distribution.

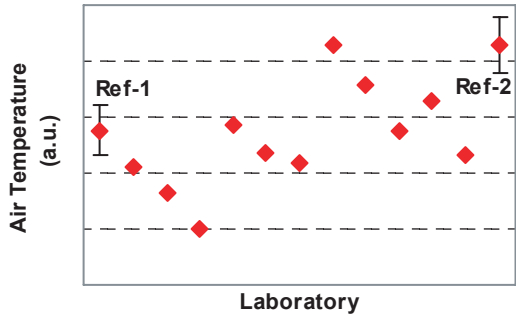


Figure 5. *Drift in a transfer instrument for which no specific correction model was applied*

Use of PT and ILC data

The interpretation of data from participation in a PT/ILC is not always that straightforward. Specific guidance exists for this purpose [16]. As already stated in the introduction, laboratories use PT/ILC data for the purpose of checking the performance of their quality system. This use is the most important reason for participation. An important second reason is to see the relative performance when comparing one's results to peers.

Often, consistently satisfactory results are a prerequisite for doing business in the first place. Good scores can furthermore be mentioned in communications to supports one's claims of providing high quality services. From a quality system point of view however, any kind of result in a PT/ILC is useful. Unsatisfactory performance will trigger the organization and lead to corrective measures. The effectiveness of these corrective measures can be challenged in a next round of the scheme. If the frequency is deemed to be too low for this purpose, an extra (bilateral) comparison can be requested.

There can also be other issues when looking at performance. In the (artificial) dataset shown in Figure 6, several cases can be distinguished. The uncertainties in the figure are given at the 95% level of confidence. Laboratory L5 is performing unsatisfactorily. The value is, in comparison with its peers, not too far off. The first option is to look at the possibility that in the calibration process something is out of control. In view of the very small uncertainty stated, it can also be that the uncertainty of measurement is severely underestimated.

Laboratory L1 (Figure 6) on the other hand has a satisfactory performance, but also the largest uncertainty of all. In some cases, this may not be an issue, because for example laboratory L1 serves a market where price is more important than accuracy. Otherwise, the laboratory may need to invest to improve its accuracy and to keep up with its peers.

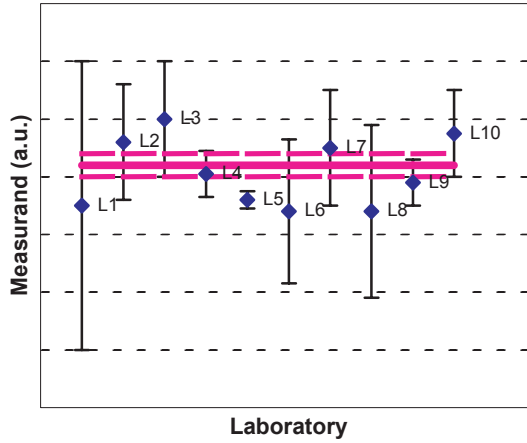


Figure 6. *Artificial dataset, illustrating different issues with laboratory performance*

As the estimation of uncertainty of measurement is still deemed to be a cumbersome process, many laboratories use their participation in PTs/ILCs to obtain an idea about the validity of their uncertainty estimate. If the reported uncertainty is part of the evaluation, as is the case when for example an E_n–score is used, then the participant receives as part of the performance rating a statement about the consistency of his result with the assigned value in view of the respective uncertainties.

Some laboratories attempt to reverse this process and try to obtain an idea about their "true" measurement uncertainty by deriving it from the deviation from the assigned value. For example Laboratory L1 in Figure 6 could be tempted to decrease its uncertainty claim based on the relatively small deviation from the reference value. Although it is possible to do so when a laboratory has participated in several (consecutive) rounds of a PT/ILC [16], it becomes dangerous if it is done on a single round.

In a case of alleged use of doping in sports, a laboratory questioned an uncertainty estimate based on the standard deviation of the laboratory results. It argued that it had practically hit the consensus value, so its uncertainty was much smaller than this standard deviation. It expressed that such a standard deviation be merely some sort of "typical uncertainty" for the whole group of laboratories. Although the latter is true, practically hitting the consensus value can be a single, lucky shot. Such an agreement should be demonstrated over a number of rounds (typically 5 or more) to obtain a realistic idea about the actual uncertainty of measurement. Another issue of course is how relevant PT–samples (artefacts) are for real life cases. This misconception has been one of the reasons to explain how

quality assurance and metrological concepts are supposed to work in this particular sector [17,18].

Another temptation that may come up when regularly participating in a PT/ILC operated with an (SI–) traceable reference value is to base the metrological traceability of the laboratory on such a reference value. Although in theory it could work, in practice it usually does not, neither is it acceptable practice. Laboratories are supposed to make their results metrologically traceable by means of a system of calibrations (ISO/IEC 17025 section 5.6 [2]) and intermediate checks (ISO/IEC 17025 section 5.5 [2]), and if there is any doubt concerning the validity of a calibration of an instrument, the instrument should be recalibrated. Such doubt can for example be raised as the consequence of participation in a PT/ILC. It is however not possible to correct the calibration of an instrument on the basis of the performance in the PT/ILC, because there can be a number of other reasons for (in particular poor) performance. Furthermore, a metrologically traceable standard should be available at all times to the laboratory, and this requirement is certainly not met by the reference value of a PT/ILC, which is established once and usually only valid for the period of time of the comparison.

Concluding remarks

There are evident advantages of using reference values in PT/ILC programs. It is much easier to demonstrate their metrological traceability and accuracy. Furthermore, the coordinator can ensure that the performance rating of a participant does not depend on the performance of the other participants. In particular in schemes with different levels of accuracy of the participants, this aspect is important.

Interpretation of PT/ILC data is quite straightforward, but there are some pitfalls. Being (once) very closely to the reference value does not imply accuracy; it does, when this agreement can be shown over a number of consecutive rounds.

The advantages of using a common approach to developing protocols for interlaboratory comparisons are evident: the model derived for a comparison is consistent with other models, and they are easy to maintain. It is also a very efficient way of dealing with a wide variety of laboratory requests.

References

[1] International Organization for Standardization, “ISO/IEC 17025 – General requirements for the competence of testing and calibration laboratories”, ISO Geneva, 1999

[2] International Organization for Standardization, “ISO/IEC 17025 – General requirements for the competence of testing and calibration laboratories”, 2nd edition, ISO Geneva, 2005

[3] International Laboratory Accreditation Cooperation, “ILAC G15 – Guidance for Accreditation to ISO/IEC 17025”, ILAC, 2002

[4] International Organization for Standardization, “ISO/IEC Guide 43-1:1997, Proficiency testing by interlaboratory comparisons - Part 1: Development and operation of proficiency testing schemes”, ISO Geneva

[5] International Organization for Standardization, “ISO 13528 – Statistical methods for use in proficiency testing by interlaboratory comparisons”, ISO Geneva, 2005

[6] International Organization for Standardization, “ISO Guide 35 – Reference Materials – General and statistical principles for certification”, 3rd edition, ISO Geneva, 2006

[7] Van der Veen A.M.H., “Determination of the certified value of a reference material appreciating the uncertainty statements obtained in the collaborative study”, in Ciarlini P., Cox M.G., Filipe E., Pavese F., Richter D., “Advanced Mathematical and Computational Tools in Metrology V”, World Scientific, Singapore, 2000, pp. 326-340

[8] Van der Veen A.M.H., Cox M.G., “Error analysis in the evaluation of measurement uncertainty”, Metrologia 40 (2003), pp. 42-50

[9] International Laboratory Accreditation Cooperation, “ILAC G17 - Introducing the Con-cept of Uncertainty of Measurement in Testing in Association with the Application of the Standard ISO/IEC 17025”, ILAC, 2002

[10] Van der Veen A.M.H., “Uncertainty evaluation in proficiency testing: state-of-the-art, challenges, and perspectives”, Accreditation and Quality Assurance 6 (2001), pp. 160-163

[11] Cox M.G., “The evaluation of key comparison data”, Metrologia 39 (2002), pp. 589-595

[12] Baldan A., Van der Veen A.M.H., Prauß D., Recknagel A., Boley N., “Economy of proficiency testing: reference versus consensus values”, Accreditation and Quality Assurance 6 (2001) pp. 164-167

[13] Van Son M., Van der Veen A.M.H., “Shell Gas Correlation Scheme – Evaluation Report Round 15”, NMi VSL Report S-CH.06.23, 13 December 2006

[14] Van Son M., Van der Veen A.M.H., “PT Natural Gas 16 – Evaluation Report”, NMi VSL Report S-CH.06.22, 5 December 2006

[15] Van der Veen A.M.H., Ziel P.R., De Leer E.W.B., Smeulders D., Besley L., Smarçao da Cunha V., Zhou Z., Qiao H., Heine H.-J., Tichy J., Lopez Esteban T., Mace T., Nagyné Szilágyi Z., Woo J.C., Bae H.-K., Perez Castorena A., Perez Urquiza M., Rangel Murillo F., Serrano Caballero V.M., Carbajal Alarcón C.E., Ramírez Nambo C., De Jesús Avila Salas M., Rakowska A., Dias F., Konopelko L.A., Tatjana A Popova T.A., Pankratov V.V., Kov-rizhnih M.A., Meshkov A.V., Efremova O.V., Kustikov Y.A., Musil S., Chromek F., Valkova M., Milton M.J.T., “Final Report on Interna-tional comparison CCQM K23ac: Natural gas types I and III”, Metrologia 44 (2007), Techn. Suppl. 08001

[16] EURACHEM, “Selection, Use and Interpretation of Proficiency Testing (PT) Schemes by Laboratories”, 2000, available at http://www.eurachem.ul.pt/

[17] Van der Veen A.M.H., “Measurement uncertainty and doping control in sport”, Accreditation and Quality Assurance 8 (2003), pp. 334-339

[18] Van der Veen A.M.H., “Measurement uncertainty and doping control in sport. 2. Metro-logical traceability issues”, Accreditation and Quality Assurance 9 (2004), pp. 311-316

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Sensory Metrology

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

A European project SysPAQ

Birgit Müller[1], Arne Dahms[1], Dirk Müller[1], Henrik N. Knudsen[2], Alireza Afshari [2], Pawel Wargocki[3], Bjarne Olesen[3], Birgitta Berglund[4], Olivier Ramalho[5], Joachim Goschnick[6], Oliver Jann[7], Wolfgang Horn[7], Daniel Nesa[8], Eric Chanie[9], Mika Ruponen[10]

[1] Technical University of Berlin, Hermann-Rietschel-Insitute
[2] Danish Building Research Institute, Aalborg University
[3] Technical University of Denmark
[4] Karolinska Institute
[5] Center Scientifique et Technique du Bâtiment
[6] Forschungszentrum Karlsruhe
[7] Federal Institute for Materials Research and Testing
[8] REGIENOV, Renault
[9] Alpha MOS
[10] Halton OY
Correspondence email: Birgit.mueller@tu-berlin.de

ABSTRACT: The European research project Innovative Sensor **Sys**tem for Measuring **P**erceived **A**ir **Q**uality and Brand Specific Odors (SysPAQ) was started under the VIth framework programme under the work programme "New and Emerging Science and Technology" (NEST PATHFINDER "Measuring the Impossible"). The project kicked off in September 2006. 10 partners (3 companies, 4 universities, 3 research institutes) from 5 countries are involved. The main goal of this project is to develop an innovative system to measure indoor air quality as it is perceived by humans to be used as an indicator and a control device for indoor air quality.

Introduction

This innovative sensor system is much demanded by the European society considering that humans spend about 90% of their time indoors, either at work, at home or when commuting between work and home. Recent data shows that improved indoor air quality will result in fewer complaints, increased comfort, less health problems and higher productivity. Consequently, quality of life will be improved.

Up to now, indoor air quality has been quantified applying three different measurement methods separately. The three methods are based on the human perception of indoor air quality, chemical measurements and sensors for specific odors. Considering measuring perceived air quality human assessments are still being superior to chemical measurements because of the unmatched sensitivity to many odorous indoor air pollutants. One of the reasons is that in most cases the chemical measurements or signals from chemical sensors designed to detect special odors could not be correlated with the assessments made by humans. They do obviously not measure the relevant indoor air pollutants triggering human sensory response. This project will build upon current knowledge on the perceptual effects of indoor air pollutants and on the experience gained in using chemical measurements and sensors for specific odors. The approach of the project will be to enhance the present state of the art of sensor systems, the perceptual methods and the software tools for modeling human response, and integrate them into one innovative system for measuring indoor air quality as it is perceived by humans. A bridge will consequently be created between the previous works in this area and progress will be achieved by integrating measurements, sensors and modeling by a holistic approach.

The main challenges are finding a perceptual space by using different reference odors, the measuring procedure for perceived air quality and a characterization method of brand specific odors. This work requires advanced knowledge in perception psychology and technical excellence in sensor system design and indoor environment research, knowledge and experience which the project partners contribute. It is still an open question how to mimic the human perception of odors and air quality. A certain risk of failure is involved in this project's unsolved problem of how to define an adequate system of reference odors for human perception of odors. In any case the project will provide new insight to the human perception of odors and indoor air quality, and will deliver a new approach to assess the perceived air quality and brand specific odors by an advanced sensor system.

Project objectives

The main goal of this project is to develop an innovative system to measure indoor air quality as it is perceived by humans based on perception modeling combining measurements of sensors and assessments of perceived air quality by sensory panels. The innovative sensor system can be used as an indicator, monitor and control device for the indoor air quality in buildings and vehicles. Furthermore, the system will be able to detect brand specific odors and it will serve as a novel interior odor design tool for the vehicle industry. The main objectives of the project are:

1. Definition of a method for measuring the perceived air quality and perceived odor intensity in buildings and vehicles. This method will be used by all different labs using sensory panels.
2. To find an advanced perception model for indoor air assessments. The model will be the major input to the software design for the innovative sensor system and it provides new insight to the human reaction to odors.
3. Development of an innovative sensor system for measuring, correspondingly, the perceived air quality and brand specific odors quality.
4. Calibration and test of the innovative sensor system for measuring, correspondingly, the perceived air quality and brand specific odors. The final version of the system is intended for the following applications:
    - monitoring of the ambient air within buildings and vehicles;
    - monitoring of the quality of the inlet air to buildings to ensure health and comfort for occupants;
    - labelling of materials based on emissions from buildings and vehicles;
    - control of the production process of building and vehicle materials.

Along with human activities, emissions from building materials, furnishing and equipment are main contributors to air pollution indoors. To reduce indoor air pollution loads it has been recommended to use low-polluting materials and to increase outdoor air supply rates. The new EU Energy Directive requires substantial reduction of energy use which may lead to the reduction of ventilation rates and increased indoor air pollution, enhancing the need for low-emission materials. In addition to measurements of indoor air quality as perceived by humans the system developed in the project can be used to control the emission rates from building materials already at the production stage. At present, manufacturers generally reduce the emissions from building materials by monitoring the emission rates of a few compounds in practice, but not necessarily the most relevant odor active compounds for perceived air quality. The pollution mixture affects the perceived air quality indoors. This is not taken into account so far. The proposed system can be used by the producers of building materials, furnishing and equipment materials to ensure that the emissions from their products would not negatively affect the perceived air quality indoors. In many countries labelling systems of building materials exist, so that the end-users can select the materials with reduced emissions. The suggested system for measuring the perceived air quality can also be used to quantify whether a material can obtain a label.

The selection of interior materials is a very important factor for the vehicle and transportation industry (train, car, boat, airplane, etc.). The goal of the selection process is to create a high-standard perceived air quality in vehicles combined with a brand specific odor impression. To meet this goal the system measuring the perceived air quality seems indispensable.

The interdisciplinary structure of the project consortium will enable innovative research and it will provide new insights to the human perception of air quality and brand specific odors. Project management will ensure a strong interaction between new perception models, hardware development as well as software design for the innovative sensor system.

Approach and methodology

The overall approach of the project and the priorities of the different work-packages are shown in Figure 1. The figure illustrates the parallelism of the human perception of air quality and brand specific odors on the left hand side and the sensor system development on the right hand side.

The two arrows in the center of Figure 1 indicate the coupling of sensory panel experiments and the sensor system hardware and software development. The innovative sensor system has to detect all relevant odor active substances. Based on the knowledge of the project partners and experiments of the project, a list of relevant substances is submitted to the two sensor specialists. The second arrow indicates the input of the new perception model of the perception specialist to the software development. The mathematical model of the new software for the sensor system will apply a reference odor system to reproduce an odor space that covers perceived air quality and brand specific odors in buildings and vehicles.

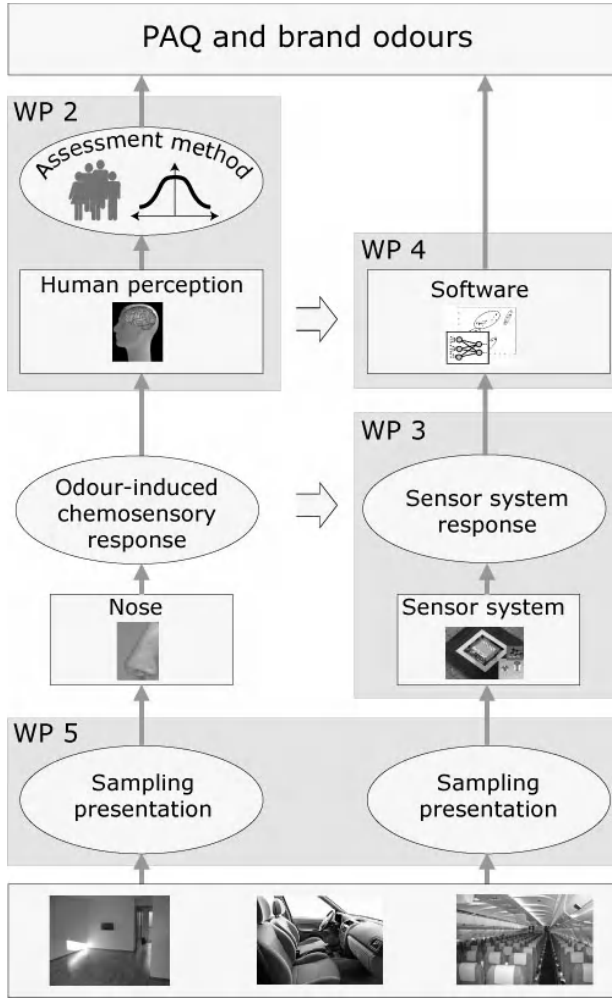


Figure 1. *Parallelism of the human perception of air quality and brand specific odors on the left hand side and the sensor system development on the right hand side*

WORKPLAN OF THE PROJECT

Introduction – general description

The work of the project is divided into two main tasks. The first task is focused on the human perception of odors. This part includes the definition of a method to measure perceived air quality and odor intensity by sensory panels. The used

assessment methods can have a major influence on the assessment results. Besides the assessment method, the main challenge of the human perception task is the definition of an odor space using reference odors. The multidimensional odor space has to account for all interior odors in buildings and vehicles.

This new odor space is the major input to the second main task of this project. This part is focused on the hard- and software development for the innovative sensor system. The contribution of two sensor specialists enables the application of the most advanced sensor technology. The sensitivity and selectivity of the sensors is adapted to odor active substances and the new odor space. The combination of different sensor types leads to a novel sensor system that covers considerably more substances than any current stand alone system. The data processing for this sensor system combines advanced statistical pattern recognition methods with the novel reference odor system. The link between the sensor signals and the perceived air quality consists of a classification process in terms of the reference odors followed by a regression method calibrated by the experimental data of the project.

These two main tasks **human perception** and **system development** break down to 4 work-packages (human perception and odor space, sensor system, data processing and pattern recognition, system calibration). The management process and the dissemination activities control and link these four work-packages, see Figure 2.

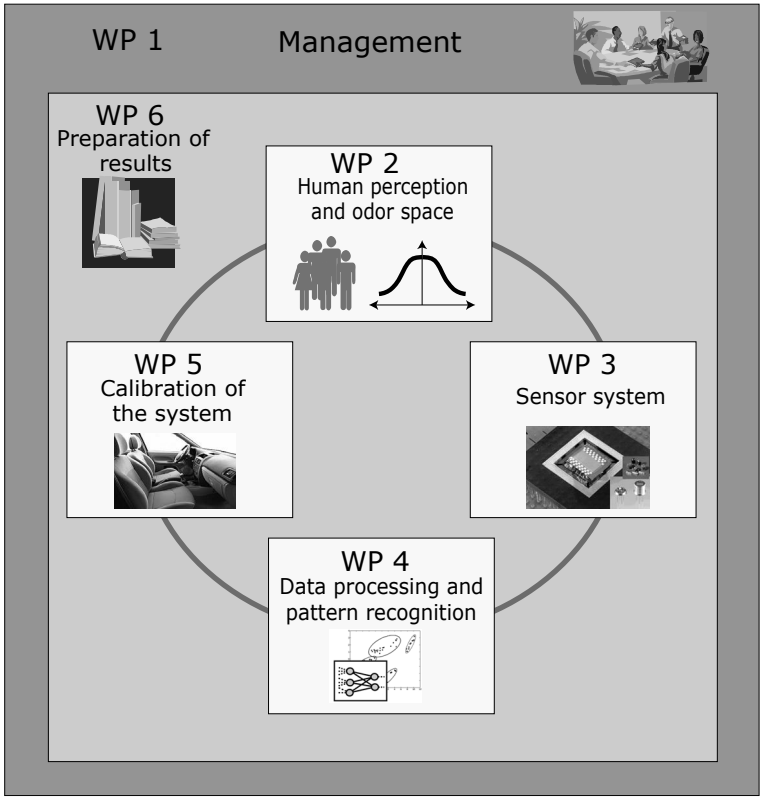


Figure 2. *Work-package structure*

Work planning

All overall **management** activities are included in WP1. The duration of the management activities covers the whole project life time. The main management activities are directed to the organization of the project meetings and the knowledge transfer between the different work-packages. The management of the project ensures a punctual delivery of all reports to the European Commission and it checks milestones and deliverables of all work-packages. The management also handles the global risk management of the project. The risk management includes the call for extra expert meetings in case of unsolved project tasks and the adjustment associated work-packages in the case of missing milestones or deliverables.

The **human perception** task of the project is mainly handled in WP2. The starting point of WP2 is the definition of a method for all sensory panel assessments of perceived air quality and odor intensity. The method is based on the knowledge of

the experimentally experienced project partner and the input of new multidimensional psychophysical perception models. This common standard for the assessment method guarantees comparable measurements at all labs for the innovative sensor system calibration data. The second and the most important part of WP2 is the definition of a new odor space based on reference odors. The new odor space has to cover all relevant odors of indoor air related environments. All project partners of WP2 have been working on the characterization of odors in the past. The work-package leader, the Karolinska Institute (KI), will provide a human perception model as a theoretical base for the new odor space. First experiments at KI will examine the useability of the new odor space based on reference odors to reproduce indoor air odors from building and vehicle environments. The quality of the odor space is the critical factor for the innovative sensor system and it will influence the software design process. A continuous knowledge transfer between WP2 and WP4 will ensure a fast translation of new findings into the software system.

WP3 covers all activities related to hardware improvements for the **innovative sensor system**. This work-package includes the improvement and adaptation of single multi-gas sensors as well as the combination of different multi-gas sensor technologies. The adaptation process and the selection of sensors are based on a literature review and some analysis of indoor air environment of buildings and vehicles in terms of odor active compounds. The sensor development and sensor combination is handled by two experienced sensor specialist Alpha MOS and Forschungszentrum Karlsruhe (AM, FZK). The two partners can provide gas chromatography/mass spectrometry combined with sniffer experiments to prove the response of the multi gas sensor system to odor relevant substances. The check and minimization of cross sensitivity (to other chemical compounds, relative humidity and temperature of sample air) of the sensors provides reproducible measurements. The sensor specialists (AM, FZK) manufacture three similar measurement devices for the calibration and validation measurements. The three systems will be jointly used by all partners. Additionally, the partner AM und FZK will provide all necessary hardware information and prototypes to the partners Center Scientifique et Technique du Bâtiment (CSTB) and Technical University of Berlin (TUB) to ensure a parallel development of the data processing and pattern recognition software (WP4).

The set-up of the general software layout for the **data processing** will be discussed in WP4. The work package leader Technical University of Berlin (TUB) contributes many years of experience in pattern recognition methods and software development to calculate the perceived odor intensity based on multi gas sensor systems. The separation of software and hardware (WP3 and WP4) enables a company independent and extendable mathematical modeling of low-concentration odor mixtures as regards sets of critical odors for building and vehicle materials and products and their combination. The task of the computational data processing is to

model the human odor perception of air samples. The method shall consider existing theories of perception and it shall link the measurements with psychophysical aspects of odor sensing. The response values of the sensor device will act as the stimuli and the software will mimic the human perception and evaluation process. The core of the data processing method will be the "memory", the calibration database. This database will have a major influence on the performance of the method, especially on the classification of the investigated odor sample. Therefore the data will span the whole odor space which is developed in WP2. The principle components of this odor space will be the basic odors. The data of the database consists of a combination of sensor response values and air quality assessments according to the methods of WP2 for a specific odor sample and concentration. The database will contain for each odor class data for different stimuli concentrations. The system shall be adaptive which requires the database to be expandable and open for new data of odor investigations. The calculation models will always refer to the current data and therefore the algorithms will be adjusted to the extended data set.

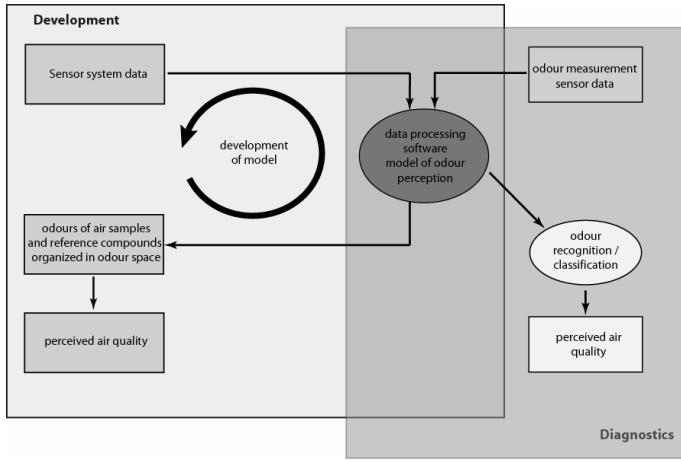


Figure 3. *The software will handle the development/calibration process and provide the final version for diagnostics of unknown odors*

The unknown investigated odor will be positioned inside the odor space which is a classification process for "odor recognition". Unknown odors will be expressed by a combination of the basic odors. Once the sample is classified, the quantitative and qualitative parameters of the perceived air quality will be estimated. This step of the data processing uses the calibration data of the basic odors, the regression and calculation algorithms for this odors as well as weighing algorithms which consider the position of the unknown odor sample in the odor space. Data of odor samples, other than the basic odors, is needed to develop, test and optimize the weighing algorithms and to estimate the perceived air quality of different odors. Most of the

odor space and combinations of basic odors should be covered. During the validation process of the software the new collected data will be included in the database which may further improve the performance of the data processing method.

The innovative sensor system has to identify all reference odors and their combination in indoor environments. Additionally, the system will link the pattern recognition method for the identification process to the sensory panel results. Due to the multidimensional character of the problem, a large effort is needed to calibrate the innovative sensor system. WP 5 handles the **collection of the calibration data**. It consists of simultaneous measurements of perceived air quality as well as the perceived odor intensity by a sensory panel and measurements with the sensor systems using the most relevant sensory methods selected in WP2 and the most promising innovative system/technical device selected in WP3. All data of WP5 iterates back to the software enhancement of WP4. The odors for calibration measurements of WP5 are produced in emission chambers. The air flow through the chambers is polluted by a series of building and cabin materials. The concentration of air pollutants will be varied within an indoor realistic range to ensure and test if the system works appropriately. The variation in concentration will be achieved by varying the material loading and the ventilation rate and by selecting high and low emitting materials. Measurements will be performed for individual materials in a laboratory setting, for combinations of materials in a full-scale setting, and in real buildings and vehicles. All data of WP5 is stored in a database. A final test run of the innovative sensor system after calibration will show the ability to predict perceived air quality and to characterize brand specific odors. During the sensor development WP5 will provide preliminary test set-ups of building materials in order to secure reasonable sensitivity and discrimination power of the novel sensor system.

The **dissemination** task of WP6 provides the publication of all project results. The main communication medium of the project is an advanced Internet portal that offers a public and a non-public information area (www.syspaq.eu). The public area publishes all non-confidential findings of the project partner during the project life time. The non-public area handles the data exchange of all project partners. All project related findings will be open to the public after the completion of the project. The project will provide a final workshop in order to discuss and disseminate all findings of the project. The consortium will prepare a brief project presentation in English which is written in a style which is accessible to non-specialists, avoiding technical language, mathematical formulae and acronyms as much as possible. Publication will be carried out via the NEST website.

More detailed information on the major work package interdependencies is given in Figure 4.

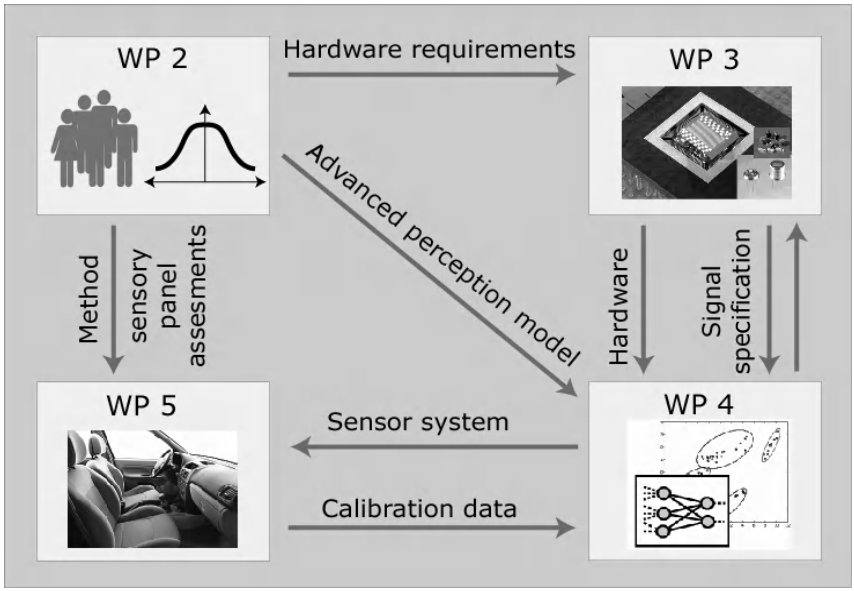


Figure 4. *Major work-package interdependencies*

Acknowledgement

SysPAQ is partly sponsored by the European Community in the Nest programme (NEST-28936) under the management of Mrs. P. Lopez. The co-ordination is performed by the Technical University Berlin. Other participants are: Danish Building Research Institute, Aalborg University; Technical University of Denmark; Karolinska Institute; Center Scientifique et Technique du Bâtiment; Forschungszentrum Karlsruhe; Federal Institute for Materials Research and Testing; REGIENOV; Alpha MOS; Halton Oy.

References

[1] Müller, B. *et al.* 2006. ANNEX I, Description of work, SysPAQ Contract.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Metrology of appearance

A. Razet, N. Pousset, G. Obein

LNE-INM/Cnam
61 Rue du Landy
F-93210 La Plaine-Saint-Denis – France

ABSTRACT: Color rendering index, calculated from the CIE publication, is not applicable correctly to light emitting diodes (LED).
The development of a new indicator for the characterization of this property appears necessary and falls within the scope of the metrology of appearance. A new project developed at LNE-INM/Cnam in this field is presented in this paper.

Introduction

Optical radiation measurement is currently undergoing an important evolution, especially in the field of applications dedicated to the satisfaction of the society's needs. These needs are as multiple as diversified. Among them, we can mention the "measurement" of the appearance of the objects and the scenes as seen by people.

The appearance of an object is defined by the Commission Internationale de l'Eclairage (CIE) as follows [1]:

"The appearance is the visual sensation by which an object is perceived to have attributes such as size, shape, color, texture, gloss, transparency, opacity, etc."

The difficulty in measuring what we see is the integration of the different elements involved in the perception of an object: the lighting, the interactions of the light with the material, the reception of the light signal by the eye and its interpretation by the visual system.

The use of new sources (LED type) in lighting as well as new materials (pearlescent type or interferential coating), generate requirements for the determination of appearance in such fields like industry, image synthesis and art conservation and restoration:

– in the industry, the control of appearance is crucial. For instance cosmetics (lipstick, hair color, etc.), car painting or the food industry express needs for measurement. As a matter of fact, visual appearance has a real impact on product sales. The measurement of appearance allows automation of the product aspect control process;

– measurements of appearance are useful in computer generated imaging. They can be applied to cinema, video games or flight simulators. All of these applications, sometimes real-time, often imply a compromise between realism, aestheticism and calculation time;

– in the field of art conservation and restoration, appearance is the raw material of artists. The measure of appearance enables curators to make choices in the exposition of artwork or its lighting. For preventive conservation, it makes it possible to follow the variations of appearance over time. Finally, the restorer can take appearance into consideration during the restoration of a pictorial layer.

The determination of the appearance requires the implementation of instrumentation and methods for the acquisition of the physical quantities associated with the sources (spectral distribution) and to the material characteristics (properties of reflection and transmission). These physical quantities could be correlated from the different models of human response obtained via psychophysical experiments with a group of people.

Currently, LNE-INM/Cnam is developing a research project in the field of the measurement of appearance, in particular on the "color rendering properties of light sources".

Color rendering properties of light sources

The use of white light-emitting diodes (LED) in lighting is currently undergoing important development. These sources are expected to lead to significant energy savings. However, the characteristics of these diodes are a problem in the color rendering field. As a matter of fact, these LEDs have extremely different spectral distributions depending on the technology used to create "white" lights.

The increasing use of LEDs in lighting generates appearance determination requirements. A good candidate might be the color rendering index (CRI), developed by the CIE to characterize the aptitude of fluorescent lights to return a color. But unfortunately because the spectral distribution of LEDs are different from those of CIE reference sources, the CRI does not apply to LEDs in a satisfactory manner. Searching for a new indicator seems necessary and perfectly fits in with the context of metrology of appearance.

Sources

Traditional LEDs are p-n junctions (juxtaposition of an n doped semi-conductor and a p doped semi-conductor) to which a tension is applied for the transmission of visible or ultraviolet radiation.

Traditional LEDs only transmit around a given wavelength, whereas solar or incandescent white light transmits over the complete range of a visible spectrum in a continuous and homogenous way. Different methods are used to create white or pseudo-white light from LEDs.

- LED color mix

  This method consists of using three different color LEDs (red, green and blue). The synthesis of the three colors leads to the production of white light, where color temperature depends on the proportion of each of the three components.

- Blue LED coupled with a phosphor

  This method is based on the fact that two photons with complementary wavelengths (λ_1 "short" and λ_2 "long") are emitted simultaneously and generate a white light sensation. The solution then is a light-emitting diode producing the short wavelength, covered by a phosphor absorbing a few "short" photons to convert them into a longer wavelength. In practice, these white LEDs are obtained by using a blue diode (InGaN) combined with a yellow phosphor (YAG:Ce cerium doped yttrium aluminum garnet).

- UV LEDs and three phosphors or more

  This method consists of using a diode producing a short wavelength (in close ultraviolet or in violet) coupled with one or more phosphors transforming UV light to visible.

Shapes of LED can be very different according to the model or technology used. LEDs can be: *5 mm round* (the most common), *3 mm round*, *7.6 mm square*, *miniature*, or even embedded on a dedicated circuit.

For all types of LED, the light semiconductor chip is completely embedded in a transparent resin box with a rounded surface; this box is used as protection while making the beam opening smaller.

Color rendering index

This color rendering index (CRI) defines the aptitude of a light source to return the different colors of the objects it lights, compared to a reference source. It varies between 0 and 100 The following are the assessments we can have from a CRI:

CRI < 50 Very bad

50 < CIR < 70 Bad

70 < CIR < 80 Average

80 < CIR < 90 Good

90 < CIR < 100 Very good

The calculation method of this rendering index is described in the 1995 CIE documentation [2].

Nowadays, according to the different types of LED, the CIR can reach values between 97 and 99 [3]. However, several results obtained in Hungary [4-6] from visual measurements have emphasized that color rendering index determined by the CIE calculation method did not seem to apply correctly to light emitting diodes. In addition, the methods used to create colored LED-based white light can deteriorate colored discrimination performances, which is difficult to forecast from CIR variations [7]. New studies are therefore recommended in order to determine a new color index taking into account the particularity of the sources using white LEDs. This is the subject of our LNE-INM/Cnam research project.

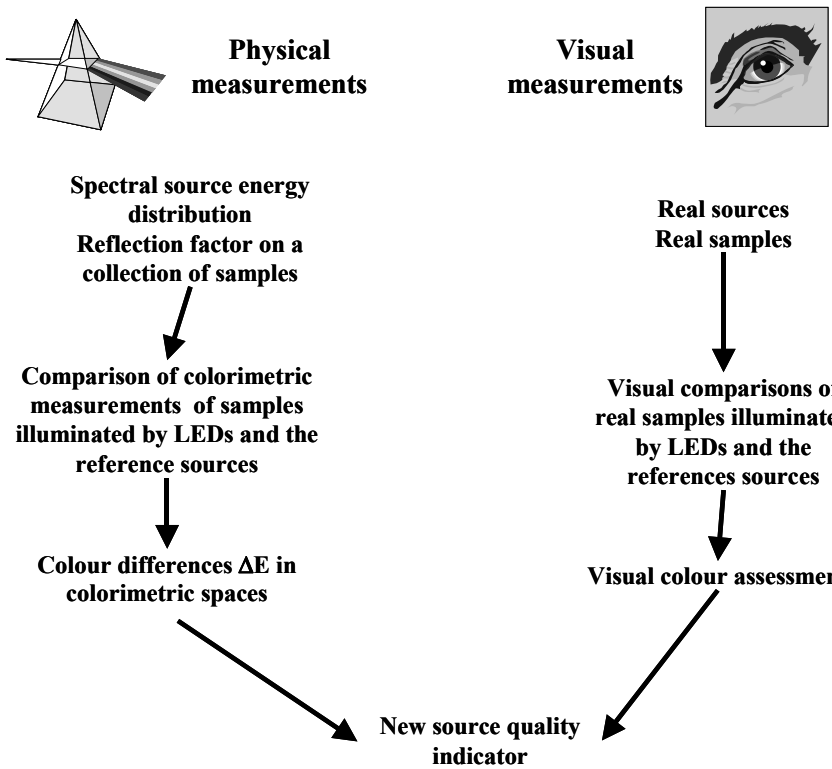


Figure 1. *Description of the color rendering properties of light sources project*

Presentation of the project

The aim of this project is to develop from physical and visual measurements a new parameter for the characterization of color rendering obtained with LED lighting.

Figure 1 presents a diagram of this project.

Spectral measurements will be carried out on the reflection of carefully selected samples, illuminated by LEDs and by CIE reference sources in order to determine color differences. The color differences will then be compared to those obtained by visual measurements in order to find a new color rendering index.

Metrological studies will be carried out on source properties used and in particular on their stability in luminance and in spectral energy distribution, which seems to be a significant parameter during visual evaluation.

Conclusion

This "color rendering properties of light sources" project has just started at LNE-INM/Cnam to define new LED-based light sources color rendering indicators.

References

[1] Commission Internationale de l'Eclairage, "A framework for the measurement of visual appearance", *CIE 175:2006*.

[2] Commission Internationale de l'Éclairage. "Method of Measuring and Specifying Colour Rendering Properties of Light Sources", *CIE* 13.3-1995.

[3] Commission Internationale de l'Éclairage, "Colour rendering of white LED light sources", CIE Technical Report, to be published.

[4] N. Sándor, P.Bodrogi, P. Csuti, B. Kránicz, J. Schanda, "Direct visual assessment of colour rendering". *Proc. 25th CIE Session. CIE* 152:2003, D1-42-D1-45.

[5] N. Sándor, P. Csuti, P. Bodrogi, J. Schanda, "Visual observation of colour rendering". *Proc. CIE Expert Symposium on LED Light Sources: Physical measurement and visual and photobiological assessment.* CIE x026:2004, pp. 16-19, 2004.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Electricity

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Measurement of the reflectivity of radio-wave absorbers

Frédéric Pythoud

Swiss Federal Office of Metrology METAS
3003 Bern-Wabern (Switzerland)

ABSTRACT: The precise knowledge of the reflectivity of radio-wave absorbers is a very important factor in the proper design of an anechoic chamber. We propose a new and simple method in order to characterize the reflectivity of absorbers at normal incidence in the frequency range from 80 MHz to more than 15 GHz. The method is derived from the "VSWR (Voltage Standing Wave Ratio) method", with numerical and technical improvements. The advantage of this method is that it does not require a dedicated infrastructure like a wave guide or a special arch. We even believe that it is more accurate than traditional methods at frequencies around 1 GHz. The method is illustrated by the measurements performed using different types of absorber technologies.

Introduction

Radio-wave absorbers are used by the industry in many emc testing laboratories (emc anechoic chambers). The quality of their absorbing properties is crucial for the quality of the chamber itself. It is therefore very important to know the reflecting properties of absorbers precisely before their installation in a chamber. Measuring the reflectivity of radio-wave absorbers may require an important infrastructure as illustrated by the following methods:

- NRL (Naval Research Laboratory) Arch Method [1] requires a good anechoic chamber with absorbers at least as good as the absorbers to measure;

- the waveguide method [1] requires a waveguide with dimensions greater than 0.6 m x 0.6 m x 4.8 m for low frequencies around 30 MHz. In this case the probe has length and width of 0.6 m.

Other methods have been developed that are easier to apply:

- the AWVSR (Advanced Voltage Standing Wave Ratio) method, which is an improvement of the VSWR method, is based on time gating algorithms included in the complex VNA (Vector Network Analyzers) method [2]. This method was basically intended to measure a whole wall of absorbers.

Method

We propose here another variation of the VSWR method that can be used to measure the reflectivity of absorbers in a wide frequency range (80 MHz to 15 GHz). This measurement method uses several principles.

Principle 1: Use of complex S-parameters

The reflectivity measurement is performed with one antenna which is used as transmitting and receiving antenna. The complex S-parameters at the antenna connector are measured with a VNA in the frequency domain. The VNA is previously calibrated with Open, Short and Load.

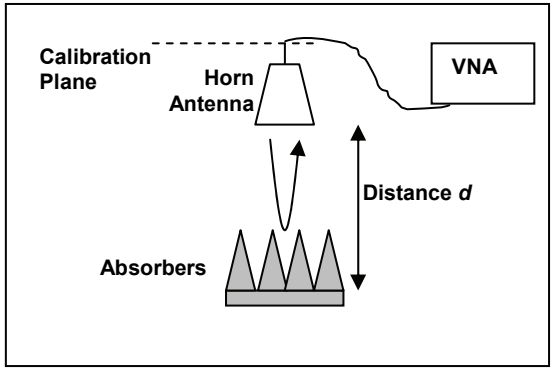


Figure 1. *S11 of the antenna is measured with a VNA*

The measured S_{11} is therefore the sum of the reflection of the antenna itself (S_{11}^{0}) with the contribution of reflections from the absorbers ($S_{11}^{absorbers}$):

$$S_{11} = S_{11}^{0} + S_{11}^{absorbers} \tag{1}$$

Based on the free-space propagation and on the equations of the transmitting and receiving antenna [3] we can write the contribution of the absorber reflections as:

$$S_{11}^{absorbers} = i \cdot \frac{\eta}{c \cdot Z_L} \cdot \frac{f}{2d} \cdot e^{-2ikd} \cdot \frac{1}{AF} \Gamma \frac{1}{AF} \tag{2}$$

where i is the imaginary number, η the intrinsic impedance of the medium (in free space, η =377 Ω), c the speed of light, Z_L=50 Ω the conventional impedance for antennae, f the frequency, $k = 2\pi f/c$ the propagation constant, d the distance to absorbers, AF the antenna factor of the antenna and Γ the absorbers reflectivity.

Principle 2: Motion of the absorbers

In order to get rid of reflections of the surroundings, the absorbers themselves are moved as shown in Figure 2 instead of moving the antenna.

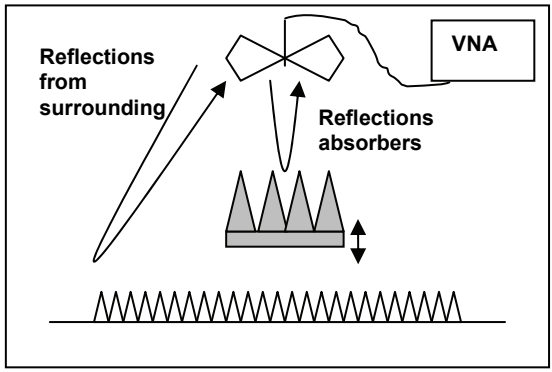


Figure 2. *The absorbers to be measured are moved, the measuring antenna is fix*

The measured $S_{11}^{(1)}$ and $S_{11}^{(2)}$ with absorbers in position 1 and 2 respectively can thus be written as:

$$S_{11}^{(i)} = S_{11}^{0} + S_{11}^{surroundings} + S_{11}^{absorbers(i)} \quad i=1..2 \tag{3}$$

The idea behind the motion of absorbers is to remove the influence of the surroundings by subtracting the measured S_{11}.

$$\begin{aligned} S_{11}^{(2)} - S_{11}^{(1)} &= S_{11}^{absorbers(2)} - S_{11}^{absorbers(1)} \\ &\cong \Gamma \cdot \frac{\eta \cdot f}{AF^2 \cdot d \cdot c \cdot Z_L} \cdot \sin\left(k \cdot dx\right) \cdot e^{-2ikd} \end{aligned} \tag{4}$$

where dx is the distance between both positions of the absorbers. The method is not limited to the use of directional antennae and can therefore be extended below 1 GHz without problems. The last equation can now be reformulated as:

$$\Gamma = const \cdot \frac{S_{11}^{(2)} - S_{11}^{(1)}}{\sin(k \cdot dx)} \tag{5}$$

with *const* depending only on the antenna factor *AF*, the distance *d*, the frequency *f* and on physical constants. The value of the reflectivity can be calculated accurately under the assumption that:

$$\sin(k \cdot dx) \neq 0 \tag{6}$$

ideally:

$$dx \cong \lambda / 4 \tag{7}$$

Principle 3: Normalization with a metal plate

In order to get rid of the constant in the previous equation, it is useful to carry out a first measurement against a metal plate:

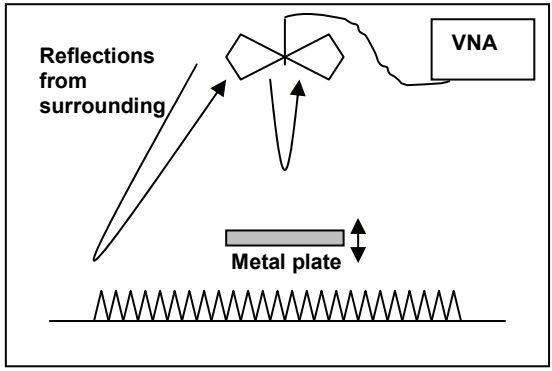


Figure 3. *Normalization is performed against a metal plate*

$$\Gamma_{metal} = const \cdot \frac{S_{11}^{(metal,2)} - S_{11}^{(metal,1)}}{\sin(k \cdot dx)} \tag{8}$$

and by dividing equations (5) and (6) we obtain:

$$\frac{\Gamma}{\Gamma_{metal}} = \frac{S_{11}^{(2)} - S_{11}^{(1)}}{S_{11}^{(metal,2)} - S_{11}^{(metal,1)}} \tag{9}$$

This normalization is not necessary if the antenna factor is well known, for example in the case of directive antennae.

Principle 4: Assume ideal metal reflections

The last step in the calculation of the reflectivity is to assume that:

$$\Gamma_{metal} \equiv 1 \tag{10}$$

This step is commonly used in absorber measurements, as for example in the NRL Arch method, despite the fact that it is a very rough approximation. This approximation is realistic for an infinite sized plane of metal. However, for an infinitesimal metal plate we almost have no reflections. The reflection from a finite surface of metal depends on its dimensions and can even be calculated accurately. This type of calculation is out of the scope of this contribution but a schematic representation of the reflection of a squared metal plate as function of its size is shown in Figure 4. In some cases, it can even bring more than 100% reflections as a parabolic mirror would do, since the reflected wave from the different points of the mirror may add constructively or in a destructive way. Before measuring absorbers, it is therefore very important to be aware of the size of the sample that is used according to the distance to the antenna and according to the wavelength.

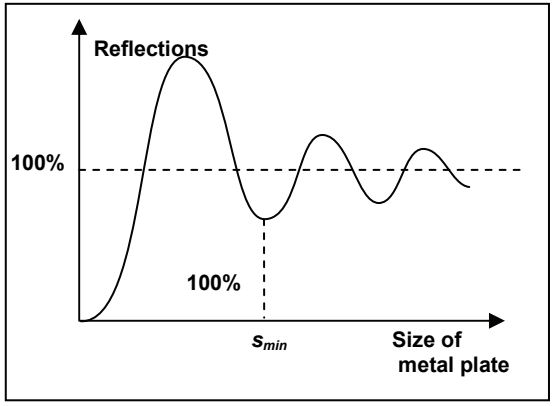


Figure 4. *The reflection from a square metal plate according to its half width (schematic representation)*

The curve shown in Figure 4 has its first minimum s_{min} given approximately by the Fresnel ellipsoid dimensions

$$s_{\min} \cong \sqrt{d \cdot \lambda} \tag{11}$$

where d is the distance to the absorbers, and λ the wavelength.

Measurements

Measurement apparatus

We were interested to compare absorbers in a wide frequency range (80 MHz to 15 GHz). We therefore divided the frequency range in 4 ranges for which we used a specific antenna as shown in Table 1.

Frequency Range	Antenna	Distance to absorbers
80 MHz to 300 MHz	Schwarzbeck VHBB 9124 (Biconical)	3.0 m
0.2 GHz to 2.8 GHz	Schwarzbeck VUSLP 9111 (Log-periodic)	2.5 m
0.7 GHz to 4.5 GHz	Schwarzbeck BBHA 9120 A (Horn)	2.0 m
1 GHz to 18 GHz	Schwarzbeck BBHA 9120 D (Horn)	1.55 m

Table 1. *Summary of antennae and distance used as function of the frequency*

The measurements have been performed with a VNA Rohde& Schwarz ZVM (20 MHz to 20 GHz).

Measurement setup

The absorbers have been moved in a vertical direction using a hydraulic lift-table shown in Figure 5. The motion of the table has been performed manually and the motion amplitude dx is reported in Table 2.

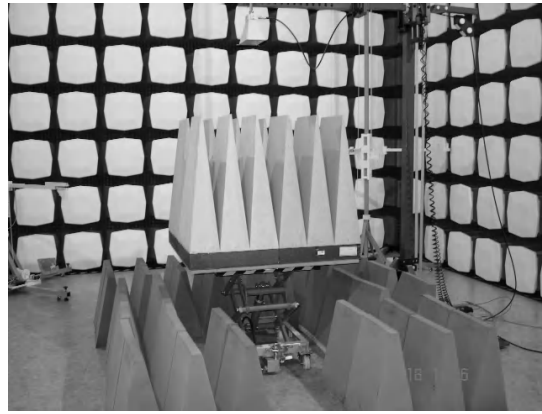


Figure 5. *Picture of the hydraulic table that was used to move the absorbers*

Motion Amplitude: dx	Valid Frequency Range
0.75 m	80 MHz to 180MHz
0.41 m	100 MHz to 300 MHz
0.15 m	200 MHz to 900 MHz
0.08 m	500 MHz to 1.5 GHz
0.05 m	900 MHz to 2.8 GHz
0.05 m	0.7 GHz to 2.5 GHz
0.03 m	2.5 GHz to 4.5 GHz
30 mm	1 GHz to 4.5 GHz
16 mm	2.5 GHz to 8 GHz
9 mm	4.5 GHz to 14 GHz
5 mm	8 GHz to 15 GHz

Table 2. *Summary of the motion amplitude used for the characterization of absorbers*

The reflection coefficient of absorbers has been defined for each frequency range separately, and finally averaged (weighted average) to provide only one curve.

Measurement material

As measurement material we used 1.2 m x 1.8 m of absorber materials from TDK, ranging from hybrid absorbers (IP 045E, IP 060BL) to microwave absorbers (IS 015A, IS 060).

Results

The results are shown in the following figure. As a comparison value, the manufacturer data are depicted in Table 3. The agreement is acceptable (deviation of the order of 5 dB) except for the hybrid IP 060 BL where we measure worse reflecting properties as the manufacturer data. Our method may not be sensitive enough to measure very good absorbers (reflectivity higher than 40 dB). However, we think that the method we use is more accurate especially around the frequency of 1 GHz since the manufacturer measurements at the frequency of 1 GHz are neither within the optimum frequencies of a wave guide (typically 20 MHz to 500 MHz) nor in the optimum frequency range of the arch method (typically above 1 GHz). Moreover, other parameters may play a role as:

– the probe size;

– the distance to the absorbers;

– the antenna beam directivity;

– the exact position of the axis of the beam with respect to the absorbers.

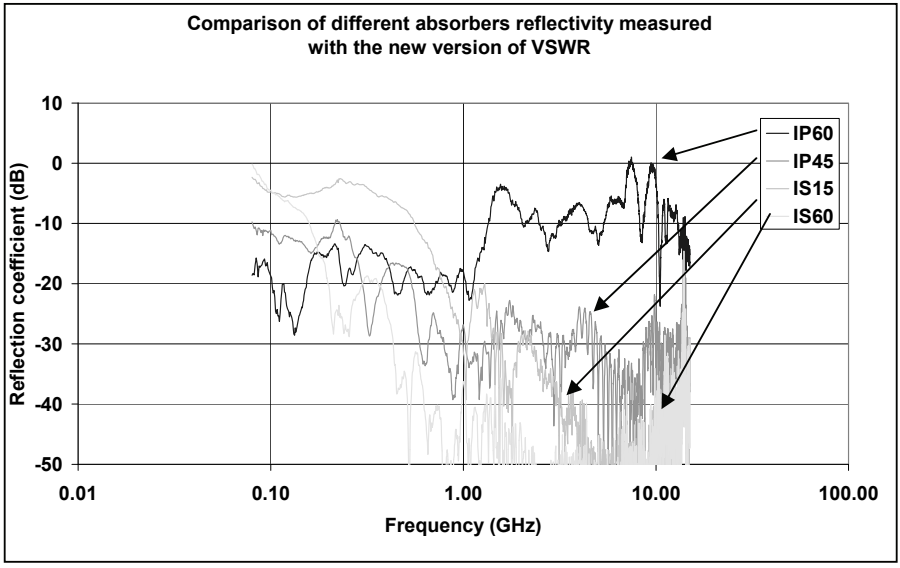


Figure 6. *Absorber reflectivity (normal incidence) as measured with a new version of the VSWR method*

	30 MHz	50 MHz	100 MHz	500 MHz	0.8 GHz	1 GHz	3 GHz	5 GHz	10 GHz	18 GHz
IP 045E	14	14	14	20	-	30	-	40	-	40
IP060BL	17	18	18	20	-	20	-	30	-	40
IS 015A	-	-	-	-	20	20	35	40	55	-
IS 060	-	-	-	32	37	42	50	55	55	-

Table 3. *Absorbers reflectivity in dB as given from the manufacturer*

Conclusion

A new variant of the VSWR method has been developed for measuring radio-wave absorbers. The method can be used in practice to compare different types of absorbers in a very wide frequency range: in our case 80 MHz to 15 GHz. The method does not require a dedicated infrastructure like a special arch or huge wave guide. Measurements can be performed in an anechoic chamber thus giving the possibility to laboratories to obtain reflectivity data that are independent from manufacturer data with ease. Moreover, we think that the method may provide more accurate results than traditional methods especially around 1 GHz. We believe that the method could be extended below 80 MHz, thus completely getting rid of the big wave guides. Care needs to be taken when considering the size of the absorber samples and the distance of measurement.

References

[1] IEEE 1128-1998: "IEEE Recommended Practice for Radio-Frequency (RF) Absorber Evaluation in the Range of 30 MHz to 5 GHz", 1998.

[2] G. Cottard, Y. Arien, "Anechoic Chamber Measurement Improvement", *Microwave Journal*, Vol. 49, Nb. 3, p. 94, March 2006.

[3] Kai Fong Lee, *Principles of Antenna Theory*, Kai Fong Lee, John Wiley and Sons, 1984.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Calibration of the spectral density of CISPR (16-1-1) impulse generators

Jürg Furrer

Federal Office of Metrology METAS
Lindenweg 50, CH-3003 Bern-Wabern, Switzerland

ABSTRACT: In the domain of electromagnetic compatibility (EMC), highly specialized and complex receivers are widely used for radio disturbance and immunity applications. In order to calibrate such receivers, specially designed test generators are often used. A short and repetitive pulse is applied to test different receiver parameters in a very efficient way. The traceable calibration of impulse generators is thus an important issue in radio frequency (rf) metrology.
Two different methods are compared: A traditional frequency domain method and an efficient time domain method, both giving good agreement. The different approaches with all necessary steps are shown, uncertainties, advantages and drawbacks are discussed.

Introduction

Measurement receivers are very important and widely used instruments. Especially in the domain of electromagnetic compatibility (EMC) highly specialized test receivers are used for radio disturbance and immunity applications. On the other side there are dedicated and specially designed impulse generators ("CISPR-generators") on the market to calibrate such receivers. A simple impulse signal allows for a fast and elegant test procedure, which investigates different receiver parameters at the same time (absolute amplitude, frequency response, linearity, detector characteristics). The IEC standardization document CISPR 16-1-1 describes these test methods. This recently updated document (2006) has now expanded the upper frequency limit from 1 GHz (CISPR Band D) previously to 18 GHz (Band E), which underlines the timeliness.

Several manufacturers offer corresponding products on the market and there is a demand in rf metrology for a traceable calibration of CISPR-16 impulse generators.

The calibration procedure can be established in different ways and we have compared a frequency domain method with a time domain method. The first and more traditional approach makes use of a transfer-receiver to compare traceable rf-power (CW) with the rf-power delivered by the impulse generator (spectral power density D for a given receiver bandwidth) as a function of frequency. The other approach uses a fast sampling oscilloscope together with an attenuator and a delayline to measure the impulse area. The frequency response of the DUT generator (device under test) is calculated from the measured time domain data by Fourier

transformation. The results of both methods should agree. Our goal was to determine the most efficient calibration procedure with the lowest uncertainties.

CISPR 16-1-1 test method

CISPR is an EMC standardization committee within IEC and the document CISPR 16-1-1 [1] describes test signals and methods for radio disturbance and immunity tests. For the four different frequency bands with upper frequency limits from 150 kHz to 1 GHz the test conditions are slightly different. For example, pulse form and repetition frequency, receiver bandwidth and the properties of the signal detectors are adapted but the principle remains the same in all bands.

This paper describes test methods for CISPR 16 generators in the highest (and most challenging) frequency band 30 – 1000 MHz (CISPR band C/D). The measurands are **impulse area *IS*** in µVs or the **spectral density *D*** in µV/Hz, which are directly related.

An ideal rectangular short pulse produces energy in its frequency spectrum which is distributed according to a sinX/X function (Figure 1a, b). The first spectral zero point depends on the pulse width δ and appears at the frequency 1 / δ. If a flat distributed pulse spectrum up to 1 GHz is required to check or calibrate a 1 GHz receiver, the pulse width δ has to be << 1 ns. On the other hand the spectral energy depends directly on the impulse area IS and it becomes clear that for the small pulse width δ, an equivalent high pulse amplitude is required.

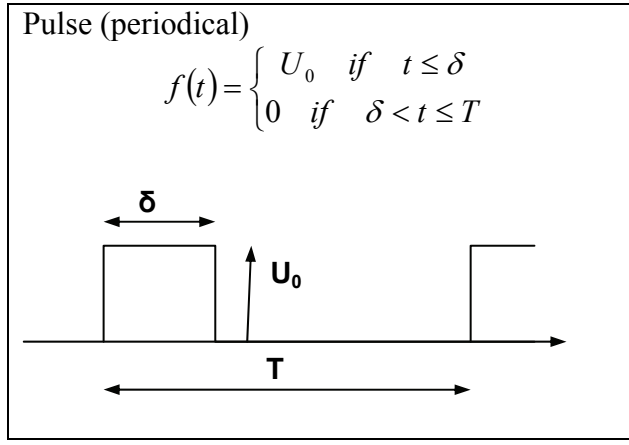


Figure 1a. *Ideal pulse (time domain)*

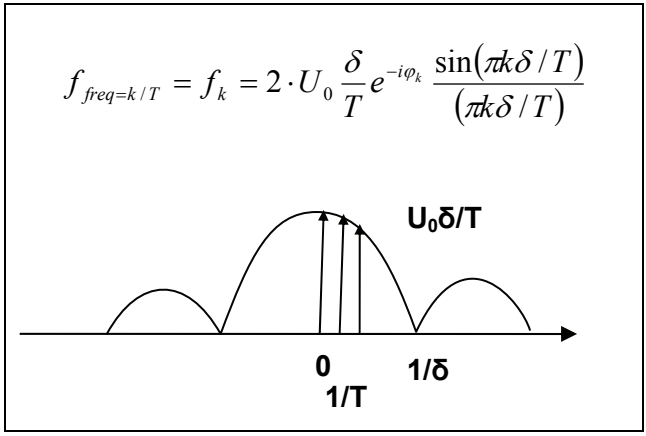


Figure 1b. *Pulse spectrum (frequency domain)*

CISPR 16 definitions

Please note that contrary to the original CISPR 16 document in the subsequent text all levels given are with respect to 50 Ω terminated circuits.

CISPR 16 states that:

The response of a measuring receiver to pulses having an impulse area of 0.022 μVs, having a uniform spectrum up to at least 1,000 MHz, repeated at a frequency of 100 Hz, shall for all tuning frequencies, be equal to the response to an unmodulated sine-wave signal at the tuned frequency having an r.m.s. value of 1 mV (60 dB(μV)). The impulse area produced by the generator shall be known with an error not greater than ± 0.5 dB. The pulse repetition frequency shall be known with an error not greater than 1%.

The impulse area *IS* is defined as the voltage-time area of a pulse defined by the integral:

$$IS = \int_{-\infty}^{+\infty} V(t)\,dt \quad \text{expressed in μVs or in dB(μVs)}$$

The spectral density *D* is related to the impulse area and expressed in μV/MHz or in dB(μV/MHz). For rectangular impulses of duration δ at frequencies $f << 1/\delta$, the relationship *D* (μV/MHz) = √2 · 10e6 *IS* (μVs) applies (The factor √2 includes the peak to r.m.s conversion).

Practical realisation of CISPR pulse generators

With a defined impulse area *IS* = 22 nVs and pulse duration δ << 1ns, the required pulse amplitude has to be in the order of 50 to 90 V depending on the pulse width as well as the rise- and fall-time of the actual signal (Figure 2).

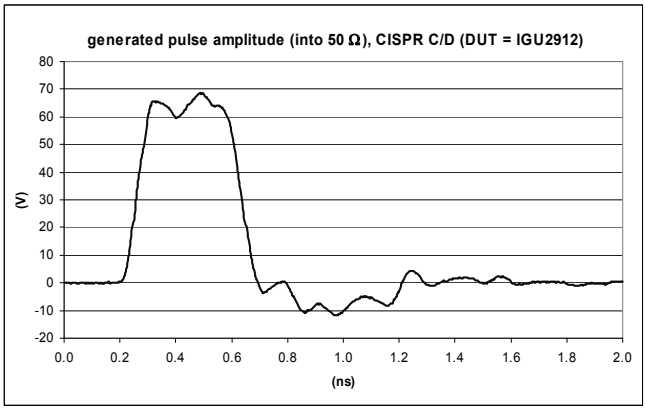


Figure 2. *Pulse amplitude (into 50 Ω), Band C/D*

Due to this required high voltage amplitude, a design of a pulse generator with semiconductor elements is not feasible. Practical realisations of CISPR16 generators still use the old but robust principle of discharging an RC-combination, where the capacitor C is realized by a short 50 Ω-coaxial line [2]. A mechanical mercury relay which is clocked with a pulse repetition frequency (*PRF*) of 100 Hz serves to discharge the line to the load R. The generated pulse length is twice the line length and the risetime is in the order of 100 ps. Furthermore, a certain degree of pulse predistortion is applied in order to optimize for a flat spectral density up to 1 GHz. The pulse amplitude can be controlled precisely over a wide range by adjusting the charging DC voltage.

Frequency domain calibration method

In this procedure the unknown spectral pulse density D_{DUT} is compared with the known CW power in three steps. A high performance EMI (electromagnetic interference) receiver is used as a transfer device in the frequency range 30 MHz to 1 GHz. According to the CISPR-16 standard, the receiver should indicate the same level 60 dB(μV) for the appropriate CW signal and for the pulse signal, if a bandwidth of 120 kHz and quasi-peak detection is used and if the *PRF* of the DUT is 100 Hz. We have chosen a slightly different method, because in case of quasi peak detection the measured spectral density depends on the detector time constants

as well as the DUT pulse repetition frequency. By using peak detection the measured spectral density no longer depends on the *PRF* because a change in *PRF* causes mutually compensating changes in amplitude and density of spectral components (Figure 1b). After the summation of all spectral components within the constant receiver bandwidth *B* in the peak detector the result depends only on the chosen bandwidth *B* (1 MHz) and on the spectral pulse density D_{DUT}.

The setup (Figure 3) consists of a thermal rf power standard, a rf power source with a monitor, a calibrated 20 dB step attenuator, the EMI receiver (ESS, R&S) and the DUT generator. In a first step, the power source is calibrated against the power standard; the step attenuator is set to 0 dB and the calibrated CW level is 3 dBm, a level which is close to the power standard calibration level 0 dBm (1 mW). The receiver calibration factor is determined in a second step with a CW power level of -17 dBm (90 dB(μV) at Z_0 = 50 Ω) by setting the step attenuator to 20 dB. In both steps, power ratios against the MON powermeter are measured.

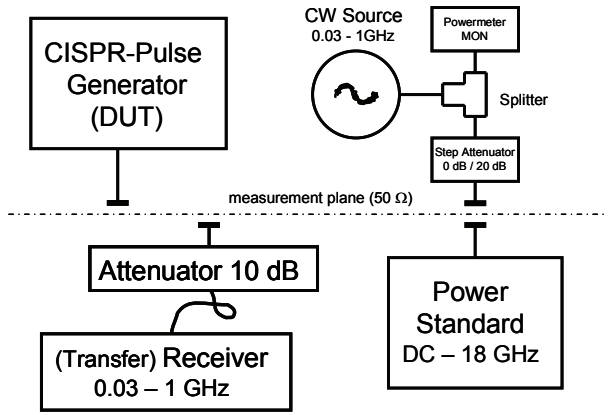


Figure 3. *Measuring setup (frequency domain method)*

In a third step the spectral pulse amplitude D_{DUT} is measured with the calibrated receiver. In a bandwith *B* of 1 MHz a nominal value of D_{nom} = 89.86 dB(μV/MHz) should arise [1], corresponding to the previously calibrated CW power level. D_{nom} is simply the peak summed voltage of all spectral components in 1 MHz bandwidth at f = 0 Hz, according to Figure 1b.

A calibration factor K_D (expressed in dB) according to frequency for the measured D_{DUT} is finally calculated:

$K_{D\_DUT} = D_{nom} - D_{DUT}$ (*D* expressed in dB(μV/MHz)).

Uncertainty contributions

It is important to note that due to the high pulse amplitude (> 60 V, Figure 2) the receiver needs to behave very linearly (CISPR16: overload factor). This is one of the two main reasons why an external 10 dB attenuator is inserted in front of the receiver, even if built-in attenuators could be activated. The other reason for this layout, where the N(male) attenuator-input-connector builds the measurement plane, is an easy way to optimize the receiver input match as close as possible to 50 Ω. This is a crucial parameter when the pulse generator amplitude needs to be assigned with sufficient low uncertainty. The reason is the unusually high output voltage reflection coefficient VRC_{DUT} of the generator (source match, $VRC_{DUT} > 0.5$) and thus an increased mismatch between generator and receiver. Due to the unavailability of all the phase information in the reflection coefficients it is generally not possible to correct for the mismatch and we need to assume an ideal (power) transmission of 1.0 with a mismatch uncertainty (MU) assigned to it, where $MU = 2 \cdot VRC_R \cdot VRC_D$ [3].

For the described setup, the input VRC_R of the receiver-attenuator combination has been evaluated with a VNA (Vector Network Analyzer) by choosing from different 10 dB attenuators in order to reach a $VRC_R < 0.01$ (> 40 dB MagLog). By changing the 10 dB attenuator we can realize different receiver *VRCs* and measure the spectral pulse density $D_{DUT}(f)$ for each matching condition (Figure 4). This allows us to analyze the impact of the DUT source match and provides confidence in the treatment of the mismatch uncertainty *MU*.

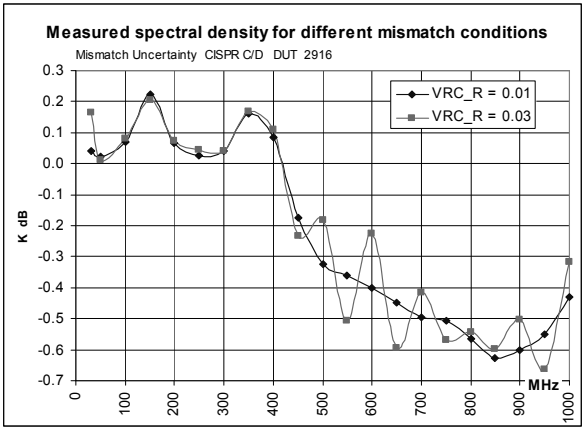


Figure 4. *Influence of measured spectral density due to different mismatch conditions*

The three mismatch uncertainties *MU* between power splitter and step attenuator and between step attenuator and power standard (or receiver) are negligible in our

setup because the *VRC* of splitter, power standard and receiver are < 0.01 at frequencies ≤ 1 GHz (*MU* < 0.003 dB).

The linearity can be checked and its uncertainty estimated by adding (or removing) additional receiver attenuation, until compression effects on the measured amplitude remain sufficiently small. However, there is a trade-off in this measurement layout: using the receiver peak detector leads to high spread in case of a decreased signal to noise ratio due to increased signal attenuation. Pulse and CW signals have to be measured with the same receiver settings. For this reason a high inherent receiver linearity and careful adjustments are of great importance.

The receiver IF filter (B = 1 MHz) does not have an ideal rectangular form. A special calibration has been performed where the measured filter area (Figure 5) has been integrated, in order to determine the equivalent rectangular noise bandwidth.

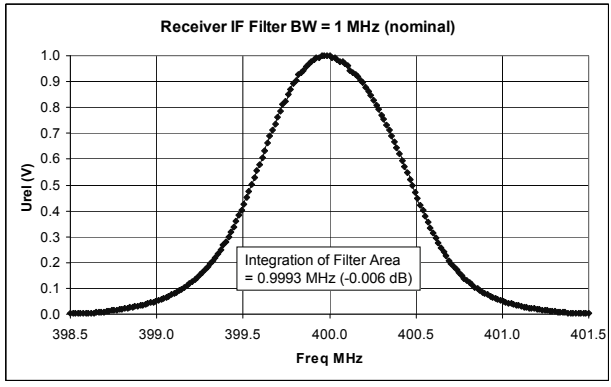


Figure 5. *Measured IF filter area of the used receiver*

The most important uncertainty contributions (k=1) are therefore:

– receiver linearity: 0.15 dB;

– receiver readings (type A): 0.12 dB;

– mismatch (DUT / Receiver): 0.05 dB;

– power standard calibration factor: 0.014 dB;

–power standard linearity: 0.003 dB;

–receiver IF bandwidth: 0.01 dB;

–20 dB step attenuator (S21): 0.004 dB.

The total expanded uncertainty (k = 2) for D_{DUT} is 0.40 dB, the result is shown in Figure 9 (compared with the result of the time domain method).

Time domain calibration method

In this procedure the unknown spectral pulse density D_{DUT} is calculated by a Fourier transformation of the directly measured impulse area. In a first step the impulse area IS_{DUT} is measured by means of a fast sampling oscilloscope (50 GHz, hp54752A). The setup (Figure 6) contains a combination of attenuation and delayline, which is needed for three reasons:

a) attenuation (46 dB) to match the max. input range (±1 V) of the scope;

b) timing: trigger needs to be ≥ 22 ns in advance to the main signal;

c) 50 Ω termination of the DUT signal with a $VRC \leq 0.01$ (40 dB return loss) to keep MU small.

A DC coupled (3 resistor type) power divider and a commercial 22 ns delayline (Picosecond Pulse Lab 5021) are used.

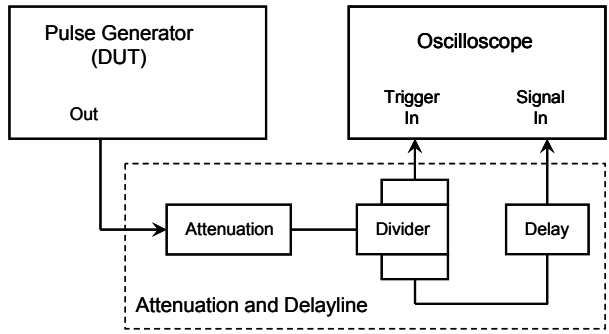


Figure 6. *Measuring setup for the time domain method*

The attenuation and delayline part is characterized as a single unit on a VNA to determine its S-parameters. As already explained in the previous section, mismatch is dealt with solely as an uncertainty without any corrections applied. Due to the very low *VRCs* of the attenuator input and of the scope signal input (≤ 0.01 up to 1 GHz) we can simply use $|1/S21|$ as a transfer function.

The unknown pulse voltage is sampled by the oscilloscope with N discrete points; for each point a new trigger impulse is required. Therefore, great care for the trigger circuit and stable signal conditions are required.

In a second step the time domain data is Fourier transformed into the frequency domain. The result is the DUT spectral pulse density at the oscilloscope input and it has to be corrected with the transfer function $|1/S21|$.

The calibration factor $K_{D\_DUT} = D_{nom} - D_{DUT}$ (D expressed in dB(μV/MHz)) is calculated. These steps are done by means of a MATLAB FFT routine.

Uncertainty contributions

The desired spectral density D_{DUT} (in V/Hz) is given by the equation $D(f_k) = |U(f_k)| \cdot dt \cdot |T(f_k)| \cdot |c_U(f_k)|$ with the Fourier transformation $U(f_k)$ of the sampled oscilloscope voltage $u(t_j)$, the sampling interval dt, the transfer function $T(f_k)$ and the scope frequency response $c_U(f_k)$.

The sampling window $t_w = N \cdot dt$ (dt = sampling interval) has to be chosen carefully. On one side it defines the frequency resolution ($1 / t_w$) but if the window is chosen too long, the measured pulse area can become biased due to DC offset. In addition to the ordinary internal offset of the scope, such an offset can also arise due to feedback from trigger input to signal input. For this reason it is advised to measure a pulse amplitude as large as possible to minimize the impact of DC bias.

The total attenuation (DUT to scope input) is therefore a trade-off between optimizing signal to noise ratio, DC-bias and DUT matching condition.

Choosing an unsuitable window length can in addition lead to spurious spectral components. This is illustrated in Figure 8, which shows the analysis of the pulse data in Figure 7 for two different window length 2.5 ns and 10 ns, the longer of which shows a superimposed oscillating pattern most likely due to additional reflections in the setup. It is generally advised to keep the window length as short as reasonably possible.

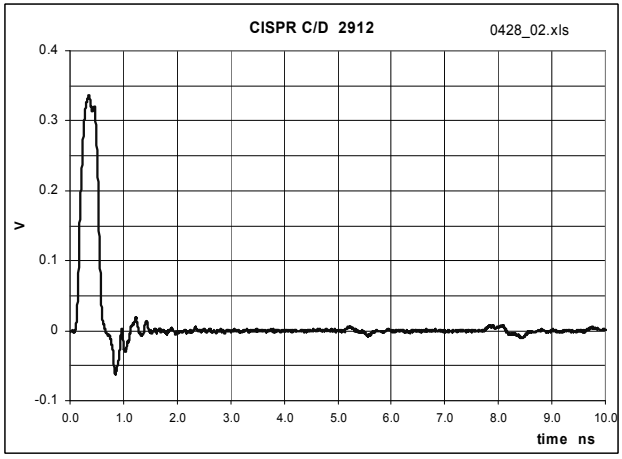


Figure 7. *DUT signal with reflections (at about 5 and 8 ns)*

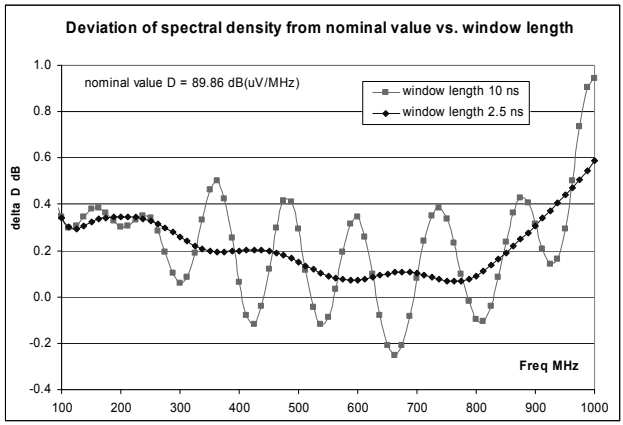


Figure 8. *Spectral density vs. frequency (analysis of sampled signal from Figure 7 for different window lengths)*

This results in a decreased frequency resolution of $1/t_{w,}$

which can be improved again by zero-padding the sampled data array.

The oscilloscope has been calibrated at DC with a DC voltage source and a precision multimeter and the noise contribution has also been estimated. The scopes frequency response $c_U(f_k)$ has been calibrated with a similar setup as shown in Figure 3, where the power splitter has been directly connected to the scope input.

The transfer function $|1/S21|$ (attenuation and delayline) and the scope frequency response $c_U(f_k)$ are the main contributing uncertainties.

For the time domain calibrating procedure the total expanded uncertainty (k = 2) for D_{DUT} is ≤ 0.25 dB and is estimated by applying the Monte Carlo method.

Results and comparison

Figure 9 presents the results from both calibration methods, which show a good agreement within measurement uncertainties.

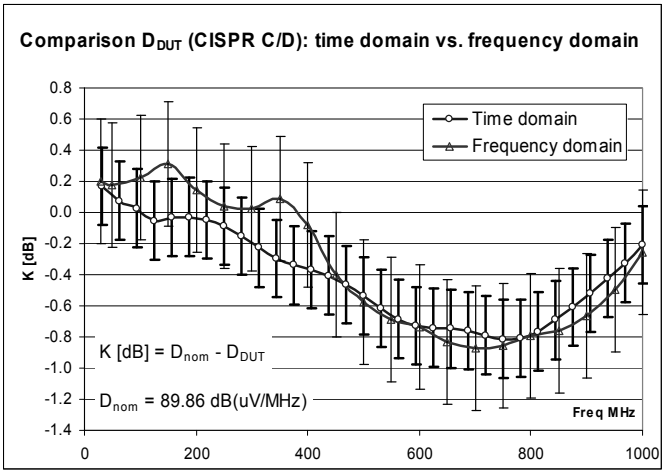


Figure 9. *Comparison of results*

Summary

Both methods are applicable with good results. The time domain method seems to be the better choice because it results in a much faster procedure and in lower uncertainties, but it requires a high precision oscilloscope.

References

[1] IEC, CISPR 16-1-1, Second edition 2006-03: Specification for radio disturbance and immunity measuring apparatus and methods – Part 1-1: Radio disturbance and immunity measuring apparatus – Comité International Spécial Des Perturbations Radioélectriques (CISPR).

[2] G. Schwarzbeck: User manual IGU2916 EMI Calibration Pulse Generator.

[3] Expression of the Uncertainty of Measurement in Calibration; EA-4/02; European co-operation for Accreditation; December 1999.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Binary Josephson array power standard

R. Behr[1], L. Palafox[1], J. M. Williams[2], S. Djordjevic[3], G. Eklund[4], H. E. van den Brom[5], B. Jeanneret[6], J. Nissilä[7], A. Katkov[8], and S. P. Benz[9]

[1]Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, D-38116 Braunschweig, Germany
[2]National Physical Laboratory (NPL), Hampton Road, Teddington, Middlesex, TW11 0LW, UK
[3]Laboratoire National de Métrologie et d'Essais (LNE), 29 avenue Roger Hennequin,
F-78197 Trappes Cedex, France
[4]SP Technical Research Institute of Sweden, Box 857, SE-501 15 Borås, Sweden
[5]NMi Van Swinden Laboratorium, P.O. Box 654, NL-2600 AR Delft, The Netherlands
[6]METAS, Swiss Federal Office of Metrology, Lindenweg 50, CH-3003 Bern-Wabern, Switzerland
[7]MIKES, Otakaari 7B, FIN-02150 Espoo, Finland
[8]Mendeleyev Institute for Metrology (VNIIM), 190005 St. Petersburg, Russia
[9]National Institute of Standards and Technology (NIST), Boulder, CO 80305 USA

ABSTRACT: In the first year of the iMERA joint research project "Binary Josephson Arrays Power Standard", important foundations have been laid towards linking power standards to a quantum-based system. The amplitude of the Josephson arrays suitable for AC applications has been increased to the 10 V level and two alternative methods for the application of Josephson standards to power calibration have been developed and preliminary tests carried out. In addition, the switching transients between quantized voltage steps have been investigated at the 10 ns level and the calculation of the uncertainty in synthesized waveforms due to these transients has been improved.

Introduction

This iMERA joint research project aims at enhancing the measurement capabilities for quality and efficiency of electrical power, which is needed as a result of the deregulated electricity distribution market. It is also important for working towards ensuring a consumer orientated, reliable and sustainable supply of electric power in the face of increasingly scarce and expensive resources. This project is one of the first steps towards implementation of industrial measurement systems that are fully based on fundamental constants and electrical quantum standards and which will enable improved production by independent, precision metrology on the workshop floor.

The principal objective of the project is to build one or more quantum-based systems for representing a power standard (120 V, 5 A) at the participating laboratories. The basis will be a programmable Josephson system working at power frequencies (16⅔, 50 and 60 Hz) for synthesizing or measuring AC signals to better than 1 μV/V.

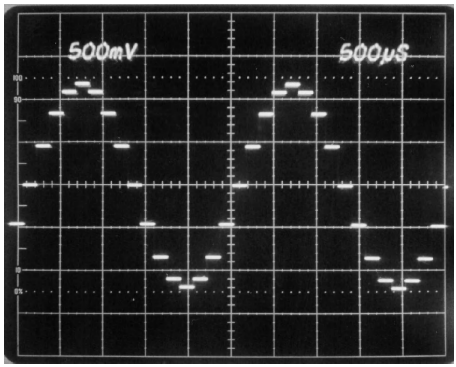


Figure 1. *Sinusoidal waveform generated using a binary Josephson array; X: 500 μs/div. and Y: 500 mV/div*

Following the development of series arrays of intrinsically shunted Josephson junctions in the mid 1990s [1-3], there has been considerable work demonstrating their use as Programmable Josephson Voltage Standards (PJVS). In addition to being essential for producing accurate and stable dc voltages, they have also been applied to ac metrology through the generation of waveforms approximated with a staircase series of constant voltages of equal duration, or samples. These AC-PJVS systems are used to produce arbitrary waveforms allowing e.g. high accuracy ac-dc difference measurements at frequencies up to 1 kHz [4, 5], and fast reversed dc comparisons between Josephson sources and thermal voltage converters [6, 7]. One example of a generated 400 Hz waveform using 64 samples is shown in Figure 1.

Although the steps are inherently quantized, the synthesized waveforms contain undesirable transients between the quantized steps. As the voltage is not known precisely during these transients, they form the major contribution to the uncertainty of the waveform. Therefore, characterization of the transients by using faster electronics and optimization by matching the characteristic impedance are important. Detailed transient measurements and uncertainty analysis for transient shapes have begun.

As the objectives of the project are very ambitious the work is divided into the following work packages (WPs): (WP1) *Transients and electronics,* (WP2) *Array Fabrication,* (WP3) *Verification of precision waveforms,* (WP4) *Sampling and synchronization methods for calibrating waveforms,* (WP5) *Combining a Josephson*

array to power standard, and (WP6) *Validation and intercomparisons.* This paper will describe the progress and goals in each of the areas of this international collaboration.

Transients and electronics

The aim of this task is to identify the limits for switching between quantized steps. The RMS values and the corresponding uncertainty for the waveforms synthesized will be estimated including the contribution from the transients previously measured. Moreover, the bias modules which drive the array segments have been adapted for higher voltages in order to drive 10 V arrays. The predicted larger capacitance presented by superconductor – insulater – normal metal – insulator – superconductor (SINIS) 10 V Josephson arrays is being investigated through further development of the electronics/bias modules.

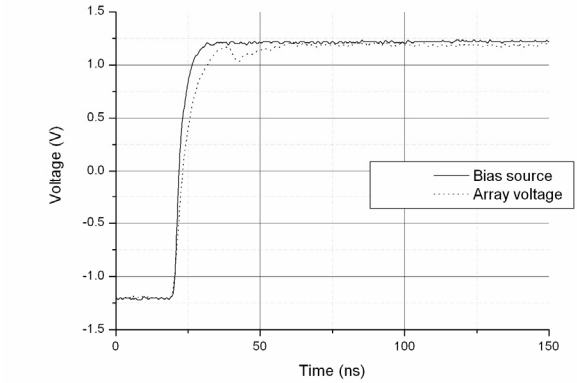


Figure 2. *Bias source output and array response for a –1 to +1 voltage step transition. Feature due to recombination of reflections occurs 20 ns after initial rise*

A schematic diagram of the fast 15-channel bias source developed at NPL and its connection to the Josephson array was presented in [8]. New output drive circuits can realize a rise time on a 50 Ω load of about 10 ns between voltage steps causing only small overshoot and ringing. The bias current can be set to a value in the range -60 mA to +60 mA using Digital to Analog Converters (DACs) with 14 bit resolution. This leads to a current resolution of 0.008 mA, sufficient for biasing an array step with a width of about 0.5 mA at a bias current in the 2 mA to 20 mA range.

A drive system for a 1 V array with 8,192 Josephson junctions has been optimized, and a cable scheme has been devised which compensates for cable reflections. Measurements of the temporal response made with this cable system are compared to an *RC* model with both 3 ns and 7 ns single time constants [9].

Figure 2 shows a transition between the –1 V and +1 V array voltage steps recorded by an oscilloscope. Two things are of interest: first, the array response is slower than the bias source output; second, there is a feature approximately 20 ns from the start of the transition. This 20 ns delay corresponds to the cable delay and equates to the time of reflections from the ends of cables.

The array is initially changing the voltage at a rate comparable with a 3 ns value, but later, it has slowed down to a rate comparable with a 7 ns value. A further experiment to measure with nanosecond accuracy the time taken to reach the quantized level is under way at NPL. In addition, SP is investigating the influence of different types of coaxial cables in the cryoprobe.

Array fabrication

Generating AC waveforms with a PJVS requires steps of constant voltage over a wide current range (> 500 μA). Presently, 1 V and 3 V arrays are routinely fabricated. A few 10 V arrays have been successfully tested, although the fabrication of these arrays remains a big challenge. However, since certain AC applications achieve their best uncertainty in the 10 V range, it is desirable to implement a 10 V AC-PJVS system as a power standard. Therefore, one aim of this WP is to develop and fabricate with reasonable yield high quality 10 V arrays with steps wider than 500 μA.

This work package covers the design and fabrication of programmable arrays for ac applications. Novel 10 V SINIS binary Josephson arrays are also required for this project.

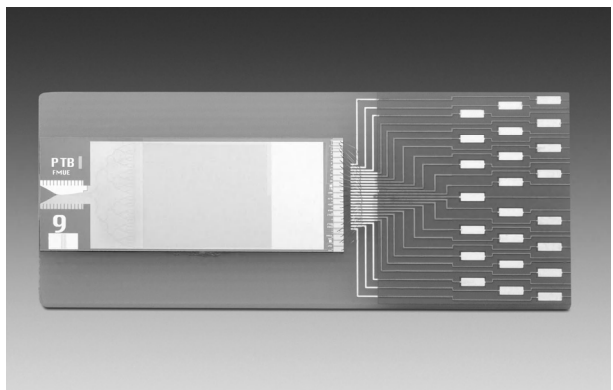


Figure 3. *Photograph of a 10 V SINIS 17-bit DAC chip for a PJVS. Electrical connections to the carrier are made by wire bonding*

Josephson circuits currently made at NIST are double stacked $MoSi_2$-junction superconductor – normal metal – superconductor (SNS) arrays containing 67,408 SNS junctions connected in series to produce a maximum ±2.5 V output voltage and divided according to a ternary weighting scheme [10]. The six least significant bits (LSBs) use a standard ternary configuration with a resolution of 612 μV (smallest array of 16 junctions biased at 18.5 GHz drive frequency), which enables the array to generate 8,427 different quantum-accurate output voltages. The sizes of the seven most significant bits (MSBs) are not a strict ternary implementation, but this enables the entire chip to have 2 mA current margins by not exceeding 8,800 junctions in any array segment. The chips are flex-mounted to ensure uniform microwave power distribution and long-term cryopackage reliability [11].

The original 10 V design was developed at PTB [12] as a 6 bit converter consisting of 69,120 junctions distributed over 64 parallel driven 70 GHz-microwave branches, each containing 1080 Josephson Junctions (JJs). Due to the high microwave attenuation for these junctions, which are embedded into a low-impedance (5 Ω) microstripline, the 10 V-steps are only 200 μA wide. For this reason and considering that the microwave power available from the 70 GHz-Gunn-diode is limited to 30 mW at the finline antenna of the chip, a new design with 128 parallel driven microstriplines was developed. As a result, the 69,632 junctions were distributed between 128 sub-arrays.

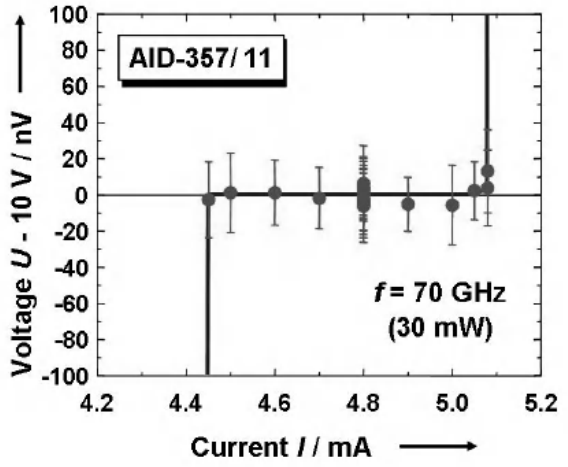


Figure 4. *Results from the first comparison between a programmable binary SINIS array and a SIS voltage standard at the 10 V level*

At National Metrology Institutes (NMIs) voltage is represented using 10 V SIS Josephson voltage standards with zero-current steps. PTB has compared a binary divided 10 V SINIS array with their Josephson voltage standard [13]. The comparison showed good agreement at the 10 V level with a difference of U_{SIS} –

U_{SINIS} = (1.8 ± 0.4) nV (N_{meas} = 64). The 1.8 nV deviation is probably due to higher than normal leakage currents in the test-cryoprobe used for the 10 V SINIS array.

Verification of precision waveforms

A Josephson waveform synthesizer uses bias electronics to drive a Josephson array as a quantum-accurate DAC. The objective of this WP is to develop a measurement routine that verifies these synthesized waveforms. This WP is directly correlated with (WP1) *Transients and electronics* as the transients mainly determine the uncertainty.

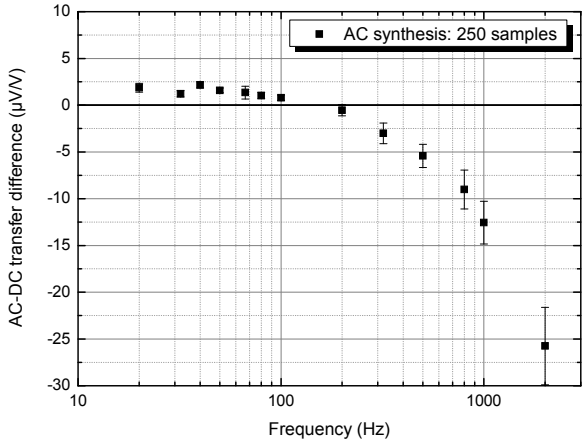


Figure 5. *AC-DC difference measurements of LNE using a Thermal Converter [15]*

In a preliminary experiment with 200 ns long transients, the uncertainty of synthesized sine waves is 0.1 – 1 μV/V below 200 Hz [4], but it is difficult and time consuming to adjust all the parameters: bias module offset correction, step bias current (for the 13 segments), load compensation current sources, synthesized frequency, number of samples, etc. This procedure needs to be validated by performing time consuming AC/DC transfer measurements with thermal converters. Informal intercomparisons using thermal converters with different resistances (90 Ω – 1.5 kΩ) have been started to test the set-up routine and investigate the influence of the additional load compensating current sources required when driving a thermal converter with a Josephson array.

At VNIIM, new models to calculate uncertainties related to different measured transient shapes and rise times have been developed [14].

Sampling and synchronization methods for calibrating waveforms

Different sampling methods are now established in several NMIs for representing their power standards. Synchronizing an AC-PJVS with a sampling method seems ideal as the transients between quantized steps can be disregarded.

One aim of this work package is to connect a sampling A/D converter as a zero detector between two synthesized waveforms, enabling a comparison between a conventional and a Josephson AC waveform. The phase of the Josephson AC waveform needs to be adjustable in order to keep the voltage at the A/D converter as small as possible. Sources to be calibrated must allow synchronization with the sampling system.

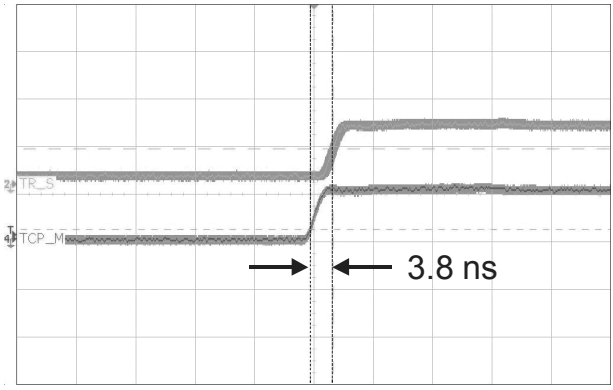


Figure 6. *Oscilloscope measurement of two synchronized AC-PJVS. The latest records are shown in color, with all the acquisitions over 2 hours shown in gray*

This procedure will require investigating various sampling methods currently employed by the partners, discussing results and optimizing methods and/or developing special sampling A/D converters. The best method for combining power measurements with Josephson standards would then be established.

At PTB the AC PJVS systems have been extended in order to enable the generation of two waveforms with a defined phase relationship. Optical coupling of the synchronization signals (clock, trigger input and output) has also been added to allow one or both sources to operate above ground potential and the differences measured using a sampling voltmeter (DVM) [16].

These extended AC PJVSs have been used to measure the difference between two synchronous waveforms with a defined phase relationship (Figure 6). We have chosen “rectangular waves” with ±1.2 V amplitude range.

The microwave frequencies and number of junctions biased in each array have been chosen to match the amplitudes to within 100 nV. Allan variance analysis indicated that the optimal sampling windows contained 128 and 256 waveform periods, depending on the integration time used. The measurements in Figure 7 were taken over 500 polarity reversals, equivalent to a measurement time of about 30 minutes. For every frequency the integration has been selected to be within the range from 10 μs at 5 kHz up to 1 ms at 70 Hz.

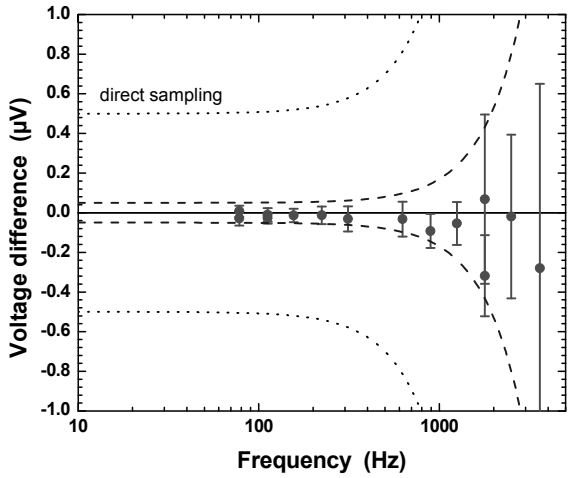


Figure 7. *Measured difference between two AC PJVS-synthesized square waves at different frequencies and at 1.2 V peak amplitude*

The type A uncertainty of this method (error bars and dashed line), is an order in magnitude better in comparison to the direct sampling method (dotted line). At frequencies below 400 Hz the uncertainty is 5×10^{-8} (60 nV for 1.2 V peak voltage). Towards higher frequency the uncertainty increases as the integration times become shorter and shorter. Recently, investigations of other zero-detectors have been started at METAS.

Combining a Josephson array with a power standard

In this work package, power standards based on Josephson AC waveforms will be compared to those presently available. Since 1 V arrays are already available, investigations will start in the 1 V range as soon as a complete system is available.

Two main possibilities for implementing a power standard based on Josephson-synthesized waveforms will be investigated:

- The synthesized waveforms can be amplified in order to obtain a calculable power standard. The gain uncertainties and limited bandwidths of the amplifiers will limit the practical uncertainty. The Josephson system as a quantum standard is expected to improve calibration of the amplifiers and dividers.
- The waveforms to be measured can be transformed using dividers and directly sampled against a Josephson system. The uncertainty introduced by the dividers has to be investigated.

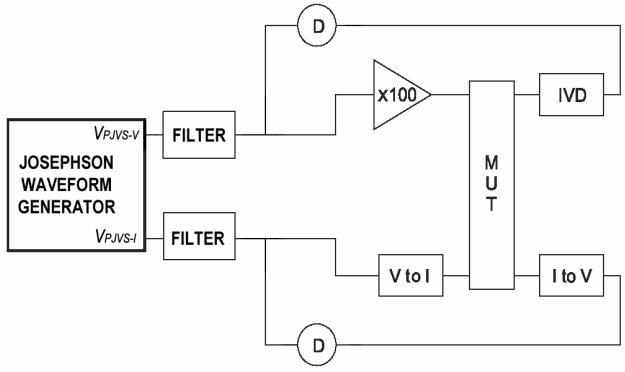


Figure 8. *Schematic of one possible implementation of the power standard with two Josephson reference voltage waveforms and other essential components. The symbol "D" denotes high precision phase-sensitive detectors*

One quantum-based power standard system proposed by NIST consists of two independent AC PJVS systems, each with separate amplitude and phase controls. Figure 8 illustrates one possible configuration of the system, where $V_{PJVS\text{-}V}$ is a 1.2 V_{rms} reference sinewave for the voltage channel and $V_{PJVS\text{-}I}$ is a 0.5 V_{rms} reference sinewave for the current channel. A 100x amplifier scales $V_{PJVS\text{-}V}$ up to 120 V_{rms}, which is applied to the power-meter under test (MUT). This same 120 V signal is divided down by a factor of 100 by use of an inductive voltage divider (IVD), and then compared in real time with $V_{PJVS\text{-}V}$ so that any gain drift in the 100x amplifier can either be calibrated or actively stabilized using servo feedback. For the MUT current channel, $V_{PJVS\text{-}I}$ is fed to a *V-to-I* converter to generate a current on the order of 5 A. A multi-stage current transformer with a precisely known burden resistance is used to measure this current so that gain drift in the *V-to-I* conversion can be measured and again either calibrated out or removed using feedback. The frequency spectrum generated can be calculated precisely. This allows uncertainty calculation of rms values at power frequencies as shown in Figure 9 [17].

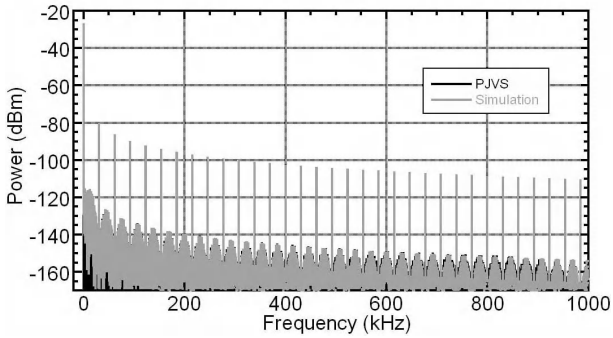


Figure 9. *Spectral measurement of a 512-sample 60 Hz sine wave generated by a Josephson array with 2.1 V amplitude zero-to-peak (1.5 V_{rms}). Measured harmonics from the Josephson array (black) is in excellent agreement with numerical simulations (gray displayed in front of black)*

The second approach is being investigated at PTB in order to reduce the dominant contribution to the uncertainty for power measurements, which comes from the sampling DVM [18]. This contribution is, in turn, dominated by the uncertainties due to gain error and drift of its internal voltage reference. It was therefore decided to investigate first the improvement which results from using a single AC PJVS, as in the block diagram in Figure 10. This arrangement makes it possible to add a DVM calibration immediately before and after the measurements of the device under test. The factory calibration, using an automated Josephson voltage standard, of the DVM employed is thus extended to characterize the dynamic contributions relevant for the sampling parameters used in the power standard.

The sampled data from $\underline{U}_1$ and $\underline{U}_2$ are used to calculate $\underline{U}$, $\underline{I}$, and the phase angle, φ, between them at the device under test. The power is then calculated and compared to the reading from the device under test.

Preliminary measurements have shown that incorporating an AC PJVS into the PTB primary power standard would simplify the traceability chain to the unit volt to the absolute minimum. More importantly, it would also reduce the dominant uncertainty contribution from 10×10^{-7} to 3.8×10^{-7} (both $k = 1$). 1 μV/V to 0.38 μV/V [19]. As a result, the uncertainty for the power standard is no longer dominated by the contribution from the sampling voltmeter. Further reductions in the uncertainty will require to improve the conversion of the voltage and current signals from the 120 V and 5 A levels to the input range of the sampling voltmeter.

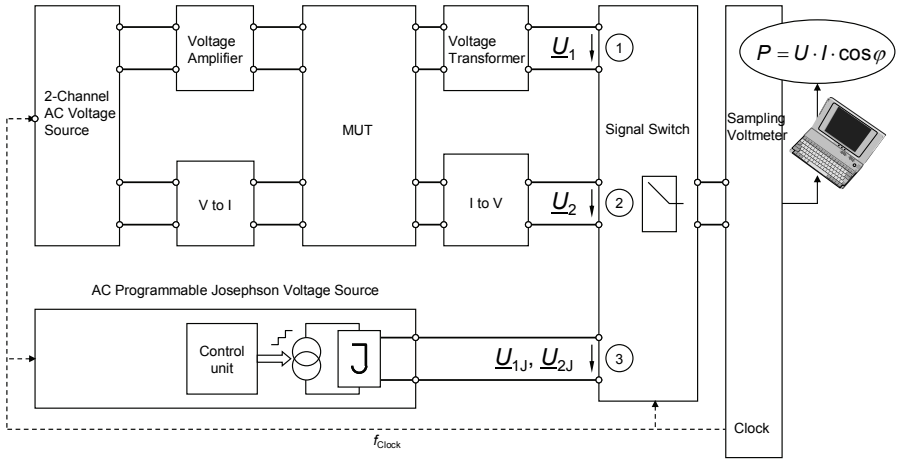


Figure 10. *Proposed primary power standard incorporating a Josephson Waveform Synthesizer to calibrate the sampling voltmeter "in-situ"*

References

[1] C. A. Hamilton, C. J. Burroughs, and R. L. Kautz, "Josephson D/A converter with fundamental accuracy", *IEEE Trans. Instrum. Meas.,* vol. 44, pp. 223-225, April 1995.

[2] S. P. Benz, C. A. Hamilton, C. J. Burroughs, T. E. Harvey, and L. A. Christian, "Stable 1-Volt programmable voltage standard", *Appl. Phys. Lett.,* vol. 71, pp. 1866-1868, September 1997.

[3] H. Schulze, R. Behr, F. Müller, J. Niemeyer, "Nb/Al/AlO/Al/AlO/Al/Nb Josephson junctions for programmable voltage standards", Appl. Phys. Lett. Vol. 73, pp. 996-998, 1998.

[4] R. Behr, J. M. Williams, P. Patel, T. J. B. M. Janssen, T. Funck and M. Klonz, "Synthesis of Precision Waveforms using a SINIS Josephson Junction Array", *IEEE Trans. Instrum. Meas.,* vol. 54, no. 2, pp. 612-615, April 2005.

[5] R. Behr, L. Palafox, J. Schurr, J. M. Williams, P. Patel, and J. Melcher, "Quantum effects as basis for impedance and power metrology", in *Proc. of the 6th International Seminar on Electrical Metrology,* pp. 11-12, September 21-23, 2005, Rio de Janeiro, Brazil.

[6] C. Burroughs, S. P. Benz, C. A. Hamilton, T. E. Harvey, J. R. Kinard, T. E. Lipe, and H. Sasaki, "Thermoelectric transfer difference of thermal converters measured with a Josephson source", *IEEE Trans. Instrum. Meas.,* vol. 48, no. 2, pp. 282-284, April 1999.

[7] T. Funck, R. Behr, and M. Klonz, "Fast reversed dc measurements on thermal converters using a SINIS Josephson junction array", *IEEE Trans. Instrum. Meas.,* vol. 50, no. 2, pp. 322-325, April 2001.

[8] J. M. Williams, P. Kleinschmidt, T. J. B. M. Janssen, P. Patel, R. Behr, F. Müller, and J. Kohlmann, "Synthesis of precision AC waveforms using a SINIS Josephson junction array", in *CPEM 2002 Conf. Dig.*, pp. 434-435, Ottawa, Canada, June 2002.

[9] J. M. Williams, D. Henderson, P. Patel, R. Behr, and L. Palafox, "Achieving sub-100-ns switching of programmable Josephson arrays", *IEEE Trans. Instrum. Meas.,* vol. 56, no. 2, pp. 651-654 April 2007.

[10] Y. Chong, C. J. Burroughs, P. D. Dresselhaus, N. Hadacek, H. Yamamori, and S. P. Benz, "Practical high resolution programmable Josephson standard using double- and triple-stacked $MoSi_2$ barrier junctions", *IEEE Trans. Appl. Supercond.,* vol. 15, no. 2, pp. 461-464, June 2005.

[11] C. J. Burroughs, P. D. Dresselhaus, Y. Chong, and H. Yamamori, "Flexible cryo-package for Josephson devices", *IEEE Trans. Appl. Supercond.,* vol. 15, no. 2, pp. 465-468, June 2005.

[12] H. Schulze, R. Behr, J. Kohlmann, F. Müller, and J. Niemeyer, "Design and fabrication of 10 V SINIS Josephson arrays for programmable voltage standards", *Supercond. Sci. Technol.,* vol. 13, pp. 1293-1295, September 2000.

[13] F. Müller, R. Behr, L. Palafox, J. Kohlmann, R. Wendisch, and I. Krasnopolin, "Improved 10 V SINIS series arrays for applications in AC metrology", to be published in *IEEE Trans. Appl. Supercond.*, vol. 17, no.2, June 2007.

[14] R. Behr, L. Palafox, T. Funck, J. M. Williams, P. Patel, and A. Katkov, "Synthesis of precision calculable AC waveforms", *IEEE Trans. Instrum. Meas.,* vol. 56, no. 2, pp. 289-294, April 2007.

[15] S. Djordjevic, A. Poletaeff, R. Behr, and L. Palafox, "AC-DC transfer measurements of AC waveform synthesized using SINIS Josephson junction arrays", in *CPEM 2006 Conf. Dig.*, pp. 372-373, Turin, Italy, July 2006.

[16] R. Behr, L. Palafox, G. Ramm, H. Moser, J. Melcher, "Direct comparison of Josephson waveforms using an ac quantum voltmeter", *IEEE Trans. Instrum. Meas*., vol. 56, no. 2, pp. 235-238, April 2007.

[17] C. J. Burroughs, et al., "Development of a 60 Hz power standard using SNS programmable Josephson voltage standards", *IEEE Trans. Instrum. Meas.,* vol. 56, no. 2, pp. 289-294, April 2007.

[18] W. G. Kürten Ihlenfeld, "Maintenance and traceability of ac voltages by synchronous digital synthesis and sampling", *PTB, Braunschweig, Germany, Rep. E-75*, 2001.

[19] L. Palafox, G. Ramm, R. Behr, W. G. Kürten Ihlenfeld, and H. Moser,"Primary ac power standard based on programmable Josephson junction arrays", *IEEE Trans. Instrum. Meas.,* vol. 56, no. 2, pp. 534-537, April 2007.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

A 100 MΩ step hamon guarded network to improve the traceability level of high resistance up to 1GΩ

P.P. Capra[1], F. Galliana[2], M. Astrua[3]
National Institute of Metrological Research
str. delle Cacce 91 – 10135 Turin, Italy
[1] Tel.: +39 011 3919424 – fax: +39 011 346384 – capra@inrim.it
[2] Tel.: +39 011 3919336 – fax: +39 011 3919334 galliana@inrim.it, flaviogal@alice.it
[3] Tel.: +39 011 3919424 – fax: +39 011 346384, m.astrua@inrim.it

ABSTRACT: At INRIM, a 100 MΩ step Hamon, guarded resistor was developed and metrologically validated. With it the traceability level of the maintained 1 GΩ at INRIM improved from $8 \cdot 10^{-6}$ to $7 \cdot 10^{-6}$. The guarding system is a chain of 10 MΩ resistors. The main 100 MΩ resistors are surrounded by two metal rings driven to guard potentials so the leakage resistance in parallel with them is fragmented in three parts. In two of these the leakage current is sensibly limited and the third portion is in parallel with the guard resistors. Details of the development of the Hamon, stability data in parallel and series configurations and a metrological validation with DMM-Calibrator-based method are given.

Introduction

The metrological chain, the devices and standards in the field of high value resistance are in continuous evolution due to the need to carry out increasingly accurate measurements, in particular due to the requests of the secondary and industrial laboratories. In this framework, a DMM-DC voltage calibrator-based measuring system for calibration of standard resistors mainly in the field 10 MΩ÷10 TΩ and for the determination of the voltage coefficient of high-value resistors under calibration, was developed and characterized. Moreover, a revision of a measurement method based on Hamon transfer boxes was also performed [1]. This method involves four Hamon resistance boxes with at least ten resistors of 10 kΩ, 100 kΩ, 1 MΩ and 10 MΩ respectively that can be configured to measure the resistors individually, in parallel or in series to extend the traceability from a high-precision 10 kΩ resistance standard, calibrated in terms of the IEN 1Ω primary group referred to the value $R_{K\text{-}'90}$ of the von Klitzing constant by means of the quantum Hall effect, up to 100 MΩ. To extend the traceability of this method and improve the traceability levels achieved with the DMM-DC voltage calibrator-based method at 1 GΩ level, a 100 MΩ step Hamon standard resistor with guarding system was projected, built and metrologically validated [2]. Moreover, the project of this Hamon-type resistor is applicable to higher value steps Hamon standards to extend traceability up to 10 TΩ and 100 TΩ.

The 100 MΩ step Hamon resistor

In Figure 1 a scheme of the Hamon resistor, is given. The main ten 100MΩ $\mathbf{R_M}$ resistors are thick film-type resistance elements. Before mounting these elements, they were accurately treated with pure isopropylic alcool to eliminate impurities. Moreover, they were individually accurately measured and they resulted in being matched within $3.4 \cdot 10^{-4}$ so, according to Hamon theory, the accuracy of the transfer 10 MΩ÷1 GΩ is of the order of $1.2 \cdot 10^{-7}$ that is obviously negligible. The passage from the configuration series and parallel is done with the simple movement of the mobile element of the Hamon box (upper part of Figure 1). All utilized connectors and cables are coaxial. The network of resistors is kept in a cylindrical box. The connectors of the series (BNC) are placed on the top, while the terminals of the parallel (binding post) are on the bottom. Inside the box, near the resistors, a 10 kΩ NTC is also placed.

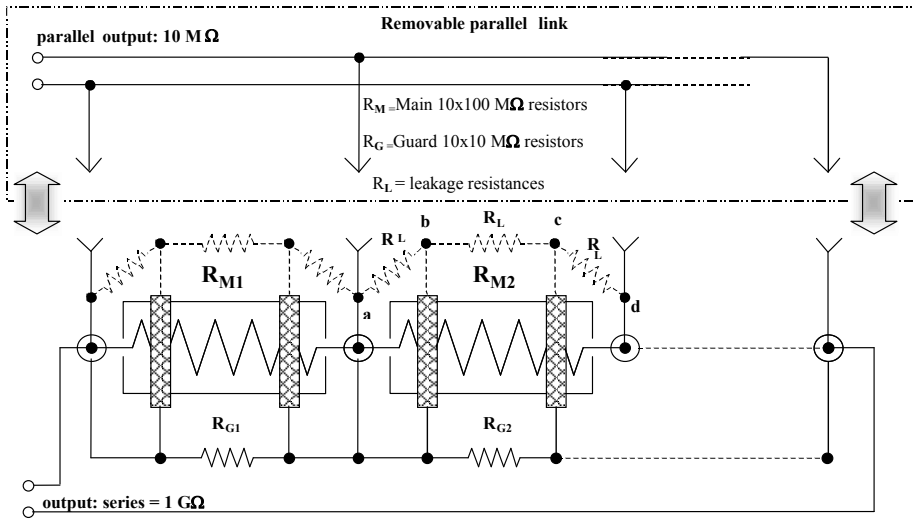


Figure 1. *Scheme of the 100 MΩ step Hamon resistor*

The guarding system

A guarding system for high resistance Hamon resistors was performed in [5]. In our resistor, the guarding chain is composed of ten resistors $\mathbf{R_G}$ with nominal value 10 MΩ. In Figure 1, coaxial terminations of the main 100 MΩ resistors are driven to a guard potential that is nearly the same potential as their junction limiting the leakage current flowing through the insulator material of the connectors. Moreover, each main resistor is surrounded by two metal rings, also driven to guard potentials.

In this way the leakage resistance, in parallel with a main resistance element, is fragmented in to three parts $\mathbf{R_L}$ (thin spotted resistances in Figure 1). The points **a**) and **b**) are nearly at the same potential limiting the leakage current flowing through this portion of the total leakage resistance. The same is valid for the portion between the points **c**) and **d**). The portion of the leakage resistance between the points **b**) and **c**) is in parallel with the guard resistor that led us to consider an additional negligible uncertainty.

Metrological validation of the realised Hamon standard

In order to evaluate the performance of the developed Hamon standard, the accuracy of its 1:100 traceability transfer was verified with the DMM-dc voltage calibrator-based method [1]. Figure 2 shows the scheme of adopted procedure. In a first step the 100 MΩ step Hamon standard, in parallel configuration, was calibrated against a Guildline 10 MΩ mod 9330 s/n 63054 maintained in a thermostatic enclosure with active regulation of temperature with a mid-term stability of (23 ± 0.01)°C, and used to maintain the resistance value at the 10 MΩ level, using a substitution method with a high precision Fluke mod 8508 DMM. Then, in a second step, the calculated 1 GΩ value Hamon standard, in series configuration, is verified with the DMM-dc voltage calibrator-based high value resistor method using the same Guildline 10 MΩ mod 9330 as standard resistor (Figure 2). The 1 GΩ values, obtained with the Hamon transfer and with the DMM-dc voltage calibrator method agreed within the uncertainties of the two methods. The evaluated differences (in 10^{-6}) was $2{,}4 \cdot 10^{-6}$. The Hamon was also compared satisfactorily with the Wheatstone bridge method using dc voltage calibrators in two arms [3, 4].

Calibration and use of the 100 MΩ step Hamon resistor to extend the high dc resistance traceability

In Figure 3, the stability behaviors in a short time measurement period typical of Hamon utilization technique of the developed Hamon resistor, either in parallel and in series configuration, are reported. While for the metrological validation it was compared against a 10 MΩ fixed standard resistor Guildline mod to take advantage from its mid-term stability, in common practice the parallel output of the 100 MΩ step Hamon is calibrated comparing it to the series output of a ESI SR1050 1 MΩ step Hamon box, as visible in the traceability chain of Figure 4.

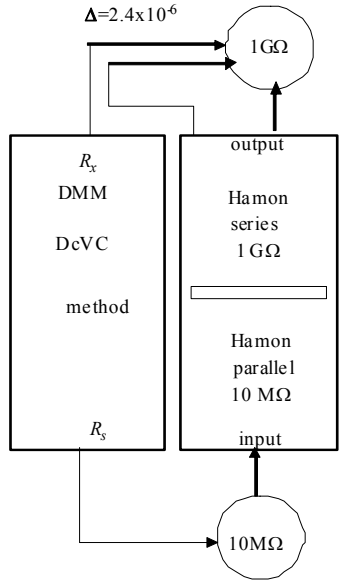


Figure 2. *Scheme of the procedure for metrological validation of the Hamon resistor*

The chain starts from the above mentioned high-precision 10 kΩ standard, with temperature coefficient $\alpha_{23}=0.054\cdot10^{-6}$/ °C, power coefficient ($<1\cdot10^{-6}$/W) and stability ($0.05\cdot10^{-6}$/year). Then, passing through a ESI SR1010 10x10 kΩ transfer box, the parallel output of a ESI SR1050 1 MΩ step Hamon box is calibrated. Finally, the parallel output of the 100 MΩ step Hamon is compared against the series output of the ESI SR1050. All these comparisons are made in 1:1 ratio directly connecting the box outputs to DMM. To transfer the traceability to a 1 GΩ Guildline standard resistor that is kept permanently in an air enclosure with stability of (23 ± 0.02)°C and used to maintain the resistance unit at 1 GΩ level, the series of the 100 MΩ step Hamon is compared with this standard by substitution using the DMM-dc voltage-based measurement method. In parallel configuration the Hamon standard is calibrated at a voltage of 10 V while in series configuration it is used at a voltage of 100V to leave the voltage across the main resistors in the transfer unchanged.

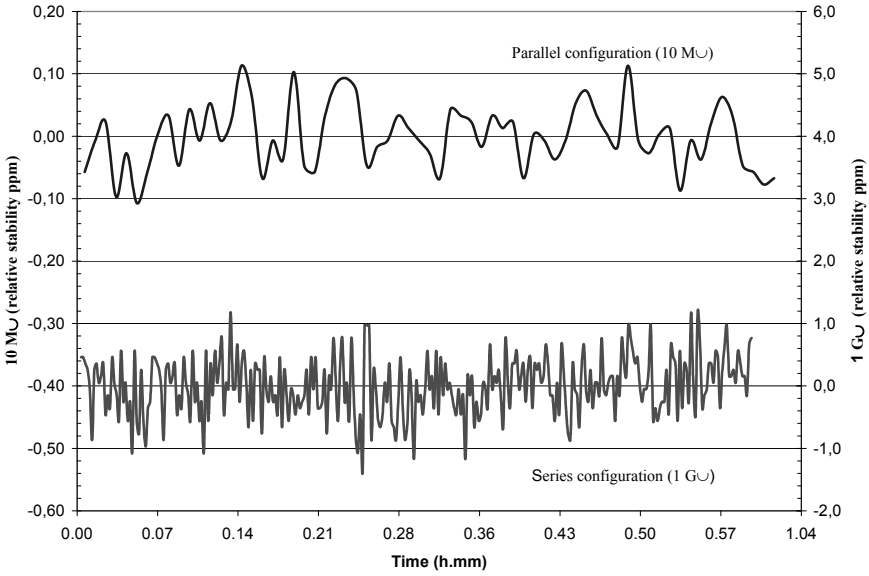


Figure 3. *Stability in parallel and series configuration of the Hamon resistor in a short time measurement period*

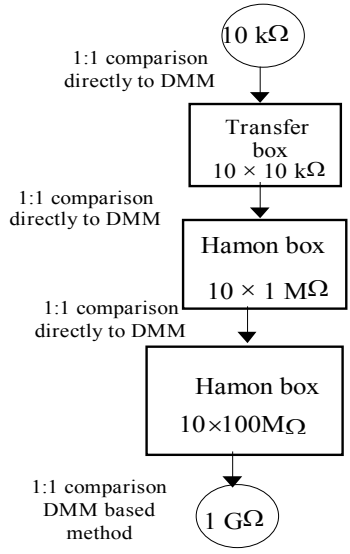


Figure 4. *Traceability chain starting from a high-precision 10 kΩ standard up to 1 GΩ*

Conclusions

As these first positive results of characterisation and of the realised 100 MΩ step Hamon resistor were positive and, as the project of this Hamon-type resistor is applicable to higher values, future aims of our work will be the development of other higher value step Hamon standards to extend traceability up to 10 TΩ and 100 TΩ. With these new standards, eventually controlling their environmental parameters [6], it will be also possible to setup a calibration system for calibration of pico-ammeters in dc current.

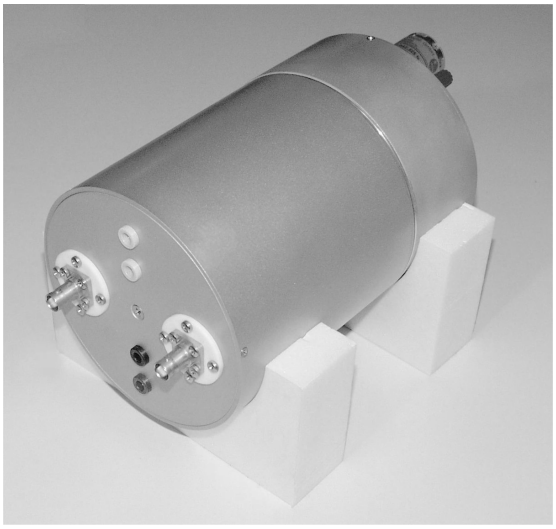


Figure 5. *View of the developed Hamon standard prototype in its final version*

References

[1] F. Galliana, G. Boella, "The electrical dc resistance scale from 100 kΩ to 1 TΩ"*, IEEE Trans. Meas., Vol. 10*, pp. 959 – 963, October 2000.

[2] P.P. Capra, F. Galliana and D.Serazio: "Realization of a 100 MΩ step HAMON guarded trasfer standard resistor to improve the traceability levels of high value resistance in dc current at INRIM, *Proceedings of the Conference on precise Electromagnetic Conference*, Turin 2006

[3] L. C.A. Henderson: "A new technique for the automatic measurement of high value resistors", *J. Phys. E. Sci. Instrum., Vol. 20, pp. 492-495, 1987.*

[4] D. Jarrett: “A new technique for the automatic measurement of high value resistors”, *IEEE Trans. Instr. Meas., Vol. 46, no. 2, pp. 325-328*, 1997.

[5] D Jarrett “Evaluation of guarded high-resistance Hamon transfer standards”, *IEEE Trans. Instr. Meas., Vol. 48, no. 2, pp. 324-328*, April 1999.

[6] F. Galliana, P.P. Capra*:* “A temperature, humidity and pressure independent 1 GΩ standard resistor: realization and preliminary results”, *Proceedings of the 12th IMEKO TC-4 International Symposium Electrical Measurements and Instrumentation*, September 25–27, 2002, Zagreb, Croatia.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Traceability of voltage and resistance measurements in Estonia

A. Pokatilov, T. Kübarsepp

Metrosert Ltd, Aru 10, 10317 Tallinn, Estonia
e-mail: andrei.pokatilov@metrosert.ee, toomas.kubarsepp@metrosert.ee

ABSTRACT: The capabilities in measurements of direct voltage and electrical resistance have been improved in Estonia. The Volt and Ohm standards are based on electronic voltage standards and standard resistors, respectively. The reference standards are traceable to Josephson voltage (JVS) and quantized Hall (QHR) effect resistance standards by means of the annual calibration of transfer standards. The estimated expanded uncertainty is 1 μV/V for 10 V output of the volt maintenance system and 1 μΩ/Ω for standard resistors in the range from 1 Ω to 10 kΩ.

Introduction

Industry and science apply a number of measurements of direct current electrical quantities. The successful development in these areas requires uniform and accurate measurements, which are ensured by firmly established traceability.

In 2006 the standards for units of voltage and electrical resistance were appointed [1, 2] in Estonia. The standards are maintained in the specially designed laboratory of Metrosert Ltd (Central Office of Metrology in Estonia) in which relative humidity and air temperature are stabilized within ranges of (45 ± 5)% and (23.0 ± 0.5)°C, respectively.

In this paper new measurement capabilities in direct voltage and electrical resistance in Estonia are described.

Voltage standards

Maintenance and traceability

Ten solid-state references (so-called Zener diodes) [3]-[5] are used in an automated system for maintenance and dissemination of the volt. Standards are divided into two modules. The transfer module includes four units which are annually calibrated against the JVS or secondary standards traceable to the JVS at 10 V and 1 V nominal values at MIKES (Center for Metrology and Accreditation, Finland) see Figure 1. After calibration, the voltage unit is transferred to the maintenance module, which is used to represent the volt in time between external

calibrations. The maintenance module incorporates a zero-detector, a built-in scanner and six Zener units connected to the hardware average (HAV). Averaging the outputs of several units is needed to reduce the influence of environmental effects [3]-[5], which gives a higher degree of confidence in predictability of the HAV output.

The voltage measurements in the range from 10 mV to 1 kV are related to 10 V HAV output by means of the high resolution digital multimeter and multifunction calibrator.

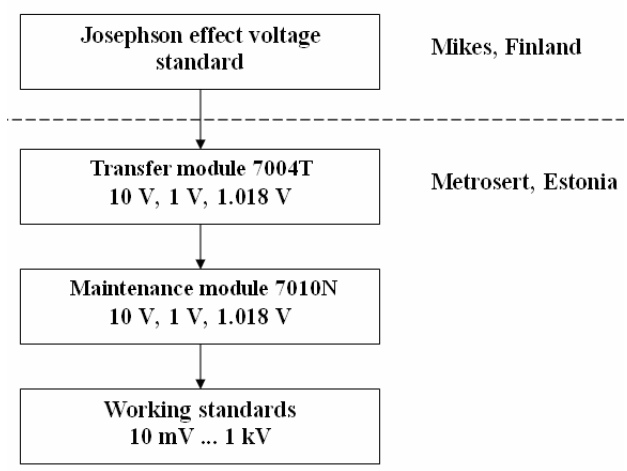


Figure 1. *Traceability of DC voltage measurements in Estonia*

Measurement results

The standards are regularly compared to the hardware average of six units to identify any unexpected changes in 10 V outputs of individual references. Typical measured differences between outputs of the HAV and a single Zener unit are shown in Figure 2. The linear regression line is fitted to the measurement data. The fluctuations near the regression line are estimated to be less than two parts in 10^7. The estimate of fluctuation is taken into account in uncertainty budget as a component due to long-term standard deviation of a regression.

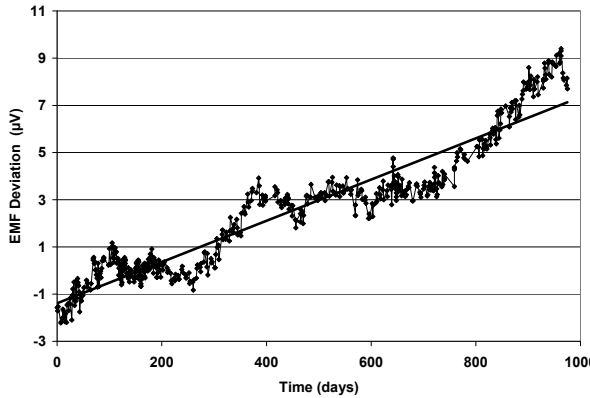


Figure 2. *Typical measured differences between 10 V outputs of the HAV and single Zener unit. The regression is fitted to measurement results*

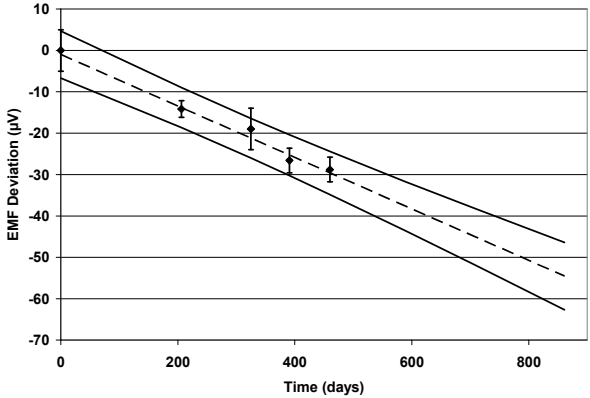


Figure 3. *The calculated regression line for the 10 V output of the hardware average*

The linear drift model is used in [3, 4] to predict the 10 V HAV output in time between calibration points. The least square estimates of regression parameters are found from calibration history (Table 1) for the HAV obtained by external calibration of the transfer module consisting of four Zener units.

Date	Value, V	Expanded uncertainty, μV/V
May 2004	10.000000	0.5
December 2004	9.999985	0.2
April 2005	9.999981	0.5
June 2005	9.999973	0.3
August 2005	9.999971	0.3

Table 1. *Calibration history for the HAV output*

The regression line (Figure 3) for the HAV 10 V output was calculated by applying the weighted least squares fitting method [6]. The associated uncertainty 1 μV/V (k = 2.67) for a 1-year calibration interval includes components due to:

– calibration;

– transportation;

– long-term drift;

– environmental effects;

– noise.

Resistance standards

Maintenance and traceability

The Estonian national standard for resistance is based on the set of standard resistors in the range from 1 mΩ to 10 kΩ. Two reference nominal values at 1 Ω and 1 kΩ are maintained with a group of three standards.

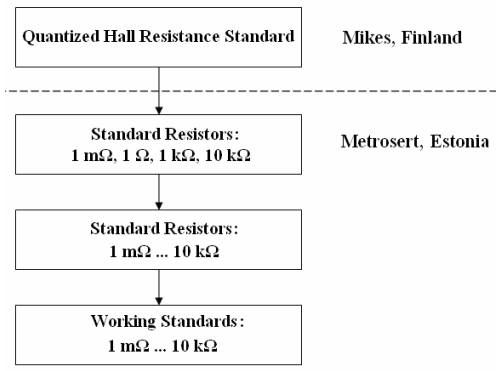


Figure 4. *Traceability of resistance measurements in Estonia*

Traceability of the standard resistors is obtained by regular calibration of four standard resistors, Figure 4. Then the maintenance group of the resistance standards is calibrated with an automated measurement system based on the direct current comparator (dcc) bridge [7], Figure 5.

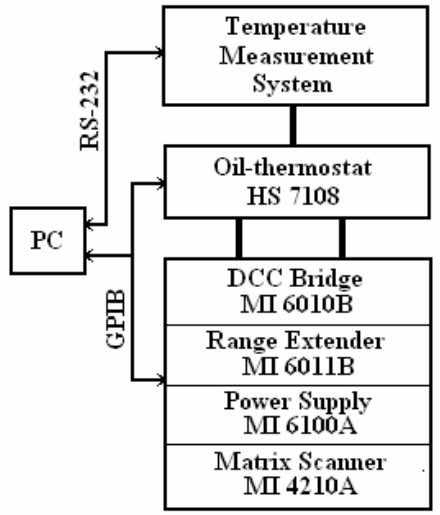


Figure 5. *The automated system for measurements of the standard resistors*

The direct current comparator bridge, range extender and matrix scanner from Measurement International (MI) are used in regular comparison of standards, which allows tracking of relative drifts of the resistance values in time between calibrations. The resistors are immersed in the oil-thermostat to maintain the standards at temperature (23.00 ± 0.05)°C. The temperature is measured from inside the resistors with 16 fast response platinum resistance thermometers (PRT). The system is fully automated with measurement software developed at Metrosert ltd.

Measurement results

The monitoring of the stability of resistance values is conducted by regular inter-comparison of the standard resistors within the range from 1 mΩ to 10 kΩ. As an example, the measurement results for the 10:1 ratio of the 100 Ω and 10 Ω standard resistors are shown in Figure 6. During the period of more than 1.5 years the drift of the ratio is less than 1 part in 10^7.

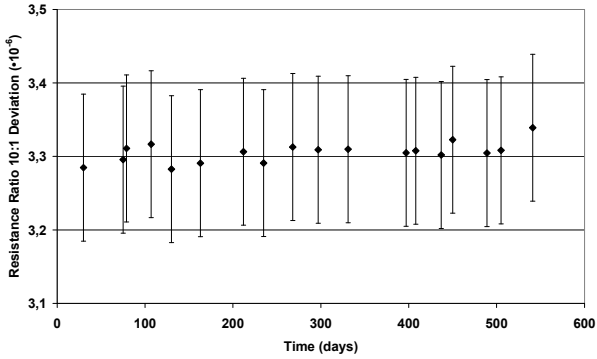


Figure 6. *Measurement results for the 10:1 ratio of the 100 Ω and 10 Ω standard resistors. Vertical lines denote accuracy specification limits of the bridge*

The long-term stability of the temperature maintained in the oil-bath can be estimated from measurement history of the standard resistors, where temperature is registered for every standard during resistance measurements. The measurement results for nine platinum resistance thermometers are given in Figure 7. A drift less than 0.02°C is observed for all PRTs. The total uncertainty in temperature measurements is 0.05°C for the one-year calibration interval.

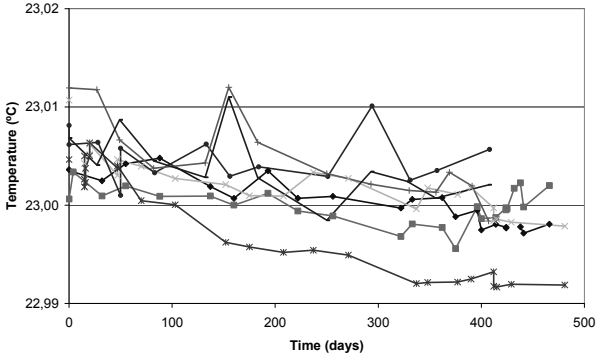


Figure 7. *Results for nine platinum resistance thermometers measured from inside of the standard resistors*

The developed measurement setup and an air-thermostat allow us to determine the temperature coefficients of the resistors which are necessary to take into account effects of temperature on resistance values.

The maintenance of resistance unit and dissemination of traceable measurements require thorough characterization of the measurement bridge. We applied

complements check and comparison methods with Hamon-type resistors [8, 9] to confirm the performance of our dcc bridge. The results of complement checks for 1:1 ratios with nominal values from 1 Ω to 10 kΩ are shown in Table 2. Two measurement series are conducted by exchanging resistors R_X and R_S. The given measurement results indicate that 1:1 resistance ratios of our bridge comply with accuracy specification of the bridge, which is 1 part in 10^7.

Nominal value, Ω	Current, mA	Relative difference in ratio, $\cdot 10^{-6}$	Standard deviation, $\cdot 10^{-6}$
1	100	-0.012	0.0007
10	30	-0.002	0.0003
100	5	-0.015	0.0011
1 k	3	-0.024	0.0020
10 k	0.3	-0.067	0.0125

Table 2. *Complements checks results for dcc bridge*

The measurements of the 10:1 ratios were performed by using Hamon-type resistors [10, 11]; see Table 3. The ratios 100 Ω: 10 Ω and 10 Ω: 1 Ω were checked with the model SR1030 transfer standard. Also, two-decade steps of the bridge in the range from 100 Ω to 1 Ω were measured with the same device. The difference between 100:1 (2^{nd} row in Table 3) and two 10:1 (3^{rd} and 4^{th} rows in Table 3) ratio measurement steps was 2 parts in 10^8.

Four (two for each range) home-made Hamon-type resistors were constructed to check the ratios 10 kΩ: 1 kΩ and 1 kΩ: 100 Ω of the bridge. The measurements of two Hamon-type resistors and 1 kΩ: 100 Ω bridge ratios were within 5 parts in 10^8.

Ratio, R_X: R_S	Relative difference in ratio, $\cdot 10^{-6}$	Relative expanded uncertainty, $\cdot 10^{-6}$
100 Ω: 1Ω	0.20	0.20
10 Ω: 1 Ω	0.09	0.20
100 Ω: 10 Ω	0.09	0.20
1 kΩ: 100 Ω	-0.05	0.20
10 kΩ: 1 kΩ	-0.61	0.30

Table 3. *Comparison results with Hamon-type resistors*

The difference between two Hamon-type resistors in the range 10 kΩ: 1 kΩ was 16 parts in 10^8, resulting in higher uncertainty of the measurement. The difference of 61 parts in 10^8, observed between the bridge and Hamon-type resistor measurement results, should be further investigated.

The typical uncertainty budget in maintenance of resistance standards includes the following components due to:

- calibration;
- transportation;
- long-term drift;
- performance of the bridge;
- connections;
- temperature and power corrections.

The expanded uncertainty ($k = 2$) is 1 μΩ/Ω for the range from 1 Ω to 10 kΩ.

Conclusions

The reliable representations of the voltage and resistance units were established at a secondary level of standards in Estonia.

The standards are traceable to Josephson voltage and quantized Hall effect resistance standards.

The automated measurement systems were developed and thoroughly characterized to ensure, that the uncertainties in maintenance of standards are 1 μV/V for 10 V and 1 μΩ/Ω for the range from 1 Ω to 10 kΩ.

Acknowledgments

The authors would like to thank Prof. T. Rang, who is the supervisor of A. Pokatilov's PhD studies.

A. Pokatilov is thankful to the Estonian Information Technology Foundation and Tiger University for the support of this research.

References

[1] A. Pokatilov, T. Kübarsepp, "Establishment of National Standard of Voltage Unit in Estonia", in *Proc. of 10th Biennial Baltic Electronics Conference BEC2006*, 2006, pp.157-160.

[2] A. Pokatilov, T. Kübarsepp, "Development of Automated Measurement Setup for Standard Resistors", in *Proc. of 10th Biennial Baltic Electronics Conference BEC2006*, 2006, pp.161-162.

[3] T. J. Witt, "Maintenance and dissemination of voltage standards by Zener-diode-based instruments," *Proc. IEE Sci. Meas. Technol.*, vol. 149, No. 6, 2002, pp. 305-312.

[4] C. A. Hamilton, L. W. Tarr, "Projecting zener dc reference performance between calibrations," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 2, pp. 454-456, Apr. 2003.

[5] O. Power, J. E. Walsh, "In-service characterization of electronic voltage standards," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 2, pp 559-562, Apr. 2005.

[6] R. J. Freund, W. J. Wilson, *Statistical Methods*, Academic Press, 1993.

[7] D. W. Braudaway, "Precision Resistors: A Review of the Techniques of Measurement, Advantages, Disadvantages, and Results", *IEEE Trans. Instrum. Meas.*, vol. 48, no. 5, pp. 884-888, October, 1999.

[8] G. F. Strouse, K. D. Hill, "Performance Assessment of Resistance Ratio Bridges Used for the Calibration of SPRTs," *Proc. of the 8th Int. Symp., "Temperature: Its Measurement and Control in Science and Industry,"* Chicago (2002), 327-332.

[9] S. Rudtsch, G. Ramm, D. Heyer, R. Vollmert, "Comparison of test and calibration methods for resistance ratio bridges," *TEMPMEKO 2004, 9th Int. Symp. on Temperature and Thermal Measurements in Industry and Science: Proc*. Vol. 2 (2005), 773-780.

[10] B. V. Hamon, "A 1-100 Ω build-up" resistor for the calibration of standard resistors", *J. Sci. Instrum.*, vol. 31, pp.450-453, 1954.

[11] J. C. Riley, "The accuracy of series and parallel connections of four-terminal resistors", *IEEE Trans. Instrum. Meas.*, vol. 1M-16, no. 3, pp. 258-268, September, 1967.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Set up and characterization of reference measuring systems for on-site live verification of HV instrument transformers

G. Crotti[a], A. Sardi[a], N. Kuljaca[b], P. Mazza[b], G. de Donà[c]

[a] Istituto Nazionale di Ricerca Metrologica - Strada delle Cacce 91, Turin (Italy)

[b] CESI RICERCA S.p.A. - via Rubattino 54, Milan (Italy)

[c] Terna S.p.A - Strada del Drosso 75, Turin (Italy)

ABSTRACT: Prototype reference measuring systems (RMS) for the on-site verification of high voltage instrument transformers for energy metering have been set up. Their use can make possible the identification, without service removal, of the current and voltage transformers which no longer comply with the accuracy requirements. Specific measurement procedures have been experimented in the characterization of the RMSs, which include as reference transducers a Rogowski coil and a resistive-capacitive divider for the current and voltage measurement respectively. The uncertainty associated with the on-site use of the RMSs is evaluated considering the operating conditions which are likely to occur. The results obtained in the first tests carried out show the feasibility of a live, on-site accuracy check of the instrument transformers.

Introduction

Calibration of high voltage (HV) instrument transformers for energy metering is usually performed only before their installation in service. Since their ageing can lead to a deterioration of their measurement accuracy, with a consequent deep economical impact [1, 2], the set-up of systems and procedures allowing the periodical check of current and voltage transformers without service removal and modification of the grid configuration is particularly interesting. The on-site verification of instrument transformers requires the availability of reference measuring systems (RMSs), which allow the determination of the ratio errors and phase displacements with an uncertainty significantly lower than the limit values specified by the relevant standards for the accuracy class of the transformers under investigation [3, 4].

To this end, suitable RMSs have been developed, whose construction characteristics, together with specific live-line working procedures, make it possible to check the HV substation instrument transformers without service outages.

The RMSs are constituted by a reference transducer (resistive-capacitive voltage divider/split-core Rogowski coil), remote units for acquisition and conversion of data coming from both the reference transducer and the transformer under test, optical links, a local unit for data recording and control and a dedicated software tool for the processing and comparison of the data coming from the RMS and the instrument transformer under test.

After a brief description of the RMS main features, the procedures adopted for their characterization and the results obtained are discussed. Taking into account the actual operating conditions, attention is focused on the determination of the RMS ratio errors and phase displacements by comparison with standard voltage and current transformers and on the evaluation of the temperature and proximity effects, with reference to the measurement in environmental situations likely to occur in the on-site operation. On the basis of the characterization outcomes, a first evaluation of the uncertainty associated with the measurement values of the instrument transformer ratio errors and phase displacements given by the RMSs when operating on-site is performed.

This work has been financed by the Italian Ministry of Economic Development with the Research Fund for the Italian Electrical System under the Contract Agreement established with the Ministry Decree of March 23, 2006.

RMS main features

Figures 1 and 2 show the RMSs configuration for the on-site check of current and voltage transformers respectively. The outputs of the RMS transducers (Rogowski coil and resistive-capacitive divider) are optically converted remotely (remote data module) and then transmitted to a local unit (local control module)

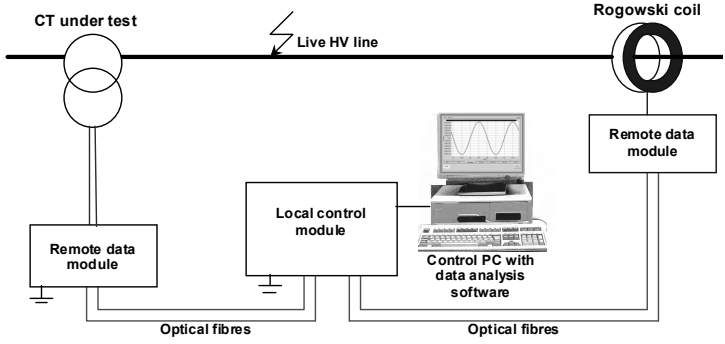


Figure 1. *Reference measuring system for the on-site check of current transformers*

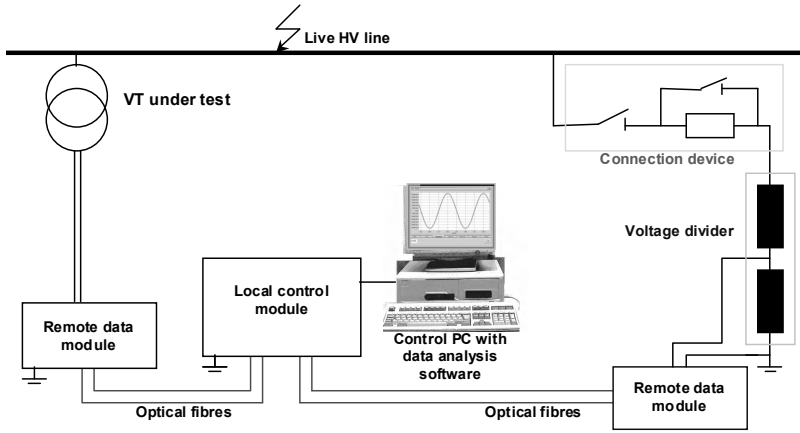


Figure 2. *Reference measuring system for the on-site check of voltage transformers*

where the comparison is performed with the simultaneously acquired outputs of the current (CT) and/or voltage (VT) transformers under test. A brief description of the components of the systems is given in the following, together with some information about the main problems encountered in live-line voltage connection.

RMS for the CT check

The current is measured by a Rogowski coil, which is an air-core mutual inductor commonly used for the measurement of high transient and alternating currents, since it is an amplitude linear device and does not suffer from saturation. The high-precision split-core Rogowski coil specifically realized [5, 6] enables the measurement of currents from 40 A to 4000 A.

For live measurements in switchgear stations, the bare coil is provided with shielding, fixations and is therefore enclosed in an external housing assuring mechanical stability; a toroidal electrode ensures the absence of corona discharges up to the line maximum rated voltage (400 kV).

The coil can be easily opened, with a limited gap between the two parts, and then safely clamped on a single conductor or on a twin bundle, according to solutions commonly used in Italian grid substations. To improve the measurement accuracy, guiding bolts in and outside the Rogowski coil ensure the correct positioning of the coil parts at the splitting areas. Live working “bare-hand” techniques are used for the installation of the Rogowski coil; Figure 3 shows its installation on a 400 kV line in the Rondissone HV substation.

Figure 3. *Installation of the Rogowski coil*

The installation is performed by a man bonded to the live conductor, on a tower insulated from earth and provided with a crane for coil lifting at conductor height.

Filtering and integration of the voltage induced across the coil, which is proportional to the time derivative of the current, is performed by the remote unit. Four attenuation levels can be selected, corresponding to full scale currents of 400 A, 800 A, 1600 A and 3200 A (plus 20% overrange). The remote unit also performs 16 bit Analog-to-Digital (A/D) conversion with 32 kilosamples/s sampling frequency. Data are then transmitted to the local control module, which is at ground potential, through a fiber optic cable by means of a light emitting diode (LED). In a similar way, the CT secondary current is synchronously acquired and converted through a second remote unit, transmitted to the local unit and then to the personal computer.

The local control system converts back the optical signals and transmits the digital data to be analyzed to the personal computer, through a duplex fiber; in addition, it provides the power, generated by an infrared semiconductor laser, to the remote modules.

Specific data analysis software is used for the calculation of the ratio errors and phase displacements of the transducers under test, in accordance with the IEC 60044 standards. The software outputs are the indications of the peak and root mean square values of the primary current, respectively given by the measurement chain including the Rogowski coil and by the one including the CT, and the phase displacement of the behavior measured by the CT with respect to that given by the RMS. The measurements can be performed in the time domain, in accordance with [3, 4], if the harmonic distortion of the signals is negligible. The influence of random noise can be significantly reduced through the application of various digital

filters and time domain averaging methods. In the case of harmonic distortion, the evaluation can be performed in the frequency domain and the ratio errors and phase displacements are obtained for each harmonic component of the signal.

The software is validated through the application of various simulated signals, with different harmonic content and random noise level.

RMS for the VT check

Voltage measurements are performed by a two stage resistive/capacitive divider. Only one stage is used to measure voltages up to 220 kV, whereas two stages are needed for the 400 kV measurements (Figure 4).

Precision low voltage dividers are used for the voltage ratio adjustment to the remote unit input at different rated voltages.

Special methodology and devices for the live-line voltage connection have been developed and experimented for the connection of the voltage divider to the HV network. A solution is adopted, which involves the use of a disconnector in series to a resistor with a suitable resistive value, whose aim is to limit the current during the divider connection phase. A remote controlled integrated rotary disconnector can then short-circuit the current limiting resistor during measurements.

As in the case of the RMS for the CT check, A/D conversion and optical transmission are performed by a remote unit, connected to the specific input of the same local unit previously described that simultaneously controls the acquisition of the data coming from the VT under check through its associated remote unit. As for the CT check, the evaluation of the ratio errors and phase displacements of the VT is performed through a software specifically developed.

Figure 4. *Reference voltage divider (two stage configuration)*

RMS characterization

A characterization of the RMSs was performed, taking into account the different operating conditions associated with the on-site operation, with particular reference to the influence of temperature and the proximity effect.

Determination of the RMS ratio errors and phase displacements

First, the ratio errors and phase displacements associated with the use of the RMSs were determined under laboratory controlled environmental conditions through the use of standard current (CT_N) and voltage (VT_N) transformers, traceable to the Istituto Nazionale di Ricerca Metrologica (I.N.RI.M.) national standards. A simplified scheme of the calibration circuits is shown in Figures 5a and 5b for the RMSs including the Rogowski coil and the resistive–capacitive divider respectively.

Taking into account that the current values to be experienced on-site can range from a few amperes up to the rated current of the CT under test and that the Rogowski coil will be used for the check of transformers of different sizes, the behavior of the Rogowski coil and its associated conversion, transmission and acquisition system (local unit, optical fiber connection, remote unit and software) was investigated from 5% to 120% of each considered rated primary current I_r (400 A, 800 A, 1600 A and 3200 A).

As for the voltage range, possible variations within 20% of the rated voltage were assumed. The ratio errors and phase displacements of the voltage divider and the associated measurement system were then assessed for voltages ranging from 80% to 120% of the considered rated VT primary voltages (130/√3 kV, 150/√3 kV, 220/√3 kV and 380/√3).

The ratio errors and phase displacements determined in the calibration phase can be used as corrections to the indications obtained by the RMSs during their on-site use for transformer checks, provided that their stability has been assessed and that, for each considered primary current or voltage, the calibration of the RMS has been performed with a standard having transformation ratios equal to those of the transformers to be checked. In this way, the errors introduced by the remote and local units associated with the transformer under check can be assumed equal to those determined in the calibration phase.

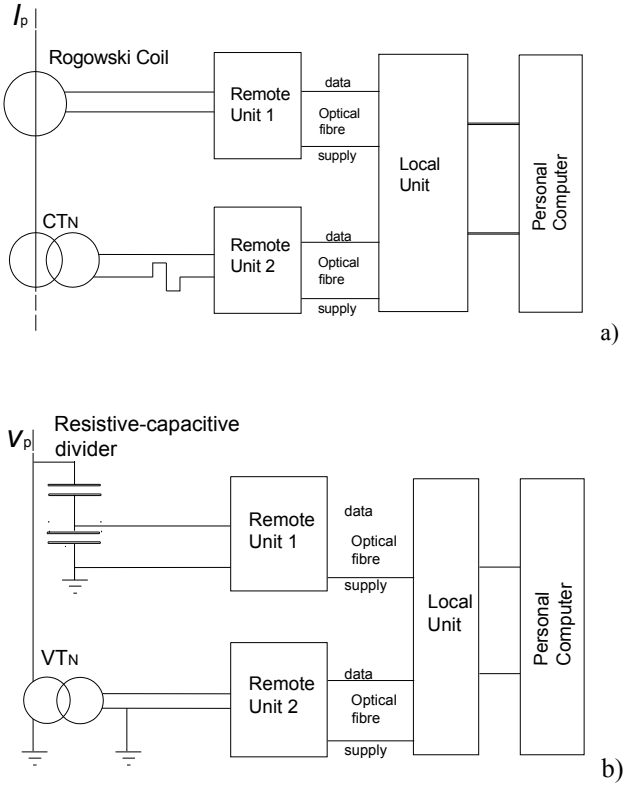


Figure 5. *Determination of the ratio errors and phase displacements of the RMSs for the CT (a) and VT (b) check by comparison with standard transformers*

Temperature effects

The temperature effect on the Rogowski coil and its associated remote unit was profoundly investigated. The coil and the remote unit were placed in a thermal cell; the ratio errors and phase displacements of a high stability current transformer (CT_S), located outside the cell and connected to the local control unit through the remote unit 2, were measured with the Rogowski coil and remote unit 1 at thermal balance with the temperature of the cell (Figure 6).

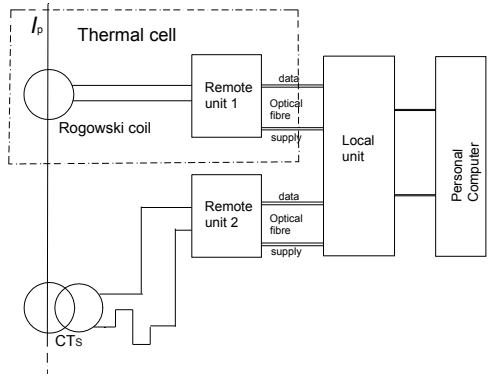


Figure 6. *Evaluation of the temperature effect on the Rogowski coil and associated remote unit 1*

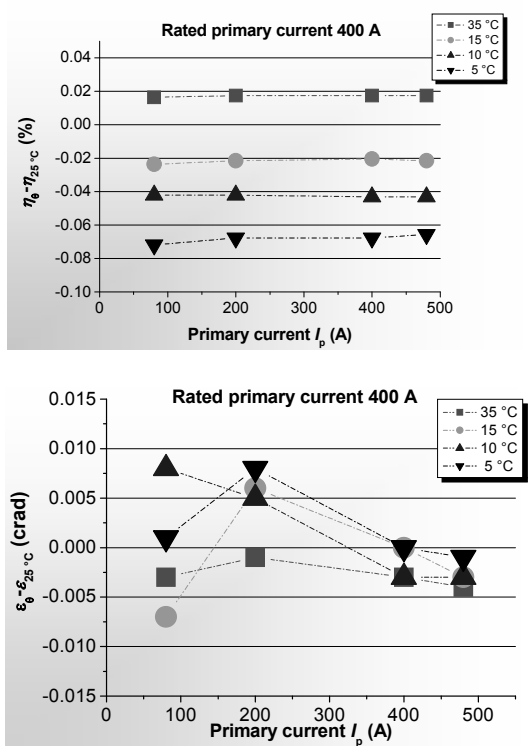


Figure 7. *Variation of the ratio error (η_θ) and phase displacement (ε_θ), with respect to the values at 25°C ($\eta_{25°C}$, $\varepsilon_{25°C}$), of a high stability current transformer measured by the RMS at four different temperature values*

The cell temperature was varied from 5°C to 35°C and monitored by means of a calibrated thermocouple. Figure 7 shows the variation of the ratio errors and phase displacements with respect to the value measured at 25°C as a function of the primary current (rated primary current 400 A), for four different temperature values.

Taking into account the results obtained for all the rated primary currents, and considering a temperature range from 10°C to 35°C, variations of the measured ratio errors with respect to those obtained at a temperature of (25 ±1.5)°C were within 0.05%, while the corresponding variations of the phase displacement were found to be within ± 0.03 crad.

In a similar way, the influence of the temperature on the reference divider and its associated remote unit was investigated. The divider, both in the single stage and two-stage configuration, and the remote unit were placed in a thermal cell; the ratio errors and phase displacements of a VT located at ambient temperature, characterized by a high stability, were measured by the RMS with the reference divider and its remote unit at thermal balance with temperatures ranging from 5°C to 40°C. Measured data showed variations in the considered temperature ranges up to some parts in 10^3, which are higher than those obtained for the Rogowski coil and associated remote unit. Further investigations are in progress to identify which components are responsible for these temperature variations.

Proximity effects

The Rogowski coil behavior can be affected by the proximity effect due to external current carrying conductors. Variations of the ratio errors and phase displacements obtained in a single-phase circuit were then investigated, increasing the distance d_1 of the return conductors from 0.6 m to 1.75 m. Two positions of the remote unit 1 with respect to the Rogowski coil were considered: a) reference position (unit placed as on-site, on the external edge of the coil, inside the corona ring); b) unit closer to the internal conductors (distance unit 1-conductors less than 0.3 m). Figure 8 shows the variations of the ratio errors with respect to those obtained in the condition assumed as reference (d_1=0.95 m and remote unit 1 in position a), measured in three different test configurations. As can be seen, the decrease of the distance to 0.61 m led to negligible variations, if the position of the remote unit is unchanged. However, variations up to 0.05% with respect to the reference conditions are found if unit 1 is placed closer to the internal conductors. Taking into account the actual measurement conditions, and on the basis of the test performed, the limit variations of the ratio errors and phase displacements have been estimated to be within 0.05% for the ratio error and 0.05 crad for the phase displacement.

As to the evaluation of the Rogowski coil position sensitivity, it must be underlined that, because of the manufacturing characteristics of the coil itself and the specific measurement conditions, the coil will be always perpendicular to the

internal conductor, with a fixed position of the split-core coil junction with respect to the internal conductor.

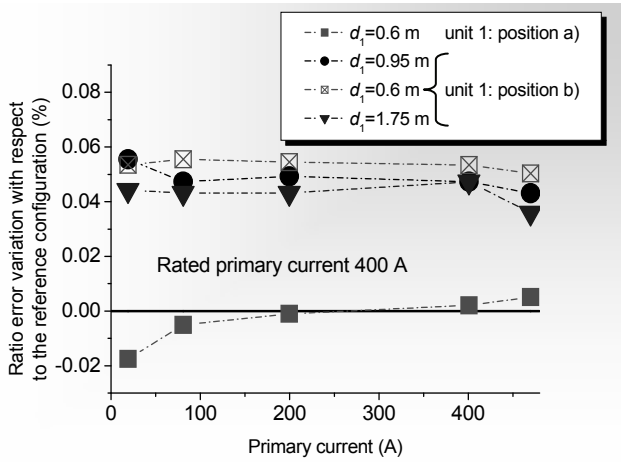


Figure 8. *Variations of the ratio error with respect to the reference configuration (d_1= 0.95 m, unit 1 in position a)*

The proximity effect due to the presence of ground potential structures as a function of the distance from the reference voltage divider has been also investigated. Measurements of the ratio errors and phase displacements of a voltage transformer, characterised by a high level of stability, were recorded varying the distance between the divider and a metallic structure at ground potential from 2.5 m to more than 10 m. With respect to the values measured with the divider 10 m distant from the ground structure, the variations of ratio error and phase displacement were found to be within 0.07% and 0.05 crad with the metallic structure 2.5 m far from the divider (380 kV configuration) and within 0.04% and 0.01 crad when the distance was increased to 5 m.

Uncertainty budget

On the basis of the data obtained up to now, a first uncertainty budget has been worked out.

The uncertainty contributions considered are due to:

a) determination of the RMS ratio errors and phase displacements in the calibration phase;

b) long term stability of the RMS;

c) proximity effect due to the presence of current carrying external conductors (for the Rogowski coil and associated remote unit) or structures at ground potential (for the voltage divider and associated remote unit);

d) temperature effect, if the RMS is used in conditions different from those occurring during the calibration (23°C±1°C);

e) variations of the ratio error and phase displacement of the instrument transformer under check due to the modification of the actual burden because of the connection to the remote unit (Figure 1);

f) in the case of the RMS including the Rogowski coil, a further component has been included, to take into consideration influence factors due to the actual measurement conditions, such as the number of internal conductors (one to three), and the presence of the corona ring externally placed with respect to the coil.

Table 1 shows the uncertainty budged associated with the ratio error and phase displacement of a 400 A/5 A current transformer (CT_X) obtained by the RMS in the case of primary current equal to 400 A. The standard uncertainty components have been estimated from the data obtained in the characterization phase, under the hypothesis of measurements performed in the range from 10°C to 35°C. By assuming uncorrelated input quantities, expanded uncertainties (confidence level 95%) of 0.1% and 0.07 crad are obtained for the ratio error and phase displacement respectively.

Comparable RMS uncertainties are obtained for the other transformation ratios, always considering operating temperatures not lower than 10°C and the remote and local unit 2 located at temperatures higher than 20°C.

Uncertainty contribution	Standard uncertainty	
	Ratio error (%)	Phase displacement (crad)
Indication of CT_x error	0.0058	0.012
RMS error	0.012	0.014
Long term stability	0.012	0.006
Temperature effect	0.012	0.006
Proximity of external conductors	0.029	0.003
Burden	0.029	0.029
Measurement conditions	0.012	0.003
Combined uncertainty	0.048	0.035

Table 1. *RMS for the CT check (transformation ratio 400 A/5 A, I_p=400 A): ratio error and phase displacement uncertainty budget*

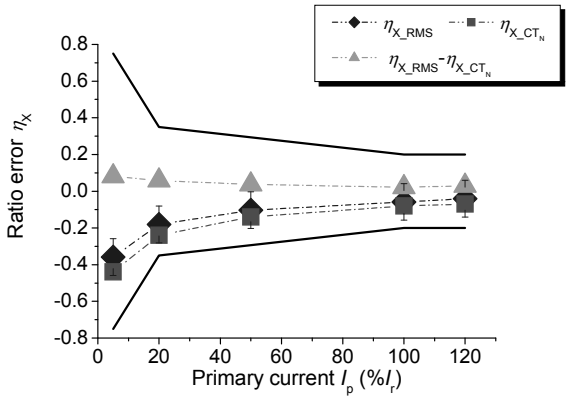
Under the assumption of measurements performed with temperatures ranging from 15°C to 35°C, the estimated uncertainties associated with the use of the RMS for the measurement of CT ratio errors and phase displacements range from 0.08% to 0.1% and from 0.06 crad to 0.07 crad, for 100% and 120% of the CT rated primary current.

Uncertainties ranging from 0.1% to 0.13% and from 0.14 crad to 0.21 crad were obtained for the RMS measurements of VT ratio errors and phase displacements, when measurements are performed at an ambient temperature of (17±3)°C.

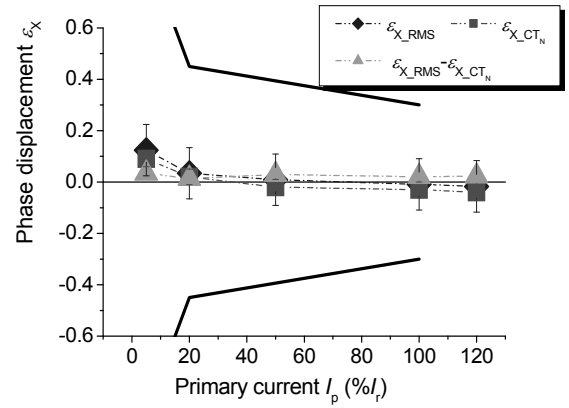
Check of a 0.2 class CT

As a final test, the ratio errors and phase displacements of a 0.2 class current transformer (CT_x) measured by the RMS have been compared with those evaluated by using the I.N.RI.M. current ratio standard (CT_N). Four commercial current transformers with transformation ratios 400 A/5 A, 800 A/5 A, 1600 A/5 and 3200 A/5 A were checked.

Figure 9 shows the ratio errors and phase displacements obtained for the 1600 A/5 A ratio when using the RMS and the I.N.RI.M. standard respectively. As can be seen, the deviation between the measured values is always less than 0.1% for the ratio error and less than 0.05 crad for the phase displacement, considering primary currents from 5% to 120% of the rated one; the limits for errors and phase displacements associated with the 0.2 class transformers [3, 4], are evidenced by the continuous lines on the same graphs.



a)



b)

Figure 9. *Ratio error η (a) and phase displacement ε (b) of a commercial transformer (CT_x) measured by the RMS and the I.N.RI.M standard (CT_N)*

Compatibility between the measurement values given by the RMS (m_{RMS}) and the I.N.RI.M standard (m_{CTN}) is evaluated according to the relationship:

$$m_{RMS} - m_{CT_N} \le U(m_{RMS} - m_{CT_N}) = \sqrt{U^2(m_{RMS}) + U^2(m_{CT_N})} \qquad (1)$$

where $U(m_{\mathrm{RMS}})$ and $\mathrm{U}(m_{\mathrm{CTN}})$ are the expanded uncertainties (confidence level 95%) associated with m_{RMS} and m_{CTN}. The correlation due to the use of the I.N.RI.M. standard in the determination of the RMS ratio errors and phase displacements is not considered, since the related uncertainty contributions (20 ppm and 0.005 crad) can

be disregarded. Relationship (1) was always found valid for all the transformers and measurement points considered.

Conclusions

The results obtained in the first tests carried out on prototypes of the RMSs both in laboratory and on-site show the feasibility of a live, accuracy check of instrument transformers.

Laboratory and field tests are in progress in order to further investigate the transducer behavior, to better identify critical points with the aim of improving the uncertainty evaluations and extending operability ranges as new data become available.

References

[1] E. Anderson, J. Karolak, J. Wroblewski, A. Hyrczak, A. Ratajczak, R. Zajac "Metrological Properties of High Voltage Instrument Transformers after Many Years' Service", CIGRE session 2004, Paper A3-113.

[2] D. F. Peelo, J. H. Sawada, B. R. Sunga, R. P. P. Smeets, J. G. Krone, L. Van Der Sluis "Current Interruption with High Voltage Air-break Disconnectors", CIGRE session 2004, Paper A3-301.

[3] IEC 60044-1 Standard "Instrument transformers – Part 1 – Current transformers", 2003.

[4] IEC 60044-2 Standard "Instrument transformers – Part 2 – Inductive voltage transformers", 2003.

[5] A. Andersson, D. Destefan, J.D. Ramboz, S. Weiss, J.M. DeHaan, "Unique EHV Current Probe for Calibration and Monitoring", Transmission and Distribution Conference and Exposition, Vol. 1, pp. 379-384, 2001.

[6] J. D. Ramboz, "Machinable Rogowski Coil, Design And Calibration", IEEE Trans. on Instrument. and Measur., Vol 45, No. 2, pp. 511-515, 1996.

[7] P. Mazza, N. Kuljaca, G. Crotti, A. Sardi, G. De Donà, U. Brand, M. Giraud, A. Andersson S. Weiss, "On-site live verification of HV instrument transformer accuracy", 2006 CIGRE' Session, Paper A3-204, Paris (France), August 2006.

[8] P. Mazza, N. Kuljaca, G. Crotti, A. Sardi, G. De Donà, U. Brand, M. Giraud, A. Andersson, S. Weiss: "Live Installation of HV Equipment for Instrument Accuracy Check. 8th Live working conference ICOLIM 2006", Session S3.10, Prague (Czech Republic), June 2006.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Testing/calibrating electrical measuring instruments under non-sinusoidal conditions at the national institute of metrology, Romania

Ioana-Izabela Odor, Dorin Flămînzeanu, Mihail Rizea, Cornel Băltăteanu

National Institute of Metrology, Romania
ioana.odor@inm.ro, doru.flamanzeanu@inm.ro

ABSTRACT: The integration of the Romanian electricity market into the regional and, subsequently, into the EU single market needs to create and implement fair and independent regulations to protect the interests of consumers and investors.
In recent years, the Electrical Measurement Laboratory of the NMI from Romania, has developed activities concerning testing and calibrating watt-hour meters and power analyzers, in accordance with the norms, prepared by IEC and CENELEC, concerning the electrical power quality.
The paper presents the development and achievements of the laboratory in the area of testing and calibrating electrical measuring instruments under non-sinusoidal conditions.

Introduction

The subject of electrical power and energy quality is under continuous development by working groups of the International Electrotechnical Commission (IEC), the International Conference of Large Electrical Networks (CIGRE) and the Institute of Electrical and Electronics Engineers (IEEE).

Power quality is becoming an important part of the international power conferences and is a specific topic of at least two international events held biannually, i.e. the International Conference on Harmonics and Quality of Power (ICHQP) and the Power Quality Assessment (PQA).

The monitoring of the quality of the electrical power and energy is of interest to manufacturers, customers and suppliers of this industrial product. There are a lot of books and papers written about this issue, i.e. [1], [2], [3].

For a long period of time, the quality of the electrical power was generally used to express the quality of the voltage (wave form, frequency, harmonic distorsions). The main deviations from the ideal waveforms being:

– disturbances (voltage dips, brief interruptions, brief voltage increases, transients, voltage notches);
– unbalance three phase voltage system;
– waveform distortion, generally discussed in terms of harmonics;
– voltage fluctuations (step voltage changes, cyclic or random voltage changes);
– flicker.

Analyzing the complexity of the power quality issue, IEC concluded to develop a series of publications around the following objectives:

– description and characterisation of the phenomena;
– major sources of power quality problems and impact on other equipment and the power system;
– mathematical description of the phenomena using indices or statistical analysis to provide a quantitative assessment of its significance;
– measurement techniques and guidelines; emission limits for different types and classes of equipment;
immunity or tolerance level of different types of equipment;
– testing methods and procedures for compliance with the limits;
– mitigation guidelines.

International standards

This series, called Electromagnetic Compatibility (EMC) Standards, is published in separate parts according to the following structures:

1 General (IEC 61000 –1-x);
2 Environment (IEC 61000 –2-x);
3 Limits (IEC 61000 –3-x);
4 Testing and Measurement Techniques
(IEC 61000 –4-x);
5 Installation and Mitigation Guidelines
(IEC 61000 –5-x);
6 Generic Standards (IEC 61000 –6-x);

The European Union's EMC Directive (89/336/EEC) addresses the need for electrical and electronic products connected to the a.c. public mains to be tested to emission and immunity standards.

The most important other organisations such as CENELEC, IEEE, ANSI developed their own standards, that are usually very much application-based or specific to a certain environment. Thus, in Europe, the CENELEC published the EN.

Romanian standards and regulations

In Romania, the quality of electrical power and energy supplied by public distribution system is the subject of following publications:

- The National Frame Contract for delivery of electrical energy, developed by the Romanian Electricity and Heat Regulatory Authority- ANRE;
- Performance Standard for the Service of delivering the electrical energy to regulated tariffs, elaborated by The Romanian Electricity and Heat Regulatory Authority- ANRE;
- SR EN 50160:1998 Voltage characteristics of electricity supplied by public distribution system.

The ANRE was established and functions as per the Government Ordinance 29/1998 (GEO 29/1998), approved by the Law 99/2000. ANRE is an autonomous public institution of national interest whose mission is to create and implement fair and independent regulations to ensure an efficient, transparent and stable functioning of the electricity and heat sector and market while protecting the interests of consumers and investors.

In accordance with these Romanian regulations, the quality of the electric power and energy delivered by the supplier is statistically checked by measuring the following quantities:

- voltage (the amplitude and symmetry of the three phase system);
- distortions (distorsion factor TDH);
- frequency of the voltage wave.

The National Frame Contract stipulates that the nonsymmetrical three-phase voltages and the distorsions will be measured in the case of existing the possibility of monitoring these quantities.

In the last 10 years, the National Institute of Metrology from Romania was frequently asked, by different companies, to calibrate watt-hour meters and power analyzers under non-sinusoidal conditions.

Testing/calibrating electrical measuring instruments under non-sinusoidal conditions

Tests/calibration performed till 2006

The supplier of electrical energy, as well as the customer, can monitor the amplitude of the voltage, the distorsions and the frequency, using more and more sophisticated devices or measuring systems. These apparatus are designed and realised to monitor the quality of the electrical power and energy supplied by public distribution systems.

As shown in [4], the Electrical Quantities Laboratory from NMI Romania, calibrated power analyzers using the block diagram presented in Figure 1.

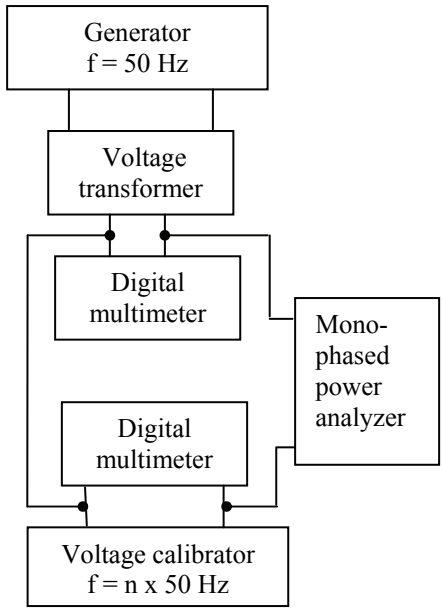


Figure 1. *Block diagram used in NMI Romania, to check the accuracy of a power analizer for voltage waveforms*

In 2003, the Laboratory from the National Institute of Metrology, Romania, was asked to make the metrological qualification of some standards, all of the same type, used for the dissemination of the electrical power unit, to the meters placed in the different points of transactions between electricity producers, electricity buyers, eligible consumers and any other party participating in the Romanian Electricity Market to enable efficient operation of such a market.

Though the reference standard of the Electrical Quantities Laboratory was traceable at SI, by calibration at PTB, under sinusoidal and non-sinusoidal conditions, the task of calibrating electrical measuring three-phase instruments under non-sinusoidal conditions was not possible to achieve, because of the missing of power sources capable of generating non-sinusoidal voltages or/and currents.

To accomplish the metrological requirements for the standards and equipment usually used to calibrate power and energy standards, as presented [5] and due to the customer's request, the calibrations were performed abroad, in the manufacturer's laboratory, using modern testing equipment (bench, three-phase power source with additional function for injection of harmonics, standard meter with traceability to SI, specialized software and PC).

The results of the calibrations performed under sinusoidal and non-sinusoidal conditions, in accordance with the Romanian and international regulations, are synthetically presented in Table 1.

Connection	Voltage [V]		Current [A]		Maximum values obtained for the errors [%]
	U_1	U_k	I_1	I_k	cosφ=1
Y_{RST0}	230	-	6	I_3=0.6	0.02
		-	80	I_3=8	0.02
		U_5=23	3	I_5=0.9	0.02
			50	I_5=15	0.02

Table 1. *Calibration made under non-sinusoidal conditions*

In 2005, The Romanian Bureau of Metrology has sustained the technical project named "The Maintenance of the National and Reference Standards from Romania" hereby the laboratory has obtained a three-phase power calibrator, traceable to SI for sinusoidal and non- sinusoidal voltages and current.

Tests/calibration performed beginning from 2006

The new and up-to-date calibrator is a three-phase power standard, supplied with voltage and current harmonic injection, and its traceability to SI is assured by calibration at the manufacturer's UKAS accredited laboratory.

From March 2006, the The National Institute of Metrology, Romania, was able to perform tests/calibrations of mono or three-phase electrical instrumentation, in

accordance with the requirements of two important international norms from the large series 61000-4-x electromagnetic compatibility (EMC):

- Part 4-7:2002 "Testing and measurement techniques - General guide on harmonics and interharmonics measurements and instrumentation, for power supply systems and equipment connected thereto", that applies to instrumentation intended for measuring spectral components in the frequency range up to 9 kHz, which are superimposed on the fundamental of the power supply systems at 50 Hz. For practical considerations, this standard distinguishes between harmonics, interharmonics and other components above the harmonic frequency range, up to 9 kHz and defines the measurement instrumentation intended for testing individual items of equipment in accordance with emission limits given in other certain standards (for example, harmonic current limits as given in IEC 61000-3-2) as well as for the measurement of harmonic currents and voltages in actual supply systems
- Part 4-30: 2003 "Testing and measurement techniques – Power quality measurement methods", that defines the methods for measurement and interpretation of results for power quality parameters in 50Hz a.c. power supply systems. Depending on the purpose of the measurement, all or a subset of the phenomena on this list may be measured. The uncertainty tests in the ranges of influence quantities in this standard determine the performance requirements.

The measurement method used to test/calibrate electrical instrumentation, under non-sinusoidal conditions, against the new power standard is the direct comparison, by which the effective values of the voltages and currents, supplied from the calibrator, are applied to the inputs of the instrument to be tested/calibrated.

In Table 2, there are presented the results obtained for the measurement point (U_{nom} = 230.0 V; U_3 = 5.0V $\geq$ 1% U_{nom}; I_{nom} = 10.0 A, I_3=1.0 A $\geq$ 3% I_{nom}, symmetric 4 wire connection), for a three-phase (A, B, C) power analyzers, of class I, according the international norm IEC 61000-4-7.

These results were obtained using a simple procedure in accordance with the ISO Guide 98:1995 "Guide to the expression of uncertainty in measurement (GUM)".

Measuring error [%]			IEC 61000-4-7 max.error [%]	Measuring error [%]			IEC 61000-4-7 max.error [%]
U_A	U_B	U_B		I_A	I_B	I_B	
1.2	1.0	1.3	$\pm$ 5	2.0	1.9		$\pm$ 5

Table 2. *Errors of a three-phase power analyzer calibrated under non-sinusoidal conditions against the power standard with voltage and current injection*

Conclusions

After 10 years of trying to respond to the different customers' request of testing/calibrating electrical measuring instruments under non-sinusoidal conditions, the Electrical Measurement Laboratory of the NMI from Romania is now able to perform measurements to characterize such instrumentation, in accordance with the international EMC norms, regarding the currents and voltages in the actual 50 Hz a.c. power supply systems.

References

[1] J. Arrillaga, N.R. Watson, and S. Chen. *Power System Quality Assessment*, Wiley, England, 2000

[2] J. Arrillaga, B. C. Smith, N. R. Watson. *Power System Harmonic Analysis*, Wiley, England, 1997

[3] C.Golovanov, M.Albu, coord.: *A Modern Approach to Power Systems,* Tehnica Press, Bucharest, 2001.

[4] I.Odor, A.Constantinescu, Problems of the metrological characterisation of instruments for estimating the quality of electrical energy, International Metrology Conference, Celebrating the Anniversary of the Romanian National Institute of Metrology, September 18-20, 2001, Bucharest.

[5] I.Odor, About the calibration of energy meters under nonsinusoidal conditions, 3-rd International Conference on Electrical and Power Engineering, Technical University, Iasi, 2004.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Characterization of high resistance standards in MIKES

A. Satrapinski, R. Rajala

MIKES, Tekniikantie 1, 02151, Espoo, Finland, email:
Alexandre.Satrapinski@mikes.fi, Risto.Rajala@mikes.fi

ABSTRACT: High resistance standards in the range 10 kOhm – 1 GOhm are characterized and the measurement results are presented. Three scaling methods were used for verification of the value of the 1 GOhm standards; substitution method with the use of the Hamon transfer standards, Direct Current Comparator (DCC) Resistance Bridge method with 1:10 scaling, and modified Wheatstone Bridge method with two calibrators, which were used as programmable voltage sources. The procedures, schemes and details of the measurements used in each method are presented. The key sources of systematic uncertainty are analyzed. Combined standard uncertainty of the 1 GOhm standard is estimated to be 4.3 parts in 10^6.

Introduction

MIKES established the realisation of a resistance unit based on the quantized Hall resistance (QHR) standard [1] and verified the resistance representation in the range 1 Ohm – 10 kOhm [2-4]. Growing needs in industry for accurate high resistance measurements at Tera-Ohm level and in the measurements of very small currents, requires further improvements in resistance scaling in the range 10 kOhm – 1 GOhm. Realisation of metrological triangle experiments [5] and participation in international high resistance comparison also require further progress in our knowledge of the behavior of high value resistance standards. Recently MIKES introduced and tested three measurement methods for calibration of the resistors up to 1 GOhm and took part in the Euromet EM-K2 high resistance comparison.

In this paper we present characteristics of the high value resistance standards used in MIKES, the measurement methods and the measurement results together with the estimated uncertainties.

Traceability of 10 kOhm – 1 GOhm standards

General traceability chain to the 1 GOhm is presented in Figure 1. Three Hamon [6] transfer standards, 100 kOhm-step, 1 MOhm-step and 100 MOhm-step and set of primary decade standards in the range 10 kOhm – 1 GOhm, are used for resistance scaling from 10 kOhm to 1 GOhm. Traceability is based on the MIKES quantized Hall resistance (QHR) standard. Two 100 Ohm standards are measured in terms of

the QHR with self-designed Cryogenic Current Comparator (CCC) Resistance Bridge [1]. That Bridge is used also for the measurements and dissemination of the ohm in 1 Ohm – 10 kOhm range. With additional re-adjustment our CCC Bridge can also be used for the measurements of 100 kOhm and 1 MOhm standards. High value resistance standards maintained now in MIKES include three Hamon transfer standards and a set of the reference resistance standards in the range 1 GOhm – 100 TOhm. In our customer and internal calibrations ranging from 10 kOhm to 1 GOhm the Guildline 6675 DCC Resistance Bridge is used as 1:1 and 1:10 scaling device. For improvement of our best measuring capabilities at this particular calibration range and for verification of the scaling of high value resistors, we implemented and studied alternative measurement methods and equipment. The tested instruments and procedures will also improve our measurement capabilities in resistance scaling up to 100 TOhm.

Methods and equipment

For the cross checking of the scaling accuracy up to 1 GOhm, we use three different methods; the substitution method with Hamon reference standards, the DCC Resistance Bridge method and the modified Wheatstone Bridge method.

Substitution method with Hamon standards

Reconnection of ten resistors of the same nominal value from parallel to series configuration provides accurate ratio of 1:100. Step-up process begins from the 10 kOhm reference resistor, R_{10k}, which is compared with the R^p_H =10 kOhm of ten, connected in parallel 100 kOhm resistors of Hamon standard. The 1:1 ratio measurement is performed with the electrometer, Keithley model 6517A, which is used as a relative resistance meter in all scaling steps. Series connection of the 10 kOhm resistors of Hamon standard, R^s_H is used as a reference for evaluation of the measured 1 MOhm resistor R_x in 1:1 measurements. The value of the unknown resistor is evaluated from: $R_x = R^s_H \cdot (R_{xr}/R_{refr})$, where R_{xr} and R_{refr} are the readings of electrometer, and $R^s_H = R^p_H \cdot 100 \cdot (1+d)$, d is the transfer error of Hamon standard. Such a procedure is used in all other ranges with the use of the Hamon transfer standards.

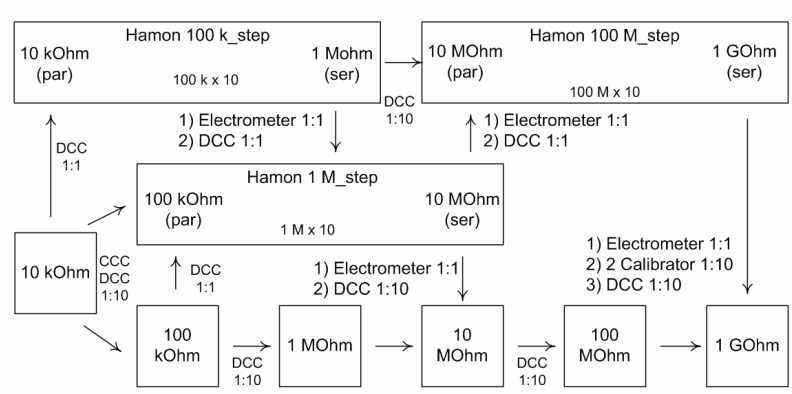


Figure 1. *Traceability chain of high resistance standards in MIKES*

DCC Bridge

Measurements with DCC Bridge are performed with 1:1 and 1:10 ratio depending on the reference used. Two terminal measurement configurations are used for the resistors higher than 100 kOhm. Uncertainty of the measurements is defined by the specified accuracy of the DCC Bridge at each measurement range.

Wheatstone Bridge

This method [7] is based on the use of a modified Wheatstone Bridge with two accurate voltage sources acting as reference of 1:10 ratio. Two calibrators, Fluke 5720 A and Fluke 5700 A, are used as programmable voltage sources to produce 10 V and 100 V. The ratio accuracy of 1:10 is based on the accuracy of the used calibrators and for these ranges is at the level of $0.5 \cdot 10^{-7}$.

Hamon transfer standards

We use Hamon standards: the 100 kOhm step standard, type SR1010, (H100k), the 1 MOhm step, type SR1050-1M, (H1M) manufactured by ESI Inc., and the 100 MOhm step, type P40114 [8] (H100M). Main characteristics of the maintained standard resistors are presented in Table 1. In table: D is the deviation from nominal, u_c – combined standards uncertainty (for the Hamon standards, D and u_c are given for series connection), TC – temperature coefficient, DR –drift rate. All values are in parts in 10^6. A detailed uncertainty budget for step up measurements is given in the text below.

Resistor Name	D	u_c, 1 σ	TC 1 / K	DR 1/day
GR10k	-0.83	0.11	0.12	0.002
G100k	38.70	0.22	2.1	0.008
G1 M	16.55	0.51	2.75	0.007
G10 M	41.21	0.68	4.57	0.021
U100 M	-89.50	2.07	-8.0	≤0.01
U1 G	95.0	4.3	≤-14	≤0.2
H100k	13.51	0.61	≤5	≤0.05
H1M	18.10	1.13	≤5	≤0.1
H100M	79.0	3.2	≤8	≤0.5

Table 1. *Main characteristics of high resistance standards*

Decade 10 MOhm - 1 GOhm resistors

For maintenance and scaling resistance in the range 100 kOhm – 1 GOhm we use decade resistors – in the range 100 kOhm - 10 MOhm, manufactured by Guildline Instr. Ltd., model 9930, and in the range 100 MOhm – 1 GOhm range, manufactured by ELIRI Inst., model P4033 and P4030, (having amber insulation). The resistors are maintained at 23.0°C, in a passive air enclosure with instability of temperature within 0.05 °C and humidity at the level of 40% (±1%). As an example, the results of long term stability of the 1 MOhm, (G1M), measured by DCC Bridge in 1:10 ratio and also measured by CCC Resistance Bridge with 1:100 ratio is presented in Figure 2. CCC Bridge measurements were performed at 0.1 Hz and with 10 V of applied voltage.

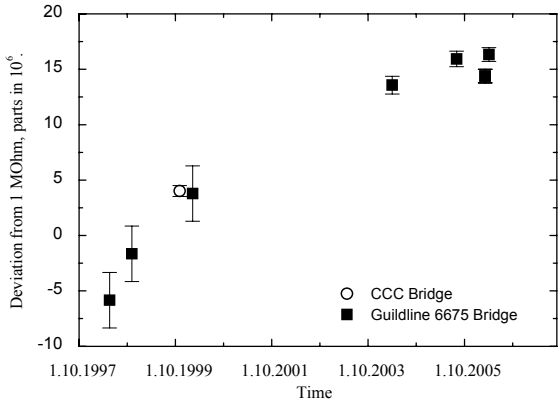


Figure 2. *Long term stability of the 1 MOhm (G1M) measured by DCC Guildline 6675 and CCC bridges*

Comparison of different methods

Substitution method

Measurement procedure for the 10 MOhm and the 1 GOhm standards consisted of direct reading of the resistance value by the electrometer, Keithley 6517A. In measurements of the 10 MOhm resistors, applied voltage is 10 V – 40 V and the used current range is 20 μA, in 1 GOhm measurements, the applied voltage is 100 V and the used current range is 200 nA. The sequence of measurements consists of applying positive and negative voltages to the unknown and reference resistors alternatively. Example of the readings of electrometer in the measurements of two 1 GOhm standards (H100M and U1G) is presented in Figure 3 and the procedure of measurement and calculation is given in the text below. As it seen from the Figure 3, there is drift in the measured signal. The linear component of that drift is of the order of $10 \cdot 10^{-6}$ / 30 min. To reduce the influence of the drift, the time for data acquisition is chosen relatively short. The readings at each polarity are taken for about 2 min and the average value is used for calculation of the relative first value of the resistor, R_{xr}^{1}. Then the measurement with the reference R_{refr}^{1} is performed and these sequences of measurements are repeated several times. To cancel the linear parts of the drift, the data from each three nearest runs ($R_{xr}^{i}/R_{refr}^{i}/R_{xr}^{i+1}$) are used for calculation of the i^{th} mean value of the ratio R^{i}_{mean}. The true value of the measured resistor, R_x, is evaluated from the measured ratio and from the value of the reference resistor, R_{ref}. The value of R_x is calculated from: $R_x = R_{ref} * \sum(R^{i}_{mean})/n$, where n is the number of R^{i}_{mean} which are used for calculation.

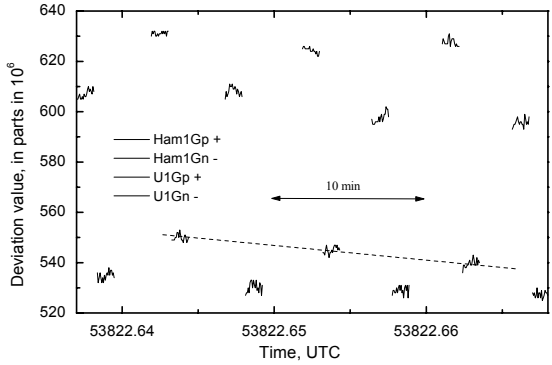


Figure 3. *Relative deviation from 1 GOhm, of two standards measured by electrometer Keithley 6517A. Influence of drifts during changes of polarity of applied voltage*

Standard deviation of the mean in the measurements at each polarity is at the level of 0.4 ppm. The main sources of systematic errors in such measurements are: non-linear drifts and temperature dependence of the electrometer input bias current, leakage currents and improper guarding in the measurement circuit.

DCC bridge measurements

In the DCC Bridge measurements, the applied voltages of 10 V for the 100 kOhm to 10 MOhm resistors, and of 100 V - 1000 V for the 1 GOhm are used. Since our three high resistance Hamon standards are used as main transfer devices for checking the 1:10 ratio of DCC Bridge, we measured their stability and temperature dependence and compared the results obtained with the DCC bridge and substitution methods. The results of time stability measurement and influence of the parallel and series connection of the 1 MOhm step Hamon transfer standard is presented in Figure 4. Parallel connection of ten 1 MOhm steps was compared with decade G100k standard, and series connection of ten 1 MOhm resistors of Hamon standard were compared with the 10 MOhm (G10M).

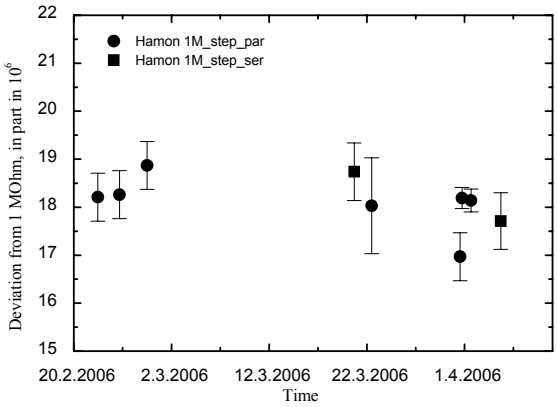


Figure 4. *Time stability and influence of reconnection from parallel to series configuration of 1 MOhm step Hamon standard*

The value of G10M, which is measured by a substitution method is agreed within combined uncertainty with the value obtained by the DCC Guildline Bridge performed in two steps of 1:10 ratio, (with the references from the G100k and the G1M). In the DCC Bridge measurements the reversal time interval is 90 sec per applied polarity of voltage. In 1:1 ratio measurements by the DCC Bridge significant error was noted depending on the position of the resistors in the measurement circuit. At 10 MOhm level this interchange error could be up to 2 parts in 10^6.

Wheatstone Bridge measurement

The measurement circuit is presented in Figure 5. The 1 GOhm resistor, R_x is measured using a bridge which consisted of two calibrators; Fluke 5700A and Fluke 5720A, and of two resistors.

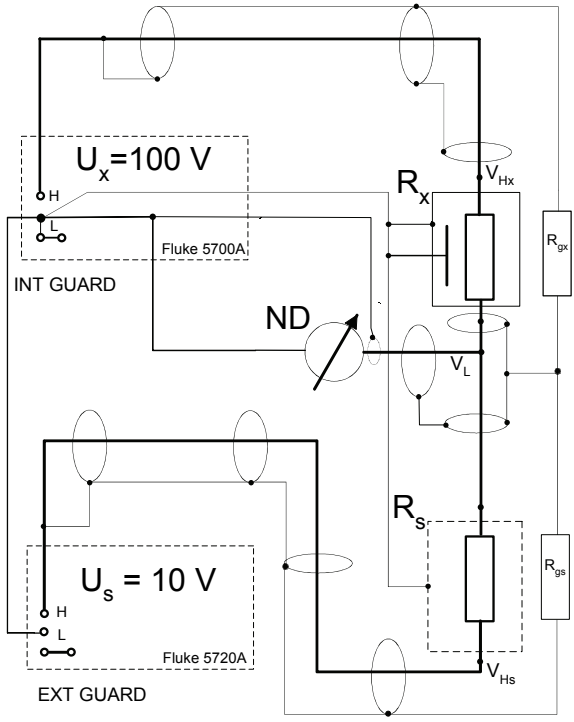


Figure 5. *Schematic diagram of measurement circuit in 1 GOhm resistor measurements by modified Wheatstone Bridge method*

The calibrators form the active arm of the bridge, and the two resistances R_x and R_s the opposite arm. The Keithley 610C analog electrometer is used as the zero detector, which is powered via an isolation transformer in order to avoid galvanic ground loops from the power line.

The voltage ratio of the calibrators is traceable to the MIKES Fluke 752A reference voltage divider, which has been characterized in Euromet EM-K8 comparison of DC voltage ratio.

The calibrator, connected to the unknown resistor R_x, was set to 100 V. The calibrator, connected to reference resistor R_s, was adjusted to obtain the best bridge balance at approximately -10 V. The ratio between the unknown and the reference resistance is obtained from the calibrator settings. The changing interval of the polarity is approximately 180 s and the consecutive average of "+" and "-" measurements are used to calculate the final result. Guarding is accomplished using the calibrator output voltages, as the guard voltage sources and the series connection of auxiliary 1 GOhm and 100 MOhm resistors, as the guarding voltage divider.

As an example, in Figure 6 the results of the 1 GOhm standard (U1G_A), measurements by three methods (Substitution, DCC and Wheatstone Bridge), at different temperatures and with 100 V applied voltage are presented.

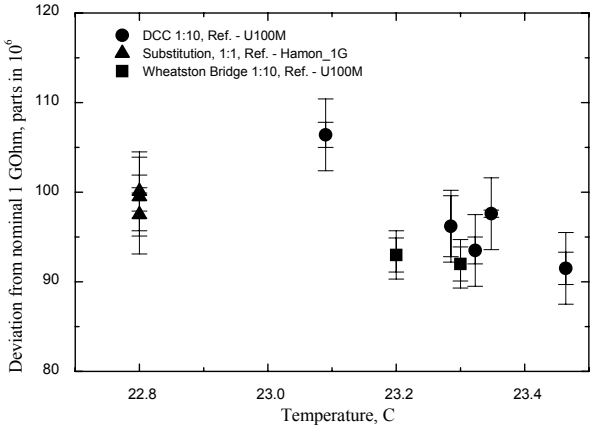


Figure 6. *1 GOhm standard (U1G_A), measured by three methods; substitution, DCC Bridge and Wheatstone Bridge*

Uncertainty estimation

In Table 2 the key components of error sources and the estimation of the uncertainties are presented for different methods in measurements of the 10 MOhm and the 1 GOhm. This table allows comparing the different contributions in uncertainty budgets and choosing the method with lower uncertainty.

In substitution method measurement of the 10 MOhm standard the biggest part in the uncertainty budget is the specified standard uncertainty of the 1:100 transfer, which is $\leq 1 \cdot 10^{6}$. In DCC bridge measurements, the maximum systematic uncertainty is due to the ratio error in the DCC bridge.

Analysis of uncertainties in the 1 GOhm measurement showed that the Wheatstone bridge method has minimum systematic uncertainty. Drifts and 1/f noise in the zero detector contribute the biggest part in the uncertainty budget and limit the combined (1 σ) uncertainty presently to 4.0 parts in 10^{6}. Type B uncertainty in the substitution method is mainly limited by specified uncertainty of the 1:100 transfer in Hamon standard. In case of the use of the P40114 standard, the combined standard uncertainty for series connection is estimated as $3.2 \cdot 10^{-6}$. In DCC bridge measurement the maximum contribution to the uncertainty budget is the standard uncertainty of 1:10 ratio.

Resistor, R_x Method	Error source	u_e, 1 σ x 10^{-6}
10 MOhm Substitution 1:1 ratio V_x = 10 V	Reference resistor (H1M_ser) temperature, drift of reference drift in detector repeatability, type A combined	1.07 0.25 0.10 0.10 1.13
10 MOhm DCC 1:10 ratio V_x = 10 V	Reference resistor, (G1M) temperature, drift of reference bridge ratio repeatability, type A combined	0.51 0.25 0.38 0.05 0.68
1 GOhm Substitution 1:1 ratio V_x = 100 V	Reference resistor (H100M_ser) temperature, drift of reference resistor settle time detector drift leakage repeatability, type A Combined	3.2 1.3 1.0 2.0 0.5 0.4 4.3
1 GOhm Wheatstone 1:10 ratio V_x = 100 V	Reference resistor (U100M) temperature, drift of reference voltage ratio resistor settle time, drift detector drift, 1/f noise leakage repeatability, type A combined	2.1 0.8 0.5 1.4 2.9 0.3 0.6 4.0
1 GOhm DCC 1:10 ratio V_x = 100 V	Reference resistor (U100M) temperature, drift of reference resistor settle time bridge ratio leakage repeatability, type A combined	2.1 0.8 1.0 4.0 0.5 1.0 5.3

Table 2. *Main error sources and estimated contributions of uncertainties in the measurements of 10 MOhm and 1 GOhm standards by different methods*

In Table 3 comparative results of summarized uncertainties assigned to different methods are presented.

The limits of the presently estimated combined 1 σ uncertainties for the used methods are; for the 10 MOhm: 0.68 μΩ/Ω – 1.13 μΩ/Ω, and for the 1 GOhm: 4.0 μΩ/Ω – 5.3 μΩ/Ω.

For evaluation of the value of 1 GOhm, we use three independent methods, each method having its own systematic uncertainty. To define the value of that standard we used the weighted mean of three results. As a Type B uncertainty of this result, it

would be possible to take the root sum square of three Type B uncertainties, but in that case we artificially increase the total uncertainty and underestimate the result. If we accept the method which has the lowest systematic uncertainty, we would limit the type B uncertainty in a narrower boundary, overestimate the value and could reduce the probability of the assigned value. In certain particular cases, with the limited number of measurements, as criteria for choosing the method with appropriate systematic uncertainty, it is possible to choose a degree of freedom. In the present estimation, for the Type B uncertainty of the mean value (obtained from the used three methods) we accept the Type B uncertainty from the method, which has maximum degrees of freedom.

R_x	Method Rref Ratio	Uncertainty			N
		Type A 1 σ	Type B 1 σ	Comb. 2 σ	
10 M	Subst. H1M 1:1	0.5	1.13	2.2	1
	DCC G1M 1:10	0.06	0.68	1.5	3
1 G	DCC U100M 1:10	1.0	5.3	11	3
	Wheat. U100M 1:10	0.5	4.0	8.1	2
	Subst. H1G 1:1	0.4	4.3	8.7	4

Table 3. *Summarized uncertainties, estimated for different measurement methods in evaluation of 10 MOhm and 1 GOhm standards. All uncertainties are in parts in 10^6, N – degree of freedom*

For the 10 MOhm, it is the DCC Bridge method, with the type B uncertainty of 0.68 μΩ/Ω and for the 1 GOhm, it is the substitution method with the Type B uncertainty of 4.3 μΩ/Ω. This uncertainty analysis is our preliminary estimation and the complete correct evaluation of all contributions will require certain time and additional test measurements. Analysis of the sources of uncertainties and the methods of their minimization will be continued.

Conclusion

Measurement equipment for realization of the link between QHR and 1 GOhm is developed in MIKES. Characterization of the high resistance standards up to

1 GOhm allowed to form the basis for reliable representation of the unit of resistance in higher resistance range. The combined standard uncertainty in evaluation of 1 GOhm standards is estimated to be 4.3 parts in 10^6. Implementation and study of different measurement methods allowed us to reduce the uncertainty of maintenance and improving the MIKES capabilities in the field of high resistance measurement.

References

[1] H. Seppä, and A. Satrapinski, "AC Resistance Bridge Based on the Cryogenic Current Comparator", IEEE Trans. Instr. and Meas., Vol. 46, No.2, pp.463–466, April 1997.

[2] A. Satrapinski, H. Seppä, B. Schumacher, P. Warnecke, F. Delahaye, W. Poirier, and F. Piquemal, "Comparison of Four QHR Systems Within One Month Using a Temperature and Pressure Stabilized 100-Ω Resistor", IEEE Transaction on Instrumentation and Measurement, Vol. 50, No.2, April 2001, pp. 238 – 241.

[3] P.O. Hertland, T.Sorsdal, G.Eklund, O.Gustavsson, M. Hammarquist, A. Satrapinski, H.D.Jensen, "Nordic comparison of the calibration of 1 Ohm and 10 kOhm resistors Euromet.EM-S18", Conference Digest, CPEM2004 27 June -2 July 2004, London, England, p. 265.

[4] A. Satrapinski, A. Rautiainen, R. Rajala, "Resistance scaling from 10 mOhm to 10 kOhm at MIKES with Cryogenic Current Comparator and Guildline Bridges", CPEM 2002, Digest, pp.56-57, Ottawa, June 2002.

[5] F. Piquemal, G. Geneves, "Argument for a direct realization of the quantum metrological triangle", Metrologia, vol 37, 2000, p.207.

[6] B.V Hamon, "A 1-100 Ohm build-up resistor for the calibration of standard resistors", J.Sci.Instr.,vol 31, Dec.1954, pp 450-453.

[7] Lesley C Henderson, "A new technique for the automated measurement of high valued resistors", J.Phys., Electron.,Sci.Instrum., vol. 20, Sept. 1987, pp.492-495.

[8] "Transfer electrical resistance standards P40111 – P40115 ",Technical Pasport, Kishinev, 1989.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Automation to guarantee traceability and improve productivity in the reference laboratory of Mexico's federal electricity commision

G. Ruiz*, B. Valera*, R. Garibay, G. Mata**, S. Ochoa***, R. Nava*, J. Sánchez*, S. Padilla***

*Center for Applied Sciences and Technological Development and
**Faculty of Engineering of the
National Automous University of Mexico
***Metrology Laboratory LAPEM
gerardo.ruiz@ccadet.unam.mx

ABSTRACT: Automation of calibration processes in an electric industry reference laboratory to assure traceable and on time calibration of standards, instruments and equipment is presented. Relationship between different organizations is emphasized to perform the project, and some expected results are shown.

Introduction

Mexico's Federal Electricity Commission (CFE) [1] has a system of 29 calibration secondary laboratories supporting the substantive tasks of generation, transmission and distribution of energy. This system is known as CFE's Metrology Institutional System (SIMCFE), having a reference laboratory, called LAPEM (Figure 1), which is responsible for providing traceability in different magnitudes to all other calibration secondary laboratories. LAPEM performs test services to national and international suppliers of electrical, metallurgical, chemical and mechanical materials, equipments and other assets required not only by CFE, but for many other users in Mexico.

Figure 1. *View of main entry to LAPEM, the CFE's reference laboratory*

Quality assurance and measurement traceability in each one of those substantive tasks, where measurements are performed, impact on some relevant electric industry indicators as availability, efficiency, energy quality, customer service, etc. Nowadays, SIMCFE carries out about 20,000 instrument calibration services per year, and all of them must be traceable and realized on time (Figure 2).

Automation of LAPEM's Metrology Laboratory has been planned in order to satisfy that demand for services in charging to the Metrology Laboratory of the Applied Sciences and Technological Development Center (CCADET) and the Faculty of Engineering (FI) of the National Autonomous University of México (UNAM) [2] to carry out this project in a period of two years, with support of CFE and Mexico's National Council of Science and Technology (CONACyT) [3].

The objective of this project is the automation of calibration processes of standards and instruments in 14 stations of LAPEM's Metrology Laboratory in areas of Electric, Time and Frequency, Temperature, Mass and Density, Pressure and Dimensional, by means of communication standards IEEE-488 and series RS-232 on a Labview® [4,5] platform, in such a way that standardized results should be obtained at the end of calibration processes, the corresponding calibration report should be generated automatically improving time response and productivity, and human errors should be diminished.

Project development

The principal financial instruments which have CONACyT to promote and grant new technological development and research projects of interest of the Federal Government are called generically *Sectorial Funds*. By means of these economic

resources, the interaction between the different departments of the Federal Government and national enterprises like CFE, on one side, and the academic institutions like UNAM, on the other side, is achieved, searching for the solution to problems of national interest. Particularly, in this case, the *Sectorial Fund for Research and Technological Development on Energy* is applied.

Initial steps of the project execution

Project's call

In the second half of 2005, CONACyT launched a call inviting all national academic institutions which could have interest in participating in the project, in order to attend a set of demands and necessities concerning CFE and LAPEM. Each one of the interested academic institutions had to present a proposal. The one presented by CCADET and FI will be discussed as the next step of the project.

Figure 2. *LAPEM´s headquarters are located at Irapuato, Gto. Another 29 CFE secondary laboratories are distributed all over the country*

A specific demand of the project's call, was that related to the automation of the calibration of standards, equipment and instruments in different areas of LAPEM's Metrology Laboratory, namely:

- Electric;
- Time and Frequency;
- Temperature;
- Mass and Density;
- Pressure;

- Dimensional.

The main objective of the project, that is, the automation of calibration processes, must be achieved in a period of two years, working on a Labview® platform and communication standards IEEE-488 and series RS-232, in such a way that as a final result, the corresponding calibration report is obtained more efficiently at the end of the process.

A further two very important goals of the project are the improvement of the time response and productivity of LAPEM's Metrology Laboratory when calibrating the standards, equipment and instruments belonging to the SIMCFE's secondary laboratories. The expected results and products which should be delivered by UNAM to LAPEM's Metrology Laboratory are:

- automatic calibration stations in each one of the areas, including a laptop computer with the IEEE-488 and series RS-232 necessary accessories for communication;
- source code of the developed software;
- executable programs, which enable the processing and manipulation of information from databases, spreadsheets and word processors, in order to obtain, as a final product, the calibration report;
- process and calculus validation of operations carried out by the software for calibration report elaboration;
- software license for using by LAPEM;
- users and manager manuals;
- training course for personnel in charged of each one of the automatic calibration stations;
- test and set up of automatic calibration stations.

UNAM's proposal

UNAM's proposal is based on the experience and skills of its academic staff, proved by means of successfully realized research projects and technological developments, particularly in the field of automation of measurement processes or standards and equipment calibration. In addition, special capacity on software development using Labview®, Visual C and other development tools to allow communication between standards and equipment with the laptop computer using interfaces GPIB (IEEE-488), Series 232 and USB, as well as the interaction of results with spreadsheets, should be shown.

Another important compromise which should be fulfilled, is the rights cession of intellectual property of the project from UNAM to CFE, and before receiving specifications and detailed information and visiting the LAPEM's Metrology Laboratory, UNAM should sign a confidence agreement, with the compromise of not divulging information of any kind coming from its relationship with CFE and LAPEM.

The main requirements which must be satisfied to present the proposal include:

- a description of how the **objectives** mentioned in the project's call should be achieved by UNAM;
- a description of the process and methodology for reaching the **expected results** of the project;
- a **delivery program** of the results of the project;
- a **Gantt diagram** and a **critic route** of the project attached to the proposal.

The Gantt diagram must include the budget of the project spreading the next items:

- human resources;
- consumption materials;
- outsource professional services;
- equipment and instrumentation.

Additionally, the expenses associated with generating academic products from UNAM's staff, as well as training courses and other academic activities, must be taken into account, namely:

- academic meetings participation;
- training for academic and professional staff;
- scholarships for students involved in the project.

Under this scheme of work, it is important to remark that neither CCADET or FI, nor academic staff, receive any additional funds or salary, except in a given percentage, only for consumption materials and equipment and instrumentation.

Agreement

Several proposals from some academic organizations were submitted to the *Sectorial Fund for Research and Technological Development on Energy* for evaluation. Once the best proposal was selected – UNAM's proposal in this case – it

should be managed electronically through CONACyT's portal in the WWW, including among others the terms of agreement between CONACyT, CFE and UNAM, participants registration and electronic signature.

Technical development of the project

The main technical aim of the project is to comply with the objectives established in the project's call in accordance with the proposal presented and the delivery of products agreed. Nevertheless, the technical development of the project could be modified while its objectives and products do not suffer changes.

The technical development of the project is accomplished in four stages:

- Information compilation. In this stage of the project, a detailed analysis of the calibration processes requirements to be automatized is carried out. With the compiled information, the current conditions of the processes are determined, and the material resources to be used are estimated. The objective in this stage is to rely on the hardware setup for the different calibration stations.
- Programming applications. This stage covers the initial steps in the development of the automation systems by means of programming applications. The aim is to attain that the associated computers control the calibration processes for their automatic execution.
- Automatized calibration systems operation. Coupling between hardware and software, protocols and algorithms for calibration stations setup, is obtained in this stage. The results should be in accordance with those coming from the former procedures previous to automation of calibration processes.
- Test and validation. Finally, as the last stage of the project, hardware and software should be installed in the instrumentation and equipment of the automatic calibration stations. Results and automatic process validation should be carried out in each of them. LAPEM's Metrology Laboratory personnel should be trained by UNAM's academic and technical staff, and delivery of results and products to LAPEM's will lead to the corresponding technological transfer and project ending.

During all the time the project lasts, a continual communication and cooperation between UNAM's and LAPEM's staffs should be maintained.

In addition to the laboratory calibration equipment, each one of the calibration stations include interface boards for communicating instruments and laptop personal computers, as well as specific software for automatized calibration process, as shown in Figure 3.

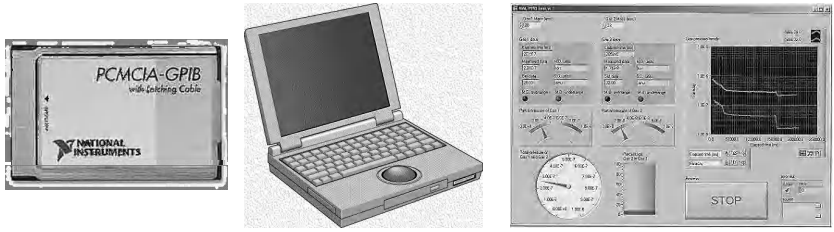


Figure 3. *Calibration station elements*

With the elements described above, a bidirectional information transmission between the instrument to be calibrated and the laptop computer is achieved. The data received in the computer are then processed and sent automatically for printing a calibration report, as can be seen in Figure 4.

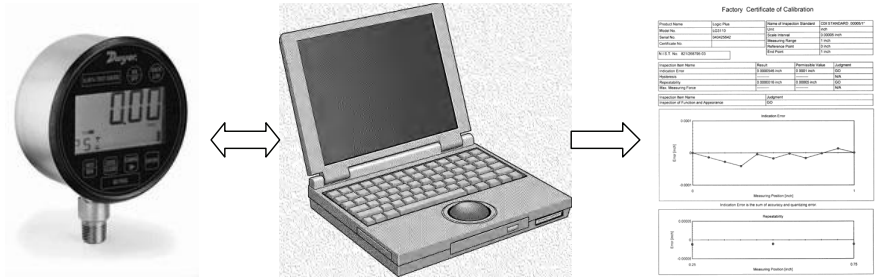


Figure 4. *Automatized calibration process and calibration report printing*

As a result, human errors and time employed in the calibration processes are diminished, and data from calibrations are stored in a data base to be managed by means of an intranet system. Calibration reports are standardized, and human resource training and infrastructure creation are reached.

Conclusions

Cooperation and technology transfer between three different organizations to solve a real problem by means of well known development tools and continual improvement processes have been shown. As a consequence, automation of LAPEM's Metrology Laboratory will lead to a productivity enhancement of its activities.

References

[1] http://www.cfe.gob.mx/en/

[2] http://www.unam.mx/english/

[3] http://www.conacyt.mx/Fondos/FondosSectoriales.html

[4] http://www.ni.com/labview/

[5] J. Travis and J. Kring, LabVIEW for everyone: graphical programming made easy and fun, NJ: Prentice-Hall/Pearson Educ., 3rd ed., 2006, 981 pp.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Legal Metrology

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

An industrial view on global acceptance of measuring instruments in legal metrology

James Matson, Jacob Freeke

GE Sensing
1100 Technology Park Dr. Billerica, MA 01821 USA

ABSTRACT: As a manufacturer of industrial gas flowmeters for the fiscal measurement of natural and other gases, the acceptance of instruments for legal metrology on a global basis is highly desirable. Having a legal metrology instrument accepted globally will bring economic efficiency, as well as uniformity of standards worldwide. This is performed by a common set of tests and certifications that can be accepted by many or all certifying organizations worldwide. The Sentinel gas flowmeter as tested by NMI for OIML R137 is an illustration of this process.

Fiscal metering

The supply of Natural Gas all over the world requires the use of flow measurement when that gas changes ownership. From the wellhead, through the storage and transmission in pipelines, to the boiler in a factory, to a turbine in a power plant or the heating of an individual home, Natural Gas must be metered precisely. Since the movement of the gas through all these steps involves a change in ownership, called Custody Transfer or CT, the exchange also involves the transaction of payment for the gas.

Figure 1. *Custody Transfer Gas Flowmeter*

When payment of money becomes involved, then so too does legal metrology become important. Both the buyer and seller of Natural Gas want the flow measurement to be, most of all, accurate. Reliability, low maintenance and cost are also considerations for measurement instruments in Custody Transfer, but accuracy is of utmost importance.

Certification

The industry faces a considerable challenge for an instrument to be accepted for legal metrology, on a global basis. For Natural Gas there are different approaches to the certification of a meter for CT.

The USA uses a recommended practice from the industry standards group, the American Gas Association (AGA), to enable buyer and seller to form a contract for custody transfer.

In Europe, Canada, Latin America and Asia the certification of a meter for custody transfer is generally by governmental authority: the legal metrology function. There are many authorities: in The Netherlands it is the Ministry of Economic Affairs; in China it is IQPRC (Pattern Approval); in Romania it is BRML; in Brazil it is INMETRO; in Canada it is Measurement Canada; in Germany it is PTB; in Malaysia it is SIRIM; in Indonesia it is MIGAS; in Russia it is GOST; in Italy it is the Ministry of Industry.

There are many more, virtually country by country, around the world. The industry must have approval from, or show compliance to, each of these authorities to be able to supply fiscal flow measurement in these countries.

Standards

In principle each authority in each of the countries involved has a standard or practice that the measurement instrument must conform to so as to be awarded the certification. In general these standards also vary from country to country.

In the USA, for Natural Gas, the practice is called AGA9 for the Ultrasonic flowmeter, like the Sentinel. Other AGA practices or standards cover other meter technologies such as turbine or orifice plates.

In the UK there are the BS or British Standards. Russia has its set of requirement as does Brazil's INMETRO. While some of the required parts of these standards are common, many parts are not.

The standards will cover many aspects of a measurement instrument. In the case of the ultrasonic flow meter like the Sentinel, these include both performance based tests, and environmental type tests.

Performance tests include:

Accuracy.

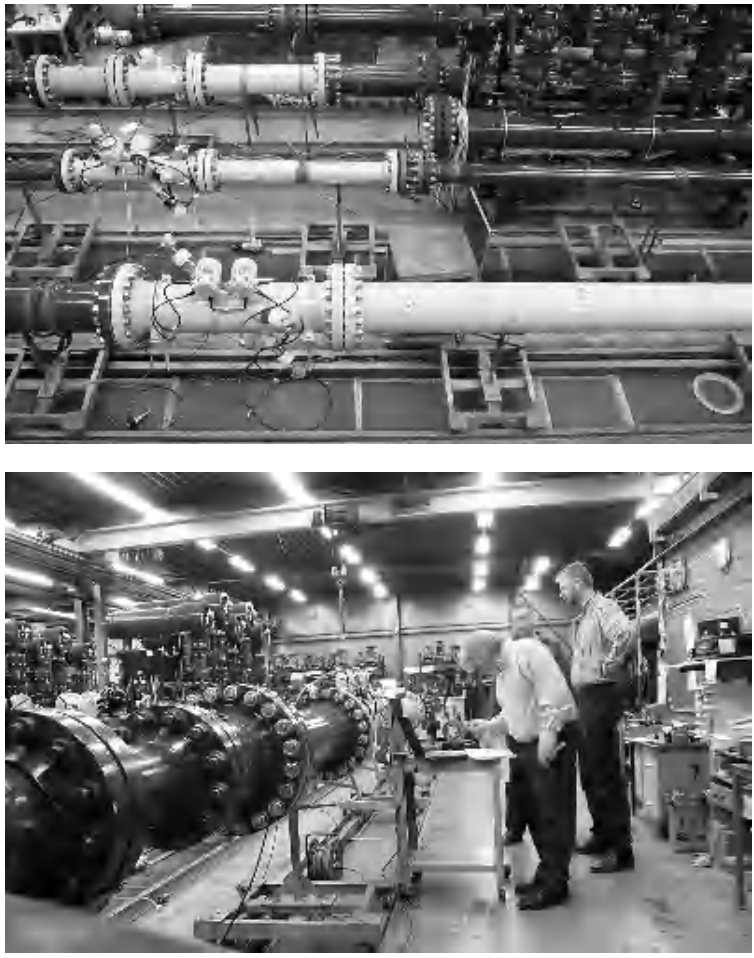


Figure 2. *Accuracy tests with pressurized Natural Gas, 6 and 12-inch flowmeters*

Reproducibility tests.
Orientation and bidirectional effects.
Different gases effect.
Different gas pressures and temperatures effect.
Interchange components evaluation.

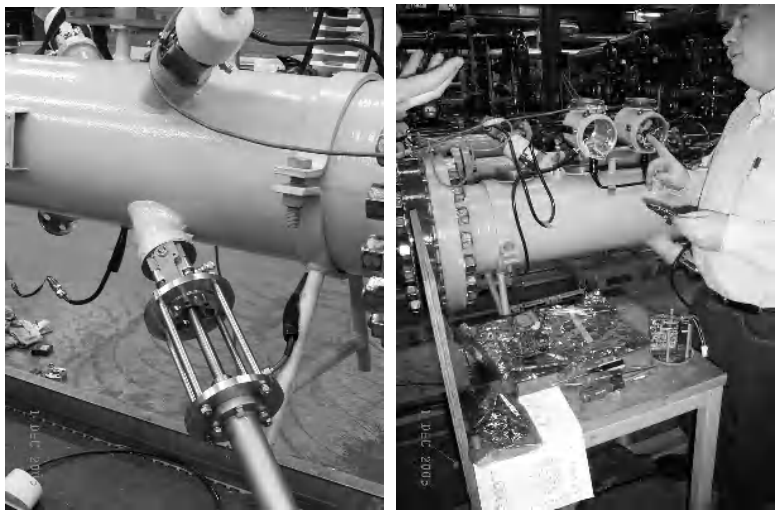


Figure 3. *Interchange of sensor and interchange of electronic power supply*

Flow disturbance tests.

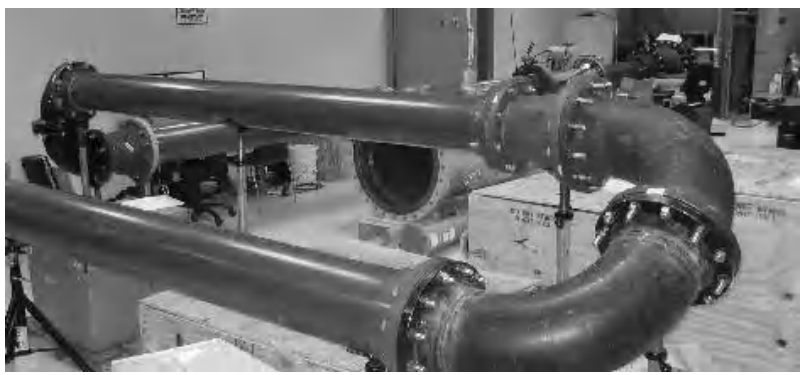


Figure 4. *Disturbed (non-straight) pipe testing*

Environmental tests include:

Temperature.
Humidity.
Electrical interference (EMC).

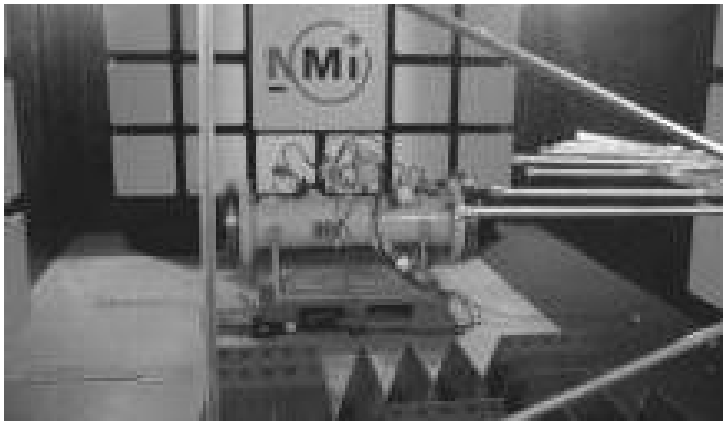


Figure 5. *EMI testing of Sentinel*

Vibration and shock

There are additional tests and evaluations that are called for in the various standards. As can be seen, some of this testing involves big pipes, high pressure natural gas, testing laboratories and other sophisticated equipment and facilities. Reproducing these tests for each standard in each country around the globe is both inefficient and extremely expensive. In fact, many countries that have a great need for Custody Transfer Natural Gas flowmeters may lack the facilities to carry out these tests.

From an industry point of view it would be a great benefit if there were a common set of standards that could be used by and applied in many, if not all, countries. The tests and evaluations would only need to be carried out once at facilities capable of fully carrying out the tests and evaluations, and even allowing the manufacturer quick and easy remedy to difficulties or improvement discovered during the evaluations.

Fortunately there are some standards today that are used by many counties throughout the world. The AGA 9 recommended practice is one and the OIML organization is another. The OIML (Organization Internationale de Métrologie Légale) is well known in the metrology field, and has a number of standards for flow measurement, including R6: General Provisions for Gas Volume Meters.

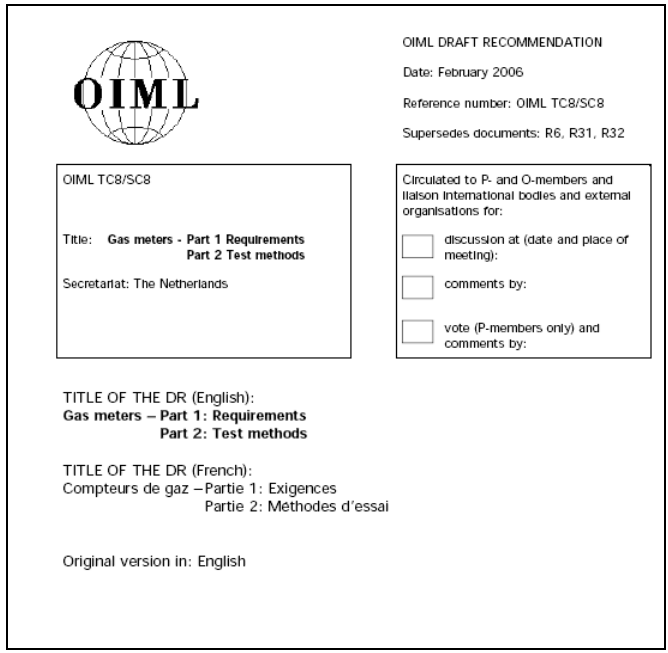
Many countries in Europe and Asia have adopted OIML standards, and even more will be added as the organizations influence grows.

Recently OIML has drafted a new recommendation for gas meters, R-137, which may be in force by time of the presentation of this paper in mid-2007.

Fortunately there are some standards today that are used by many counties throughout the world. The AGA 9 recommended practice is one and the OIML organization is another. The OIML (Organization Internationale de Métrologie Légale) is well known in the metrology field, and has a number of standards for flow measurement, including R6: General Provisions for Gas Volume Meters.

Many countries in Europe and Asia have adopted OIML standards, and even more will be added as the organization's influence grows.

Recently OIML has drafted a new recommendation for gas meters, R-137, which may be in force by the time of the presentation of this paper in mid-2007.


OIML

OIML DRAFT RECOMMENDATION

Date: February 2006

Reference number: OIML TC8/SC8

Supersedes documents: R6, R31, R32

OIML TC8/SC8

Title: **Gas meters - Part 1 Requirements**
Part 2 Test methods

Secretariat: The Netherlands

Circulated to P- and O-members and liaison international bodies and external organisations for:

☐ discussion at (date and place of meeting):

☐ comments by:

☐ vote (P-members only) and comments by:

TITLE OF THE DR (English):
Gas meters – Part 1: Requirements
Part 2: Test methods

TITLE OF THE DR (French):
Compteurs de gaz – Partie 1: Exigences
Partie 2: Méthodes d'essai

Original version in: English

Figure 6. *Draft cover page for R-137*

Evaluations

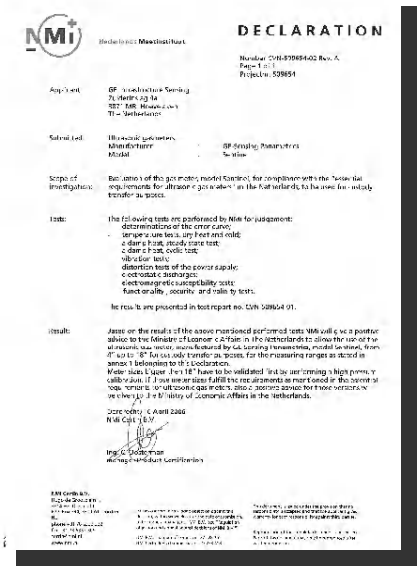
Having a common standard is only part of the task for a globally accepted legal metrology. The evaluation and testing of a measurement instrument to that standard must also be globally accepted. As mentioned, different countries may require different evaluations, tests or even levels of the same test (e.g. vibration) as part of their certification of a CT meter.

The well-known Netherlands Measurement institute (NMi) has the facility to test to the new OIML standards. In addition, NMi has global recognition agreements and other mechanisms for sharing evaluation and test data/results with standards organizations in many counties throughout the world. As a supplier to the world market our company is using the services of NMi to reach as many standards organizations in as many countries as possible, with one set of tests and evaluations.

The Sentinel gas flowmeter is the first meter of its kind to be tested in accordance to the OIML R-137 standards. NMi is able to carry out all of the requirements and tests in Holland where high pressure Natural Gas is available as well as all the other testing facilities.

Results

Through NMi testing to the OIML standards, the Sentinel Natural Gas flowmeter has been issued legal metrology approval for the Netherlands under number B37 by the Dutch Ministry of Economic Affairs based on the NMi type approval.



DECLARATION

Number: CVN-509654-02 Rev. A
Page 1 of 3
Projectnr. 509654

Applicant: GE Investments Sensing
Zug
Switzerland

Submitted: Ultrasonic gas meters
Manufacturer: GE Sensing, Panametrics
Model: Sentinel

Scope of investigation: Declaration of the gas meter, model Sentinel, for compliance with the essential requirements for ultrasonic gas meters in the Netherlands, to be used for custody transfer purposes.

Tests: The following tests are performed by NMi for judgement:
- determinations of the error curve;
- temperature tests, dry heat and cold;
- damp heat, steady state test;
- damp heat, cyclic test;
- vibration tests;
- distortion tests of the power supply;
- electrostatic discharges;
- electromagnetic susceptibility tests;
- functionality, security and reliability tests.

The results are presented in test report no. CVN 509654 01.

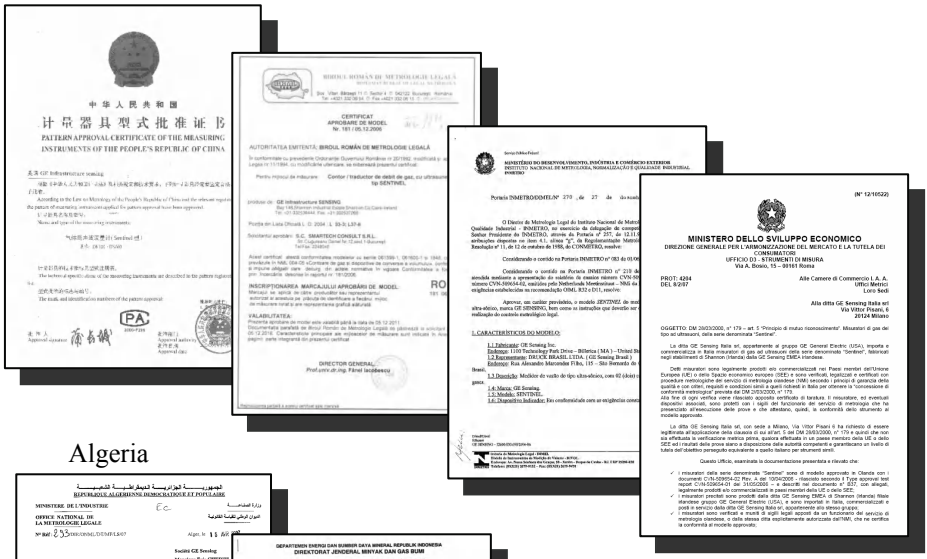
Result: Based on the results of the above mentioned performed tests NMi will give a positive advice to the Ministry of Economic Affairs in the Netherlands to allow the use of the ultrasonic gas meter, manufactured by GE Sensing Panametrics, model Sentinel, from Ø inches 8" for use as custody transfer purposes, for the measuring ranges as stated in annex 1 belonging to this Declaration.
Meter sizes Upper Limit 16" have to be validated first by performing a high pressure calibration. If the meter sizes fulfill the requirements as mentioned in the essential requirements for ultrasonic gas meters, also positive advice for those versions will be given by the Ministry of Economic Affairs in the Netherlands.

Dordrecht, 6 April 2006
NMi Certin B.V.

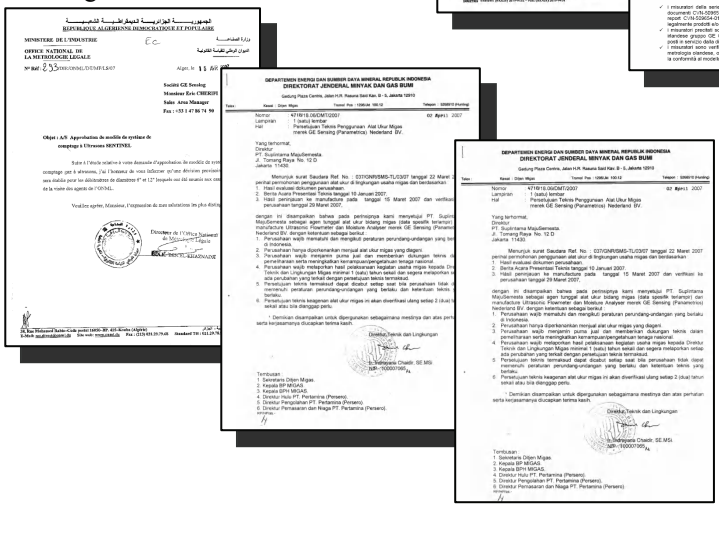
Ing. C. Dodemont
Manager Product Certification

The test results, reports and documentation, and of course the efforts of communication and cooperation by NMi with GE Sensing has made possible these other country certifications for Sentinel, based largely on the OIML standards:

China



Algeria



Conclusion

The flow measurement industry can greatly benefit from a global acceptance of standards and certifications from a small number of sources. When each country had to define their own standards, and conduct their own testing, certification was very slow and the ability to adopt a new, improved technology was impaired. Legal metrology requires qualified meters to stringent standards, and new uniform standards, and global standards testing organizations have made acceptance on these flowmeters much easier and better. There is more work to be done with even more countries to partake of the global effort, but the results to date have shown it is possible and inevitable.

References

[1] Henri Schouten "NMi type approval for the Sentinel ultrasonic gas meter", *NMi Certin B.V.*

[2] OIML TC8/SC8 Draft Recommendation, Feb. 2006

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Views from a notified body towards global acceptance

Pieter van Breugel MSc.

Director NMi Certin B.V.
P.O. box 304, 3300 AJ Dordrecht, the Netherlands
pvanbreugel@nmi.nl

ABSTRACT: Global acceptance is one of the most important topics in Legal Metrology to facilitate global trade. This paper gives the view of a leading Notified Body in Europe on global acceptance. What needs to be organized and by whom? Which problems arise today? Will the future bring a system of one stop testing for manufacturers, resulting in approvals all around the world?

Introduction

The globalization of the economy forces us to also develop our Legal Metrology system towards a global system too. Today, this is absolutely not the case. Although most countries signed the Technical Barrier to Trade Treaty of the World Trade Organization (TBT treaty of WTO), we still notify technical barriers for manufacturers when they intend to export to the global market. The main question today is: ”Do we want to solve this, and how?” As a leading test and certification institute, Notified Body and OIML Issuing Authority, NMi keeps close contact with both manufacturers and regulators around the globe. In this presentation, I will explain my views in relation to a real global acceptance of type approval tests for measuring & weighing instruments.

Who is responsible for global acceptance?

In my opinion, manufacturers should focus on their main business, namely to design and produce instruments. All unnecessary barriers for gaining approvals should be taken away by governments together with metrology institutes. In principle, it is the task of governments to facilitate global trade: they signed for it as a member of the WTO. As a testing and certification laboratory, we are willing to support this. To put it briefly, governments set up the rules and conditions for the acceptance of test reports and the test laboratories follow these rules. At NMi, we test the products and absolutely don’t waste any time on discussing the regulations. We must not overestimate our role. In most countries, the national government is responsible for issuing the type approvals. Getting test reports accepted is possible in most countries. However, the final decision takes the national authority. In the EU, one laboratory can issue approvals that are accepted in all EU countries. Such

exceptions are the result of a general political decision. Metrology has little influence on this.

Let us focus on the common situation: a manufacturer needs to test his product once, and he wants approvals for several countries without testing again. How can we realize this?

What are the reasons for non-acceptance?

Most manufacturers have bad experiences in getting their products approved, even if all tests have been performed successfully. To understand how to realize global acceptance, we first need to know why test reports are rejected. The reasons are:

1. The test report is not accepted/mistrusted.
2. The test report is not understood, or not drawn up in common format.
3. Not all requested tests are performed.
4. The test method is not accepted.
5. Misunderstanding in interpretation of the regulations for a specific country.
6. The exact test sample used for testing is not documented.
7. The documentation is not complete.
8. The conclusion/evaluation is not accepted.
9. Politics/protectionism/formal issues.

Except for the last reason, all these problems can be solved. If a country does not want to accept and due to national legislation, we could try and send a complaint to the WTO.

Basic philosophy in acceptance

To organize global acceptance of type approval tests, we first need to define what kind of documents have to be accepted. This seems to be obvious, but it is not. At NMi, we limit the acceptance documents to **test data and relevant documentation of the test sample.** This means that we don't argue with states about their laws. We don't argue if the laws are not harmonized with OIML recommendations. We don't argue about their test methods whatsoever. We simply want to know their requirements, their test methods and criteria. This implies that an approval certificate is not part of the acceptance documents. The final decision, expressed by a certificate, only has a function in the country itself, and is based on the test data and relevant documentation.

Conditions for acceptance

If we agree on the objective, as mentioned above, then the following steps must be taken.

Transparent testing

The test laboratory is the key player and organizes the acceptance. It is essential to demonstrate competence of the staff, traceable test facilities and transparent methods & procedures. Accreditation (by an organisation recognized by ILAC) according to ISO 17025 is the first proof at NMi. In most cases this is sufficient. Some countries do not trust accreditation only. These countries normally send over experts to check procedures and methods, sometimes as a member of the official audit team. In addition, we agree on the format of the test report and all tests to be performed. As a standard we always present reports based on the OIML recommendations, but if anything else is necessary, we supply the format we agreed on. By acting this way, reasons 1 to 5 to reject the test report are solved. It is important to keep in touch with the authorities in case of questions or doubts.

Acceptance protocol

The next step is to supply the relevant documentation in addition to the test report. Our bilateral agreements include specific lists of documents that need to be presented together with the test report. This enables us to guarantee that the receiving country can always check if the product presented by the manufacturer is exactly the same as tested in the laboratory. Some countries ask for a sample, which they can evaluate together with our test report and documentation. Some countries ask NMi to send the proven copies of the original documents. They don't trust anything else. We have now covered reasons 6 and 7 of the list.

Evaluation & decision

In my opinion, the evaluation of test reports and documentation is always the task of the country that intends to accept. So, we do not cover this in the acceptance protocol. It would be a mistake to include this in the process of acceptance. Even in Europe, a Notified Body is not allowed to delegate these tasks.

Thanks to this approach, refusal on reason 8 is not possible, because we don't ask for acceptance of our certificate / conclusion, we ask them to do that themselves. Therefore, if a country still does not issue the approval, it must be based on reason 9.

Experiences so far

NMi has several bilateral agreements and not-formalized procedures based on the described method. Of course, most important is that the government officials who evaluate our test reports and related documents know the relevant NMi engineers. There is frequent email contact if explanation is needed. We have very good experiences in Asia (including China and India), Australia and New Zealand, South Africa and Russia. It becomes more complex in the Americas, depending on the type of instrument and applicable regulations. Especially the development of the OIML MAA should cover this region (a latter presentation).

The 27 EU countries and the 3 EFTA countries form a single market and a single approval is satisfactory. In Europe, acceptance of test reports between countries is no longer an issue for measuring instruments covered by the MID and NAWI Directives.

The Future

In the short term, acceptance is expected to improve day by day. However, the next huge step is to harmonize the legal systems. This process will take decades. Therefore, NMi is working on global acceptance of test reports. In a relatively short time, we have achieved a reasonable worldwide acceptance of test reports.

The next step could be the development of the OIML MAA. In the first phase, we can try to harmonize all existing bilateral agreements on acceptance. In the second phase, it should increase value by attracting more and more countries, leading to *one global system* of acceptance of test reports. I will not explain the OIML MAA system in detail, but refer to one of the following lectures today.

During the Milestones in Metrology congress in May 2006 [1], a voting session on the future showed that 70% agreed that the OIML MAA should have a chance by implementing at least one category of instruments and build up experience. Also 70% were happy with the current OIML Certification system. This means that the acceptance based on OIML reports is working well, and the OIML MAA could be a further improvement of it.

References

[1] Voting the Future, congress Milestones in Metrology, Groningen 2006.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

High resolution modeling for evaluation of measurement uncertainty

Marco Wolf, Martin Müller, Dr. Matthias Rösslein, Prof. Walter Gander ETH

Zürich, Switzerland; Empa St. Gallen, Switzerland mawolf@inf.ethz.ch

ABSTRACT: The determination of the measurement uncertainty is a rather tedious task. The *Guide to the Expression of Uncertainty in Measurement* introduced an evaluation method for the uncertainty. Usually we choose a model of a perfect world that behaves preferably in a linear way and includes as few dependencies between the influences as possible to keep the effort for the calculation small. Using the Monte Carlo method that will be propagated in the first supplement of the GUM we allow with our simulation software a more detailed view of the uncertainty calculation model. We try to be as close to the real world as possible.

Introduction

If we want to publish results of a measurement nowadays additional information about the reliability of the results has to be provided. This tells readers something about the precision and quality of the measurement. The *Guide to the Expression of Uncertainty in Measurement* (GUM) [1] is a key document for the measurement uncertainty evaluation. It introduces a framework for the evaluation of the measurement uncertainty called GUM uncertainty framework (GUF). Using a Taylor series approximation the method of the framework propagates statistical information of the input quantities through a model equation that is usually reduced to the five or six most essential influences and has to be sufficiently linear. Usually a perfect world model is used for the sake of simplicity that behaves from the beginning preferably in a linear manner and includes none or at least as few dependencies, i.e. correlations, between the input quantities as possible. The result is a standard deviation, called a combined standard uncertainty that expresses in one number to some degree how reliable the result is. This approach is applicable and sufficient for numerous scenarios where the requirements hold that the model behaves linear and the input distributions can be adequately described using the first two moments. Therefore, an additional document to the GUM is currently in progress and has reached the status of a final draft, the *first supplement to the GUM* (GS1) [2]. It introduces a new method for the evaluation of the measurement uncertainty using a Monte Carlo method. Utilizing the complete information provided for all possible input quantities, it evaluates the measurement uncertainty. Furthermore, it is not restricted to linear models. Due to those advantages we are currently developing a software package called *MUSE* [3] that is able to calculate the measurement uncertainty applying MCM. As the model equations describe the

real world more realistically now, they are becoming quite complex and hard to manage. We support the user of the software *MUSE* to build realistic models by introducing basic models for measurement equipment and splitting the model equation into smaller units. These units can be defined individually and then combined to the final model equation for the measurement uncertainty. We start this paper with a short comparison of the GUM uncertainty framework and the Monte Carlo method from GS1. Afterwards we show how highly detailed models can be built and give examples of some different distributions, followed by the application of the technique to a practical example illustrated in the GUM that also occurs in GS1 and in a descriptive example from chemistry.

From GUF to MCM

The GUM uncertainty framework is widely used and is applicable to a lot of scenarios as already mentioned. The problem here is that the GUF uses only the expectation value and the standard deviation of the input quantities for the propagation of uncertainties. Additional information like the skewness and form of the distributions is omitted for the evaluation of the measurement uncertainty. Furthermore the model equation has to be linear or at least adequately linearizable, as the GUF uses only the first term of a Taylor series approximation for the calculation.

Due to these limitations of the classical GUM approach, GS1 is currently under development and soon to be released by the *Joint Committee for Guides in Metrology* (JCGM). It recommends a MCM for the evaluation of the measurement uncertainty as already mentioned in the introduction. MCM gives good results also in cases where GUF cannot be applied, as it uses the full information provided for the input quantities. Regarding software development for measurement uncertainty evaluation there is another source of information [4] with recommendations and practical tips for software developers implementing MCM as well as GUF.

In short, the simulation procedure of MCM can be summarized in the following three steps:

1. Select number of Monte Carlo trials M.
2. Generate M vectors corresponding to the PDFs provided for the input quantities.
3. Calculate the result of the measurement equation for all M vectors.

As it is not very intuitive how to select a number of Monte Carlo trials to get to a wanted accuracy as required in step 1, GS1 describes an algorithm for an adaptive approach that continues the simulation until some statistical parameters reach a fairly stable result.

We can then apply statistical methods to analyze the simulation data by calculating the expectation value, standard deviation and shortest convergence intervals for some ranges.

Our software package *MUSE* uses MCM for the evaluation of the measurement uncertainty, supports adaptive Monte Carlo and includes a tool to analyze the resulting data files, where it calculates mean, standard deviation and the shortest 95%, 99% and 99.9% coverage intervals as well as summarizing histogram data. As we also provide the complete data sets of the simulation, one can use any statistical software package like R [5] for the analysis if it is able to cope with such large data files. So we can provide an overall solution for the calculation of the measurement uncertainty according to GS1.

High resolution modeling

As MCM uses the full information provided for the input quantities, the quantification of the influences to the measurement uncertainty is becoming more important. The model will be closer to a real world situation in a more detailed resolution. That implies that the equations are becoming more complex and it is harder to keep an overview. We support users of our software project MUSE with a structured and manageable modeling schema. For this we subdivide the modeling process itself into three main parts as can be seen in Figure 1. First there is the definition of basic models for modeling measurement equipment abstractly. The equation of the measurand is defined using instances of these basic models with individual parameter settings. In the calculation section parameters related to the simulation, e.g. number of simulations or selection of random number generator, are set.

We use a meta language defined in XML (extensible markup language) for the definition of basic models, processes and calculation parameters. That allows us to provide an easy-to-understand, structured and to many people well known syntax. It also reduces the risk of wrong definitions in the files, because of the possibility of validation of the measurement model files using a document type definition file.

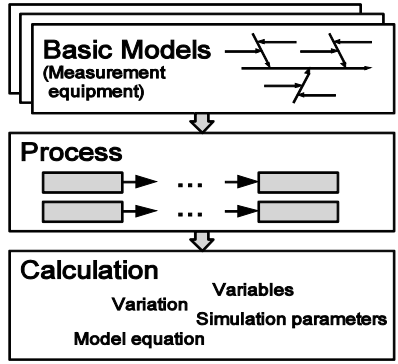


Figure 1. *Subdivision of modeling process into basic models, processes and a calculation section*

Basic models

Basic models represent an abstraction of measurement equipment. Each basic model uses one definition file that can be instantiated arbitrarily often for the calculation with individual parameter settings. MUSE already supports all distributions for the definition of influences mentioned in GS1 (Gauss, rectangle, triangle, trapezoidal, U-shaped, etc.) as well as some special distributions. It is possible to use simulation results from preceding measurement uncertainty calculations directly. We then use the data as presampled input distributions. Sampling can start at a random position in the list of numbers or just from the beginning for reproducible results.

Another special feature for input distributions is the support for a program widely spread in the Bayesian statistics community called WinBugs [6]. We can define parts of the measurement scenarios there and directly import the obtained simulation data in MUSE. The fundamental idea is to not only fit parameters of a distribution to current measurement data, but to additionally use knowledge from experts or from proceeding measurements to define a new distribution. The concept is presented in Figure 2. The more current data points are available the more important they become in the evaluation.

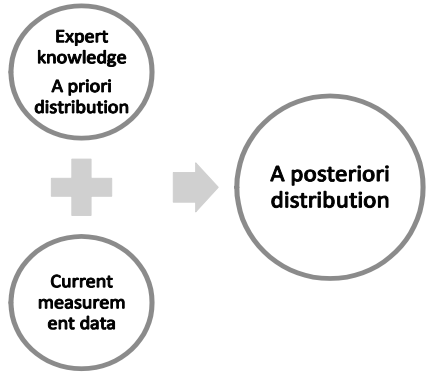


Figure 2. *Supporting OpenBugs [6] it is possible to combine expert knowledge and current measurement data into an a posteriori distribution*

A third group of special distributions can be used to model distributions of influences near physical limits. For example, there can never be a length with a value below zero or a concentration with more than 100%. The classical GUM approach uses only the expectation value and the standard deviation for the evaluation of the measurement uncertainty. That does not allow us to take such information into account. As we use MCM for the evaluation, distributions can be easily cut at their physical limits. We support two kinds of cuts here. The first one discards all values sampled over the limit, more precisely it resamples values until they are within the limits. The second method resets all values outside of the limit to the value of the limit. S. Cowen and S. Ellison have analyzed the modeling of natural limits [7]. Supporting the modeling of measurement scenarios with different aspects of this paper we make it possible to get a deeper insight into the behavior of complex measurement systems and to get a feeling for the most important influences in the measurement procedures.

A long term target is to provide a freely accessible database of basic models for measurement equipment in laboratories. Only if people use the same or at least comparable models as basis for their measurement uncertainty evaluation, a validation and comparison of results is possible.

Processes

A measurement procedure includes a lot of smaller measurement activities. A process in MUSE encapsulates such a step into a separate unit. Processes can then be combined to calculate the measurement uncertainty for bigger systems. For instance, the sample preparation can practically always be separated from the actual measurement procedure. The subdividing allows individual handling of such units with its own sets of parameters and variables. Processes can use an arbitrary number

of instances of basic models as well as the results of other processes as input for the evaluation of the measurement uncertainty. What follows is a type of hierarchical system and the subdivision into processes allows a better overview for a specific measurement scenario.

Another big advantage here is the flexibility of the modeling. Parts of the measurement sequence can easily be removed, exchanged or plugged in. Different measurement settings or replacements of parts of the measurement equipment can be analyzed without much effort. We can decide then for instance if an improvement of the measurement uncertainty takes place if more precise, but maybe even more expensive equipment is used.

Calculation

Finally the processes are combined to evaluate the measurement uncertainty for the whole system. In the calculation section it is possible to set parameters for the simulation, like the number of trials, the type of random number generator, the parameters for creating and analyzing data files, etc.

We also allow defining global variables (constant values or random numbers drawn from a specific distribution) and variation loops in this section. The loops enable us to start a simulation system only once for the evaluation of a set of distinct parameter sets. The conventional variation loop corresponds to *for*-loops best known from common programming languages. A specified iteration variable is increased or decreased for a constant value at a time until the variable exceeds a given limit. There is the limitation that all permutations of possible values of the iteration variables are taken into account if more than one variable should be varied, because more than one *for*-loop is interpreted as nested loops. A more sophisticated loop tackling that problem is the *variation-set*-loop. Here not only one variation variable is set at a time, but a complete set of variables for each step. The variables are not automatically increased or decreased, but they are defined explicitly for every step. Therefore we gain the full control for the permutations of values and it is also possible to define loops with variables in non-equidistant intervals.

Using this set of techniques together makes it possible to evaluate measurement scenarios with different parameter settings in one run of the simulation system of *MUSE*. It also means that a direct comparison of various measurement systems is possible and the whole measurement uncertainty evaluation becomes a lot easier to manage.

A word about dependencies

Causal dependencies between influences appear quite often, even more in more complex measurement scenarios. Usually we try to keep the model as simple as possible and therefore omit dependencies. We support the approach from GS1 if there are two or more normally distributed correlated input quantities, but preferably we recommend a refinement of the model equation to define causal dependencies due to manageability and lucidity. We should try to extract the quantity that causes the dependency between two other influences and model it as an explicit influence with its own distribution and model equation. That may not be possible in many cases because of a lack of knowledge and due to time and cost issues, etc., but if applicable it allows a more accurate and concise modeling.

Example: Gauge block calibration

We use the example of a gauge block calibration from appendix H.1 of GUM and section 9.5 of GS1, where the length of a nominally 50mm gauge block is determined by comparing it with a known reference standard of the same length. For the evaluation of the measurement uncertainty they use the model equation

$$\delta L = \frac{L_s\left[1+\alpha_s\left(\theta-\delta\theta\right)\right]+d}{1+\left(\alpha_s+\delta\alpha\right)\delta}$$

whereas the difference d in the lengths of the gauge block being calibrated and the reference standard is a summary of the subformula $d=D+d_1+d_2$. D is the average length difference; d_1 and d_2 are quantities describing the random and systematic effects associated with the comparator. The quantity θ represents the deviation of the temperature from 20°C of the gauge block and can be expressed as $\theta=\theta_0+\Delta$. θ_0 represents the average temperature deviation and Δ the cyclic variation of the temperature deviation from θ_0. Furthermore we have the following influences:

- Length L_s of the reference standard;
- Thermal expansion coefficient α_s;
- Difference $\delta\alpha$ in expansion coefficient; and
- Difference $\delta\theta$ in temperature.

For a more detailed explanation of the measurement system and different settings of distributions see Table 10 in GS1 [2].

In *MUSE* it is quite comfortable to model subformulae of the equation individually, and we can maintain a better overview of the main formula. The example uses different distributions for the parameters, namely normal, student-t,

rectangular, U-shaped and trapezoidal distributions. As the classical GUM approach forces a linearized model equation and a reduction of input distributions to expectation value and standard deviation, this quite simple example already shows how MCM enhances the process of measurement uncertainty calculation. MCM uses all the information provided as-is for a simulation of the system and delivers a more reliable result for the measurement uncertainty. It is also possible to calculate proper shortest confidence intervals for different probabilities.

Method	Number of simulations	Mean	Standard deviation	Shortest 95% interval	Shortest 99% interval	Shortest 99.9% interval
GUF		838	32		[746,930]	
MCM (GS1)	$1.36*10^6$	838	36		[745,931]	
MCM (MUSE)	10^4	838	36	[707,959]	[744,933]	[766,908]
MCM (MUSE)	10^5	838	36	[718,958]	[743,931]	[768,909]
MCM (MUSE)	10^6	838	36	[718,959]	[745,931]	[768,908]
MCM (MUSE)	10^7	838	36	[718,959]	[745,931]	[768,908]
MCM (MUSE)	10^8	838	36	[718,958]	[745,931]	[768,908]

Table 1. *Results of gauge block calibration of GS1 and MUSE*

In Table 1 the results of the measurement uncertainty evaluation are given in GS1 for this example compared to the results of the simulation with MUSE. We obtain the same results and we can see that at least for this example MCM delivers quite stable results. Even the 99.9% coverage interval does not change a lot with 106 simulations and more.

In the next section we will show how an advanced example with different steps can be managed using *MUSE*.

Example: modeling a chemical process

Let us use a common measurement scenario from chemistry to show how modeling complex measurement system works with MUSE. We weigh some reference material on a scale and afterwards put the substance into a flask for a first solution, the stock solution. After that we use a pipette to obtain a certain amount of

the liquid from the flask to create another solution in a second flask. This procedure is repeated once more to obtain to a third and final solution. The model equation for this measurement looks like this:

$$c = \frac{m * purity}{V_{Flask,s}} * \frac{V_{Pipette,1}}{V_{Flask,1}} * \frac{V_{Pipette,2}}{V_{Flask,2}}$$

The idea now is to split up the model equation into logical parts, so-called processes. We can divide the formula into three more or less independent processes as shown in Figure 3.

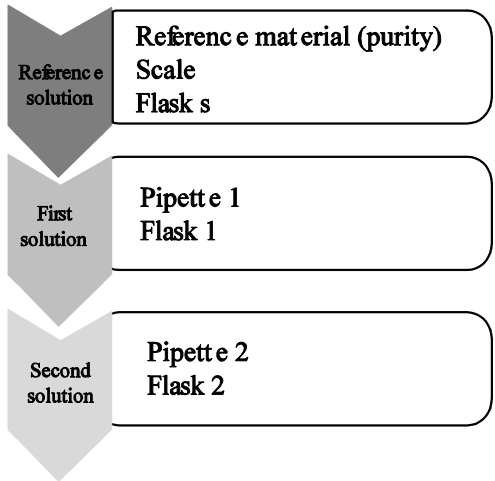


Figure 3. *We separate the model equation into three parts for the evaluation of the measurement uncertainty: reference solution, first solution, second solution*

Each of these processes can be managed individually. That means in the definition file we can use separate variables and influences for each process. The results of one or more processes can be used as input for succeeding processes.

It can be useful to allow instances of basic models to be used in different processes. For example the pipettes 1 and 2 from both solution steps need not be distinct. After a first solution the pipette may be cleaned and reused. MUSE makes it possible to use instances of influences in different processes. Another useful feature is the keyword static of the modeling language. An instance of a basic model marked with the keyword static is precalculated and keeps its value during the whole evaluation of the model equation for one pass of the simulation. Like this the values of a whole instance of a basic model or just a part of it can be kept constant for a simulation run, if a device is used more than once. To show the difference in using one or two pipettes we assume for all the influences normal distributions with the following settings:

- Purity in g/g N(0.99,0.008);
- Scale in g N(0.1,0.0001);
- Pipette 1 & 2 in ml N(1,0.061);
- Flask s, 1 & 2 in ml N(100,0.15).

Pipettes	Mean	Standard deviation	Shortest 95% interval	Shortest 99% interval	Shortest 99.9% interval
1	9.94	1.21	[7.61,12.33]	[6.98,13.17]	[6.32,14.13]
2	9.90	0.86	[8.24,11.59]	[7.77,12.16]	[7.28,12.82]

Table 2. *Comparison of results of two measurement scenarios. First line represents measurement with one pipette for both solutions. Second line shows results of measurement uncertainty evaluation using two pipettes. All values are to be taken* 10^{-4}

Table 2 shows the results of a simulation with 107 trials. Even for this strongly simplified scenario where we have only normal distributions we obtain conspicuously different results. We see here quite a high potential to obtain more sophisticated models that represent the real world much better than those using strongly idealized model equations and information reduced input distributions.

Summary

Usually we try to use a simple model to keep the effort for the uncertainty calculation as small as possible. The five or six most important influences are chosen, they are quantified using normal distributions and put into a linearized model equation to evaluate an estimation for the measurement uncertainty. In some cases this maybe an accurate choice to keep time expenses and costs as small as possible. Using existing software packages for the GUM framework simplifies the work once more.

However, we think a more realistic modeling can enhance the quality of measurement results a lot more. Highly detailed, more realistic models can lead definitely to a better inside to the mechanisms of configuration parameters of measurements and therefore help to increase the reliability of results. MUSE supports the user of the software to model the measurement process and automates the simulation runs and analyzing data. Dividing the model equation into different parts that use different sets of variables and parameters allows a very well arranged modeling process. We showed first in an example of GUM and GS1 that MUSE arrives at the same results and allows a logical structuring of the model equation. In a rather small example from chemistry we then explained how our method fits the needs of the user. Different scenarios can be evaluated quite easily and a direct comparison of results using various parameter sets is possible.

It is hard to decide in advance which influences are the most important to a measurement scenario due to the measurement uncertainty. We support the user of our software in finding the most important influences to the measurement, as we make it possible to test different settings with very small effort, for example in using different kinds of loops for repeated simulations with different parameter sets. Whole parts of the measurement can easily be replaced, edited or deleted.

All in all we hope to support and automate the user in enhancing the measurement uncertainty evaluation task with our software package *MUSE*.

References

[1] Guide to the Expression of Uncertainty in Measurement (GUM), first edition, 1993, corrected and reprinted 1995, International Organization for Standardization (ISO), Geneva, 1993.

[2] Draft-Guide to the Expression of Uncertainty in Measurement Supplement 1: Numerical Methods for the Propagation of Distributions. JCGM Working Group on the Expression of Uncertainty in Measurement, 2006.

[3] MUSE – Measurement Uncertainty Simulation and Evaluation, software package, ETH Zürich, http://www.mu.ethz.ch.

[4] Cox, M. G.; Harris, P. M.: Software Specification for Uncertainty Evaluation, Report to the National Measurement System Directorate, Department of Trade and Industry, NPL Report CMSC 40/04 (CMSC 10/01 revised), Crown, 2004.

[5] R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2006.

[6] Thomas, A.; OHara, B.; Ligges, U.; Sturtz S.: Making BUGS Open. R News, Volume 6. The University of Auckland, New Zealand, 2006, S. 12–17.

[7] Cowen, S.; Ellison, S.: Reporting measurement uncertainty and coverage intervals near natural limits. The Analyst, 131. Royal Society of Chemistry, 2006, S. 710–717.

Appendix

The software project MUSE is currently developed by the Measurement Uncertainty Research Group (MURG) at ETH Zurich/Switzerland in cooperation with the Empa St. Gallen/Switzerland. MUSE realizes most aspects of GS1 as well as a lot of additional functionalities. The application and modeling language is

intended to be used by interested experts to obtain a better insight into the behavior of measurement systems and the corresponding measurement uncertainty. An early version of the program is available for free at www.mu.ethz.ch for download and testing.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Evaluation of uncertainty using Monte Carlo simulations

M. Désenfant *, N. Fischer *,
B. Blanquart **, N. Bédiat**

*Laboratoire national de métrologie et d'essais (LNE)
** Centre Technique des Industries Aérauliques et Thermiques (CETIAT)

ABSTRACT: The supplement 1 to the GUM, dealing with the evaluation of uncertainty using the propagation of distributions, will be published this year. This supplement, based on a Monte Carlo method (MCM), allows us to deal with the evaluation of uncertainty even when the hypotheses of the law of propagation of uncertainty (LPU) are not verified. We describe the different steps of the MCM and its advantages. Then to illustrate the method, we present two examples in the fields of metrology and testing and we compare the results given by both methods.

Introduction

The guide to the expression of uncertainty in measurement (GUM) is based on the law of propagation of uncertainty (LPU) for the calculation of the combined standard uncertainty. It is estimated by Taylor development applied to the measurement process model; calculation then requires the validity of a certain number of mathematical hypotheses. First, the model must not present significant non-linearity. Observed deviations must be low for each variable of the measurement process, comparable in size and must present symmetric distributions. Finally, the distribution of the model output variable must present a Gaussian distribution to easily calculate and interpret the value of the coverage factor *k*.

From a strictly operational point of view, the application of the law of propagation of uncertainty requires a model diversion of the measurement process in relation to each input variable in order to estimate sensitivity coefficients .

Besides the mathematical hypotheses of the diversion model, its application can turn out to be tricky in a manufacturing context. It can slow the calculation process and be a potential source of error.

An alternative to this calculation is described in supplement 1 to the GUM, published in 2007 [1]. The metrologist can have the use of numeric simulation tools, particularly the Monte Carlo method, to propagate not only two statistics (average and variance) but also the whole distribution describing the measurement process.

As an illustration, both estimation methods of the combined standard uncertainty were applied to two measurement processes in order to highlight the strong points of each one and the problems that can occur during their implementation.

The first example comes from the performance qualification activity of thermal components carried out by the Centre Technique des Industries Aérauliques et Thermiques. It is the process of determination for the thermal power provided for example by a boiler where the uncertainty is traditionally obtained with the law of propagation.

The second example is taken from mass metrology; it involves the determination of conventional mass of a body.

Propagation of distributions by the Monte Carlo method

Because of the limits and constraints of the LPU method described in the GUM to evaluate measurement uncertainty, an alternative approach was developed. This approach known as the Monte Carlo method (MCM) is the subject of supplement 1 to the GUM soon to be published. It is important to note that this supplement is not intended to replace the NF-ISO 13005 standard (GUM) but to complete it.

The principle of this method is no longer to propagate uncertainty via the model, but the probability density function (PDF) of input variables in order to obtain the PDF associated with the measurand. Since the PDF for each input variable is known, the PDF for measurand Y can be analytically obtained by the Markov formula. In practice, except for very simple models, the multiple integral cannot be analytically evaluated. Supplement 1 to the GUM provides a numeric method that determines distribution propagation by using a Monte Carlo method (MCM).

This alternative approach can be summarized step by step, see Figure 1, according to the following process:

1. Define the measurand, measurement process, influence factors and express mathematically the relationship between the measurand and the input quantity. This vital step is actually common to all uncertainty evaluation methods.
2. Associate with each input variable a (average, rectangular, etc.) distribution or a joint distribution in the case of correlated variables. This choice must be made by taking into account available information and according to the maximum entropy principle, i.e. choosing the PDF g which maximizes entropy S:

$$S[g] = -\int g_X(\xi) \ln g_X(\xi) d\xi$$

3. Carry out M trials of each input variable by drawings in their PDF. To achieve these simulations, it is necessary to have a pseudo-random number generator

very efficient (it must go through a series of randomness tests). Supplement 1 to the GUM suggests $M = 10^6$ draws (minimum).

4. Calculate via the mathematical model the M values obtained from the output variable, making it possible to build the empirical measurand distribution.
5. Synthesize the information obtained on the output quantity by sending back:
   a. the average;
   b. the standard deviation;
   c. the shortest interval for the probability specified (often 95%).

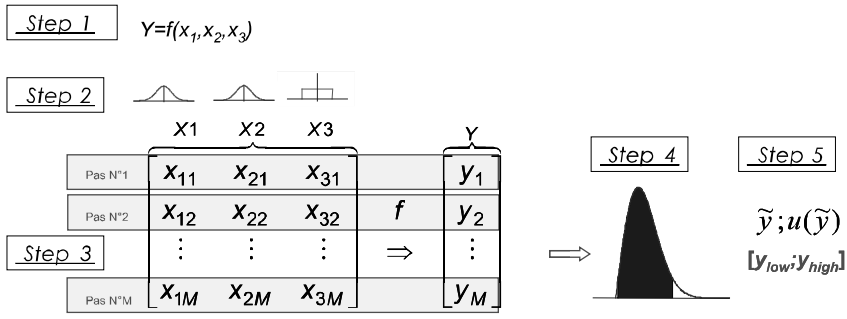


Figure 1. *The Monte Carlo method steps*

The evaluation method of the uncertainty by propagation of distributions presents several advantages compared to the traditional approach based on the law of propagation of uncertainty:

– calculation of partial derivative is no longer necessary;

– no more problems because of non-linearity and/or "strong" uncertainties over input variables;

– no more problem with the choice of the coverage factor.

In addition according to supplement 1 to the GUM (sec. 8.1.1), the MCM validity context is larger than with the LPU method. Consequently, using MCM results to validate the results obtained by the LPU method is recommended. A simple validation procedure is described in the supplement.

First it is necessary to determine tolerance δ associated with the calculated z result. It is expressed in the form $z = c.10^l$ where c is the value of the last digit, $n_{dig,}$ deemed significant for z. We will then write: $\delta = 0.5.10^l$. In the *u(y)* uncertainty estimation example of a mass where only one significant number is retained. This

means that, n_{dig} = 1, $u(y)$ = 0,0006 g, i.e. c = 6 and l = -4. The tolerance then equals: δ = 0,5 10-5 g.

The second step of the validation process comes down to comparing expanded uncertainties obtained by each method. We form differences:

$$d_{low} = \left| y_{lowLPU} - y_{lowMCM} \right|$$
$$d_{high} = \left| y_{highLPU} - y_{highMCM} \right|$$

If these differences are both lower than the tolerance previously determined, the results obtained by the LPU method are validated. Otherwise, we have to use results obtained by the MCM method.

Application in testing

The Centre Technique des Industries Aérauliques et Thermiques (CETIAT), carries out thermal performance tests for components (boilers, air conditioners, etc.). A frequent measurement process is the thermal power determination of a component. It is determined by a series of measures (input temperature, output temperature, flow) and calculations (heat content, power).

Description of measurement process

The experimental assembly implemented during thermal power determination recovered at the secondary circuit of a heat exchanger is represented in Figure 2.

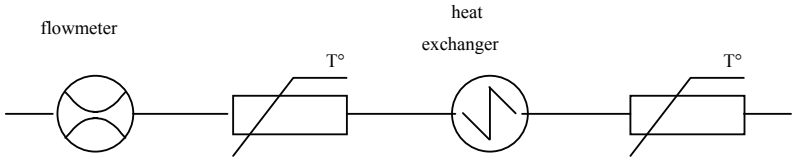


Figure 2. *Measurement device for thermal power*

Thermal power is the product of the mass water flow value by the mass heat content difference measured at the latter's limits, based on the relation:

$$P = q_m \cdot \left[h(T_s) - h(T_e) \right]$$

with

q_m : mass water flow in the exchanger, in kg.s^{-1}

T_s : water temperature at exchanger output, in °C

T_e : water temperature at exchanger input, in °C

$h(T)$: heat content of water at temperature T, in kJ.kg^{-1}

The flow is determined by a Coriolis mass flowmeter which metrologic characteristics are followed by periodic calibration. Water temperatures at exchanger input and output are measured with the help of Pt100, periodically calibrated. Sources of error associated with the measurements are linked to properties of instrumentation (accuracy, repeatability) and to implementation conditions (stability, influence of environmental conditions, temperature homogeneity in the section, etc.).

The heat content is calculated from the temperature with a smoothing method that presents a smoothing error linked to accuracy of reference values and to quality of interpolation. It is made in the form $h(T) = h(T) + \alpha$, where α represents smoothing correction, zero on average with an added uncertainty.

Application of the propagation law

In the absence of correlations between variable estimations, the application of the propagation law of uncertainty leads to the following equation:

$$u_c^2(P) = \left[\frac{\partial P}{\partial q_m}\right]^2 \cdot u^2(q_m) + \left[\frac{\partial P}{\partial t_s}\right]^2 \cdot u^2(T_s)$$
$$+ \left[\frac{\partial P}{\partial \alpha_s}\right]^2 \cdot u^2(\alpha_s) + \left[\frac{\partial P}{\partial t_e}\right]^2 \cdot u^2(T_e) + \left[\frac{\partial P}{\partial \alpha_e}\right]^2 \cdot u^2(\alpha_e)$$

In practice, this calculation is programmed directly in the result processing software specific to testing platforms in an Excel tab. Information relating to type B standard uncertainties are updated according to information for the network asset management software whereas type A standard uncertainties are calculated from measurements from each test.

Application of the Monte Carlo method

The application of the Monte Carlo method requires a specific calculator to obtain a sufficient number of random draws (>10^6), with a generator having a longer

period than the number of draws. The calculation is done in a specific software package (BOOST) previously validated by the application of examples presented in supplement 1 to the GUM. This software package is independent from the results processing software package in test platforms.

The BOOST software makes it possible to define a module for each model quantity (input temperature, output temperature, flow, etc.) and then to assemble them into a "power" meta-module. For each quantity, the user writes the model in great detail, in C, *via* a data entry interface. Finally, he or she chooses the numeric value of each model term, its associated uncertainty and probability function.

Sensitivity factors of the measurement process model in relation to each variable can then be estimated (except for the sign) according to the method described in annex B of supplement 1 to the GUM, by completing simulations where the values of the other parameters are set.

Comparison of both methods

The Monte Carlo method is used to confirm hypotheses of the application of the law of propagation of uncertainties to the measurement process considered. The comparison of both calculation methods can only be achieved by calculations using the same input values (example in Table 1) and the same hypotheses of variable density functions.

The calculation was done for a finite number of cases, corresponding to the most frequently encountered situations during tests. It shows that uncertainty values obtained by both methods are in agreement. In fact, gaps observed between both methods are in accordance with the validity criterion proposed in supplement 1 to the GUM and are minimal compared to calculation sensitivity to variations of an input variable.

The probability density of thermal power P is represented in Figure 3 for 10^6 random draws corresponding to values in Table 1. The Gaussian gradient with the same standard deviation, represented by a black line, overlaps with the histogram, showing that the hypothesis of considering power P PDF as Gaussian is acceptable.

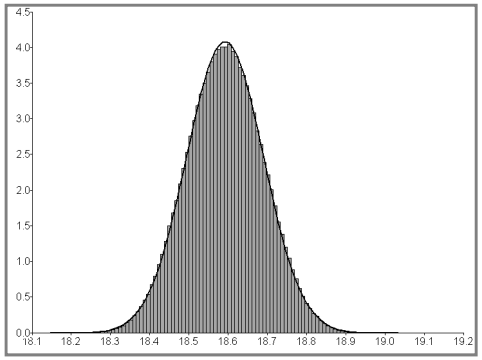


Figure 3. *Power P PDF*

			Law of propagation		Monte Carlo	
Dimension	**Value**	**Unit**	**Typical uncertainty**	**Sensitivity factor**	**Typical uncertainty**	**Sensitivity factor**
Mass flow q_m	804.6	kg/h	2.9625	$2.31.10^{-2}$	2.9595	$2.31.10^{-2}$
Input heat content $h(T_e)$	251.39	kJ/kg	0.2091	$2.24.10^{-1}$	0.2171	$2.23.10^{-1}$
Output heat content $h(T_s)$	334.57	kJ/kg	0.2098	$2.24.10^{-1}$	0.2218	$2.23.10^{-1}$
		Unit	**Value**	**Expanded uncertainty (*k*=2)**	**Value**	**Expanded uncertainty (*k*=2)**
Power *P*		kW	18.59	0.19	18.59	0.20

Table 1. *Example of compared results for one data set*

Conclusion

In practice, the law of propagation of uncertainty is used to obtain a result in the test software, as soon as the data is available. This calculation is validated by the comparison with the application of the Monte Carlo method.

However, beyond the choice of validity criterion, one of the validation limits is its representative character compared with the field of variable variation and of the model and standard uncertainties associated with each component.

In fact, the application of the Monte Carlo method for the validation of the application of the law of propagation into a single value is not sufficient. For a

complete validation, we need to carry out a minimum number of simulations by exploring two dimensions for each variable of the model: the measurement range on one hand and the standard uncertainty values on the other hand.

The problem to treat becomes multidimensional. Since calculation time is a significant parameter, we need to treat this case with the application of experimental designs to explore the whole surface of response and compare it to the response surface of the application of the law of propagation. Beyond the understanding of metrology elements for creating the uncertainty assessment, this validation requires the understanding of experimental designs combined with an understanding of numeric simulation.

Application in metrology

The application we are presenting comes from the Mass technical unit of the centre de métrologie scientifique et industrielle of the LNE. It is conventional mass calibration of a class F1 standard from a class E2 standard [4]. This calibration is particular in that it takes air buoyancy correction into consideration. The example presented in this study is similar to the one developed in part 9 of supplement 1 to the GUM.

Measurand and measurement process

The goal of calibration is the determination of conventional mass determination. The measurand is therefore conventional mass C_M, of a mass M. This conventional mass of body is equal to the mass of an 8000 $kg.m^{-3}$ volume mass standard which balances this matter in an 1.2 $kg.m^{-3}$ volume mass air, the operation takes place at 20°C and the parameters are defined at this temperature.

Calibration of a mass standard (M) consists of the determination of its mass by comparison with the mass (E) of a mass standard. This comparison is made with the help of a mass comparator based on a doubled BORDA type substitution diagram.

Mathematical model and and input variables

Value (C_M) of the conventional mass is given by the following relation:

$$C_M = C_{Eét} + C_{Ep} + \Delta X + \delta + R + \frac{(a-a_0)(q-r)}{(q-a_0)(r-a)} \times (C_{Eét} + C_{Ep})$$

with the following symbols:

C_M : conventional mass of mass M

C_{Eet} : calibration of standard E

C_{Ep} : permanence of standard E

ΔX : repeatability of comparison results

δ : bias from the comparator for the difference of indication (Δ) between M and E

R : resolution of the comparator (no quantification of indications)

q : density of standard E

r : density of mass M

a : air density during comparison

a_o : conventional value of air volume mass = 1.2 kg.m^{-3}

The distributions associated with input variables and the parameters of distributions are summarized in Table 2.

X_i	**Distribution**	**Parameters**			
		Expectation μ	Standard deviation σ	Expectation $x = (a + b)/2$	Half range $(b - a)/2$
C_{Et}	N(μ, σ^2)	100,000.150 mg	0.025 mg		
C_{Ep}	R(a, b)			0	4,33.10^{-2} mg
q	N(μ, σ^2)	8,000 kg/m^3	15 kg/m^3		
r	N(μ, σ^2)	8,000 kg/m^3	335 kg/m^3		
R	T(a,b)			0	0.01 mg
δ	R(a, b)			0	1,118.10^{-2} mg
ΔX	N(μ, σ^2)	0.3 mg	0.018 mg		
a	N(μ, σ^2)	1.16 kg/m^3	0.058 kg/m^3		
a_0	constant	1.2 kg/m^3			

N(μ, σ^2) *means* μ *average Gaussian law and* σ *standard deviation*
R(a, b) *means rectangular law on segment* $[a,b]$
T(a,b) *means isosceles triangle law on segment* $[a,b]$

Table 2. *Distributions of input variables*

Propagation and results

Once the laws associated with input variables are defined, we carry out Monte Carlo simulations to obtain the empirical distribution of the output variable. We have

proceeded to 2.10^5 draws, sufficient number in this application to ensure convergence of results (mathematical expectation, standard deviation and interval at 95%).

In the case of the LPU method, a coverage factor equal to 2 was chosen to calculate the expanded uncertainty. Intervals in Table 3 were calculated on this basis.

In this same table, results obtained by both methods are presented.

In order to compare both methods, MCM and LPU, we used the validation objective criterion defined in [2]. After discussion with metrologists, we considered that the first non-zero number of the typical uncertainty on the conventional mass was the only one that was significant. Consequently, we have deduced the associated tolerance δ value. We then compared this value to distances d_{low} and d_{high} between expanded intervals obtained by both methods.

If we are limited to a first order development of the propagation of uncertainty formula (LPU1), results will not be validated. On the other hand if we consider the first order terms and the fact that we use the propagation of uncertainty formula at order 2, the new results obtained (LPU2) for their part are validated by the Monte Carlo method.

This example thus makes it possible to illustrate one of the main limits of the LPU method; when the mathematical model is not linear, the approximation of development at first order can turn out to be too rough and lead to incorrect results. In this case, it then appears necessary for correct results to develop the propagation formula at order 2 (which requires tedious partial derivative calculations) or to use the Monte Carlo method.

Method	C_M g	$u(C_M)$ g	Smallest interval of confidence at 95% g	d_{low} g	d_{high} g	Validated LPU? ($\delta = 5.10^{-6}$)
LPU1	100.000450	$4.5.10^{-5}$	[100.000359; 100.000541]	$1.6.10^{-5}$	$1.8.10^{-5}$	No
MCM	100.000492	$5.8.10^{-5}$	[100.000345; 100.000559]			
LPU2	100.000450	$5.5.10^{-5}$	[100.000341; 100.000559]	$2.5.10^{-6}$	$2.3.10^{-8}$	Yes

Table 3. *Comparison of results of LPU and MCM methods*

Conclusions

We have list the major principles of measurement uncertainty evaluation and evoked limits and problems encountered during the use of the law of propagation of uncertainties method. Supplement 1 to the GUM provides an alternative method based on propagation of distributions using Monte Carlo simulations. We presented the main steps in this method as well as some of its advantages. An example of

uncertainty evaluation on a thermal power measurement was exposed and has enabled us to see that the numeric Monte Carlo method can validate the results obtained by the law of propagation. We then presented an evaluation example of the uncertainty in mass metrology for which we obtain different results between both methods. We were thus able to illustrate the necessity of using the Monte Carlo method when one or more application conditions of the law of propagation of uncertainty are no longer present.

Acknowledgements

The authors wish to thank Gérald Perrin, author of the BOOST software, for his contribution to this study.

References

[1] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML, *Guide to the Expression of Uncertainty in Measurement*, 2nd ed., ISBN 92-67-10188-9, 1995.

[2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and *OIML, Evaluation of measurement data – Supplement 1 to the Guide to the Expression of Uncertainty in Measurement-Propagation of distributions using a Monte Carlo method, draft*.

[3] Bédiat N., "Méthode numérique de propagation des incertitudes de mesure (Méthode de Monte Carlo)", NTV 06/022, Internal Technical note CETIAT, 2006, 19 pages.

[4] Gosset A., "Comparaison de masses étalons par substitution. Formules et incertitudes", Internal 02 Technical note LNE\CMSI

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Monte Carlo

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Limits of the uncertainty propagation: Examples and solutions using the Monte Carlo Method

Martin Müller, Marco Wolf,
Dr. Matthias Rösslein, Prof. Walter Gander

ETH Zürich, Switzerland; Empa St. Gallen, Switzerland
muellema@inf.ethz.ch

ABSTRACT: The Guide to the Expression of Uncertainty in Measurement (GUM) [1] provides internationally agreed recommendations for the evaluation of uncertainties. It introduces the *Law of Propagation of Uncertainty* to determine the combined standard uncertainty. In a number of cases the application of this method can lead to physically senseless results. Therefore the first supplement to the GUM recommends the use of the Monte Carlo Method. We compare these two methods by means of two examples and show in addition how correlations should be handled using the Monte Carlo Method.

Introduction

The GUM uncertainty framework (GUF) [1] is nowadays the standard procedure to calculate uncertainties of measurements. Within that framework, the *Law of Propagation of Uncertainty* is thus applied using a linear approximation of the measurement equation to determine an estimate and its associated standard uncertainty for the value of the measurand. The procedure is based on representing influence quantities of a model in terms of their best expectations and standard deviations and is suitable for numerous, but by far not all, uncertainty evaluation tasks met in practice. As the requirements for the usage of the GUF are often not fulfilled the Joint Committee for Guides in Metrology (JCGM) has drafted a first supplement to the GUM (GS1) [2]. In this supplement the Monte Carlo Method (MCM) is recommended to propagate the distributions through the equation of the measurement even if the model is strongly non linear, or the input distributions are not well represented by their first two moments. In such cases the usual uncertainty propagation can lead to useless results, whereas the MCM uses the full information to calculate the probability density function of the output quantity. Example 1 of this paper illustrates these facts by using a simple model equation with two input quantities to calculate the uncertainty and it shows how the uncertainty propagation fails, whereas the MCM provides a reasonable probability density function. In a second example we perform a multiplication of two complex valued quantities and point out how correlations between the real and imaginary part of the output quantity arises. Based on this example we explain the difficulties of using the correlation coefficient for building up dependencies in

measurements and explain the terms, logical and statistical correlation. In the end we show how statistical correlations can be described through logical correlations and therefore avoid the creation of correlated random numbers for the use of MCM.

Basics of calculating measurement uncertainty

Aim of the calculation of the measurement uncertainty is to determine a probability density function (PDF) for the output quantity given the PDFs of the input quantities and the measurement formula. From the PDF of the output quantity characteristic values like the mean value or the standard deviation can be calculated. Figure 1 illustrates this procedure for three input quantities.

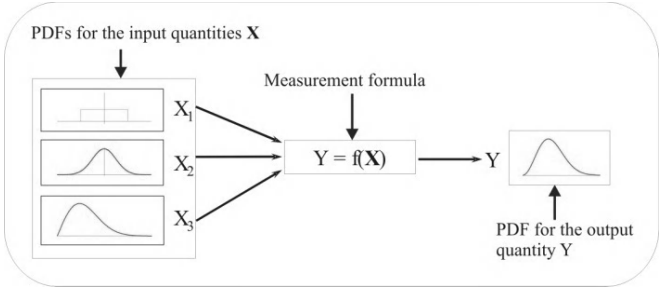


Figure 1. *Measurement process*

Writing this relation as a formula we receive the Markov equation:

$$g_{Y(\eta)} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_{X_1,\ldots,X_N}(\xi_1, \ldots, \xi_N) * \delta\big(\eta - f(\xi_1, \cdots, \xi_N)\big) d\xi_1, \cdots, \xi_N. \tag{1}$$

Formula (1) represents the calculation of the PDF for the output quantity $g_Y(\eta)$ which we obtain by summing all values of the joint distribution of the input quantities $g_{X_1,\ldots,X_N}$ that correspond to the model formula $f(\xi_1, \ldots, \xi_N)$. The values according to the model formula are selected by the Dirac function δ. The calculated PDF $g_Y(\eta)$ represents all the given information about the measurement process. From this PDF the best estimate y and the associated standard uncertainty $u(y)$ can be calculated:

$$y = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_{X_1,\ldots,X_N}(\xi_1, \ldots, \xi_N) * f(\xi_1, \cdots, \xi_N) d\xi_1, \cdots, \xi_N \tag{2}$$

$$u^2(y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_{X_1,\ldots,X_N}(\xi_1, \ldots, \xi_N) * (f(\xi_1, \cdots, \xi_N) - y)^2 d\xi_1, \cdots, \xi_N \tag{3}$$

As the multiple integral (1) cannot be evaluated analytically in most instances, different methods are used to approximate it. The two most important, the GUM Uncertainty Framework and the Monte Carlo Method, are presented and compared in the next sections.

GUM Uncertainty Framework (GUF)

In 1995 the International Organization for Standardization (ISO) published the Guide to the Expression of Uncertainty in Measurement (GUM) [1]. Within that guide a step by step procedure for the evaluation of the measurement uncertainty is given called the GUM Uncertainty Framework (GUF). This procedure is based on the so called *Law of Propagation of Uncertainty* or uncertainty propagation. The law uses a first order Taylor series expansion of the model equation and propagates the best estimates and standard uncertainties of the involved quantities through this linearized model. Therefore the result of the uncertainty propagation is again a best estimate and its associated standard uncertainty.

The first order Taylor series expansion of the model formula is

$$\begin{aligned} f(\xi_1, \cdots, \xi_N) &\cong f(x_1, \cdots, x_N) + \sum_{i=1}^{N} c_i * (\xi_i - x_i) \\ &= f(\boldsymbol{x}) + \boldsymbol{c}^T * (\boldsymbol{\xi} - \boldsymbol{x}) \end{aligned} \tag{4}$$

where the sensitivity coefficients c_i are given by the gradient of the model formula

$$c_i = \frac{\partial f(X_1,\ldots,X_N)}{\partial X_i}.$$

By inserting the Taylor series expansion of the model formula into (3) we find

$$\begin{aligned} u^2(y) = &\sum_{i=1}^{N}\sum_{j=1}^{N} c_i u(x_i, x_j) c_j = \sum_{i=1}^{N} c_i^2 u^2(x_i) \\ &+ 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} c_i u(x_i) r(x_i, x_j) u(x_j) c_j \\ &= \boldsymbol{c}^T \boldsymbol{V} \boldsymbol{c} \end{aligned} \tag{5}$$

i.e. the known formula to calculate the combined standard uncertainty presented in the GUM [1, § 5.2.2]. The degree of correlation between two input quantities x_i and x_j in this formula is characterized by the correlation coefficient

$$r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)} \tag{6}$$

The GUF is based on two main assumptions:

1. the first order Taylor series is an appropriate approximation of the model formula;
2. all the input quantities can be expressed in terms of their first two moments, i.e. the mean value and the standard deviation.

If one of this assumptions is not fulfilled the GUF can lead to physically senseless results and therefore to wrong estimates of the uncertainty budget. As this procedure is suitable for numerous, but by far not all, uncertainty evaluation tasks encountered in practice, the Joint Committee for Guides in Metrology (JCGM) has drafted the first supplement to the GUM to take these problems into account. This supplement recommends the Monte Carlo Method to propagate the distributions through the equation of the measurement.

Monte Carlo Method (MCM)

The Monte Carlo Method is a tool for the propagation of distributions by performing random sampling from the given PDFs of the input quantities. As the PDFs are propagated through the model function, the result of the MCM is again a PDF. This PDF describes the knowledge of the output quantity, based on the knowledge of the input quantities, as described by the PDFs assigned to them. The MCM is based on the repetition of the following two steps:

1. Generate a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ by sampling from the assigned PDFs, as realizations of the input quantities X_i.
2. Evaluate the model formula for the generated random vector to receive one value for the PDF of the output quantity Y.

By repeating these two steps n times we receive n possible values for the output quantity from which we can build an approximated PDF. The quality of the approximation of the PDF depends on the number of runs. Once the PDF for the output quantity is available, that quantity can be summarized by its expectation, taken as an estimate of the quantity, and its standard deviation, taken as the standard uncertainty associated with the estimate.

The MCM uses the full information of the input quantities and the model formula to calculate the PDF of the output quantity. Therefore the MCM can be designated as optimal in a sense that all the given information is used for the determination of the uncertainty budget.

In the following two examples we use the software *MUSE* [3], which we are developing at the ETH Zürich together with the Empa St. Gallen, for the application of the MCM. It is possible to model large measurement processes by defining basic models that represents measurement equipment within that software. Then these models can be combined in a measurement formula to calculate the overall measurement uncertainty. *MUSE* has implemented all distributions described in the GS1. The calculation with complex valued quantities is also possible as the usage of the same distributions in different parts of the calculation to receive a correlation of 100% between these two quantities. For a closer description or a download of the software please visit our project homepage [7].

Example 1. Distance of a point to the origin

In the first example we calculate the distance R of a point given in Cartesian coordinates $\boldsymbol{x} = (X_1, X_2)$ to the origin. The coordinates of $\boldsymbol{x}$ are normally distributed with $X_1 \sim N(0,1)$ and $X_2 = N(1,1)$. The model function of this example is therefore $R = \sqrt{(X_1^2 + X_2^2)}$ where $X_1 = g(\xi_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi_1^2}$ and $X_2 = g(\xi_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\xi_2 - 1)^2}$.

Analytical solution

For this simple example we can calculate the PDF for the output quantity R by calculating the integral

$$
\begin{aligned}
g_R(\eta) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{X_1,X_2}(\xi_1, \xi_2) \\
& * \delta\big(\eta - f(\xi_1, \xi_2)\big) d\xi_1 d\xi_2 \\
= & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi_1^2} * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\xi_2 - 1)^2} \\
& * \delta\left(\eta - \sqrt{(\xi_1^2 + \xi_2^2)}\right) d\xi_1 d\xi_2.
\end{aligned}
$$

The best expectation of this PDF is

$$r = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi_1^2} * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\xi_2-1)^2} * \sqrt{(\xi_1^2+\xi_2^2)} d\xi_1 d\xi_2 = 1.548572$$

with a combined variance of

$$u^2(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi_1^2} * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\xi_2-1)^2} * \left(\sqrt{(\xi_1^2+\xi_2^2)} - r\right)^2 d\xi_1 d\xi_2 = 0.601923.$$

The combined standard uncertainty of r is therefore $u(r) = \sqrt{0.601923} = 0.775837$.

GUM Uncertainty Framework

To calculate the combined standard uncertainty using the GUF we first need to calculate the best estimates of the two involved quantities and their sensitivity coefficients. As the input quantities are both normal distributed their best estimates are their mean values and their standard uncertainties the square root of their variance. Hence $x_1 = 0$, $x_2 = 1$, $u^2(x_1) = 1$ and $u^2(x_2) = 1$. The sensitivity coefficients are $c_1 = \frac{x_1}{\sqrt{x_1^2+x_2^2}} = 0$ and $c_2 = \frac{x_2}{\sqrt{x_1^2+x_2^2}} = 1$. The best estimate for the distance is according to the GUM-Method $r = \sqrt{x_1 + x_2} = 1$. The combined variance of r is $u^2(r) = c_1^2 * u^2(x_1) + c_2^2 * u^2(x_2) = 1$ and thus the combined standard uncertainty $u(r) = 1$.

Monte Carlo Method

The input files for the calculation of the distance of a point to the origin by the aim of the software *MUSE* [3] can be downloaded from our project page [7]. We use the language and environment for statistical computing R [4] for the interpretation of the results. The number of Monte Carlo runs for this calculation is set to 10^7. The expectation value of the resulting PDF, which holds all the information on the measurement result, is 1.548567 with an associated standard uncertainty of 0.775834. Thus the MCM delivers nearly the same values as the analytical calculation.

Quantity	Expectation	Standard uncertainty
X_1	0	1
X_2	1	1
Analytical $\sqrt{X_1 + X_2}$	1.548572	0.775837
GUF $\sqrt{X_1 + X_2}$	1	1
MCM $\sqrt{X_1 + X_2}$	1.548567	0.775834

Table 1. *Result: distance of a point to the origin*

The results of the three calculation methods are summarized in Table 1. In Figure 2 we can see the resulting PDFs of the different methods. As described in the GUM we assign a normal distribution to the output quantity calculated by the GUF. Therefore we receive a PDF that allows negative values, i.e. negative distances of a point to the origin. The resulting PDF of the analytical calculation and the MCM is a non symmetric distribution. This distribution does not allow negative values and therefore represents a physically meaningful description of the output quantity.

This example demonstrates one of the main advantages of the MCM versus the GUF very well. As the result of the GUF is a best estimate with an associated standard deviation, the MCM uses the full information of the given input quantities and the model formula to calculate the PDF of the output quantity.

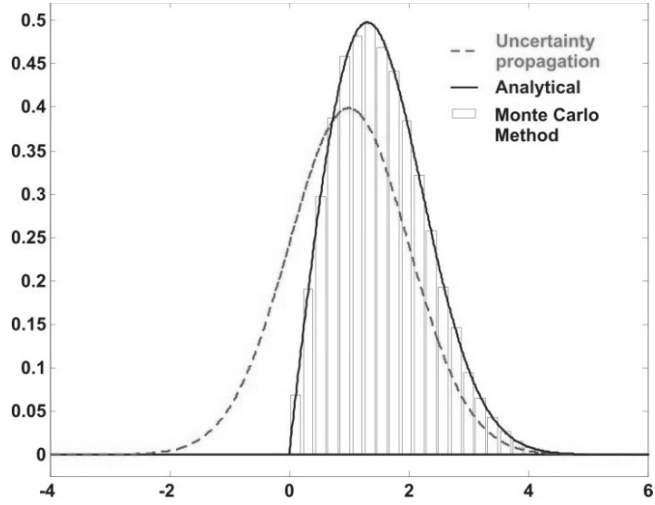


Figure 2. *Resulting PDFs of the calculation of the distance of a point to the origin*

Example 2. Multiplication of complex valued quantities

We perform the multiplication of two complex valued quantities in the second example. A complex valued quantity is described by a two dimensional PDF where one marginal distribution defines the real and the other one the imaginary part. If x is a complex valued quantity, $re(x)$, defines the PDF of the real part and $im(x)$ the PDF of the imaginary part of x.

Let a and b be two complex valued quantities with the marginal distributions $re(a) \sim N(10,1)$, $im(a) \sim N(1,0.1)$, $re(b) \sim N(10,1)$ and $im(b) \sim N(1,0.1)$. In the following we will calculate the uncertainty budget of the product of these two quantities $c = a * b$ with the aim of the uncertainty propagation as well as the MCM.

We use the software GUM++ [5] for the calculation of the uncertainty budget of $a * b$ using the GUF and again *MUSE* for the MCM with 10^7 runs. The results are summarized in Table 2. The calculated mean values of the two methods are very close together. The standard uncertainty obtained by the MCM is slightly larger than that calculated by the GUF. The reason for the difference is caused by the fact that the resulting PDFs for the real and imaginary part given by the MCM are again skew distributions because the operator in the measurement equation is a multiplication. The approximated PDFs are displayed in Figure 3.

An important aspect of this example lies in the fact that the complex multiplication mixes the real and imaginary parts of the input quantities. As the imaginary and real parts of the two input quantities effect the real as well as the imaginary part of the resulting PDFs, we receive a correlation between the real and imaginary part in the result. The covariance calculated by the GUF is 19.8 where the MCM delivers a covariance of 19.80917 which corresponds to a correlation coefficient of approximately 0.6966, i.e. a correlation of nearly 70%. In the next section we will have a closer look on how to handle correlation coefficients using the MCM.

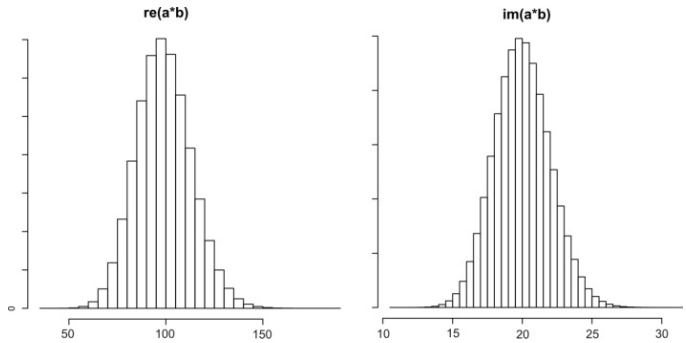


Figure 3. *Resulting distributions of complex valued multiplication calculated by MCM*

Correlated quantities using MCM

In many measurement processes the involved quantities are not independent. In the GUF this fact is taken into account by the correlation coefficient. In the formula to calculate the combined standard uncertainty (5) presented in the GUM the correlation between the input quantities x_i and x_j is expressed by the term $r(x_i, x_j) = \frac{u(x_i,x_j)}{u(x_i)*u(x_j)}$, where $u(x_i, x_j)$ describes the covariance between the quantities and is calculated by

$$u(x,y) = cov(x,y) = \frac{1}{n(n-1)} * \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) . \quad (7)$$

If we want to use correlated input quantities using MCM we need to calculate correlated random numbers. Let us assume that the only information about the measurement process is given by the distributions of the quantities and the dependency of some of the involved quantities is described by a correlation coefficient. Then there exist an arbitrary number of joint distributions that accomplish these settings. Figure 4 shows two possible joint distributions of the same settings.

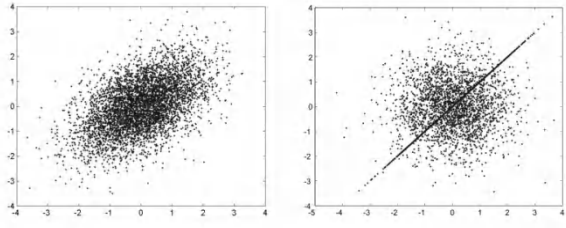


Figure 4. *Correlated random numbers generated by a special procedure described in GS1 (left) and 50% same values (right)*

Quantity	Real part		Imaginary part	
	Expectation	Standard uncertainty	Expectation	Standard uncertainty
a	10	1	1	0.1
b	10	1	1	0.1
$a * b$ (GUF)	99	14.14284	20	2
$a * b$ (MCM)	99.00989	14.17973	20.00089	2.005369

Table 2. *Expectations and standard uncertainties of the input and output quantities of example 2*

The two marginal distributions are both normally distributed with a mean value of 0 and a standard deviation of 1. The correlation between the distributions is 0.5. Thus we have created two sets of random numbers that accomplish the same boundary conditions. The set displayed on the left implies a reasonable dependency between the two quantities whereas the right one does not express a logical dependency but also satisfies the given specifications.

The correlated random numbers displayed in the left picture are generated by the help of a special procedure described in GS1. The details of this procedure can be found in [2, section 6.4.8]. The points plotted on the right hand side of Figure 4 can be subdivided into two parts. One half of the points has a correlation of 0 whereas the other part has a correlation of 1. Therefore the overall correlation of the two fragments is 0.5, i.e. the same correlation as in the set of points displayed on the left. This simple example demonstrates that the correlation coefficient just describes the relationship between two statistical variables. It is not, and should therefore not be taken as, a degree of the causal relationship of two quantities.

In the following we will distinguish between two types of correlations, the logical correlation and the statistical correlation. A statistical correlation arises from multiple measurements of two input quantities where the correlation coefficient between these quantities is calculated as printed in formula (7). If the involved quantities of a statistical correlation are normal distributed we can use a special procedure to calculate correlated random numbers that implies a reasonable causal relationship. If they are not, the only feasible way to generate the needed correlated random numbers lies in a closer description of the causal relationship by a refinement of the model equation. Thereby we can generate a logical correlation out of a statistical correlation. A logical correlation arises from the usage of equal measurement equipment, reference materials or environmental influences in different input quantities. If we take these equal influences in the measurement formula into account the involved input quantities of the model formula are no longer correlated. In general we can write two correlated quantities X_1 and X_2 as two functions of L uncorrelated quantities $\boldsymbol{Q} = (Q_1, \dots, Q_L)^T$. The quantities X_1 and X_2 then depend on some, or maybe all, quantities in $\boldsymbol{Q}$ and we can write them as $X_1 = f_1(\boldsymbol{Q})$ and $X_2 = f_2(\boldsymbol{Q})$.

As an example we use this notation to describe the correlation between the real and imaginary part in the result of the complex valued multiplication of example 2.

Then the uncorrelated quantities are

$$\boldsymbol{Q} = \big(re(a), im(a), re(b), im(b)\big)^T$$

and the functions that describe the causal relationship between the real and imaginary part of the quantity $c = a * b$ are given by

$$f_1(\boldsymbol{Q}) = re(c) = re(a) * re(b) - im(a) * im(b)$$

$$f_2(\boldsymbol{Q}) = im(c) = re(a) * im(b) + im(a) * re(b).$$

Using this procedure we have build up the correlation between the real and imaginary part through uncorrelated quantities.

If we like to use the quantity c of example 2 as an input for a new calculation we can do this by taking the representatives of $re(c)$ and $im(c)$, i.e. the means and standard deviations and the covariance between them. Another, and likely better alternative is to use the previously generated values as input for the new calculation. *MUSE* supports this using a special distribution called *density*. With the help of this distribution it is possible to use previously calculated quantities that are represented by a data file as input quantities. If we do this *MUSE* will read in the saved values in the order they are stored. Therefore we will not lose the correlation between quantities if we reuse them.

Summary

The calculation of the distance of a point to the origin is an example that demonstrates how the inappropriate approximation of the model equation by a Taylor series of order one can lead to physical senseless results using the *Law of Propagation of Uncertainties*. As the result of the Monte Carlo Method is an approximated probability density function it is possible to avoid such misleading results and help therefore to be more accurate in the documentation of the measurement uncertainty.

Building up physical dependencies by the aim of a correlation coefficient can lead to misunderstandings as the correlation coefficient does not describe a causal relationship of quantities. In this paper we distinguish between statistical correlations, i.e. correlations arising out of multiple measurements of two input quantities, and logical correlations, which accrue from the usage of equal influences in different input quantities. As correlations between two quantities can always be described by a number of uncorrelated quantities and two functions that build up the correlated quantities, it is throughout possible to describe statistical correlations by logical correlations. This process demands a deep knowledge about the measurement process and can be extensive, but it is the only way to take the causal relationship implied by the correlation coefficient in a correct manner into account.

With the aid of the software package *MUSE* it is possible to build up physical dependencies by using identical distributions, i.e. same random values in one Monte

Carlo run, in different parts of the modelled measurement system. *MUSE* also supports the user by allowing a logical structuring of the model equation. Therefore it is becoming easier to model large measurement processes and identify dependencies within that processes.

References

[1] Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization, 1995.

[2] Draft-Guide to the Expression of Uncertainty in Measurement Supplement 1: Numerical Methods for the Propagation of Distributions, JCGM Working Group of the Expression of Uncertainty in Measurement, 2006.

[3] *MUSE* – Measurement Uncertainty Simulation and Evaluation, software package, ETH Zürich and Empa St. Gallen, http://www.mu.ethz.ch/muse.

[4] R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2006.

[5] Hall B. D.: *GUM++*: A tool for calculation measurement uncertainty in C++, Industrial Research Limited Report 1303, 2005.

[6] K. –D. Sommer and B. R. L. Siebert, "Korrelation und Messunsicherheit", in *PTB-Mitteilungen*, 116, pp. 129-139, 2006.

[7] MURG - Measurement Uncertainty Research Group, ETH Zürich and Empa St. Gallen, murg@inf.ethz.ch htttp://www.mu.ethz.ch.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Mass

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Weighing small samples on laboratory balances

Arthur Reichmuth

Mettler-Toledo AG
Greifensee, Switzerland

ABSTRACT: There is a great need to weigh very low sample masses due to cost reduction and environmental protection requirements, namely the reduction of waste and toxic substances. During weighing, interactions take place between the balance, the object to be weighed, the environment and the operator. Many of these factors are detrimental to weighing accuracy. Unfortunately, the lower the sample mass, the more noticeable these influences become. This paper presents solutions to overcome or substantially reduce these shortcomings. Consequently, it is possible to reduce the weighing uncertainty for a given sample, or to weigh lower sample masses at a given level of uncertainty.

Introduction

Weighing is about determining the mass of an object. Weighing supports many analytical processes, such as sample preparation, the determination of concentration, or the calibration of comparative analytical instrumentation, such as HPLC[1] or titration.

Various constraints are driving the tendency to prepare and use ever smaller quantities of substances. "Green Chemistry", for example, is a program originally initiated to address pollution by reducing waste, toxic and hazardous substances[2]. While excess or used liquids can often be recycled, excess solids generally cannot, and must therefore be disposed of. Another reason to use small amounts is unstable compounds, so that only the amount actually required for a process should be prepared immediately prior to its use. Yet another reason is the preparation of solutions with concentrations in the sub-milligram per milliliter range. If only a few milliliters of such a solution are required, a very small amount of solute is needed to prepare it. Last but not least, primary substances may be expensive. For example, a few milligrams of a certified reference substance may well cost several thousand dollars. An economic use of such substances is prudent.

1 High performance liquid chromatography.

2 www.epa.gov/greenchemistry.

Limits of Weighing

Depending on the application, balances with various capacities and readabilities are used in the laboratory. Precision balances with readabilities of 1 mg or larger have capacities from several hundred grams up to kilograms. The traditional (macro–)analytical balance with a readability of 0.1 mg and a nominal capacity of 200 g is the workhorse of laboratory weighing. In recent years, the semi-micro balance, with a readability of 0.01 mg, has become more popular (Fig. 1). In the wake of ever smaller amounts, the same can be said of the micro balance with a readability of 1 µg and a capacity of several tens of grams (Fig. 2). These balances with their high relative resolution of ten or more million digits allow very small amounts to be weighed directly in large containers.

Figure 1. *Semi-micro balance (model XP205). Readability: 0.01 mg; capacity: 200 g; resolution: 20 million digits (0.05 ppm per digit); repeatability: < 0.03 mg at full capacity*

Weighing Accuracy

Weighing uncertainty is mainly caused by the finite repeatability of a balance. Nonlinearity [3] and eccentricity [4] also contribute to uncertainty, although generally at a lower level. Weighing bias is mainly brought about by air buoyancy [5], and possibly, but to a much lesser degree, by the sensitivity offset of the balance [6].

3 Nonlinearity is the deviation of the characteristic curve of the balance from the straight line through the points of zero load and full capacity.

4 Eccentricity, or eccentric load deviation, is the difference between the readings for the same load when it is placed eccentrically on the weighing pan and in the center.

5 The buoyant force is caused by, and equal to, the weight of air which is displaced by the object being weighed.

6 For a comprehensive discussion of weighing accuracy see [1].

Most of today's laboratory balances have a built-in reference mechanism which allows the sensitivity to be adjusted easily on-site without the intervention of the operator. Air buoyancy is more of a problem. One way to deal with it is to perform the weighing in a vacuum – hardly the method of choice for most applications. Another way is to correct for it. To do this, the densities of the object being weighed and the air must be known[7]. To give an idea of the effect of buoyancy: an aqueous solution being weighed suffers a weighing bias of about -0.1%; a substance with a density of 2 g/cm^3 about -0.05%. The weighing uncertainty and bias that can typically be achieved with semi-micro and micro balances are shown in Figure 3.

Figure 2. *Micro balance (model XP56). Readability: 1 µg; capacity: 50 g; resolution: 50 million digits (0.02 ppm per digit); repeatability: < 6 µg at full capacity*

When weighing small samples, the weighing accuracy is essentially determined by the sample repeatability of the balance; depending on the density of the sample and the accuracy required, air buoyancy may be ignored. Sometimes, the accuracy for certain weighing procedures is prescribed by regulations such as the United States Pharmacopeia (USP), or limits described in standard operating procedures of

7 Air buoyancy is dependent on the density of the object ρ and the density of air ρ_a (default value 1.2 kg/m^3). From the weighing value W displayed by the balance, the mass m of the object can be calculated with the formula $m = \frac{1-\rho_a/\rho_c}{1-\rho_a/\rho} W$, where ρ_c is the density of the reference weight used to adjust the sensitivity of the balance (usually 8000 kg/m^3). See also [2].

GLP (Good Laboratory Practice) regulated industries. The lowest sample mass that can be weighed on a balance, while ensuring that the weighing result still complies with the required accuracy, is called the 'minimum weight'. This amount depends on the required uncertainty and the repeatability of the weighing[8]; it can also be derived from Figure 3. A small standard deviation of repeatability is equivalent to a low minimum weight.

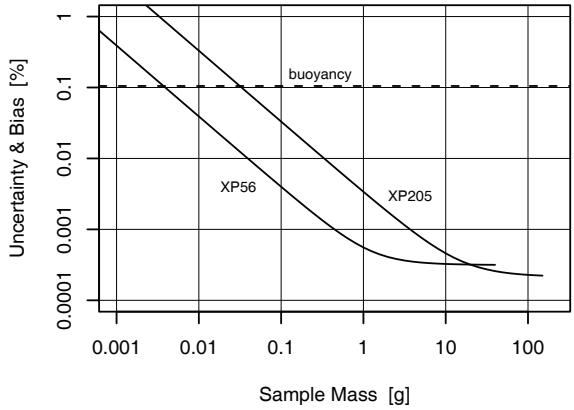


Figure 3. *Relative weighing uncertainty versus sample mass, estimated from balance specifications. The diagram shows typical sample uncertainties (expanded by k = 3; valid for compact objects only) for a semi-micro balance (model XP205, with 50 g tare load) and a micro balance (model XP56, with 10 g tare load). It also shows the magnitude of the relative bias caused by air buoyancy for samples with a density of 1 g/cm^3 (dashed line). For denser samples, the bias decreases approximately by the factor by which the sample is denser*

Weighing Shortcomings

During weighing, the balance, the object, the environment and the operator interact in various ways. This interaction may influence the weighing result. The lower the sample mass, the more significant this influence becomes. Although the prime interest of weighing is to determine the mass of an object, a balance measures the weight force of the object instead. Unfortunately, in a practical weighing setup, there will be spurious forces present which are also measured by the balance and are inadvertently and erroneously interpreted as weight forces. This impairs the weighing result. As a consequence, weighing small amounts on balances with readabilities in the microgram range is a challenge in itself.

8 $m_{\min} = \frac{k}{U} s_{\mathrm{RP}}$

U required (expanded) uncertainty
k expansion factor
s_{RP} repeatability (standard deviation)

Air Drafts

The most obvious environmental influence, and the most prevalent, is air drafts. Air velocities in the sub-centimeter per second range exert forces onto the weighing pan and the object being weighed which are sufficient to affect the weighing. Voluminous objects or objects with a large surface area are especially prone to this effect. Analytical balances with a readability of 0.1 mg and lower are therefore equipped with a draft shield to reduce air velocity. Some balances have grid-shaped weighing pans to reduce their sensitivity to air drafts (Figure 4). The insensitivity to air drafts becomes even more important when a balance is operated in a fume hood or safety workbench with forced air flow.

Figure 4. *Grid-shaped weighing pan. Due to the reduced surface area, air drafts exert less drag on the weighing pan*

Temperature Difference

A temperature difference between the object being weighed and the ambient air is another influence. A warmer object eventually warms up its surrounding air. In turn, this warmer air layer around the object flows upward along the object, creating a spurious drag which reduces the weight reading. If the warmer object is a vessel, the air contained inside the vessel is lighter than the surrounding air. This causes a buoyant force which further reduces the reading. Both influences affect the weighing accuracy, i.e., they introduce bias (drift) and the repeatability of the reading deteriorates. The larger the object or the temperature difference, the greater the effect will be. Hence, a correct reading is unlikely to be obtained when weighing an unacclimatized object.

Power Dissipation

The inevitable power dissipated within an electronic balance warms up the air contained within the draft shield by several tenths of a degree Celsius. The draft shield is opened when the object to be weighed is loaded onto the balance. Hence, the warm air escapes from the weighing chamber and is replaced by cooler air

streaming in. This air draft causes an uplift, reducing the weight reading. Moreover, cooler air is denser. This increases the buoyant force on the object being weighed, thereby lowering the weight reading. Once the draft shield is closed again, the air inside the draft shield slowly warms up, causing the air buoyancy to decrease. Depending on the situation, and especially if large volume containers are involved, this puts the weight reading into a transient with an amplitude up to a milligram, from which it will recover only after considerable time (Figure 5).

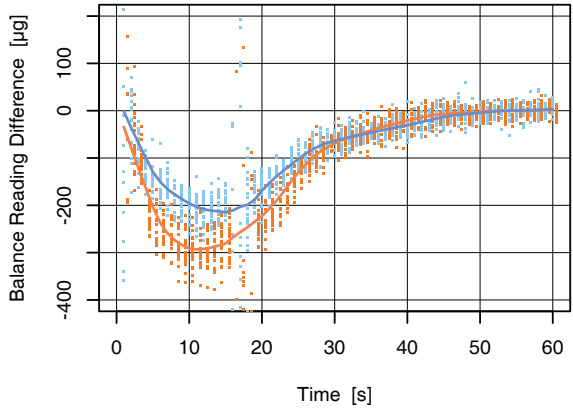


Figure 5. *Semi-micro balance (model XP205) with a 100 ml volumetric flask on the weighing pan. The change in the reading with the draft shield open (0…18 s) and closed (18…60 s) is shown. Multiple sequences are overlaid: the lower dots are readings obtained with the Peltier cooler inactive, the upper dots are with the cooler active (Fig. 10). The lower and upper line represent the corresponding smoothed values. The negative drift is caused mainly by air draft and increased air buoyancy, since cooler air flows into the weighing chamber during the period when the draft shield is open, causing additional uplift. Once the draft shield is closed again, this effect dissipates*

Heat Radiation

Another influence related to temperature is heat radiation. Both the object being weighed and the balance should be protected from direct heat radiation, such as incandescent lamps or other heat sources, as this may warm up the air inside the draft shield or the object to be weighed. Even the radiation from an operator's body can be sufficient to affect the weighing value of balances with readabilities in the microgram range and below (Figure 6).

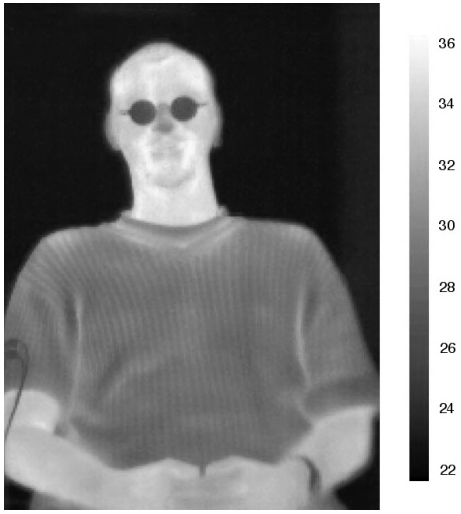


Figure 6. *Surface temperature of the upper part of a human body (temperature scale: lower end 22°C, upper end 36°C)*

Electrostatic Influence

Electrostatic charges present another challenge in the context of high-resolution weighing. Containers with low electrical conductivity, such as laboratory vessels made from borosilicate glass or plastic materials, are notorious for picking up electrostatic charge when they are handled. By way of electrostatic induction, this charge gives rise to an induced charge which accumulates on the surface of the balance console[9]. As the polarity of this induced charge is always of the opposite polarity from the originating charge, they mutually attract one another (Figure 7). Depending on the situation, the spurious electrostatic force can range from parts of a milligram to parts of a gram. This may well thoroughly corrupt the weighing result (Figure 8).

9 I.e., the non-movable, electrically conductive parts of the balance.

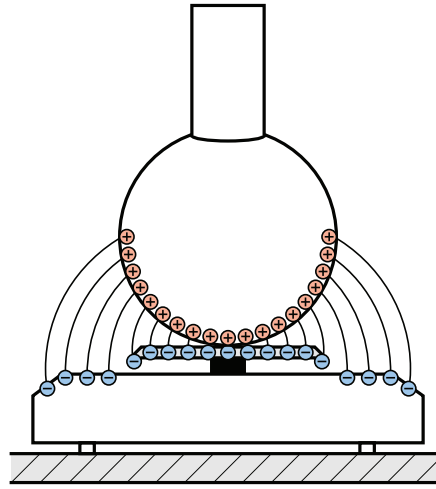


Figure 7. *Electrostatic charge on the vessel accumulates a charge of opposite polarity on the weighing pan and the console (so called induced charge). Therefore, the two charges attract one another. While the field lines ending on the weighing pan do not impair the weighing, those ending on the balance console introduce a spurious force which affects the reading*

Weighing Vessels

The magnitude of weighing inaccuracy caused by air draft and temperature difference depends on several properties of the object to be weighed. As already stated, voluminous objects with large surface areas and low densities tend to be influenced to a greater extent than small, compact and dense objects. Figures 9a and 9b show the results of a series of tests carried out with various standard glass vessels. Sample masses were weighed together with volumetric flasks and glass beakers of different volumes on a semi-micro and a micro balance. These measurements reveal a positive correlation between the repeatabilities of the sample value, and the surface area and volume of the vessels. As a rule of thumb, weighings should thus always be carried out with the smallest vessel possible.

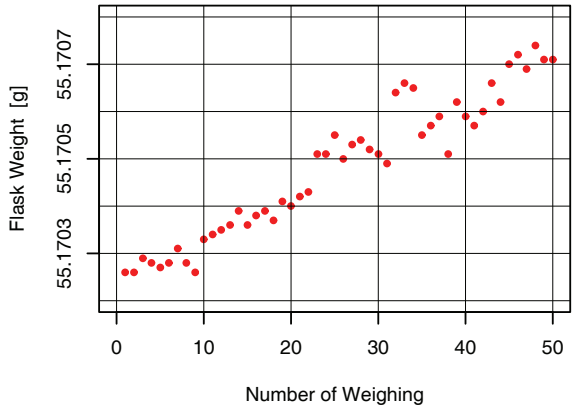


Figure 8. *Multiple weighing results obtained with a 100 ml volumetric flask on a semi-micro balance. The flask, initially uncharged, picks up electrostatic charge during repeated weighings. Putting the flask onto and removing it from the balance causes a continuous buildup of electrostatic charge on the base of the flask, where it touches the workbench and the weighing pan (see Figure 7). (No measures against electrostatic charge were taken during these test weighings.) The resulting electrostatic attraction between the flask and the balance console leads to a drift in the reading*

Pushing the Limits of Weighing

Air movements near the object to be weighed and the weighing pan can be reduced in several ways. One is to keep the object to be weighed in a confined space of minimal volume, thereby reducing air currents in its vicinity. Measures to achieve this are to fit a second, inner draft shield, as shown in Figure 2, or to partition the volume inside the draft shield with a glass plate, as shown in Figure 1. These accessories also allow objects to be stored within the draft shield, which is a convenient way to acclimatize them before they are weighed.

The lower the temperature increase of the air inside the draft shield relative to the ambient air, the better the weighing performance. The draft shield on many semi-micro balances, and most micro balances, is coated with a heat-blocking layer to reduce the influence of external heat radiation. In addition, high performance balances are sometimes equipped with a cooler (Peltier element) mounted on the rear (Figure 10), extracting the majority of the dissipated heat. This lowers the internal temperature rise of the balance. Weighings with a smaller bias, a lower standard deviation of repeatability and a shorter settling time are thus possible.

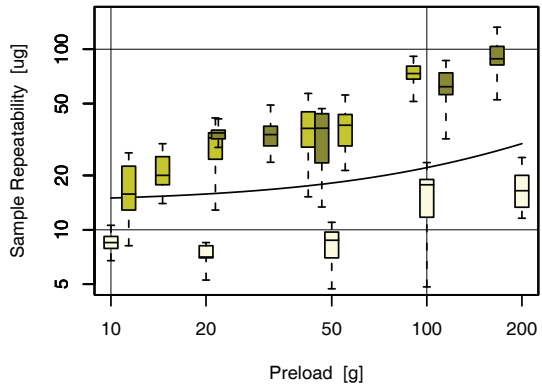


Figure 9a. *Boxplot* [10]*, showing the repeatability of samples, weighed together with vessels on the weighing pan, versus the tare mass of the vessels (preload). Data was obtained from 50 weighings with each type of vessel and a 10 g weight as sample substitute on a semi-micro balance (model XP205, see Figure 1). The following preloads were used: volumetric flasks of 5, 10, 25, 50, 100 and 250 ml (, from left to right), glass beakers of 25, 50, 100, 250 and 600 ml (). The repeatability with compact preloads, i.e., weights of 10, 20, 50, 100 and 200 g (), as well as the guaranteed repeatability specification (solid line), are shown as a reference*

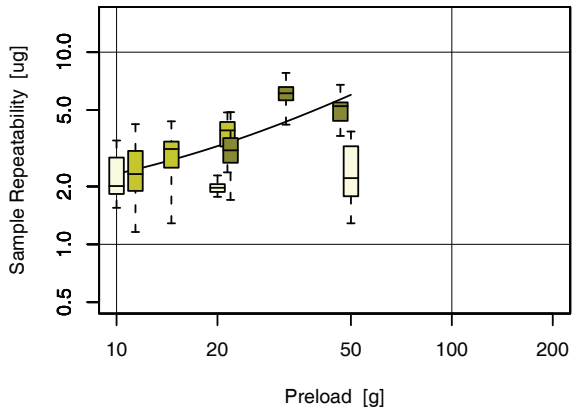


Figure 9b. *Boxplot, showing the repeatability of samples, weighed together with vessels on the weighing pan, versus the tare mass of the vessels (preload). Data was obtained from 50 weighings with each type of vessel and a 1 g weight as sample substitute on a micro balance (model XP56, see Figure 2). The following preloads were used: volumetric flasks of 5, 10 and 25 ml (, from left to right), glass beakers of 25, 50 and 100 ml (). The repeatability with compact preloads, i.e., weights of 10, 20 and 50 g (), as well as the guaranteed repeatability specification (solid line), are shown as a reference*

10 A boxplot depicts the median of a sample, its first and third quartile (box), as well as the (approx.) 4th and 96th percentile (tails).

While the sample repeatabilities obtained in the presence of compact preloads are well below the guaranteed specification, which is valid for compact loads, the repeatabilities in the presence of vessels are much larger, caused by their large size and volume, as stated in the text.

Direct Sample Weighing into Vessels

There are standard practices for weighing solid or powdery substances, namely to use a weighing paper or a small cup. These methods are advantageous from the point of view of obtaining a high level of weighing accuracy. Their disadvantage lies in the potential sample transfer error. To avoid it, the paper must be weighed a second time after the sample has been transferred in order to detect any residue. Similarly, a cup must be rinsed several times to assure the complete transfer of the sample. Both methods are quite elaborate and not necessarily foolproof. Weighing the sample directly into the final container makes these operations obsolete; by definition, there is neither transfer loss, nor contamination. However, the size of the container is detrimental to the weighing accuracy, as demonstrated.

Figure 10. *Peltier unit of an analytical balance. This unit cools the back of the balance, thereby reducing the increase in air temperature inside the draft shield which is caused by the electric power dissipated within the balance. See also Figure 5*

To successfully weigh directly into a vessel and obtain a high degree of weighing accuracy, the influence of the vessel must be reduced. One way to achieve this is to keep the draft shield closed once the vessel has been placed onto the weighing pan. Instead, the sample is passed through a small, adjustable opening in the draft shield. The neck of the vessel is held near this opening with the aid of an adjustable weighing cradle. This allows the sample to be introduced directly into the vessel without opening the draft shield (Figure 11).

Figure 11. *Introduction of a sample directly into a vessel on a micro balance. The powdery substance is introduced into the volumetric flask through a small, adjustable opening (slot on the right-hand side) of the dedicated draft shield ("MinWeigh Door")*

This has several advantages. Firstly, there are virtually no air drafts present inside the draft shield when weighing the sample, as they have already dissipated. Therefore, despite the size and large surface area of the vessel, the resulting standard deviation of the sample re- peatability is considerably lower than the one obtained with the conventional procedure by which the entire draft shield is opened to add the sample (Figure 12a and 12b). Secondly, the sample weighing time is reduced for the same reason. Consequently, weighings will not only be more accurate, but they can also be carried out in less time.

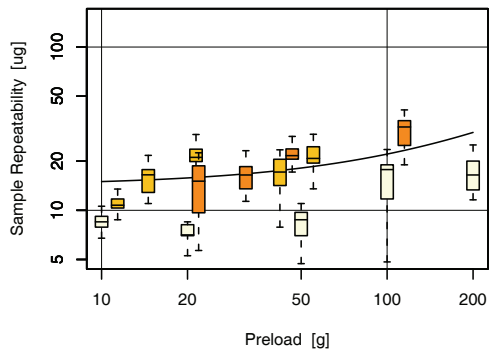


Figure 12a. *Boxplot, showing the repeatability of samples, weighed together with vessels on the weighing pan, versus the tare mass of the vessels (preload). Data were obtained from fifty weighings with each type of vessel and a 10 g weight as sample substitute on a semi-micro balance (model XP205, see Fig. 1) equipped with a draft shield dedicated to weigh directly into vessels ("MinWeigh Door", similar to the one shown in Fig. 11). The following preloads were used: volumetric flasks of 5, 10, 25, 50 and 100 ml (250 ml is too large to fit) (*, *from left to right), glass beakers of 25, 50, 100 and 250 ml (600 ml is too large) (*). *The repeatability with compact preloads, i.e., weights of 10, 20, 50, 100 and 200 g (*), *as well as the guaranteed repeatability specification (solid line), are shown as a reference*

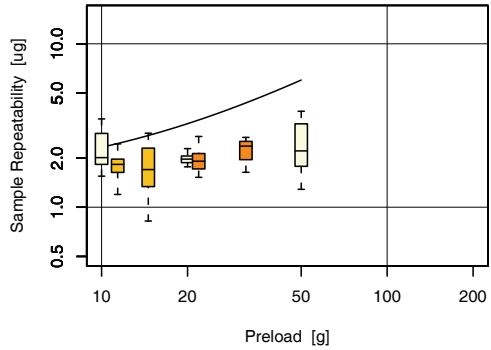


Figure 12b. *Boxplot, showing the repeatability of samples, weighed together with vessels on the weighing pan, versus the tare mass of the vessels (preload). Data were obtained from fifty weighings with each type of vessel and a 1 g weight as sample substitute on a micro balance (model XP56, see Fig. 2) equipped with a draft shield dedicated to weigh directly into vessels ("MinWeigh Door", see Fig. 11). The following preloads were used: volumetric flasks of 5 and 10 ml (25 ml is too large to fit) (, from left to right), glass beakers of 25 and 50 ml (100 ml is too large) (). The repeatability with compact preloads, i.e., weights of 10, 20 and 50 g (bright yellow), as well as the guaranteed repeatability specification (solid line), are shown as a reference.*

These test measurements show that although the samples are weighed directly into (large) vessels, it is nevertheless possible to obtain considerably lower repeatabilities when the draft shield is kept closed while the sample is introduced and weighed. For instance, take the 100 ml volumetric flask which has a mass of slightly over 50 g. Weighing the sample conventionally by opening and closing the draft shield leads to a median value of the repeatability in the order of 40 µg (Fig. 9a), while using the "MinWeigh Door" setup reduces this value to around 20 µg (Fig. 12a). Similar reductions are obtained with the other vessels.

Weighing Accessories

Various shapes of containers and vessels are used in everyday laboratory life, some of which are difficult to place onto the weighing pan. A series of adapters which can be fixed onto the weighing grid simplify the placement of such containers. Test tubes, for example, can be placed in small baskets (Figure 13) that hold tubes and similar vessels securely on the weighing pan. Should the application require the withdrawal of the tube between the tare and gross weighing, the basket will keep the tube in the same position relative to the weighing pan once the tube is replaced. This prevents potential eccentric load errors. Moreover, since the tubes are kept at an angle, the apparent opening of the tube is larger than with a vertically-placed tube. This makes it easier to dispense low density powders.

Figure 13. *Metal basket, clipped onto the weighing pan, both serves as a holder for test tubes or similar glassware, and acts as a (partial) Faraday cage to shield spurious forces stemming from electrostatic charges on the glass*

Since the basket is made of metal, it is electrically conductive and acts as a partial Faraday cage. Thus, the electric field of charged vessels is mostly shielded by the basket. This avoids spurious forces originating from the electrostatic charge. Another way of combatting electrostatic charges when dosing the sample is to use air ionizers. Such devices ionize air molecules, thereby producing positively and negatively charged air ions which neutralize electrostatically charged objects in their vicinity. A convenient way to ionize air is by means of a high voltage corona discharge (Figure 14).

Conclusion

Generally, analytical weighings are affected by influences that can introduce uncertainty and bias into the result. Such influences include air drafts, temperature differences, electrostatic charges or non-ideal behavior of the balance. The lower the sample mass, the more pronounced the effect on the weighing result.

Figure 14. *Add-on ionizer to eliminate electrostatic charge on weighing objects. A pin electrode (dark area on the right-hand side, half way up) that emits electrically charged air molecules (ions) is incorporated into the back wall of the draft shield. These ions neutralize the trapped charge on the surface of the glass vessel*

From a point of view of working efficiency and weighing accuracy, it is advisable to weigh samples directly into the final container, instead of using an interim receptacle. The direct method comprises fewer handling steps as there is no need to transfer the sample into the final container, and a higher sample accuracy can be achieved as, by definition, no transfer loss occurs. However, large containers have a greater interaction with their environment, which is detrimental to the weighing accuracy of the sample.

Weighing directly into containers and obtaining a high sample accuracy is nevertheless feasible if proper measures are taken. Using a dedicated draft shield with a small opening allows the sample to be introduced without opening the draft shield. Excessive air drafts affecting the container are thus prevented, and the standard deviation of the sample repeatability, as well as any potential bias, are significantly reduced, approaching levels normally observed for weighings with compact weights. A lower standard deviation of repeatability provides a lower minimum weight, which allows the amounts of substances involved in processes to be reduced, resulting in lower costs and less waste material.

Care should be taken when weighing electrically non-conductive objects as they are prone to picking up an electrostatic charge when being handled. This may seriously impair the weighing result. Proper measures to counter this effect are electric shielding with suitable weighing accessories, or the use of ionized air.

References

[1] Arthur Reichmuth: Weighing Accuracy With Laboratory Balances. Proc. 4th. Biennial Conference of Metrology Society of Australia, Oct. 2-4, 2001, Broadbeach (QLD, Au), pp. 38-44.

[2] Randall M. Schoonover, Frank E. Jones: Air Buoyancy Correction in High-Accuracy Weighing on Analytical Balances. Analytical Chemistry, May 1981, Vol. 53, pp. 900-902.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Design and performance of the new Sartorius 1kg-prototype mass comparator for high precision mass determination and research applications

Thomas Fehling[1], Thomas Fröhlich[1], Detlef Heydenbluth[2]

[1]Sartorius AG Göttingen, Germany
[2]Technical University of Ilmenau, Germany

ABSTRACT: The 1kg-prototype mass comparator is a result of the technical collaboration between Sartorius AG/Germany and BIPM/France. Construction and functionality of the 8-position load alternator are based on the known BIPM FB2-technology. The new 1 kg mass comparator has been especially developed for mass comparisons with the national kilogram prototype, and for creating and maintaining the national mass scale with the highest possible accuracy. It permits not only mass comparisons at the 1 kg level but also comparisons of relatively large objects such as 1 kg silicon spheres or buoyancy artefacts.
The complete measurement device is installed inside an enclosed airtight aluminium chamber which can be evacuated to primary vacuum. The weight changing mechanism can be loaded comfortably from an external quick load-lock device. Additional standard vacuum flanges are freely available for measuring sensors, control purposes and electronic or other connectors. The control unit running the control software allows flexible and easy programming of the required measuring sequences. Routine mass calibration as well as sophisticated weighing series can be performed.
A detailed description of the technical and metrological parameters as well as measurement results under airtight an vacuum conditions will be presented. The function of the load alternator and the load-lock device is depicted [5].

Introduction

The development of the 1kg-prototype balance is a collaboration between BIPM/France and Sartorius AG/Germany. When signing the contract, Sartorius intended to include its collaboration partner of many years in the field of basic research in weighing technology, the Technical University of Ilmenau/Germany (TUI). This was a good choice, due to the fact that necessary tasks in research works, in construction and design and even the provision of a suitable mass laboratory have taken place in Ilmenau/Thuringia.

When starting the project in 2000 the main task was to reproduce the FB2 prototype balance from the mass section of BIPM [1][2]. Different NMI's from all over the world would like to use a prototype balance for their own high accuracy mass determination and research applications similar or equal the well known FB2

which is unique. However, what does a reproduction mean, what are the main difficulties, what should be complemented and what did Sartorius expect to do in a different way than the people from BIPM? When discussing all details in the mass lab at BIPM two things quickly became clear. First this was the realization that the 8-position loading device was designed perfectly to obtain the expected performance, but needed to be expanded for the application of measuring 1kg silicon spheres as used for example in the Avogadro experiment. Second, Sartorius as a seller and manufacturer for analytical balances and high accuracy mass comparators should face the challenge to insert their own weighing system specified for the task of a prototype balance like FB2.

The new 1kg-prototype balance is built for use in the following applications:

- use as a prototype mass comparator for connection of primary standards of national metrology institutes (called NMIs for short in the following) to national mass standards (kilogram prototype) of the various countries;
- dissemination of the mass scale of NMIs in the range from 1 kg to 100 g;
- measurement of mass (100 g to 1 kg) within the scope of international comparison;
- measurements (key comparisons) and calibrations for government institutes, calibration laboratories and industry according to the attainable uncertainties of measurement given in the CMC tables of the BIPM;
- experimental determination of the air density by comparative weighing of special buoyancy artefacts in air and under vacuum;
- measurement of the mass of 1-kg silicon spheres (also within the scope of the Avogadro project to determine the Avogadro constant and redefine the kilogram as a unit of mass) [4];
- experiments to study the influence of cleaning procedures and of the effects of sorption and convection on the measurement of mass and to study the long-term stability of mass standards.

The construction of the whole unit is a modular one. The 1 kg-prototype balance consists of 3 main parts:

1. enclosure with integrated load alternator and mass comparator;
2. control unit with PC and drive unit;
3. vacuum pumping unit with turbo-molecular pump.

The design of load alternator and measuring system allows easy exchange of the installed weighing system. This has a favorable affect in case of service or scientific research. The different parts of the 1kg-prototype balance will be presented in detail in the following sections.

Enclosure for airtight and vacuum conditions

Designing the enclosure, several things had to be taken into accounts which are in contrast to the enclosure of the FB2. First of all the whole equipment of load alternator, weighing cell and special loading device had to be included. The diameter of the designed turntable of the load alternator increased to 555 mm due to the fact that silicone spheres of about 94 mm diameter have to be carried. If we follow the design of a usual vacuum enclosure which is typical a cylindrical one, we would get an unacceptable outer diameter of the whole bell. This leads to the chosen square of 840 to 940 mm.

The enclosure consists of four main parts, which are the bottom plate, the base chamber in the middle, the top cover plate and the cover (tower) for the weighing cell. The four parts are connected to each by another HV-compliant with a vacuum seal made of Viton. These parts of the whole enclosure do not need to be dismounted during normal work after first installation. The special type of upper tower is equipped with a 160 mm square quick-lock door to add or remove substitution weights on the upper weighing pan of the balance. The base chamber in the middle is equipped with the same 160 mm square quick-lock device, for loading the weights onto the turntable. For this purpose a special external quick load device has been designed (Figure 1). This construction allows easy loading of weights for fast performance of mass comparisons under airtight and especially vacuum conditions. A sensitive handling of expensive and valuable weights during the loading procedure without any risk of damage is obtained.

The quick load device is equipped with 3 strain gauge systems set up to take the weight after manual loading on the main carrier. The center of gravity of the weight in relation to a centric position on the carousel can be determined with the 3 mass values from the strain gauge systems. During automatic loading procedure the eccentricity of the weight will be adjusted in the x-axis by the movement of the carrier in direction to the carousel, in the y-axis by the rotation of the carousel itself. The effect is a pre-centred weight on the carousel ready for mass comparison procedure.

The enclosure is equipped with several standard vacuum flanges on the base chamber for any type of electrical or mechanical feed through. This allows us to implement additional sensors for measuring temperature or pressure for instance. Furthermore, the enclosure is equipped with a DN100 vacuum flange for connecting the turbo molecular pump in case of evacuation.

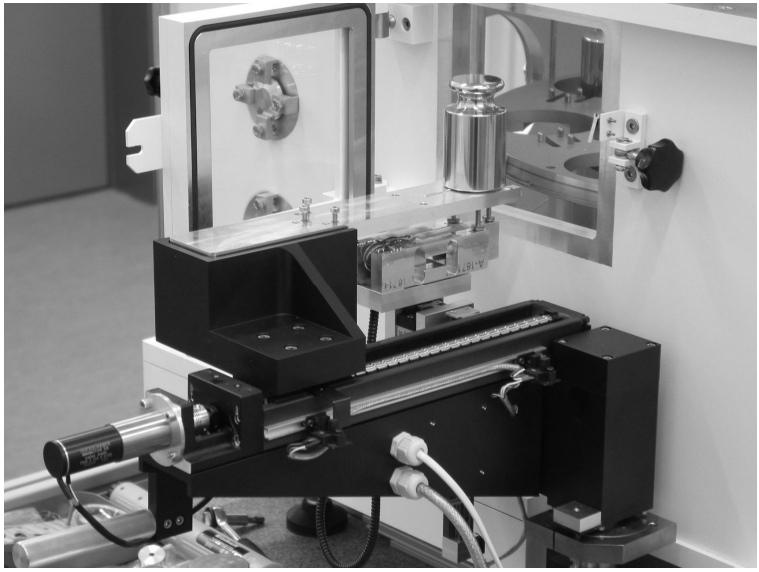


Figure 1. *Quick load device with 1kg weight in front of the vacuum chamber*

The complete enclosure was tested from the manufacturer for high vacuum application. The Helium leakage rate is $\leq 10^{-7}$ (mbar * l) / s. The chosen material for the enclosure is aluminum. This material has the advantage of being absolutely non-magnetic and having better thermal conductivity than stainless steel. For having a stable mechanic setup inside the enclosure even under vacuum conditions we needed to calculate the thickness of the base plate accurate. To approve all calculations and being sure that no influence due to deflection could cause difficulties during measurement, a finite elements analysis (FEA) calculation has been carried out at the TUI.

These FEA calculations have been performed to ensure the mechanical stability of the enclosure under vacuum conditions a well as the temperature gradients inside the chamber under some different thermal disturbances. Inside the chamber, gradients are not larger than some mK in this case. The bending of the base plate under vacuum conditions is shown in Figure 2.

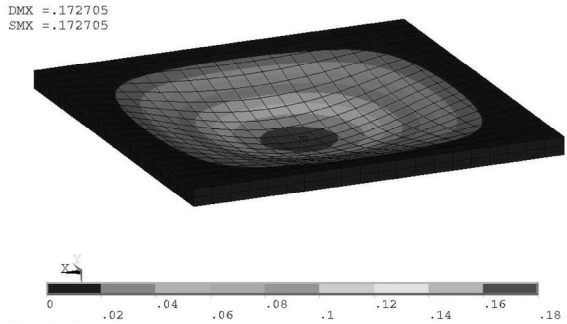


Figure 2. *Bending of the base plate of the chamber under vacuum conditions*

The load alternator

The load alternator (Figure 3) is designed similar to the original FB2 alternator with the main difference in diameter of the turntable. This important increase up to 555 mm is necessary to fulfill the stipulations made to carry spherical weights up to 100 mm diameter on each position. The demand on providing a place for 2 reference weights, 2 buoyancy artifacts and at least 4 test weights is fulfilled with the 8 positions on the turntable. All electrical heat sourcing driving mechanisms of the load alternator are located outside the enclosure. The movement of the turntable is performed by a thin tungsten wire with a diameter of 0.125 mm which is wound up at the end face of the turntable 4 times. This is sufficient for all necessary movements during mass comparison between each of the 8 positions.

Figure 3. *Opened base chamber with turntable*

The used balance

For the required specifications for mass comparison as described in section 6 we need a balance for 1 kg maximum load with a readability of 0.1 µg. The used balance which has to be developed is based on the commercial 1 kg Sartorius Mass Comparator CC1000S-L. This instrument gives us already a readability of 1 µg with a reproducibility of less than 1 µg (Std.-deviation from 6ABBA-cycles). Several things had to be modified for using this electromagnetic force compensation system for the prototype balance. The weighing cell had to be changed for more sensitivity at the electronic sampling device, given by the optical position sensor of the beam. This has been realized by decreasing the spring constant of the used flexures in the universal joint (Figure 5). A special lever (Figure 4) was designed using the knowledge of Sartorius monolithic manufacturing process. This leads to a minimum of movement of the lever caused by material tension with the result of a reproducible positioning after load change.

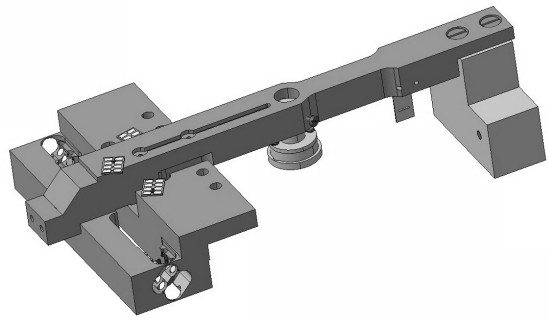


Figure 4. *Lever with coil and counterweight*

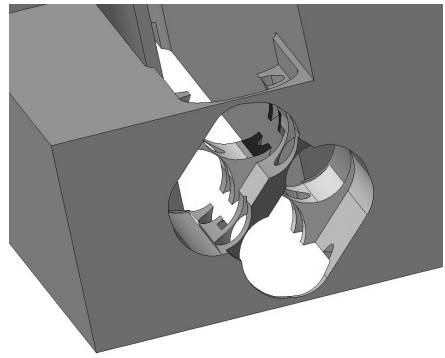


Figure 5. *Monolithic universal joint (fixed)*

The weighing cell also had to be modified for use under vacuum conditions. This means for example that no blind holes are allowed and any kind of electric wire has to be changed into a suitable one with special isolation made of PTFE or Kapton.

Knowing the perfect performance of the FB2 equal arm balance [3] which is more or less depending on a constant load device based on the piezoelectric effect, we decided to install such a constant load device into the Sartorius weighing cell, too. The piezoelectric transducer is directly connected to the vertical weighing axis of the balance coupling device. A location where any force due to stressing by weights is been guided to the balance beam. When changing the weights by the load alternator, the feedback loop of the balance controls the installed piezoelectric transducer in between a small range of the electronic weighing range which is shown in Figure 6. As the desired result the beam is under the same mechanical stress each time. This causes less creeping effects of the material shown by unstable mass values after load change.

Another important difference to the commercial mass comparator is the change from a top loading balance back to traditional suspension device. The used Centermatic pan from the CC1000S-L was no longer necessary. The centering of the weights is performed by the suspension device designed like a pendulum. Special efforts had to be carried out to design a pendulum suspension which works nearly in absence of feedback caused by eccentric loads, otherwise needs to be damped in order to obtain stable weighing results. A special conical coupling device with a flexion spring was designed.

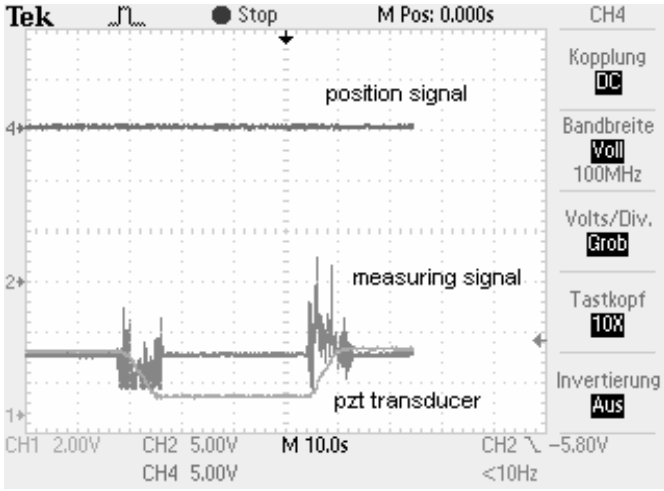


Figure 6. *PZT-transducer signal*

The control unit with application software

The control unit consists of 3 main parts, which are installed in a separate rack. These are the driver motors for all necessary movements of the turntable and the lifting device of the load changer, the high voltage driver unit for the piezo transducer, and the PC equipped with the operating software. The design of the rack can be extended for customer-specific applications so that it is possible to integrate further modules.

The operating software of the prototype mass comparator consists of software for controlling the sequence of mass comparisons and of evaluation software for calculating the measurement results using all the features necessary for this. The structure of the evaluation software is essentially comprised of the following modules:

– determination of the conventional mass by substitution weighing;

– highly accurate mass determination by substitution weighing;

– dissemination of the mass scale.

Specifications

The prototype mass comparator is equipped with a fully automatic load alternator with 8 alternating positions. All 8 alternating positions are suitable without further modification for mass standards, buoyancy artefacts or spherical test objects (e.g., silicon spheres).

The maximum diameter of the weights is 95 mm for cylindrical and for spherical objects on each alternating position. It is possible to position spherical objects with a maximum diameter of 110 mm in each 2nd alternating position if the spherical objects with a diameter ≤ 80 mm are used in the remaining positions on the load alternator.

The maximum height of the weights is 110 mm (also in the form of combinations).

It is possible to load platinum-iridium kilogram prototypes or steel kilogram standards (according to OIML R 111) without tilting or turning them. A novel, integrated loading device on the vacuum chamber makes this possible.

The mechanical effect of the load alternator on the surface of platinum iridium kilogram prototypes, steel kilogram standards and 1 kg silicon spheres is metrologically negligible.

The vacuum systems consists of a vacuum housing and an oil-free pump stand (consisting of a turbo-molecular pump, fore-pump, valves, power supplies and control electronics/control unit). It is suitable for a pressure range of 0.1 Pa to standard barometric pressure (10^5 Pa).

The pressure stability of the vacuum system in the entire pressure range is better than 2 Pa in 24 hrs (with the vacuum pump shut off and all valves closed), disregarding desorption processes within the vacuum chamber.

The metrological requirements and specifications are as follows.

The allowable pressure range in the vacuum housing for maintaining specifications is between 0.1 Pa and standard barometric pressure (10^3 hPa).

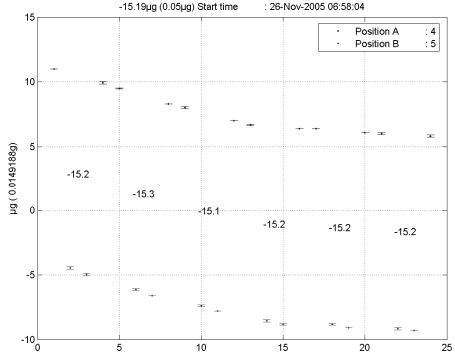
The maximum capacity of the prototype mass comparator is 1,011 g (1 Kg).

The use of external substitution weights enables the mass of any values to be determined in a range from 1 g up to the mass comparator's maximum capacity.

- The electrical weighing range is 2 g.
- The readability is 0.1 µg.
- The repeatability is ≤ 0.1 µg.
- The linearity in the electrical weighing range is ≤ 1.5 µ.
- The stabilization time is ≤ 60 s.
- The sensitivity drift is ≤ 2 ppm/°C.

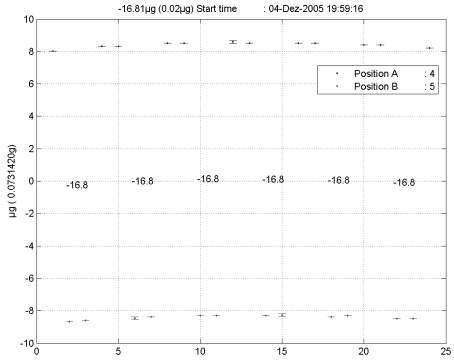
Results

Measuring results are taken under airtight and vacuum conditions as follows:
a) airtight conditions (20°C, 45% r.H., 860mbar)



Mean = -15.19µg Std.Dev. = 0.05µg

b) High vacuum conditions (20°C, $4 \cdot 10^{-6}$mbar)



Mean = -16.81µg, Std.Dev. = 0.02µg

Summary

The new Sartorius 1 kg prototype balance is designed to fulfil high accuracy mass determination between the Pt-Ir reference weight and standard steel weights as well as silicone spheres. The measurements can be carried out under airtight or vacuum conditions. The chosen material of aluminium guarantees a maximum of insensitiveness against magnetic influence on the measurement caused by the enclosure. The enclosure is designed to allow fast measurement procedures after load changes due to small load-lock ports. A special load device takes care of the expensive and valuable weights during the loading procedure.

Acknowledgements

The authors thank the colleagues from BIPM, Dr. Richard Davis and Mr. Alain Picard, for their support and assistance during the long development of a new type of prototype balance as presented. As our cooperation partner these colleagues helped us in a very efficient and successful way to solve all arising mechanical or electronically problems during valuable discussions in Sévres or at the TUI laboratory in Ilmenau with all their excellent knowledge on mass metrology.

References

[1] Picard A., “The BIPM flexure-strip balance FB-2”, *Metrologia* 41 (2004) 319–329.

[2] Quinn T. J., Speake C. C. and Davis R. S., “A 1 kg mass comparator using flexure-strip suspensions: prelimary results”, *Metrologia* 23 87–100.

[3] Quinn T. J., “The beam balance as an instrument for very precise weighing”, *Meas. Sci. Technol*. 3 (1992) 141–59.

[4] Picard, A., Primary mass calibration of silicon spheres *Meas. Sci. Technol*. 17 2540-2544 2006.

[5] Fehling, T., Fröhlich, T., Heydenbluth, D., “The New Sartorius 1Kg-Prototype Balance for High Precision Mass determination”, *NCSLI Conference*, Washington (USA), August 2005

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Series of pressure balances for industrial use

Marcello Caravaggio [1], Gianfranco Molinar Min Beciet [2], Paolo De Maria [3]

[1] Dott. Ing. SCANDURA & FEM S.r.l. E-mail: caravaggio@scandura.it
[2] Istituto Nazionale di Ricerca Metrologica, INRIM. E-mail: G.Molinar@inrim.it
[3] Istituto Nazionale di Ricerca Metrologica, INRIM. E-mail: p.demaria@inrim.it

Introduction

The design and development of a series of pressure balances, operating in liquid media up to 120 MPa and in gas media up to 12 MPa, started during a recent co-operation between Italian company SCANDURA & FEM and INRIM (Italian National Research Institute of Metrology). The details of the project design for the pressure balance in liquid media up to 120 MPa are here presented with particular emphasis on the piston motor drive. The main metrological characteristics of some piston-cylinder units were experimentally determined and some key results are here reported.

Figure 1. *MPA pressure balance*

Design of the pressure balance

The 120 MPa full scale balance is the first of a series of four different units, with different full scales: 120 MPa, 60 MPa, 36 MPa and 12 MPa. For the 120 MPa model a tungsten carbide piston-cylinder was used. The piston-cylinder unit has been inserted into a stainless steel body where two mechanical holes have been foreseen, one for a temperature probe for measuring the piston-cylinder temperature, and the other one to purge the measuring fluid.

The piston-cylinder unit is of the "free deformation" type. The top part of the piston has been integrated into another stainless steel component that supports the mass carrier. After the insertion of the bottom part of the piston into the cylinder, the same is screwed to the body containing the cylinder. In this way, a mechanical assembly, with the following main tasks, will be achieved:

- to allow the piston to flow across the cylinder defining at the same time the full piston stroke;
- to define the points of the force application on the assembly;
- to guarantee the assembly verticality;
- to support the masses carrier;
- to purge the fluid that slowly flows through the piston-cylinder gap;
- to use a probe for the temperature measurement of the piston-cylinder unit.

A stainless steel masses carrier is put on the mechanical component containing the piston and its surface wraps up this component without any mechanical interference with the same. A small extension, at the bottom part of the mass carrier, supports the masses inserted from the top and lying on each others in order to achieve the desired total mass.

The stainless steel masses are of different sizes and shapes in order to realize all the pressure steps between the minimum and the full scale values with a resolution according to the smallest mass in use. In the central part of each mass, a cavity, closed by a lid, has been realized, in which different stainless steel spheres, with different diameters, can be inserted to adjust the mass value to the nominal one. Different pressure balance models will be available; from the simplest model, totally mechanical, to the complete model, with the motor to drive the piston rotation and with all the sensors needed to measure and to compensate errors due to the main influence variables. In this case a microprocessor based on an electrical board collects the data from the different sensors and, combining these data with the constants of the system, gives the pressure value according to the pre-defined

mathematical model. The total mass value is calculated from each mass value, available by the specific calibration certificates and also given to the memory of the electrical board. A barcode identifies each mass and can be read at the time of mass use during the measurements.

The integrated sensors are: a barometer for atmospheric pressure measurement, an ambient temperature probe, a relative humidity sensor, a thermo-resistance for the measurement of the piston-cylinder temperature and of a proximity sensor to collect information about the floating level of the piston.

The motor drive system is designed in such a way that the application of the rotation to the piston is given only when the piston is at its lowest position. When pressure is increased the piston moves up to reach its floating position, where the piston is in its free rotation. To avoid a useless increment of temperature, a proximity sensor detects the piston floating position and conveys this information to an electronic card that automatically switches off the motor. Moreover, to avoid any mechanical interference between the motor system and the piston-cylinder unit, these two parts are totally independent and not in mechanical contact with each other during measurements. The rotation is transferred to the piston by friction of three rubber wheels with the mass carrier. The wheels are put into rotation by a motor, connected to a series of gears and a driving belt. A series of eccentrics allows the positioning of the wheels and maintains the driving belt at the correct tension.

In Figure 2 the motor drive system is shown:

1) motor
2) eccentric system
3) wheel with gears
4) driving belt
5) proximity sensor

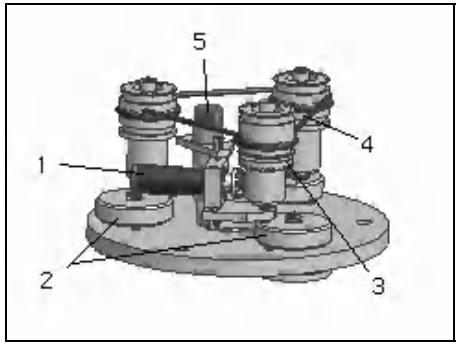


Figure 2. *Motor drive system*

The masses carrier changes its vertical position between the three wheels and this position is detected by the proximity sensor. The information about the vertical position of the piston is also used to calculate corrections due to the difference in height between the reference level and the reference floating level.

Metrological characteristics

After the initial evaluation on three piston cylinder units [4], a more accurate investigation has been carried out on a piston-cylinder unit, here called SCAND3, one of the different assemblies available.

For this piston-cylinder unit the following tests have been performed:

– piston fall rate as a function of pressure,

– rotation speed of the piston and mass assembly as a function of time,

– effective area of the assembly as a function of pressure,

– sensitivity of the cross-floating of two pressure balances.

All the tests related to the piston fall rate and assembly rotation speeds have been performed with the balance isolated on its own.

All the piston fall rate data are collected from a capacitive sensor and have been managed by a LabView software program prepared at INRIM.

Different tests, in different conditions, have been performed in order to experimentally evaluate the piston fall rate and effective area as a function of pressure, considering the relevant uncertainties.

In Figure 3 the behavior of the piston fall rate according to pressure is shown. The expanded uncertainties related to these measurements are considered to be equal to 7 $\mu m \cdot s^{-1}$, a high value that also takes into account the contribution due to temperature variation of ± 3°C around a 20°C reference value.

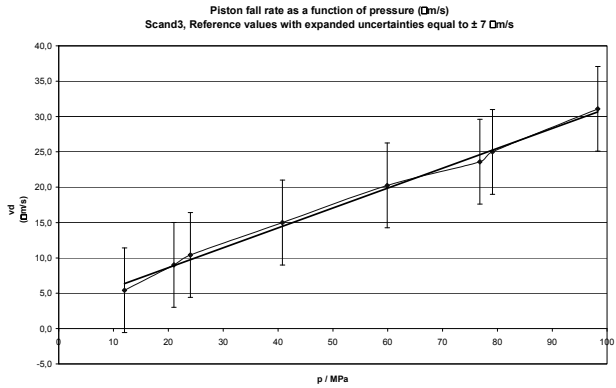


Figure 3. *Behavior of the piston fall rate according to pressure*

For this assembly the behavior of the piston fall rate is linear according to the pressure and so an eventual extrapolation from 100 to 120 MPa would certainly be fully reliable.

The behavior of the reduction of rotation speed is also extremely regular according to time. After 30 minutes the reduction of rotation speed starting from 33.3 rpm is only 25.0 rpm for SCAND3.

This qualitatively represents a proof of the regularity of pressure distribution in the gap of the SCAND3 piston-cylinder and the fact that there is not a large amount of mechanical friction.

In Figure 4 the behavior of rotation speed according to time, for the SCAND3 assembly, is shown.

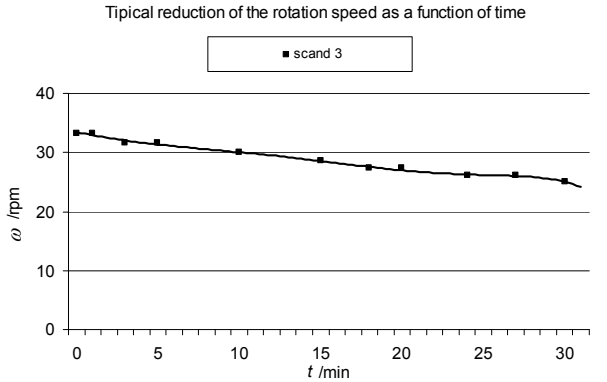


Figure 4. *Behavior of rotation speed according to time*

The effective area of the piston-cylinder assembly has been determined from pressure cross-floating with one of INRIM. pressure national standards and taking into account each influence quantity during the experimental cross-floating.

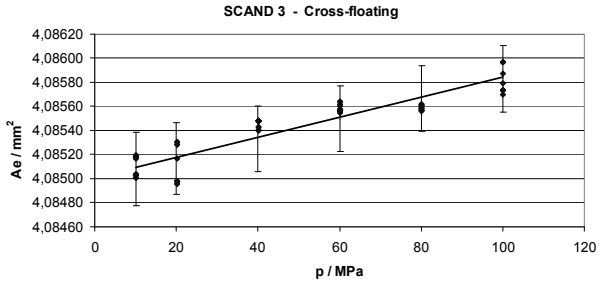


Figure 5. *Behavior of $A_e = f(p)$*

The behavior of $A_e = f(p)$, shown in Figure 5, points out a good linearity of the effective area according to pressure, as expected for a free deformation unit.

The mean value of effective area of the piston-cylinder unit at atmospheric pressure and at a reference temperature of 20°C, is equal to 4.08501 mm^2.

This value, considering its estimated expanded uncertainty of the order of 80 ppm, is in agreement with the effective area value derived from dimensional measurements [1].

The value of pressure distortion coefficient λ is equal to $2.0 \cdot 10^{-6}$ MPa^{-1} and this value has to be confirmed by using the finite element calculation method FEM as previously performed for other pressure balances [2][3]; this work is actually in progress with the support of the Cassino University.

The sensitivity tests have been performed, by cross-floating, comparing the pressures measured by the national standard pressure balance and by the piston-cylinder assembly SCAND3. For this assembly the sensitivity value is always smaller than 12 ppm over the full measuring pressure range (10 to 100 MPa).

Conclusions

– The piston-cylinder unit here presented has an acceptable geometrical quality for piston-cylinders of pressure balances for industrial uses (the average radial clearance span from 6 μm to 0.5 μm in its central part);

– the piston fall rate, measured at the pressure of 100 MPa, using as pressurized medium the di-ethyl-hexyl sebacate, is about 30 $\mu m \cdot s^{-1}$;

– the quality of the piston-cylinder assembly is confirmed by the regular behaviors of rotation speed according to time. Starting at approximately 30 revolutions per minutes (rpm), after 30 minute, the rotation speed of the assembly is 25 rpm;

– the effective area behaviors (nominal value is 4 mm^2) according to pressure are strictly linear as expected for a free deformation piston-cylinder unit;

– the pressure sensitivity, derived from cross-floating against a national standard, was found to be better than 12 ppm in all pressure range from 10 MPa to 100 MPa.

The relative expanded uncertainty (k = 2) of pressure measurements in all the measurement range of this type of pressure balances, is always between 70 and 100 ppm (1 ppm = 10^{-6}) of the applied pressure.

References

[1] G. Molinar *et al*., "Calculation of effective area A_0 for six piston-cylinder assemblies of pressure balances. Results of the EUROMET Project 740", *Metrologia*, **42** (2005), S197-S201.

[2] F. Pavese and GF. Molinar, *Modern Gas-Based Temperature and Pressure Measurements*, Plenum Press, New York (USA), 1992

[3] Buonanno G, Dell'Isola M, Iagulli G and Molinar G, "A finite element method to evaluate the pressure distortion coefficient in pressure balances" *High Temp. – High Pressures* **31** 131-43, 1999.

[4] Caravaggio M, Molinar Min Beciet GF, De Maria P, Pressure balance for industrial application up to 120 MPa, AdMet06, Delhi (India), 2006, paper # 44 on CD proceedings, 456-459

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Fully automatic mass laboratory from 1 mg up to 50 kg – robots perform high precision mass determination

Christian Buchner

BEV- Bundesamt für Eich- und Vermessungswesen
(Federal Office of Metrology and Surveying), Austria
christian.buchner@bev.gv.at

ABSTRACT: In order to meet the requirements for a state-of-the-art metrology institute and in an endeavour to build a fully automatic mass laboratory, the Austrian Federal Office of Metrology and Surveying (BEV) has developed and realized in cooperation with Sartorius AG in Goettingen and the Vienna University of Technology three different handling systems for automatic testing weights on high-precision mass comparators.
Fully automatic system for alternating weights on mass comparators with different weighing capacity have been realized; one from 1 mg up to 10 g, one from 10 g up to 1 kg an the third one from 1 kg up to 20 (50) kg using two comparators. The resulting load alternator "robot" holds between 10 and 80 test weights and a corresponding number of reference standard weights. All weights are transported without pallets. The systems are based on computer control and linear drive trains, and are used at the BEV to disseminate mass and for the calibration and verification of weights.

Initial situation

In 2003, the BEV had a rather unsatisfactorily situation about available equipment in the mass laboratory. There was a new CC1000S-L comparator from Sartorius with a 4-weight load alternator (resolution 1 µg) for the dissemination of mass in the range from 100 g up to 1 kg. For the lower range of the mass scale the BEV have had one aging 4 g mass comparator scale (resolution 0.1 µg), one traditional 100g mass comparator (resolution 1 µg) and a standard 100 g balance for the dissemination of mass and for calibrating and adjusting small weights up to 100g. For the range above 1 kg there was only a C10000S mass comparator (resolution 10 µg) with a 2-weight load alternator and a conventional 30 kg mass comparator for comparing 20 kg weights (resolution 10 mg) available. With the exception of the mass comparators including the changer, all weights were positioned manually.

In order to meet the appropriate standards for a National Metrology Institute respectively to meet the requirements for a state-of-the-art mass laboratory, it was necessary to build up new equipment. When considering the requirements for the dissemination of mass and the expected number of weight calibrations to be handled, traditional mass comparators did not appear to be sufficient for the job.

Moreover, to be able to offer high precision mass determination it was self-evident to build up automatic systems. Due to the circumstances that for the intended requirement of the BEV there have not been really fitting handling systems on the market there was the decision to built their own system

Introduction

The objective of the BEV was to develop and realize handling systems for loading and alternating weights on high-precision mass comparators. The overall process includes retrieving weights of 1 mg up to 20 kg from a repository unit, determining the load's center of gravity for correct positioning, and placing loads on the mass comparator in sequence. Each completed measurement routine encompasses a comparison of test weights to reference standard weights. The systems have to fulfill all requirements, in view of the strict demands placed on the dissemination of mass standards and the number of weight calibrations expected to be performed at the BEV

Building on years of excellent cooperation, the BEV and the Institute of Production Engineering (IFT) of the Vienna University of Technology, together with Sartorius AG in Goettingen, have realized these three projects.

Robot performing high-precision mass determination on weights in the range from 1 kg up to 20 kg

Requirements and objectives

The objective was to design and manufacture an automatic load alternator (the "handling system" referred to in this report) for weights of up to 20 kg, for use with both bar weights and cylindrical knob weights. The handling system would have be capable of working with 20 kg consisting sets of 10 kg, 5 kg or smaller weights, as well as with 10 kg and 5 kg consisting sets of smaller weights. A further requirement was that the system held up to 10 test weights and 2 reference standard weights in a changer magazine.

Furthermore, the handling system had to be designed to enable any type of mass comparator to be used and several different types to be operated at a time and so it must be possible to enlarge the possible weighing range up to 50 kg.

Figure 1. *System implemented at BEV*

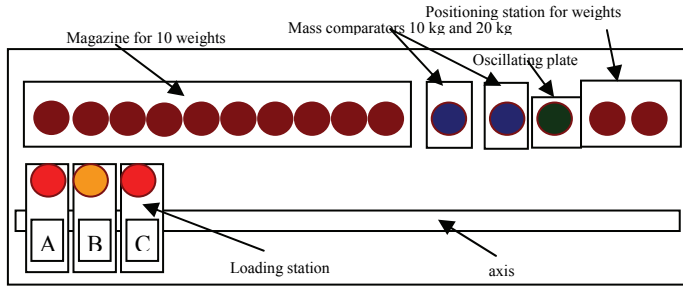


Figure 2. *Basic concept*

The system should have an external control unit with options for user-definable programming of comparison routines for all weights in the magazine. The control program should also need to have at least one standard program for fully automatic calibration of weights (comparison of weights) in accordance with the applicable standard operating instructions, and a fully automatic system for inserting and centering the required weights. In other words, the weights should be placed on an insertion device, assigned a position in the changer magazine, automatically centered, and then inserted into the magazine. The placement of the weights on the mass comparator, the comparison, and the return of the weights to the magazine would also have to be fully automatic, and the weights should be lifted directly, without the aid of pallets.

The handling system should need a number of control and monitoring systems to prevent potentially damaging errors, such as double-loading of a magazine slot or of a mass comparator. The design of the handling system should have to rule out interference with the measuring process.

Design

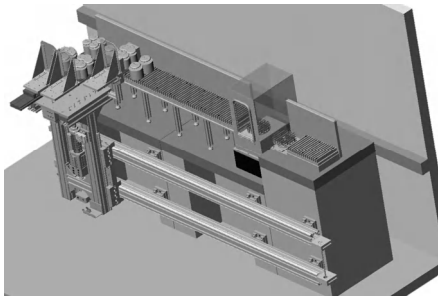


Figure 3. *CAD design overview*

The handling system has a positionable axis that moves the transport unit longitudinally. The loading station on the transport unit are extended and retracted by pneumatics. The transport unit is designed in such a way that the three receptacles – A, B and C – can be extended separately and raised simultaneously. As an added safety feature, a spring is provided for lifting the receptacles; weights are lowered pneumatically against the spring force. The system is designed for computer control, and the linear drive train has a motion controller with point-memory for precision control.

Function

Figure 4. *Load platform with axes*

At the time weights of 20 kg max. are introduced into the system via the repository unit on the platform, and data concerning the weights are entered on the computer. The handling system retrieves a specified weight and places it on an oscillating plate to center the load. The weight is lifted and lowered repeatedly to adjust its center of gravity to a predefined position, after which the weight is inserted in the assigned magazine slot.

This preparatory weight positioning procedure is repeated until all of the required weights are in the magazine. A similar procedure is follows when positioning the reference standard weights, with the exception that these are placed on loading station A and C rather than in the magazine. The comparison sequence is as follows: loading station B retrieves a weight from the magazine, and this weight is compared to that on receptacle A or C by alternating the loads on the mass comparator in accordance with the appropriate measurement procedure (e.g., A-B-B-A). This comparison can be repeated, placing the weights on the mass comparator in any order, as often as desired. At the end of the measurement routine, the transport unit returns the test weight to the repository unit and retrieves the next weight to be tested.

Implementation

The mass comparator, the magazines and the repository unit are mounted on a granite slab with concrete pylons; the guides are installed on the front of the pylons. The axes are aligned with the load platform using adjustment plates. A stepper motor moves the longitudinal axis, and an error detection function has been integrated to ensure the highest possible safety of operation. The mechanical drive of the horizontal axis is operated by toothed belts using a gantry design. This drive has a brake that is set automatically when the drive comes to a halt, so the voltage can be switched off to prevent vibration or electrical interference during measurement.

The transport unit consists of a load platform that has a travel distance of about 50 mm. For optimum safety, upward movement is implemented using a spring; in the event of a failure in the pneumatics, the platform moves to the uppermost position. The platform is lowered pneumatically against the force of the spring. There are three positions on the platform; i.e., the loading station, from which weights can be transported off the platform. These receptacles are equipped with precision guides and are pneumatically driven. The control system runs on a commercially available personal computer (PC) that addresses the load alternator hardware over an I/O card. The predefined motion sequences and constant optical monitoring of the entire system ensure that weights cannot collide.

To eliminate negative ambient influences, the mass comparator has a protective enclosure that shuts automatically as soon as a weight is positioned for

measurement. Additionally, the load alternator system is encased in a barrier housing that is not in contact with the measuring device. Both, the control unit and the power electronics are installed outside the equipment housing to keep the effects of thermically induced currents to a minimum. Additionally, the system is made for the most part from non-magnetic and/or antistatic materials to preclude magnetic disturbances.

Figure 5. *Measuring the test weight on the below weighing system. On the right there is the precentering plate; below the weight holder a "liquid brake" to avoid pendulation can be seen*

The measuring instruments are a Sartorius CC20000S-L (resolution 100 µg) and a Sartorius CC10000U-L mass comparator (resolution 10 µg) both specially adapted for this purpose by the manufacturer. Both mass comparators have been rebuilt for a weighing position below the mass comparator

This eliminates off-center loading problems on the weigh cell. The specially developed load alternator makes it possible to insert the 1 kg to 20 kg weights with no rearrangement required, and to position them precisely in the suspension device of the mass comparator.

The system can be adapted to special ambient conditions and metrological requirements at any time. Several mass comparators can be used simultaneously with this system, which permits testing of a wide range of weights. Experiments have shown that weights from 1 kg to 50 kg can be transported by this conveyor. Since 2004 the handling system has worked with the CC20000S-L for 10 kg and 20 kg weights. At present, the BEV install a the second comparator in the system for the range from 1 kg to 10 kg.

Test sequence

Weights are stored on a positioning device in the repository unit. The identification number, nominal mass value and type of weight (i.e., test or reference standard weight) are entered in the controller PC. Once the input is confirmed, the weights are centered and inserted in the magazine.

Preparation for measurement includes entering the selected test routine, choosing the tests weights and defining the number of measurement repetitions. Once this input has been confirmed, the weights are centered and positioned automatically by the handling system and the measurements performed in sequence. The preparatory steps can be carried out for all test weights at one time. Coordination of weights with different nominal mass values and the two mass comparators and also the mass comparator's internal substitution weight is automatic, as are the measurements. Once tested, weights are moved to the repository unit; a number of safety mechanisms have been integrated to make sure one weight is not mistaken for another. The result of each measurement series is a calibration certificate that is issued for the weight tested.

Figure 6. *Routine testing: measuring a 20-kg bar weight, a 20 kg block weight and a 10 kg bar weight(test weight); in the background are two 10 kg weights and two 5 kg weights (reference weight). A sensor with its opposing reflector foil for weight position determination and the automatic protective enclosure for the mass comparator can be seen clearly*

Measurements and results

A series of measurements has been carried out at BEV since May 2004 to provide the required mathematical basis. In these measurements, weights values were measured and recorded at 250 ms intervals, together with data pertaining to various environmental parameters, and the results evaluated. To validate the system for observance of the BEV's legal requirements, approximately 1,000 individual measurements were performed during this phase, each consisting of ten A-B-B-A cycles, and evaluated. Comparisons were performed subsequently using weights with nominal mass values of 10 and 20 kg. After implementation of the system with the weighing position below the mass comparator, the standard deviation achieved for 20 kg mass comparator for the 10 kg weights between 0.04 mg and 0.10 mg; for the 20 kg weights, the standard deviation was between 0.05 mg and 0.12 mg. The repeatability of results was between 0.03 mg and 0.3 mg for the higher range.

Since December 2004 this system has been implemented at BEV for the dissemination of mass standards for a range up to 500 kg, as well as for internal and external calibrations and verifications of weights with nominal mass values of up to 20 kg. At the present the range from 1 kg up to 5 kg is implemented.

The second robot performs high-precision mass determination on weights in the range from 1 mg up to 10 g

A further development of the BEV realize a handling system to supply a new high precision mass comparator with weights. Weights from 1 mg to 10 g are retrieved from a magazine and then placed on the mass comparator. Each completed measurement routine comprises a comparison of test weights to reference weights.

Specifications and objective

The objective, therefore, was to design and build an automatic handling system for weights with a mass of between 1 mg and 10 g. The specification included the use of weights in the form of both wires and flat polygonal sheets and without a weight transportation unit. The load alternator should guarantee the application of several weights at once (groups of weights on loading), namely for the dissemination of mass up to 10 g but also for calibrating special weights. In doing so it would be necessary to place at least four complete sets of weights of up to 10 g in the alternator magazine. Control should take place via an external control unit and should allow any desired programmable alternation of all weights in the magazine. The control programme should contain at least one standard programme for fully automatic calibration (comparison of weights) in accordance with the relevant operating instructions. After introducing the weights into the magazine they should be assigned to their positions by means of software. The placing of the weights on to

the mass comparator, the mass comparison and the return of the weights to the magazine should take place automatically. The weights would thus be picked up directly (without weight transportation units).

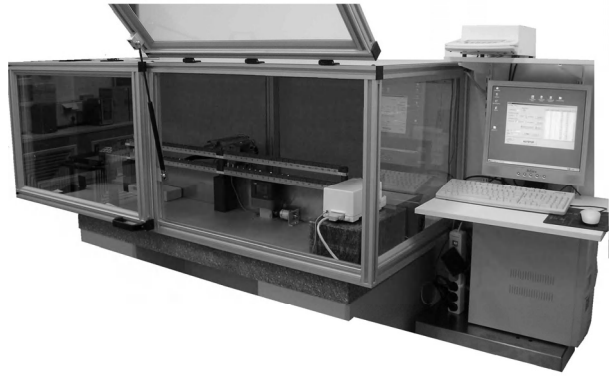


Figure 7. *The handling system for micro weights at the BEV*

Also this system should have several suitable control and monitoring devices to prevent double loading of magazine slots or the scale in addition to other types of damage and errors. Furthermore, the handling system should be designed in such a way as to exclude the possibility of compromising the measurement.

The handling system should also be designed to be suitable for operating with a variety of mass comparators.

Concept

The handling system has an adjustable axis that moves the transport unit horizontally. With a concept differing slightly to that of the big robot, this small robot should have only two weight carrier units on the loading station extended and retracted pneumatically. The reason for this is the greater number of standards weights; these weights are also stored in the magazine.

The loading station is designed in such a way that the two weight carrier units can be extended separately and raised simultaneously. The vertical movement of the weight carriers is carried out pneumatically. The complete control of the system is based on a personal computer solution and takes place via digital inputs and outputs. The linear axis has a servo drive. The dimensions of the complete handling system were restricted to a maximum of 2,000 mm x 800 mm x 800 mm thus making the handling system suitable to fit on a conventional weighing table.

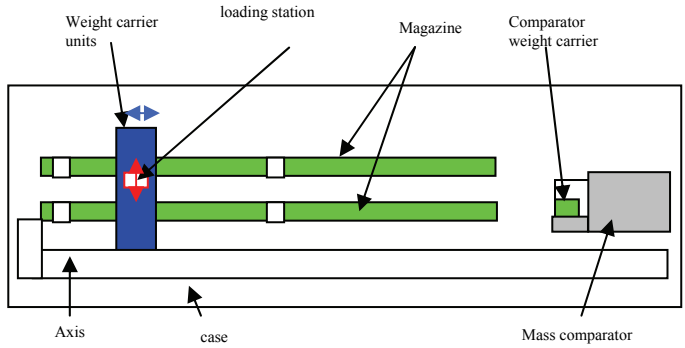


Figure 8. *The basic concept*

Method of operation

The weights – 1 mg to 10 g – or groups of weights are introduced by hand into the correct magazine slot and the weight data is entered into the control PC. The same procedure is also followed for the reference weights.

The comparison measurement now takes place, whereby the loading station retrieves one weight and one reference weight from the magazine and places these alternately on the scale. This comparison measurement can now be repeated as often as necessary and in any desired sequence. When the measurement routine has been completed the loading station places the tested weight and the reference weight back into their corresponding slots in the magazine.

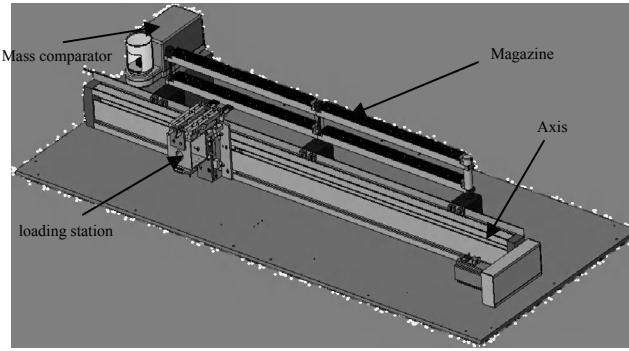


Figure 9. *Design overview*

System design

The system was designed on a linear basis due to the positive experience gained from the previous system and easy, universal manageability – but also for production cost reasons. This makes it possible to use a controlled axis e.g. by a servo drive. All other movements are pneumatically driven. The magazine capacity has been designed at 80 slots on two levels, which may be loaded with reference weights and test weights as desired.

The heart of the system is the pneumatically actuated loading station which can be driven easily to four vertical and two by two horizontal positions. A comb-type grabber system was installed in order to meet the design requirement that weights should always be picked up without the aid of a transportation unit. However, in order to be able to use 1 mg weights, the required tolerances between magazine and finger grabber are extremely tight. This means that the grabber must be able to position itself at any point with an accuracy of better than 0.05 mm. This is guaranteed by means of a servo-driven precision spindle along the entire movement path of 1800 mm.

The pneumatic drive provides "total" positionability due to the mechanical end stop at both end positions and the absence of interference from electromagnetic fields. By contrast, the use of electrical actuators to realise the lifting movements would only be possible at higher cost and greater complexity. The system is controlled by a personal computer which communicates with the DC-servo controller via Profibus-DP. This is a very safe, proven industrial solution.

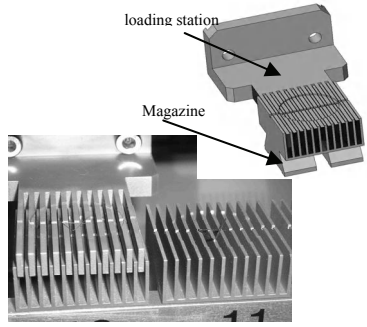


Figure 10. *Weight carrier – design and reality (carrying 1 mg and 2 mg wire weights)*

The whole system stands on a granite plate, while the comparator itself is mounted on an additional granite slab. In order to prevent disturbing environmental effects, the comparator has a wind shield which closes during measurement. In addition, the whole system is housed in a protective enclosure which is detached from the measurement unit. Both the control and power electronics were placed

outside the enclosure in order to keep the effects of thermically induced currents to a minimum. The whole system is constructed mainly from non-magnetic and/or antistatic materials to prevent disturbances from magnetic fields.

The mass comparator which is used in the system is a modified Sartorius CCE6 mass comparator to 10 g maximum load which has a resolution of 0.1 µg. This was specially adapted by Sartorius for use in handling systems. In addition, the weighing pan for weights was completely redesigned.

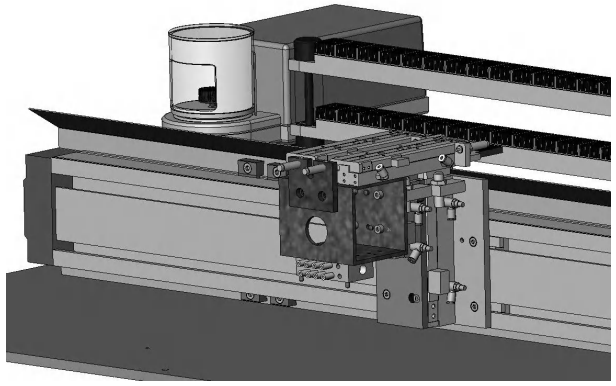


Figure 11. *Loading station with pneumatic drive*

Test procedure

The weights are placed in the magazine slots. The identification number, nominal value and type of weight (test or reference weight) are then entered into the software on the control PC and the corresponding magazine slot is allocated.

Preparation for measurement takes place by entering the preselected test routine, selecting the weights or entire magazine bank and the number of repeat measurements. After confirmation the handling system automatically retrieves the individual weights from the magazine and the measurement is processed. All weights are prepared at the same time. The coordination of the weights of different nominal values and the substitution weights internal to the mass comparator – as well as the measurements themselves – take place automatically. On completion of the test routine the weights are removed from the magazine slots and the corresponding data are deleted from the database. Reference weights can remain in the magazine.

Figure 12. *The CCE 6 weight carrier*

Due to continuous logging of all programme steps and movements it is possible to request the position of a weight at any time – even in the event of a complete system crash.

The recording of the ambient parameters for the determination of air density is automatically documented for each weight measurement. The documentation relating to the measurements and the calibration certificate of the measured weight are printed out as the result.

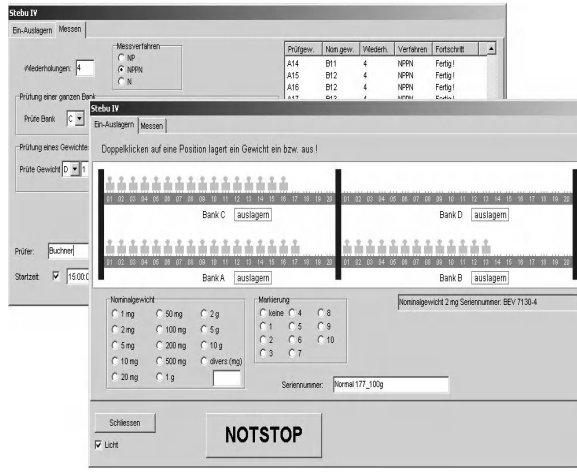


Figure 13. *Input form for weights of the software*

Measurement results

In order to validate the system, starting in Summer 2006 onwards measurement sequences were carried out using already calibrated BEV weights. The BEV also conducted internal comparison measurements using alternative procedures for mass determination. Previously, a special calibration of the mass comparator and the development of the test software were necessary. After placing the weights in position, measurements were taken at intervals of 500 ms and with a series of measurement results of environment parameters documented.

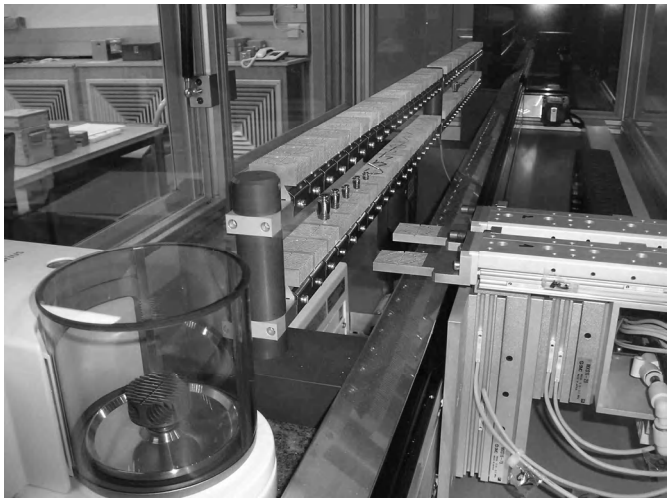


Figure 14. *Direct comparison of two sets of weights*

The measurements recorded are averaged into "measurement blocks" and taken for further analysis. For the purpose of evaluation of the mass comparator, several measurements are carried out on E1 and E2 sets of weights and evaluated in measurement cycles of 10 ABBA or RTTR (Reference – Test – Test – Reference, i.e. the sequence in which the weights are alternately weighed). This includes comparison with weights with a nominal value from 1 mg to 10 g. The standard deviations of the average values obtained were less than 0.2 µg for the weights used. The repeatability of the average values of the measurement results was for the measurement cycles less than 0.1 µg for the smaller weights and less than 0.2 µg for the larger weights.

The system was installed at BEV in Autumn 2006 for the dissemination of mass in the range up to 10 g as well as for internal and external calibrations and verifications of weights.

The BEV is currently building an additional robot for weights up to 1 kg. A 1 kg load cell from Sartorius with a resolution of 1 x 10^{-6} is being used here. This equipment will be ready by the first quarter of 2007. Thus the BEV will have completed its plan to fully automate its mass laboratory. In future all measurements for calibration and adjustment and a large part of the measurements for the dissemination of mass will be carried out by three "weighing robots".

The third robot performs high-precision mass determination on weights in the range from 10 g up to 1 kg

The last target in the area of mass measurement was to fill the gap, in the most accuracy range of the mass scale between 10 g and 1 kg. Therefore, the BEV has developed a further handling system to supply a high precision 1 kg mass comparator with weights. Weights from 10 g to 1 kg are retrieved from a magazine and then placed on the mass comparator as with the other systems before a measurement routine is completed, when a comparison of test weights to reference weights has been carried out.

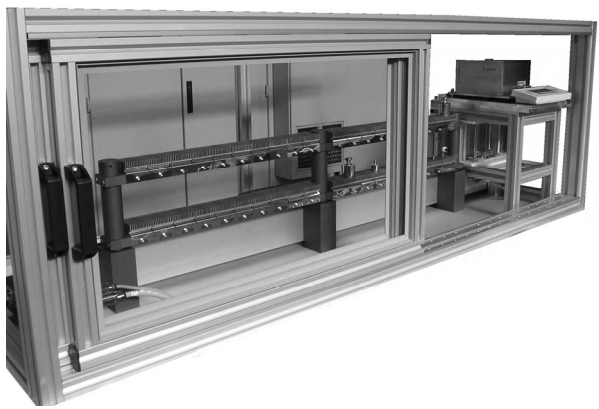


Figure 15. *The handling system for medium weights at the BEV*

Specifications, objective and concept

The aim was to design and build an automatic handling system for weights on equal terms as the handling system for the micro weights i.e. the mass of between 1 mg and 10 g. The load alternator should guarantee the application of several weights at once (groups of weights on loading), namely for the dissemination of mass up to 1 kg but also for calibrating special weights. The number of alternator

magazine places should be in the size of at least four complete sets of weights from 10 g up to 1 kg. Due to the excellent experience with the performance of the other facilities the same control mechanism and procedure also the program should be used. This system should also have the suitable control and monitoring devices to prevent double loading of magazine slots or the scale in addition to other types of damage and errors. Furthermore, the handling system should be designed in such a way as to exclude the possibility of compromising the measurement.

The concept for the realisation is the same as the system for the micro weights. There are only two significant differences: on one hand due the size of the weights there are only 40 magazine places and on the other hand the weights are placed below the mass comparator while they are measured.

Method of operation

All test weights and reference weights from 10 g to 1 kg or groups of weights are introduced by hand into the correct magazine slot and the weight data is entered into the control PC. For pre-centering the weights in the magazine, a stencil is used for bringing in the weights. The comparison measurement procedure and the calculation is the same as it is used at the other system.

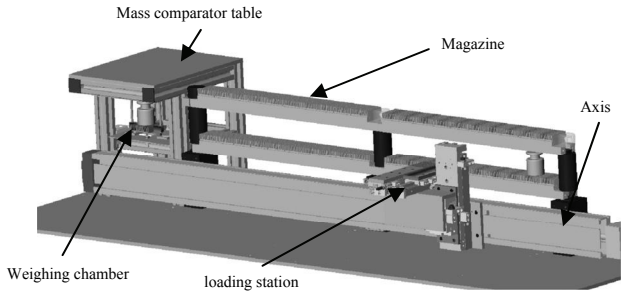


Figure 16. *Design overview; view from the back*

System design and test procedure

Due to the positive experience gained from the previous system the system was designed on the same linear basis that is easy, universal and well manageable. All other movements are also pneumatically driven. The magazine capacity was designated at 40 slots on two levels, which may be loaded with reference weights and test weights as desired.

The pneumatically actuated loading station can be driven to four vertical and two by two horizontal positions. A comb-type grabber system was installed in order to meet the design requirement that weights should always be picked up without the aid of a transportation unit.

The system is controlled by a personal computer which communicates with the DC-servo controller via Profibus-DP. This is a very safe, proven industrial solution.

Figure 17. *The CC 1000S-L weight carrier with a 500 g weight in the weighing chamber; below the weight holder a "brake" to avoid pendulation can be seen*

The whole system is placed on a granite plate, while the comparator itself is mounted on a table. For the weights, a position below the mass comparator was chosen. This eliminates off-center loading problems on the weigh cell. The measurement unit used is a modified Sartorius CC1000S-L mass comparator which has a resolution of 1 μg at a maximum load of 1 kg. The cell was specially adapted by Sartorius for use in this handling system and to fix the below weight holder. The specially developed comb-type grabber system makes it possible to insert the 10 g to 1 kg weights with no rearrangement required, and to position them precisely in the suspension device of the mass comparator.

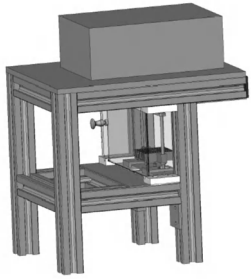


Figure 18. *CAD model: mass comparator table with the below weighing chamber*

During handing over the weights and when the loading station is moving the hole suspension device of the mass comparator is fixed. Thus, only a very small gap is allowed between the fingers of the comb-type grabber and the weight holder of the comparator. During the measurement the fixation is unfastened.

To prevent disturbing environmental effects, the weighing chamber has a class box – wind shield which closes during measurement. In addition, the whole system is housed in a protective enclosure, which is detached from the measurement unit. Like the others, the whole system is constructed mainly from non-magnetic and/or antistatic materials to prevent disturbances from magnetic fields.

Building on the excellent experience the same evaluation logic and the same calculation procedure is used as in the smaller system.

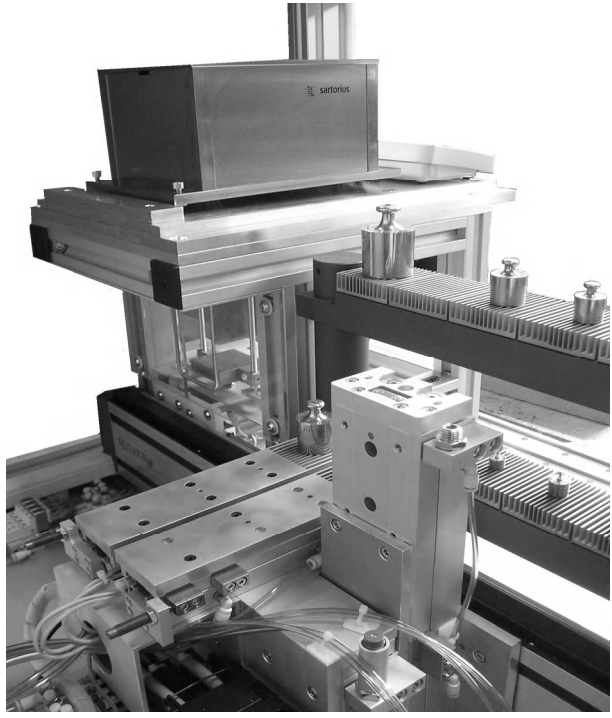


Figure 19. *Foreground: Loading station; Mass comparator table with the below weighing chamber in the background*

Measurement results

In order to validate the system, at the present, further measurement sequences are carried out using already calibrated BEV weights. The BEV also conducts internal comparison measurements using alternative procedures for mass determination.

For the purpose of evaluation of the system, several measurements are carried out on E1 and E2 sets of weights and evaluated in measurement cycles of 10 ABBA. This includes comparisons with weights with a nominal value from 10 g to 1 kg. The first results bring excellent standard deviations and also the repeatabilities of the measurement results are remarkably.

Figure 20. *Direct comparison of two 500 g weights*

The system was installed at the BEV in March 2007 just like the others for the dissemination of mass in the range up to 1 kg for internal as well as for external calibrations and verifications of weights.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Interlaboratory comparison of seven standard weights in several Romanian laboratories

Adriana VÂLCU, George Florian POPA, Sterica BAICU

National Institute of Metrology, Bucharest, Romania, Sos. Vitan Barzesti nr. 11, sect. 4, Bucharest
adriana.valcu@inm.ro; george.popa@inm.ro; sterica.baicu@inm.ro

ABSTRACT: This paper reports the results of measurements performed on seven standard weights by 15 metrology laboratories throughout Romania. The goal of the Inter-laboratory's measurements was to provide verification of each participating laboratory's measurement capability by obtaining a measurement that agrees with the pilot laboratory.
The comparison began in 2005 and concluded in 2006, the National Institute of Metrology (INM) acting as a pilot (coordinating) laboratory for the programme.
Seven mass standards (nominal values: 10 kg, 1 kg, 500 g, 200 g, 100 g, 20 g and 100 mg) were sent to the participants.
The results are analyzed using E_n values.
The laboratory's results are presented for each weight, uncertainty declared and the errors normalized by each laboratory, with respect to the INM.

Introduction

An interlaboratory comparison of seven mass standards was carried out between 15 metrology laboratories throughout Romania, during the period 15.08.2005 – 10.05.2006. The transfer standards used were carefully selected by the pilot laboratory, LP, and the comparison scheme was chosen to minimize the influence of any instability in their mass.

The 15 participants in the comparison were: Târgovişte, Piteşti, Ploieşti, Bacău, Iaşi, Botoşani, Piatra Neamţ, Braşov, Târgu Mureş, Sibiu, Bucureşti, Buzău, Brăila, Timişoara. Each laboratory has been assigned a code (for the confidentiality's results), LP having the code "1".

Circulation Scheme

The artifacts were initially calibrated by LP and then circulated between participant laboratories in two "petals". At the end of each petal, the artifacts were

returned to LP for re-calibration, before being sent out to the participating laboratories in the next petal.

The drift of the mass standards, estimated using the difference between the initial and final LP measurements, is negligible compared with the associated uncertainty.

The transfer standards (10 kg, 1 kg, 500 g, 200 g, 100 g, 20 g) were stainless steel weights of class E2. The 100 mg weight was a nickel silver polygonal sheet of E_2 class.

The density of the weights was provided by LP as follows:

– from 10 kg to 20g: 7950 kg/m^3, U=140 kg/m^3;
– for 100 mg weight: 8600 kg/m^3, U=170 kg/m^3.

Measurement instructions

To calculate the buoyancy correction, the densities of the weights was given by the LP. The participants carried out the calibrations without re-determining the density of the weights.

The following information about the transfer standards was given in advance to the participants: nominal masses, densities and their uncertainties and magnetic proprieties. In addition, instructions were given on how to handle, store and transport the weights. For each laboratory the measurement time was two weeks. The participants were asked to send their results to the LP within two weeks after the end of measurements.

Following the intercomparison practice the laboratories were given figure codes (1…15).

No detailed calibration instructions were given to the laboratories.

Tasks

It was the participant's task to determine the mass of the standards with an uncertainty corresponding to their capability. The nominal values of the weights were selected such that the weighing instruments and mass standards of the participants could be tested within a wide range. The evaluation by the participant was to furnish the following information:

- mass and uncertainty of the 7 mass standards;
- traceability of reference standards used;
- physical properties of the reference standards used
- specifications on the measuring instruments used (weighing instruments, barometers, hygrometers, thermometers);
- ambient atmospheric conditions at the time of each measurement.

The participants were requested to specify the uncertainty budget in sufficient detail.

Results

A full calibration report with all relevant data and uncertainty estimates based on recommendation was requested to be sent to the LP. All 14 participants were able to perform measurements and submit the measurement results to the LP.

The results of the intercomparisons are summarized in Table 1.

Each laboratory reported the measured mass value that is assigned to each of the seven artifacts, together with an expanded uncertainty for each weight. For all laboratories, the coverage factor was 2.

The measurements results of the interlaboratory comparison are summarized in Table 1.

All data are reported on the sample as received. The results are presented exactly that were sent by the participants.

Laboratory's	10 kg		1 kg		500 g		200*g		100 g		20* g		100*mg	
code	*E*	*U*	*E*	*U*	*E*	*U*	*E*	*U*	*E*	*U*	*E*	*U*	*E*	*U*
	mg	mg	mg	mg	mg	mg	mg	mg	mg	mg	mg	mg	mg	mg
1	11027.1	2.4	617.27	0.12	32.00	0.07	16.20	0.04	17.97	0.04	2.226	0.010	-0.056	0.002
5	11040	21	616.7	1.4	31.1	1.3	6.16	0.09	8.01	0.06	2.240	0.019	-0.057	0.005
2	10889.583	121.140	617.375	5.904	34.125	7.677	16.158	0.192	17.998	0.139	2.238	0.025	-0.05733	0.013
12	10893.433	13	616.770	1.7	31.851	0.3	16.377	0.08	17.577	0.06	2.374	0.013	-0.05105	0.005
11	11021	25.4	616.75	2.4	31.52	2.3	16.36	0.3	17.85	0.3	2.25	0.03	-0.048	0.01
7	10999.7	21.33	701.7	2.66	37.6	2.10	20.9	0.263	20.6	0.166	2.3	0.132	0	0.102
13	11032.43	6.90	617.479	1.54	31.933	0.32	16.266	0.07	17.987	0.10	2.246	0.02	-0.057	0.044
3	11040	30	617	2	32.1	1.5	16.22	0.26	18.10	0.14	2.250	0.038	-0.049	0.013
10	11020	300	617.0	3.4	31.5	3.6	16.3	0.6	18.0	0.4	2.4	0.2	-0.04	0.19
8	215.7	10.52	616.3	1.44	31.84	1.22	16.38	0.40	18.15	0.54	2.19	0.13	**-0.058**	0.04
1	11028.4	2.4	617.34	0.16	31.96	0.07	16.16	0.04	17.96	0.04	2.220	0.011	-0.061	0.002
6	11030	10	617.0	1.5	31.9	0.8	16.2	0.3	18.00	0.15	2.250	0.025	-0.063	0.005
9	11030	20	617.3	1.0	31.7	0.8	16.2	0.2	18.02	0.15	2.254	0.015	-0.065	0.008
4	11047	28	616.8	2.3	32.4	1.6	16.17	2.0	18.31	0.13	2.25	0.06	-0.06	0.01
15	11032	35	617.1	2.9	31.2	2.8	16.3	0.2	18.10	0.16	**2.24**	0.04	-0.053	0.004
14	11050	6	617	1	32	1	16.2	0.13	18.1	0.1	2.32	0.03	-0.06	0.01
1	11028.2	2.4	617.14	0.16	31.97	0.07	16.17	0.04	17.98	0.04	2.227	0.01	-0.061	0.002

Table 1. *Deviation from nominal mass (E) and expanded uncertainty (U) for the corresponding values*

A tool often used in analyzing the results from interlaboratory comparisons is the normalized error E_n, which takes into account both the result and its uncertainty. The normalized error En is given as:

$$E_n = \frac{x_{lab} - x_{ref}}{\sqrt{U_{lab}^2 + U_{ref}^2}}$$

where:

E_n = normalized error;

x_{lab} = result of measurement carried out by a participant laboratory;

x_{ref} = the comparison reference value of LP;

U_{ref} = measurement uncertainty of LP;

U_{lab} = measurement uncertainty reported by participant laboratory.

Using this formula, an acceptable measurement and reported uncertainty would result in an E_n, value of between -1 and +1, with a desired value close to zero. The E_n data for each laboratory is presented in Table 2.

This calculation provides supplemental information concerning the measurement capability of the participating laboratories.

The x_{ref} and U_{ref} used for the calculations was the mean of the opening and closing LP measurements.

Figures 1 to 7 present the differences between the participants' results and the reference value, and uncertainty (k=2) for all the weights.

LABORATORY'S CODE	Nominal value						
	10 kg	**1 kg**	**500 g**	**200g**	**100 g**	**20 g**	**100 mg**
Xref	11027.9	617.25	31.98	16.18	17.97	2.224	-0.059
Uref	3.2	0.26	0.09	0.06	0.05	0.014	0.005
Normalized deviations E_n							
1	-0.20	0.07	0.20	0.31	0.00	0.10	0.63
5	**0.57**	**-0.39**	**-0.67**	-91.13	-125.91	**0.66**	**0.33**
2	-1.14	**0.02**	**0.28**	**-0.09**	**0.19**	**0.47**	**0.14**
12	-10.05	**-0.28**	**-0.40**	1.97	-4.97	7.72	1.18
11	**-0.27**	**-0.21**	**-0.20**	**0.60**	**-0.39**	**0.77**	**1.02**
7	-1.31	31.59	2.68	17.46	15.13	**0.57**	**0.58**
13	**0.60**	**0.15**	**-0.13**	**0.95**	**0.15**	**0.88**	**0.05**
3	**0.40**	**-0.12**	**0.08**	**0.16**	**0.87**	**0.63**	**0.74**
10	**-0.03**	**-0.07**	**-0.13**	**0.20**	**0.07**	**0.88**	**0.10**
8	-984.57	**-0.65**	**-0.11**	**0.50**	**0.33**	**-0.26**	**0.03**
1	0.13	0.29	-0.14	-0.22	-0.15	-0.24	-0.32
6	**0.20**	**-0.16**	**-0.10**	**0.08**	**0.19**	**0.89**	**-0.52**
9	**0.10**	**0.05**	**-0.34**	**0.11**	**0.32**	1.43	**-0.60**
4	**0.68**	**-0.19**	**0.26**	**0.00**	2.43	**0.42**	**-0.06**
15	**0.12**	**-0.05**	**-0.28**	**0.59**	**0.77**	**0.37**	**1.00**
14	3.26	**-0.24**	**0.02**	**0.16**	1.16	2.88	**-0.06**
1	0.08	-0.36	-0.06	-0.09	0.15	0.15	-0.32
Normalized deviations E_n from reference values							
					$E_n > 1$		
					$E_n < -1$		

Table 2. *Normalized deviations E_n from reference values*

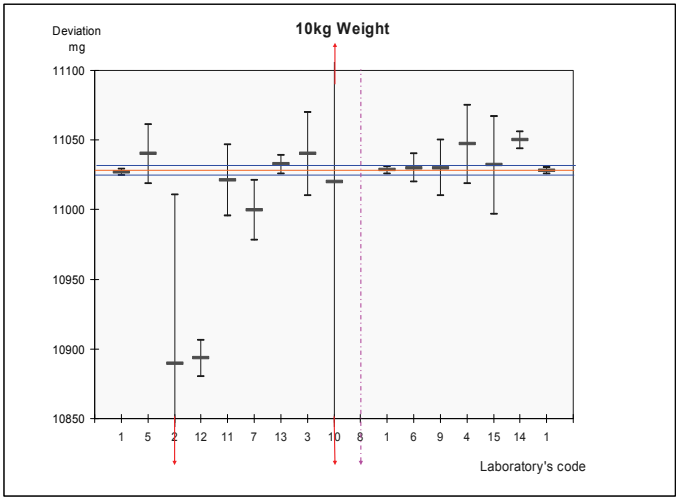


Figure 1. *The difference between participants' results and reference value and uncertainty for 10 kg weight*

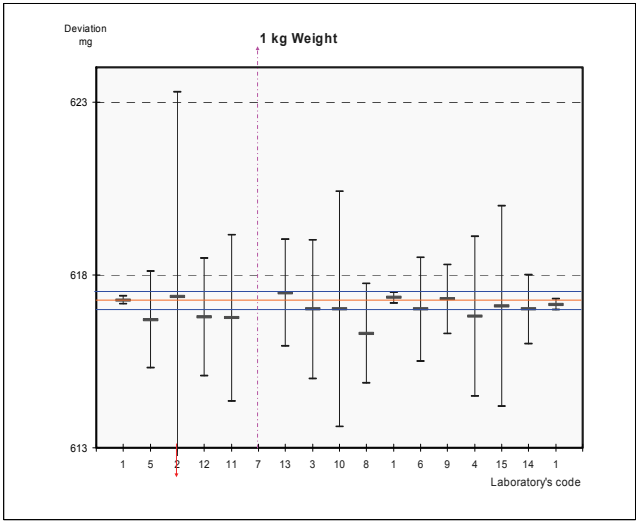


Figure 2. *The difference between participants' results and reference value and uncertainty for 1 kg weight*

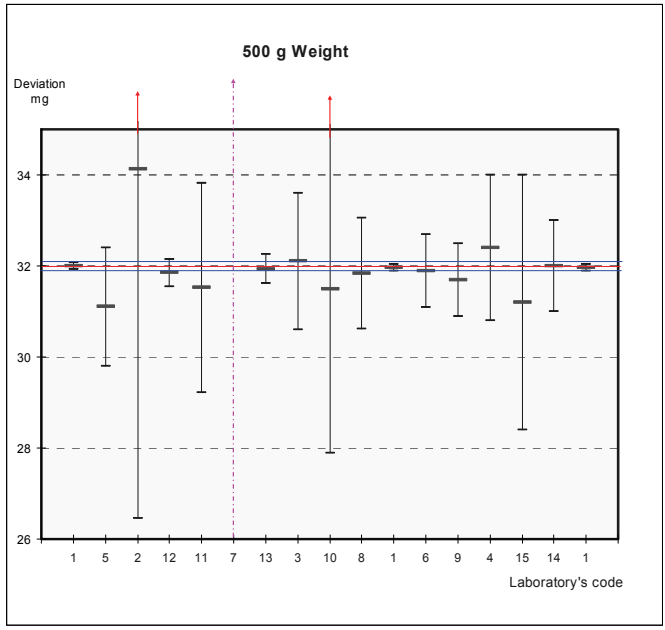


Figure 3. *The difference between participants' results and reference value and uncertainty for 500 g weight*

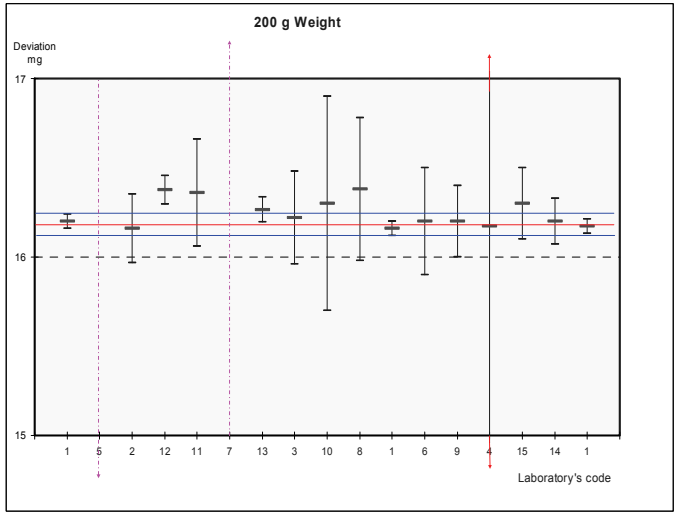


Figure 4. *The difference between participants' results and reference value and uncertainty for 200 g weight*

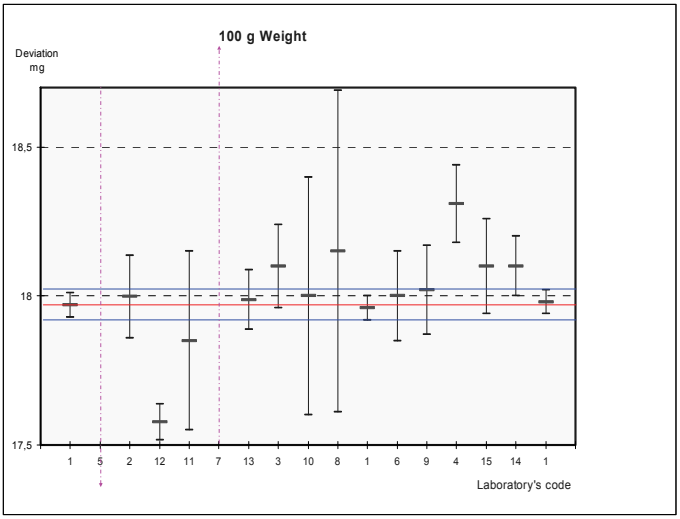


Figure 5. *The difference between participants' results and reference value and uncertainty for 100 g weight*

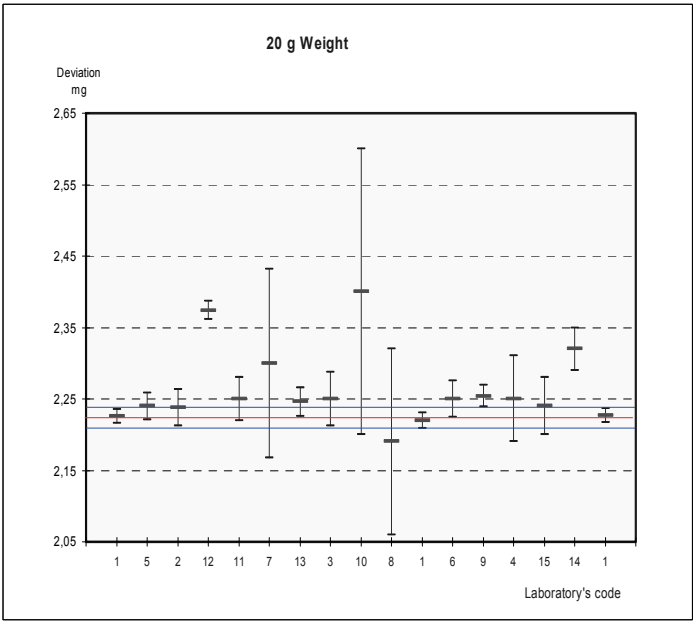


Figure 6. *The difference between participants' results and reference value and uncertainty for 20 g weight*

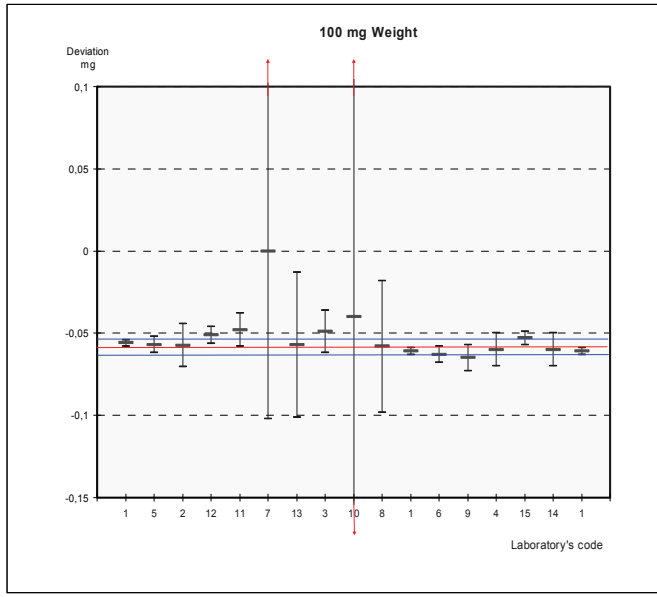


Figure 7. *The difference between participants' results and reference value and uncertainty for 100 mg weight*

Reference value

Measurement uncertainty $U(k=2)$ of LP

Result of measurement carried out by a participating laboratory

Measurement uncertainty $U(k=2)$ reported by a participating laboratory

Discussion

– Six participating laboratories (3, 6, 10, 11, 13, 15) obtained compatible results with that of LP for all the weights;

– for 10 kg weight, five participating laboratories (2, 7, 8, 12, and 14) differ from the LP results;

– for 1 kg weight, one participating laboratory (7) differs from the LP results;

– for 500 g weight, one participating laboratory (7) differs from the LP results;

– for 200 g weight, three participating laboratories (5, 7, and 12) differ from the LP results;

– for 100 g weight, five participating laboratories (4, 5, 7, 12, and 14) differ from the LP results;

– for 20 g weight, three participating laboratories (9, 12, and 14) differ from the LP results;

– for 100 mg weight, one participating laboratory (12) differ from the results of LP.

LP asked participants to review their results for confirmation. Five participating laboratories replied to the LP. One participant made a little change in its calculations;

Eight participating laboratories (2, 5, 8, 9, 10, 11, 12, 13) have taken into account uncertainty due to the eccentricity even the balances have a suspended load receptor;

Three participants have no taken into account in their uncertainty budgets the contribution due to the difference between standard and test. Most laboratories have taken this into account minimizing the problem using additional weights;

Six participants have not taken into account air buoyancy effect in their uncertainty budgets.

The LP sent out the summary results in a draft report to all participants, with the code number in place of the laboratory name. Each laboratory would see all results but would not know the owners of results other than its own.

Conclusions

Analyzing the results of interlaboratory comparison it can be seen that 19% from the total results are discrepant.

Four from the 14 laboratories have two or more results with "E_n" number larger than one.

The results obtained can be used to demonstrate the participating laboratory's measurement capability. The participants who obtained results greater than [-1, +1] should analyze the reasons in order to remedy and correct them.

– by analyzing the results, the following corrective measures are proposed:

– it is advisable that some participants must review the uncertainty analysis;

– a better control and monitoring of environmental conditions;

– a controlled access of the personnel in the laboratory during the calibrations;

– in the case of a big difference between standard and test it is advisable to use additional weights so that this difference is as small as possible. Otherwise it should be increased uncertainty with this component;

– further qualification of the personnel in calibrating and estimating uncertainty.

References

[1] BRML: PML-5-03 “Comparări Interlaboratoare”, 2002

[2] Adriana Vâlcu; Mr. George Popa; Mr. Sterica Baicu: “Determinarea masei şi evaluarea incertitudiniide măsurare pentru măsurile etalon”, 2006

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Automatic testing facility for determining the density of liquids and solids; and determining the volume of E1 weights

Christian Buchner

BEV- Bundesamt für Eich- und Vermessungswesen,
(Federal Office of Metrology and Surveying), Austria
christian.buchner@bev.gv.at

Summary

Fundamental aperture for density determination:

The Austrian Federal Office of Metrology and Surveying (BEV) has developed, in cooperation with Sartorius AG in Goettingen and the Vienna University of Technology, a test system for hydrostatic weighing and determining the density of liquids and solids, whereby mass comparators are automatically loaded alternately with weights and submerged plummets. With the load alternator for weights and plummets, the density of a liquid on one side and the density of a submerged solid (and therefore its volume) on the other side can be determined hydrostatically via direct comparison with standard plummets and the applied substitution weights.

Volume comparator:

To determine consecutively the volume of E1 weights and plummets, additional a fully automatic handling system for a 1 kg mass comparator, a volume comparator was developed. Thus, the hydrostatic weighing principle can be applied for serial volume determination on masses from 1 g to 1 kg by direct comparison with a single volume reference (e.g. a silicon sphere). This principle is implemented with a newly developed, fully automatic insertion mechanism for both completely submerged weights and substitution weights.

Both systems are used for the dissemination of density and volume by BEV as well as for testing and calibrating of liquids and plummets and determining the volume of weights in connection with the determination of mass scale.

The volume comparator is also used in other metrology institutes and in accredited calibration laboratories.

Introduction

In metrological affairs, high-precision determination of the density of liquids and solids, and therefore their volume, is often necessary for further measurements, calculations or for information about material characteristics.

For example, the determination of mass using conventional testing facilities is generally carried out in air under normal conditions. The mass of air, however, causes a systematic error in the procedure. Every solid body is affected by buoyancy in relation to the amount of air it displaces (Archimedes' Law). Air buoyancy makes a 1 kg stainless steel weight appear to be approx. 0.15 g lighter than it is. If the test piece has the same volume as its mass reference, the air buoyancy can be discounted, because they will have the same buoyancy. But if the test piece has a different density the air buoyancy have to be corrected in the calculation for the mass.

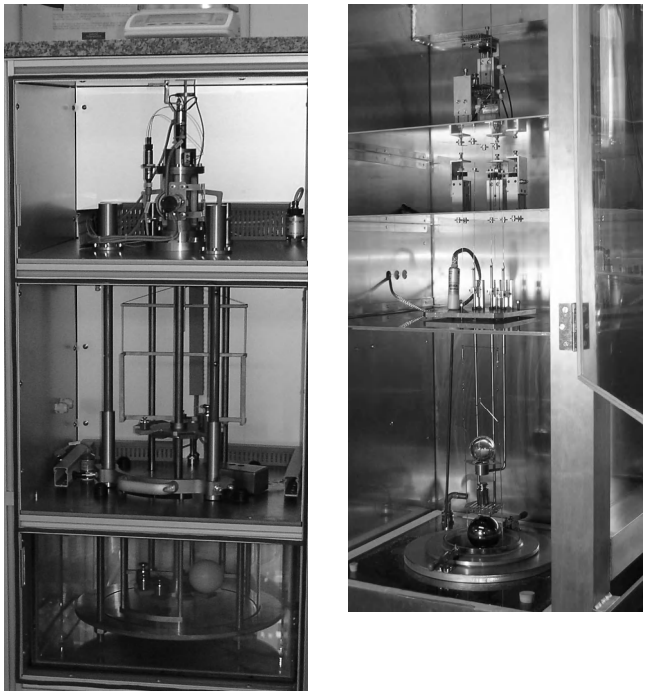


Figure 1. *The new volume-comparator (left) based on the new fundamental density aperture for solids (right) in the BEV; both are hydrostatic-below-weighing systems from 1 g up to 1 kg; left: in the volume comparator the magazine changer, the suspension system and the magazines for weights, the standard sinker and the test weights can be seen; right: on the test tower for measuring the density of solids the cylinder for lifting weights, standard sinker holder and test object holder can be seen*

Determination of the volume of weights is necessary for the correction of air buoyancy, and thus is an essential capability for today's metrology institutes.

This is why determination of density is indispensable for precise determination of mass, which is essential for metrology institutes, accredited calibration services and similar laboratories. This is reflected in international regulations, such as OIML recommendation R111 for a 1 kg E1 weight, which allows a mass tolerance of only 0.5 mg. Such precision cannot be attained in the determination of mass without determining the density of the weights and making the necessary correction for buoyancy.

Initial situation

Up until 2004, BEV had only a simple test assembly for hydrostatic weighing, in which the density of submerged test objects was calculated in individual weighings in an open thermostat. The temperature measurement, thermal stratification, stability of the mass comparator and the required volume of liquid were all considered insufficient for the standards required in this area. In order to meet the appropriate standards for a national metrology institute, the complete reconfiguration of the test assembly and new construction of a test system were essential.

Specifications and objectives

The main objective was to find an option for hydrostatically determining the density of less than one liter of liquid. For this measurement and for the hydrostatic determination of the density of solids, the primary aim was direct comparison of the test object with standard plummets or weights. In order to achieve the greatest possible reproducibility and minimize measurement uncertainty, it was specified that one or more load alternators were to be constructed for weight measurement below the mass comparator.

The entire measurement process up to and including evaluation of the measurement results was to be fully automatized. Particular consideration was given to the use of different mass comparators and to thermal stratification and other influences, such as surface tension, magnetic fields, etc.

Building on years of excellent cooperation, the BEV and the Institute of Production Engineering (IFT) of the Vienna University of Technology, together with Sartorius AG in Goettingen, have developed and built a fully automatic fundamental aperture for the determination of density.

Concept of the fundamental aperture

The system consists of an insulated, thermally controllable, double-walled liquid tank that can be raised using a lifting table. In the lower position, the test station can be equipped easily with weights, standard plummets and test objects. By raising the liquid tank, the test objects and weights suspended from a mass comparator can be immersed in the liquid and thermostatized. During actual measurement, the individual objects and weights can be raised and lowered using lifting rods that are moved pneumatically. This places and replaces the mass on the suspension system of the mass comparator. System control and data documentation are carried out via a PC.

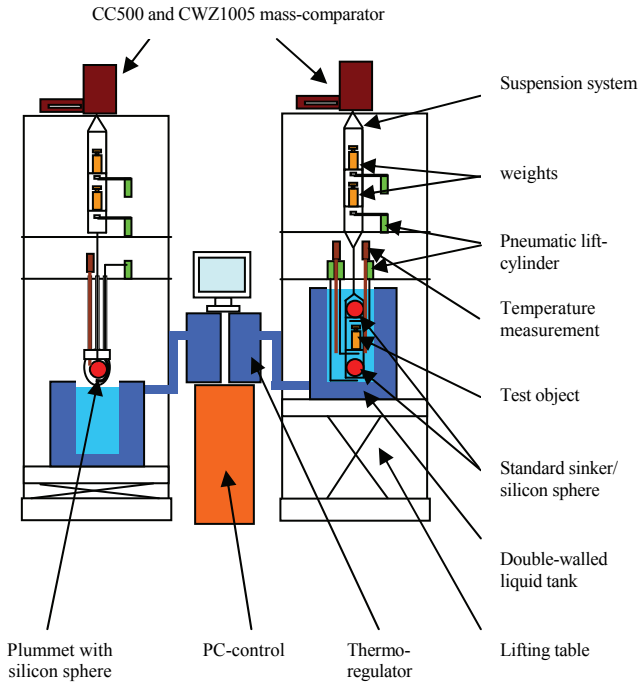


Figure 2. *Left: test station for measuring the density of liquids; right: test station for measuring the density of solids*

Functionality and procedure of the fundamental aperture

The entire measurement system consists of two separate units (towers), whereby each tower can be considered an independent test assembly. The first tower is used to determine the density of liquids; the second, the density of solids. This second tower can also be used universally for processes such as determining the density of relatively large volumes of liquid or the calibration of hydrometers or immersed objects.

Figure 3. *Test station for measuring the density of solids in CAD*

Each tower, consisting of an aluminum frame with movable shelves, has a mass comparator on top. The mass comparator specifications correspond to the range of the test objects to be measured. In addition, each tower has a lifting table with a thermostat reservoir and the appropriate number of pneumatically operated lifting units for the specific application. In order to keep heat-related currents as low as possible, the entire control system, power electronics and thermostat are located outside the enclosures. The measuring tower interiors are also provided with several partition plates in order to minimize convection currents in higher temperature measurement ranges.

Once the test objects and weights and plummet(s) have been positioned in one of the two towers, measurement can begin after thermostatization. Density can be determined in a fully automatic process by running the user-defined program steps of the evaluation software, and is based on the weight values and other relevant parameters that are measured by positioning the test objects and weights at different heights. The result of one measurement series is the calculated density or volume along with an extensive report that documents the individual measurements.

The measurement units used to determine the weight values based on a modified Sartorius CC500 mass comparator for the liquid density and a modified Sartorius CWZ1005 mass comparators for the solid density. Both were adapted for this specific purpose by Sartorius, whereby the mass comparator for density determination of solids was subject to particular adaptation so that a weight value of

up to 1 kg can be measured with a resolution of 10 µg. The liquid temperature is determined via two resistance bridges, each with two 25-ohm platinum resistance sensors set at 0.1 mK. The weights are positioned directly (without a pallet) on the suspension system or the lift cylinder rack.

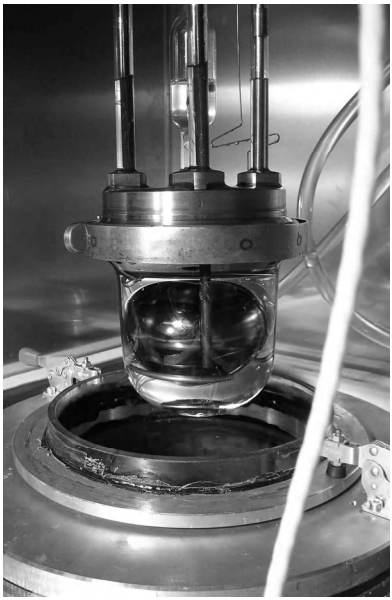


Figure 4. *measurement vessel with a bell-shaped chamber holding the silicon sphere before the thermostatized bath is raised*

A barred, bell-shaped measuring vessel with a standard- silicon sphere inside a suspension system is the main component used to determine the density of liquids. It has a capacity of 650 ml and, for insulation reasons, is connected to the outer environment only by glass tubes. These tubes are used for temperature measurement, suspending and manipulating the silicon sphere inside and filling/emptying the bell chamber. Individual screw fasteners for the glass tubes have been specially developed for this purpose. In addition, a method had to be found that would ensure the reproducible build-up of the liquid meniscus around the suspension wire of the mass comparator. The temperature range in which density can be determined is from 5°C to 60°C. Particular consideration was given to the problems of thermal stratification and condensation.

Figure 5. *Weight holder: lifting cylinder with mass comparator suspension system*

Test process of the fundamental aperture

The standard weights, standard plummets and test objects are manually positioned in the appropriate areas. To measure the density of a liquid, the plummet with the bell-shaped chamber is closed and filled with the test liquid. The corresponding thermostatized bath is raised, and the system is thermostatized. All necessary basic data and the preselected test routine are entered on the computer in preparation for measurement.

If the system is balanced sufficiently, measurement may begin. Only those measurement series that are determined under optimum conditions are used for final evaluation; here, the standard deviations obtained for the weight values and temperature measurement are compared with predefined tolerance levels. All individual measurements are recorded along with the results for each measurement series, and test/calibration certificates are recorded and printed for the test object measured.

Measurement results with the fundamental aperture

Measurement series with already calibrated BEV test weights and plummets have been carried out at different temperatures since summer 2005 to validate the system. Excellent results have been achieved in terms of reproducibility and standard deviation in international comparisons. For final validation, international comparative measurements have been taking place on a bilateral basis with the national metrology institutes of Hungary and Germany (MKEH (former OMH) and PTB). The system is already being used for disseminating density and volume at

BEV for liquids and solids up to 1 kg, it is used continuously in internal and external tests and calibrations and has been implemented in the BEV quality system.

Figure 6. *System installed at BEV for the fundamental density determination*

Development of a volume comparator

High-precision measurement of the density of solids through comparative measurements with density standards, such as those carried out by the BEV with the fundamental aperture or with similar procedures by national metrology institutes are generally too time-consuming for a large number of samples, due to the complicated preparation and installation in measurement cells.

Beginning with September 2005, the information and knowledge collected from the project of the fundamental aperture have been used to formulate the basis of a similar density application concept for serial determination of the volume of weights with a mass ranging from 1 g up to 1 kg.

Once again it was possible to engage in a competent partnership with the IFT and Sartorius AG for the implementation of this project.

Figure 7. *Volume comparator VD1005: prototype at BEV*

Requirements and objectives

In this application, a load alternator is to be provided to enable the volume of test objects to be determined by direct comparison with standard weights and a plummet.

The aim was to enable hydrostatic density determination in accordance with OIML R111 (class E1 for multiple solid bodies or weights). The procedure is based on mass comparison, as opposed to weighing, with specific steps taken to ensure that the temperature in the measurement chamber is as stable as possible. The result is a testing facility that analyzes the density of up to eight weights ranging from 1 g to 1 kg. With this approach, uncertainty of measurement is not related to mechanical factors, and thus depends on ambient conditions and the quality of the measuring instruments.

System design

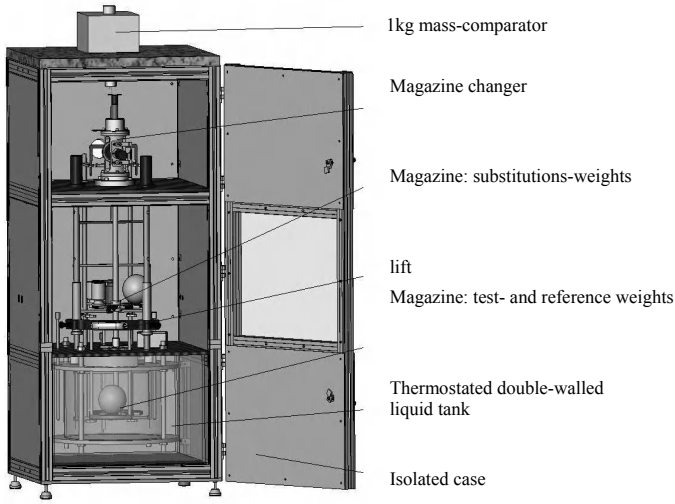


Figure 8. *CAD model of the volume comparator*

The testing facility is constructed in the form of a tower. A double-walled storage container for liquids is installed on the lowest level. The outer layer contains normal water and serves to regulate the temperature; it is supplied and regulated by an external thermostat. Baffle plates provide for a homogenous flow of water around the internal container at all times. A stable temperature is ensured by means of a 40 mm layer of insulation around the entire system and the use of insulated metal inserters. Additionally, the insulation of the individual levels of the testing facility prevents convection currents around the mass comparator. The inner vessel – the measurement cell – has been kept as small as possible in order to use as little as possible of the highly purified measuring liquid and, with the exception of a small entry portal, it is completely surrounded by the liquid in the outer layer. Even regulation of the temperature on all sides makes it possible to reduce the formation of temperature layers to a negligible level. The temperature of the measuring liquid is monitored by two diagonally mounted high-precision sensors (25 ohm standard platinum-resistant thermometers (SPRT) in compliance with ITS 90).

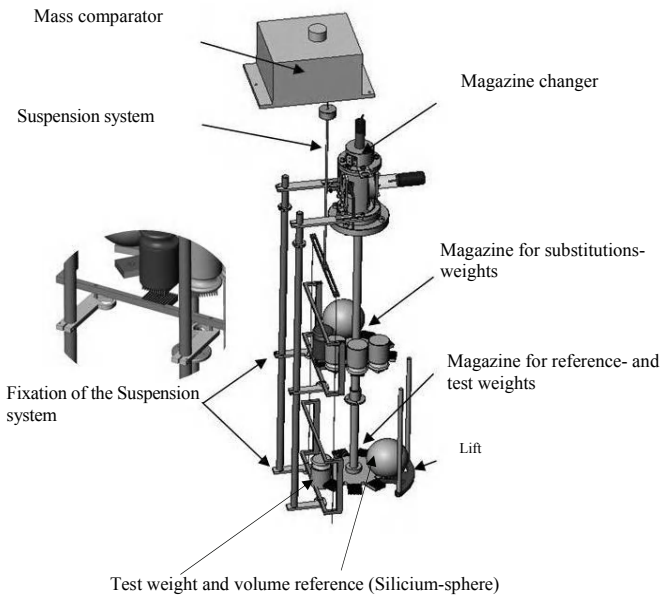


Figure 9. *CAD model of the transport unit*

The used mass comparator is the well-tried modified Sartorius CWZ1005 with a possible dead load of up to 400 g. The maximum load is 1 kg with a resolution of 10 µg.

For both the substitution weights and the test weight a position below the mass comparator was chosen. This eliminates off-center loading problems on the weigh cell. The specially developed load alternator makes it possible to insert the 1 g to 1 kg weights and density reference spheres with no rearrangement required, and to position them precisely in the suspension device of the mass comparator. The insertion mechanism positions the weights on the magazine spaces in the measuring fluid vessel.

The entire control system as well as all electronic components are housed in a control cabinet. The system is controlled by a computer, which is also used to evaluate the data. In order to determine environmental parameters the unit has sensors for air pressure, humidity and air temperature. A number of additional sensors are provided in order to exclude errors in the measuring process.

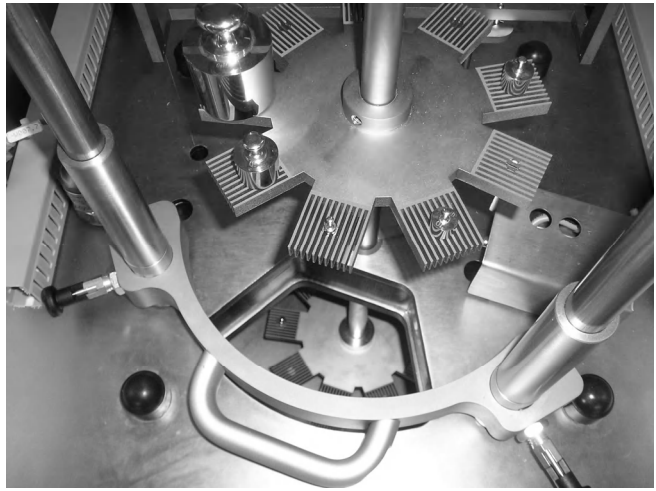


Figure 10. *Magazine for substitution weights, insertion mechanism, and view from above to the magazine in the weighing chamber*

Measurement procedure

Unlike other systems, this unit uses a 1 kg mass comparator to compare each test weight directly with only one volume reference. Each measurement run is used to compare the weight value of the test weight with the weight value of the standard weight. This means, however, that either the test weights must have a reference of their own with a similar mass, or the measured weight of each object must be corrected through reference to a different mass by using substitution weights in air. This unit realizes both methods described above. Thus it is possible to adapt test weights with different masses directly to a volume reference (e.g., a silicon sphere). Although the density of the measurement fluid as a transmitter is constantly monitored, it is of secondary importance due to this mass dissemination capability. It is important, however, to make sure that the density remains constant during the course of a measurement (approximately 2 minutes).

The software's user interface prompts the operator to prepare the measurement inserting the reference weights, the test weights and the appropriate substitution weights whereas a database can be used. This entails use of the weight insertion mechanism; it is important at this point to make sure there are no air bubbles on the weights once they are immersed in the liquid. A total of nine spaces are available for the reference and test weights. Once the sequence and number of cycles have been entered, the measurement can begin.

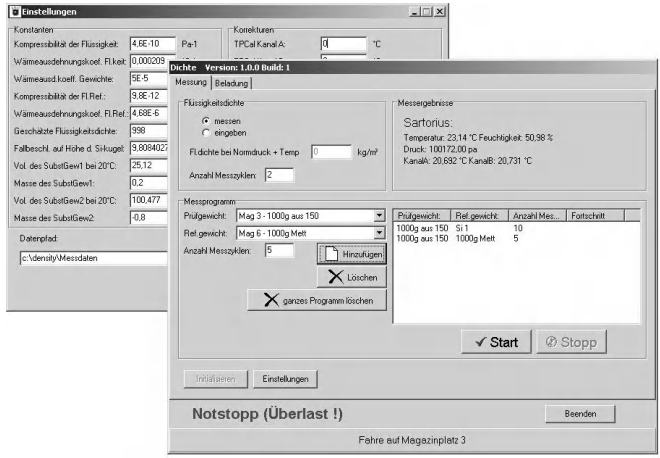


Figure 11. *Windows surface of the software*

Software

The software assists the operator in the preparation and execution of the measurements. In addition to the fully automatic measurement program, operators have the option of performing individual steps separately (single-step mode). In addition, the program shows the operator all the measurement data from the sensors in real time as well as the current progress of the measurement.

The result of each measurement is the volume, the density and the mass of each test weight and the standard deviation of the results, with complete documentation of the conditions of measurement and default values. Measurement values and all associated data can be provided as raw value output or on printouts that contain a complete evaluation of the measurement data.

In validation and testing of the initial system at BEV, in-house comparative measurements of the density of solid bodies – in this case with a sinker (Pyrex sphere; mass approx. 119 g, volume approx. 97 cm³) – with the new testing facility showed a uncertainty in volume of less than 0.9 mm³ and in mass of less than 0.07 mg. Comparisons with reference weights at BEV also showed excellent results. The prototype unit is already being used to measure the volume of weights with a mass of 1 g to 1 kg in the scope of mass determination at BEV, for ongoing in-house and external calibrations and it has been implemented in the BEV quality system.

Further units have been built for other metrology institutes and for accredited calibration laboratories whereas every unit is given an examination by the Physico-Technical Testing Service of the BEV,

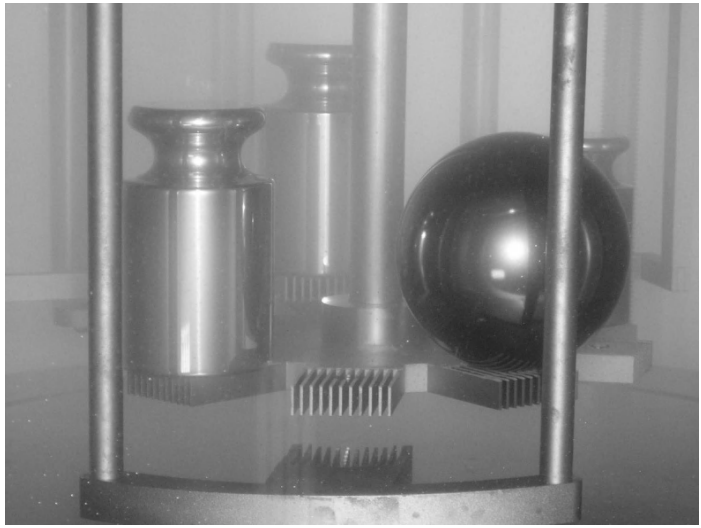


Figure 12. *Calibration of a client volume comparator in the BEV: comparison of a 1 kg, 500 g and 200 g test-weight with a national Volume standard (500 g silicon sphere Nr. 2) of the BEV*

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Use of balance calibration certificate to minimize measurement uncertainty in mass determinations performed in physico-chemical laboratories

Andrei Hoisescu M.Sc. Physics

National Institute of Metrology, Bucharest, Romania
Sos. Vitan Barzesti 11, sector 4, Bucharest
ahoisescu@yahoo.com

ABSTRACT: This article proposes a way of estimating measurement uncertainty for direct weighing using an electronic balance (in the physico-chemical field) by resorting to the general information provided by the calibration certificates, in order to achieve the most accurate results, with a minimum associated uncertainty. Adapting a formula to specific conditions of the laboratory can be a useful approach when the maximum possible accuracy is required.

Introduction

Mass determinations are very frequent procedures performed in physico-chemical laboratories, and the instruments most often used for these measurements are electronic balances.

Taking into account the wide area of applications in the "mass determination" field and the multiple use of weighing instruments, their calibration certificates provide generic information.

This turns over to the laboratory the responsibility of estimating weighing uncertainty for specific applications. Since the implementation of quality management systems, measuring uncertainty has become a key-element in demonstrating the laboratory's capability to provide reliable services and products that would meet the customer's demands.

Assumptions

The important assumption made is the fact that the certificates are released by accredited laboratories that follow the EA-10/18 guidelines. In this case, the certificates would respect a certain pattern and would provide the laboratories with a

formula for estimating the global uncertainty of weighing results. This formula covers several factors that may influence the accuracy of the measurement's result.

However, the certificate also provides the measurement results (usually repeatability, linearity and eccentricity), that the laboratory may use in order to create a formula that covers their specific conditions. Using correction factors and adapting a new formula would allow the laboratory to perform a more sensible evaluation of the weighing uncertainty.

Case report

In order to better illustrate this idea, we have resorted to an example [1]. The chosen instrument is a frequently used type of electronic balance (non-automatic weighing instrument) with a 200 g capacity and a scale interval of 0.1 mg.

The basic information provided by a certificate developed according to EA-10/18 guidelines contains the measurement results and the formula leading to an estimation of the value of direct measurement uncertainty.

The following tables are what is expected to be found in a calibration certificate related to the type of instrument described above.

Linearity / Errors of indication			
test loads each applied once; discontinuous loading only upwards, indication at no load reset to zero, where necessary; all loads in center of load receptor			
Load	Indication	*E*	*u*(E)*
30 g	30.0001 g	0.1 mg	0.16 mg
60 g	60.0003 g	0.3 mg	0.17 mg
100 g	100.0004 g	0.4 mg	0.17 mg
150 g	150.0006 g	0.6 mg	0.21 mg
200 g	200.0009 g	0.9 mg	0.23 mg

*or, in other cases, *U*(E) accompanied by the indication of the expansion factor, *k*

Repeatability	
Load = 100 g	
Measuring	Indication
No. 1	100.0002 g
No. 2	99.9999 g
No. 3	100.0001 g
No. 4	100.0000 g
No. 5	100.0002 g
No. 6	100.0002 g
Repeatability s = 0.13 mg	

The formula provided is, in general, the global uncertainty of weighing result without correction to the reading, as in the following example:

$$U_{gl}(\mathrm{W}) = 0.27\ \mathrm{mg} + 7.13\times10^{-6}R$$

at a level of confidence of 95% (k = 2), where R is the reading (in mg), without correction.

Considering loads of 100 g and 200 g, the following uncertainties are obtained.

Load	U(W), k=2
100 g	1.0 mg
200 g	1.7 mg

This formula includes several possible sources of uncertainty, covering all important aspects, which makes it extremely practical and secure.

The use of this formula is essential if we take into account the fact that during the normal usage of an instrument, the situation may differ from that at calibration in some of the following aspects [2]:

– the indications obtained for weighing bodies are not the ones at calibration;
– the weighing process may be different from the procedure at calibration:
  a. only one reading for each load, instead of several readings to obtain a mean value;
  b. reading to the scale interval d of the instrument, not with higher resolution;
  c. loading up and down, not only upwards;
  d. load kept on load receptor for a longer time, not unloading for each loading step, or vice versa;
  e. eccentric application of the load;
  f. use of tare balancing device, etc.;
– the environment (temperature, barometric pressure, etc.) may be different;
– on instruments that are not readjusted regularly (e.g. by use of a built-in device), the adjustment may have changed, due to aging or to wear and tear.

However, there are situations in which the results obtained by applying this formula are not sufficient, or a better uncertainty value has to be obtained. In this case, the laboratory may use the measurement results provided by the certificate in order to create its own formula, adapted to its working conditions.

Dispensable uncertainty sources

The potential uncertainty sources we can attempt to eliminate are some of those mentioned above.

Elimination of these sources may be performed in some cases with a minimum effort:

– Ensuring that the load to be weighed is applied in the center of the load receptor would eliminate the uncertainty due to the eccentric loading.

Other factors that may be ruled out are environmental factors:

– Uncertainties related to variations of temperature can be omitted if the instrument's adjustment is performed before the measurement. This adjustment will also eliminate any changes that may have occurred due to aging or wear and tear.

– Also, related to environmental factors, no correction should be applied for buoyancy, unless a variability of the air density larger than that at calibration is expected [3].

– Possible effects of creep and hysteresis are uncertainty sources when the loading at calibration was continuously upwards, or continuously up and downwards, so that the load remained on the load receptor for a certain period of time.
These effects can be ignored if the calibration is performed respecting the pattern by which the instrument is used (e.g. the test for errors of indication is done by applying the test loads increasingly by steps with unloading between the separate steps – corresponding to the majority of uses of the instruments for weighing single loads).

– Another possible source of uncertainties can be ruled out by giving up the use of the *Tare* function, when possible.

An important step in this approach is applying the correction to the reading. Therefore the weighing result is:

$$W = R - E,$$

where R is the reading (indication obtained) and E is the error of indication given in the calibration certificate.

Results

Assuming that all conditions are met to allow the sources above to be eliminated, the formula to assess the uncertainty of the weighing result would become:

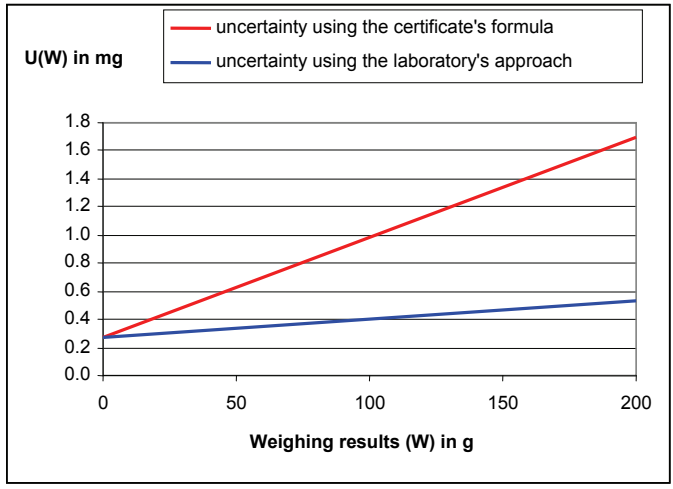
$$u(W)=\sqrt{\{u^2(R) + u^2(E)\}},$$

where $u^2(R) = d^2/6 + s^2$ and $u^2(E)$ is calculated from the information given in the calibration certificate.

Applying this to our example, the following results are obtained.

Load	U(W), k=2
100 g	0.4 mg
200 g	0.5 mg

The following graph shows the differences between applying the global uncertainty formula provided by the calibration certificate and the use of a formula adapted to certain laboratories' special conditions.



Conclusions

While the global uncertainty formula is a very practical and useful approach, under certain circumstances its results may not be accurate enough.

In these situations, based on information from the calibration certificate, laboratories may create their own specific formula, depending on the conditions they

have complied with, that could result in a more sensible value of estimated uncertainty.

An essential aspect of this process is calibrating the weighing instrument in a manner as similar as possible to the way that it is regularly being used.

References

[1] EA-10/18 EA Guidelines on the calibration of non-automatic weighing instruments, 2005, App G, pp. 73-76.

[2] EA-10/18 EA Guidelines on the calibration of non-automatic weighing instruments, 2005, ch. 7., pp. 30-31.

[3] EURACHEM/CITAC Guide CG4 Quantifying Uncertainty in Analytical Measurement, Second Edition, 2000, pp 37, 115.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Optic – Time Frequency

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Experimental study of the homogenity of a polychromatic light spectrum

J.E. Rakotoniaina, R. Malfara and Y. Bailly

Institut FEMTO ST CREST, CNRS UMR 6174
Parc Technologique, 2 avenue Jean Moulin-90000 Belfort-France

ABSTRACT: This study is a contribution to the development of an optical three-dimensional flow visualization and velocimetry technique using a color coding process. Previous works have been carried out in order to characterize the polychromatic light volume required to the use of the topical measurement method. The present paper is intended to initiate a study allowing us to search for further information about the description of the light spectrum, above all, the homogenity of the spectral distribution.

Introduction

This study is a contribution to the development of an optical three-dimensional flow visualization and velocimetry technique. Flow measurements usually require seeding of the flow with tracer-particles in order to describe its behavior. Designed to detect these particles in an illuminated measurement scene, the topical technique relies first on a standard optical observation system using a single camera which provides a 2D coordinate position. Then, the third spatial coordinate, the depth coordinate, in addition to the two first planar coordinates, is defined thanks to a color coding process using a polychromatic light spectrum. Thus, the crucial feature of this optical measurement method, then, called "RVV" or "Rainbow Volumic Velocimetry", consists of the fact that a 2D2C configuration results in a 3D3C one.

During the first years of development of the RVV technique, much of the work has been done. As a matter of fact, the experimental set-up is well established, above all, the for ready-to-use optical line dedicated to the generation of the light spectrum [1]. Then, the measurement calibration, central to the correspondence between colorimetric information and spatial depth coordinate, was carried out [2], [3]. Besides the experimental process, automated and optimized image processing tools and exploitation method have been established and have proved to be effective for investigating in flow study [4], [5].

However, even if these previous works are on the way to provide definitively good results, more improvements would still lead to an increase in the accuracy of the RVV technique. In this order, the present paper goes further than the previous ones into the description and the spatial stability of the spectral distribution of the polychromatic light volume. It is organized as follows. The first part will briefly review the principle of the characterization process and the resulting description of

the light spectrum, based on a color-space correlation model. The second part will feature the influence of the position (horizontal) along the light propagation direction, meaning also the width of the rainbow light volume, on the spectral distribution. The closing section will focus on the prospects and open questions.

Light spectrum characteristics

Given the characterization of the light spectrum has been undertaken previously, this section intends only to recall the operation philosophy and the typical description results.

As is mentioned in the above presentation, the major feature of the RVV technique is the detection of the particle depth-position (z) by means of color, produced by a polychromatic illumination, and, therefore, the concept of "color-meter" by setting a direct correspondence between chromatic and spatial data (Figure 1). Therefore, in the practical purpose, the observation direction, directed towards the depth axis (z), is parallel to the wavelength gradient, that is, the color variation axis of the rainbow light volume (Figure 1) is put forward; thus, if (x, y) represent the coordinates in the observation plane (camera), we define the orthonormal reference based (x, y, z) coordinate system.

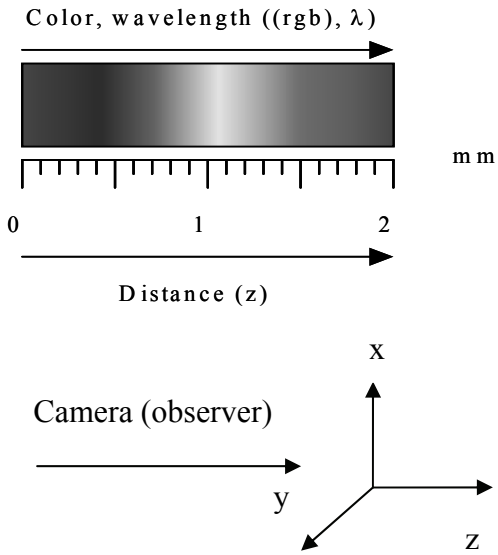


Figure 1. *Principle of the color coding measurement*

Both the x-y spatial planar coordinates and the colorimetric information (R, G, B) of the reflected light from a given particle are recorded by a 3CCD camera which is at the end of the optical measurement line. The latter is also composed with a white light source, a light dispersion device and a beam light spreading-adjusting optical system generating the rainbow light volume [1].

We must note that in the experimental characterization process, particles, instead of being introduced in a flow, are fixed into a transparent geometrical object. Then, by moving this latter, a chosen particle goes through the rainbow light volume along the color variation axis step by step, leading to a set of reference colorimetric coordinates corresponding to depth position values. For a suitable use of colorimetric values regarding the subsequent flow measurement, the (R, G, B) values are replaced by the normalized (r, g, b) ones.

When plotted against the spatial depth position (z), the colorimetric coordinates (r, g, b) show a Gaussian trend (Figure 2).

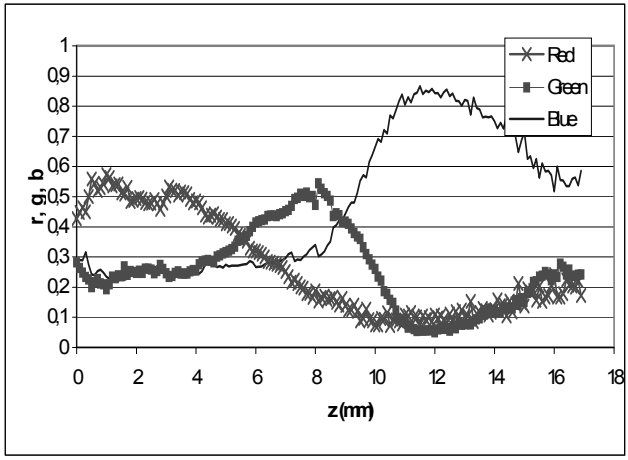


Figure 2. *(r, g, b) variation with respect to the depth coordinate (z)*

Since disturbance phenomenon (noise, etc.) appears at the two ends of the depth axis of the light spectrum, the region of interest is restricted to that comprised between the red and blue curves peaks, covering an interval of about 9 mm.

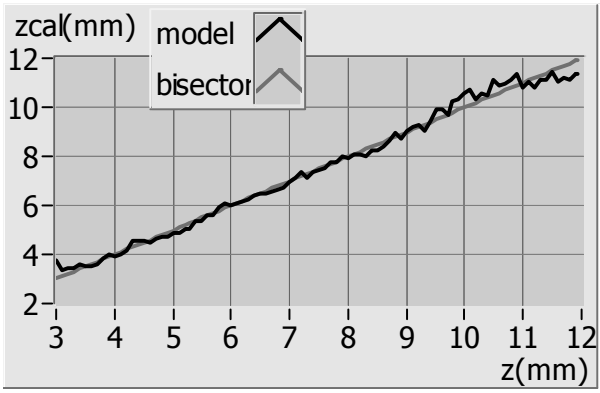


Figure 3. *Approximative depth coordinate values (model of equation 1) against actual values and the bisector curve along one depth axis*

A simple linear model has been introduced as a possible representation of the relationship between colorimetric and spatial depth data (equation (1)):

$$z = a_0 + a_1 r + a_2 g + a_3 b \tag{1}$$

where a_0, a_1, a_2 and a_3 are least squares fitting coefficients.

The results obtained by this empirical law are shown below and approach the bisector curve which represent the actual values of depth position with a mean relative error of 4% (Figure 3).

If we conjecture that the continuous illumination spectrum has a rigorously uni-dimensional chromaticity, then, the above model (equation 1) will yield unique coefficients for determining the depth position (z), whatever the position on x or y axis in the rainbow light volume. Actually, the spectral distribution in the polychromatic light volume turns out to display a non-strictly homogeneous feature as is illustrated in the next section.

Spatial stability of the spectral distribution in the rainbow light

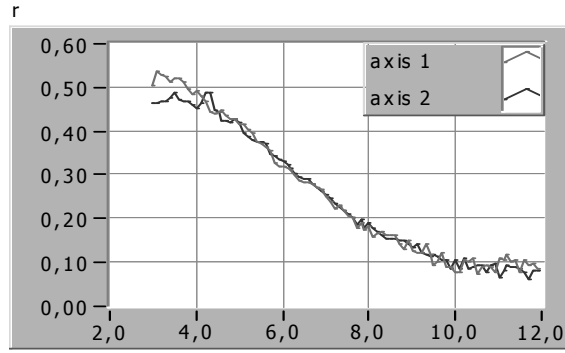
A relevant question arises from the exploration of the light spectrum: is the colorimetric information along the depth axis sensible to horizontal (y) or vertical (x) position ?

Full information about this subject cannot be given herein but the current section, focusing on the influence of horizontal position (y-axis) in the rainbow light, which is related to its width, proposes an initiating study.

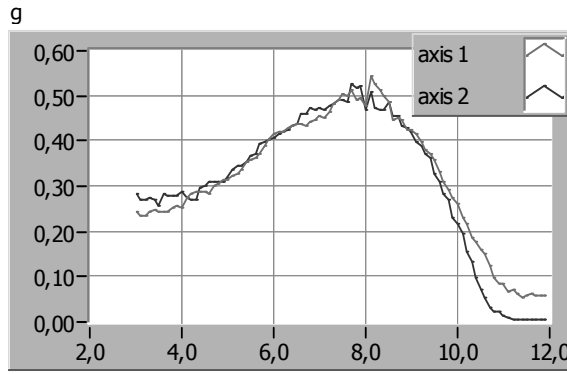
The suggested approach falls into three progressive levels.

Firstly, the issue of the homogenity of the spectrum is described by direct observation of the (r, g, b) values behavior. An immediate illustration of the influence of the horizontal y-axis position is delineated in Figure 4 (a, b, c), showing values recorded along two depth-axes of the rainbow light; the second axis is located at 1,4 mm from the first one, which is the reference axis of the light volume. It appears that at a same given value of the depth position, the corresponding colorimetric coordinates relative to the two axes remain close to each other over about 66% of the considered region, whereas a variation occurs at the two extreme areas, with an absolute difference mounting up to 0.11 for the blue colorimetric coordinate curve. The detected variation can be attributed to, among other possible causes, the fact that the reference axis is also the optical one of the vision system and it may cause light intensity reception to vary with angular position.

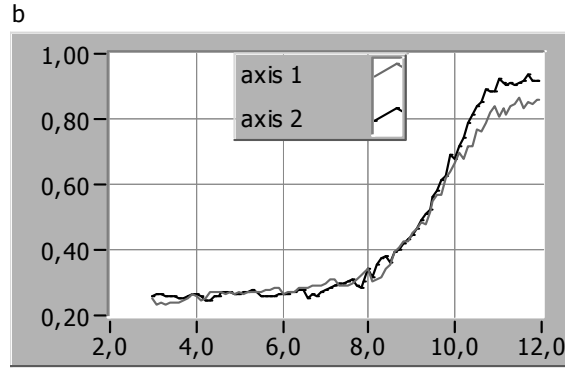
The above observation provides the undisputed evidence that if the light spectrum is seemingly stable over a major part of the color variation axis, it is horizontal position-dependent (y-axis).



4-a



4-b



4-c

Figure 4. *r, g, b values versus depth position (z) for two depth-axes with 1.4 mm inter-axial distance(Δy)*

Secondly, the influence of the horizontal position (y) on the spectral distribution can be revealed through the direct effect of (r, g, b) value variation on other colorimetric parameters such as the Euclidian colorimetric distance. Indeed, this parameter is defined as the distance between two points in the (r, g, b) color spatial representation system, and is formulated as follows (equation (2)):

$$E.D. = ((r_2 - r_1)^2 + (g_2 - g_1)^2 + (b_2 - b_1)^2)^{0.5} \quad (2)$$
$$E.D. = (\Delta r^2 + \Delta g^2 + \Delta b^2)^{0.5}$$

As is clearly put, this presentation appears to be an appropriate illustration of the issue of the homogenity of the rainbow light since it expresses a variation notion and, by so doing, allows us to simultaneously take into account the variation of the three pieces of chromatic data.

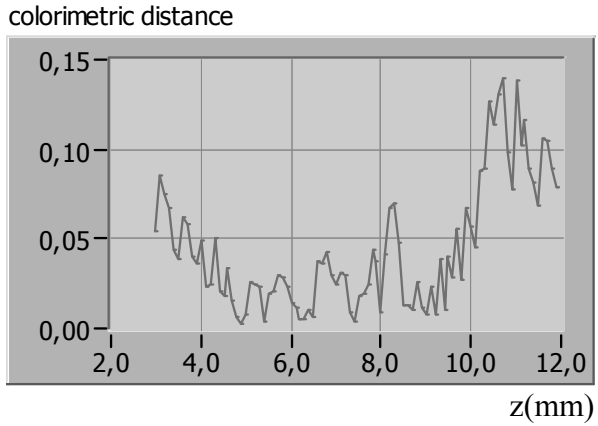


Figure 5. *Euclidian colorimetric distance between two depth-axes along the color axis variation*

Figure 5 features the results of the application of this approach to the previous depth-axes. It confirms and summarizes what is separately shown in the first presentation level: the colorimetric distance remains low on a major part of the region of interest and then increases markedly at its ends, the maximum occurring in the "blue region" (right-end of the depth-axis).

Figure 6 illustrates the influence of the horizontal y-axis position or the width of the rainbow light volume on the colorimetric distance (mean and maximum value).

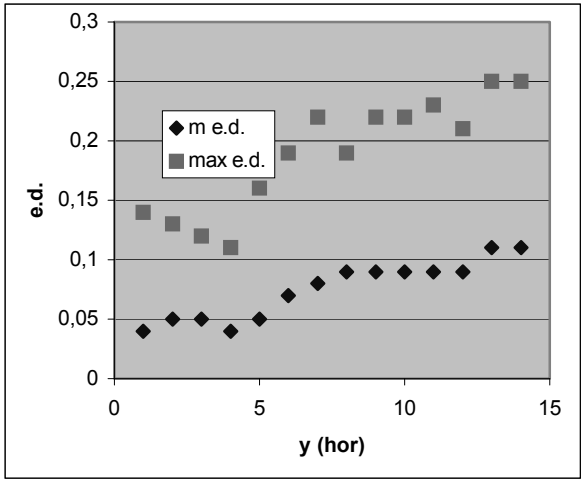


Figure 6. *Colorimetric distance versus depth-axis number along the horizontal y-axis*

The depth-axis position is obtained by multiplying its number on the y-axis by the inter-axial distance (Δy=1.4 mm). We can note the colorimetric distance rises with the increase of the value of y position; the maximum value rises by about 78% when the depth axis is located at 20 mm from the reference axis (at y = 0 mm).

Figure 6 also shows a possible subdivision of the exploration domain into almost four regions: 0 mm<y<7 mm (slightly stable spectral distribution), 7 mm<y<11 mm (transition and variation), 11 mm<y<17 mm (steady), and 17 mm<y<20 mm (quite steady).

This second observation based on the colorimetric distance demonstrates conclusively the inhomogenous feature of the polychromatic light spectrum.

The last indication level of the influence of the horizontal y-axis position on the spectral distribution is that inspired of the model expressed in equation 1. This model allows us to relate the spatial depth coordinates to the colorimetric coordinates by means of a linear relationship.

Hence, variation of (r, g, b) values along the horizontal y-axis gives rise directly to that of the depth coordinate ones. When applied to the two previous depth-axes, with an inter-axial distance (Δy) of 1.4 mm, this model (equation 1) gives a mean variation of 5% (Figure 7). The same feature as in Figures 5 and 6 is also illustrated: variation occurs at the ends of the depth axis.

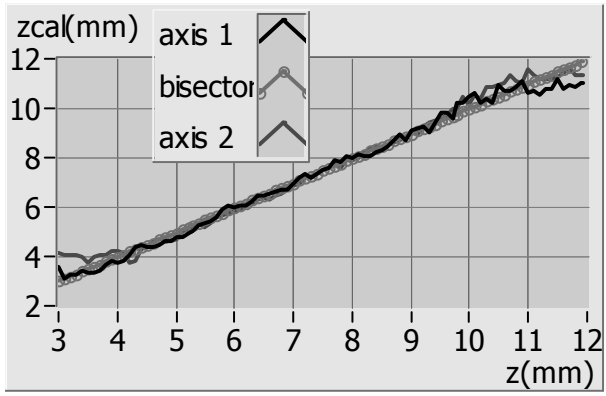


Figure 7. *Approximative depth coordinate values (model of equation 1) along two depth axes (Δy=1.4 mm) against actual values and the bisector curve*

Conclusion

This study shows that the continuous polychromatic light spectrum turns to display an inhomogenous color distribution along the light propagation direction. In addition to the description only based on colorimetric representation, it introduces the empirical model relating chromatic coordinates to spatial depth coordinates which is instrumental in delineating the limits of the stability region of the spectral distribution. However, some other aspects of this subject still have yet to be answered. How about the influence of the height of the rainbow light upon the spectral distribution, the adequacy of the suggested approach when changing light intensity, etc.?

References

[1] M. Gbamelé, Analyse quantitative d'écoulements complexes par imagerie polychromatique, Ph.D. Thesis, University de Franche-Comté at Belfort, 2001.

[2] J.E Rakotoniaina., D.Zibret and Y. Bailly, "Caractérisation tridimensionnelle d'un spectre de lumière polychromatique continu", in *Actes du colloque interdisciplinaire en instrumentation*, ENS Cachan, 2004, pp. 99-106.

[3] J.E Rakotoniaina., D.Zibret, Y. Bailly and R. Malfara, "Calibration technique of a color coding optical 3D measurement", in *CD-ROM of Proceedings of 7th Conference on Optical 3D measurement technique*, Vienna, 2005.

[4] R. Malfara, Y. Bailly, D. Zibret, C. Cudel, J.P. Prenel, "Utilisation d'un algorithme de Ransac pour le traitement des images obtenues par vélocimétrie arc-en-ciel", *5th Symposium of the francophone club SFO/CMOI*, St-Etienne, France, 2004.

[5] D. Zibret, Y. Bailly, R. Malfara, J.P. Prenel, "Unsteady flow in the wake behind a circular cylinder by means of rainbow volumic velocimetry", *11th International Symposium of Flow Visualization*, Notre-Dame, USA, 9-12 August 2004.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Statistics

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

An Innovative Method for the Comparison of Measurement Results

Alessandro Ferrero and Simona Salicone

Dipartimento di Elettrotecnica - Politecnico di Milano
Piazza Leonardo da Vinci 32 - 20133 Milan - Italy

ABSTRACT. Measurement results are very often used as the starting point for decision-making processes, whenever a suitable action must be pursued according to the measured values. This requires the comparison of measurement results with each other or with a given threshold. The result of this comparison is useful only if measurement uncertainty is considered. Unfortunately, few and unsatisfactory indications are available on how to take into account uncertainty in a comparison. This paper proposes an original approach, based on the Theory of Evidence, for comparing measurement results when they are expressed in terms of Random-Fuzzy Variables.

Introduction

It is well known that any measurement result is only an approximation of the value of the measurand, and can therefore be usefully employed only if a quantitative indication is given about how approximate it is.

To this purpose, the IEC-ISO Guide to the Expression of Uncertainty in Measurement (GUM) [IEC 92] states that a parameter for qualifying the measurement result must be always provided, associated with the measurement result itself. This parameter is the well known standard uncertainty and represents "*the dispersion of the values that could reasonably be attributed to the measurand*" [IEC 92].

One of the most important fields in which uncertainty plays a critical role is that of decision-making processes. Such processes are, in general, the final goal of every measurement procedure as, for instance, in the final tests of industrial products, to assess whether they stay within the design specifications or not, or in biomedical analysis, to assess whether a person is affected by a pathology or not, or in environmental analysis, to assess whether an area is safe for human beings or not.

All above actions require us to compare measurement results with each other or with a given threshold. This comparison is quite immediate if measurement

uncertainty is not taken into account, but, in this case, it might lead to incorrect decisions that, in critical applications, such as biomedical or environmental analysis, may even endanger life.

Despite decisions are almost always performed at the end of each measurement process, sometimes they could also be required within the measurement process itself. In fact, the techniques of digital signal processing (DSP) allow the implementation of measurement algorithms which may also contain *if … then … else …* structures. In this case, the *if …* step requires us to perform a comparison: if *measurement A is greater than measurement B*, do something, otherwise, do something else.

Very few indications are presently available on how to consider uncertainty in measurement result comparison. Recommendations on how to prove conformance or non-conformance with specifications are given in ISO EN 14253 Standard [ISO 98]. However, this document considers only the particular case of the comparison of a measurement result with a threshold, but does not consider the general case of the comparison of two measurement results. Even in this particular case, the considered solution is not totally satisfactory, since an ambiguity zone is left, whose width depends on the expanded uncertainty, for which no decision can be taken.

More recently, several authors have proposed to represent measurement results, together with the associated uncertainty, with fuzzy [URB 03, MAU 00, MAU 01, MAU 06] or random-fuzzy [FER 04, FER 06, FER 05, SAL 07] variables. One of the advantages of this approach is that it opens the door to an interesting method for comparing measurement results in the presence of uncertainty. This paper, after briefly recalling the fundamentals of the fuzzy approach, will analyze this new comparison method and propose a possible decision rule.

Fuzzy and Random-Fuzzy Variables

Fuzzy (FV) and Random-Fuzzy (RFV) variables can be defined, within the Theory of Evidence, in different equivalent ways [SAL 07, FER 05]. The most useful way to define FVs and RFVs, when they are used for representing measurement results, is in terms of a set of confidence intervals, for all possible levels of confidence.

In fact, the membership function (MF) $\mu_X(x)$ of a FV can be defined [SAL 07] in terms of its α-cuts:

$$X_\alpha = \left\{ x \mid \mu_X(x) \geq \alpha \right\} \tag{1}$$

It has been proved that each α-cut represents a confidence interval of type 1, with a level of confidence $p = 1 - \alpha$ [SAL 07]. Therefore, a FV defines the intervals of confidence at all possible levels of confidence in [0,1], within which the measurand value is supposed to lie.

Different methods are available to obtain such intervals. For example, following the GUM approach [IEC 92], that requires us to identify and compensate all systematic effects, the measurement result is supposed to be affected only by random contributions to uncertainty and can be represented by a probability density function. Consequently, each confidence interval can be determined by applying a suitable coverage factor to the standard uncertainty associated with the measurement result [IEC 92].

On the other hand, non-random contributions to uncertainty also cause the measurand value to lie in an interval about the measurement result, and a suitable confidence level can again be estimated for each given interval.

According to the different contributions represented by the defined confidence intervals, different mathematical approaches must be followed to combine them, as proved in [SAL 07]. The limit of FVs, when they are used to represent measurement results, is that they allow us to process and combine different measurement results only when they are affected by the same kind of contributions to uncertainty (random or non-random) [SAL 07, FER 06, FER 05].

Since, in general, random and non-random contributions are simultaneously present, RFVs have been defined to correctly take into account both phenomena. Similarly to FVs, an RFV can be defined as a set of intervals of confidence of type 2 [SAL 07]. These intervals are confidence intervals, whose bounds are uncertain, and are therefore represented by closed intervals themselves, instead of ordinary numbers. Each α-cut B_α of an RFV is therefore given by four ordered numbers:

$$B_\alpha = \left[b_1^\alpha, b_2^\alpha, b_3^\alpha, b_4^\alpha\right] \tag{2}$$

that obey the following constraints [SAL 07]:

- $b_1^\alpha \leq b_2^\alpha \leq b_3^\alpha \leq b_4^\alpha, \forall\alpha$;
- the sequences of intervals of confidence of type 1 $[b_1^\alpha, b_4^\alpha]$ and $[b_2^\alpha, b_3^\alpha]$ generate two membership functions (MF) that are normal and convex;
- $\forall\alpha, \alpha'$ in the range $[0, 1]$:

$$\alpha' > \alpha \Longrightarrow \begin{cases} \left[b_1^{\alpha'}, b_3^{\alpha'}\right] \subset \left[b_1^\alpha, b_3^\alpha\right] \\ \left[b_2^{\alpha'}, b_4^{\alpha'}\right] \subset \left[b_2^\alpha, b_4^\alpha\right] \end{cases} ;$$

- $[b_2^{\alpha=1}, b_3^{\alpha=1}] \equiv [b_1^{\alpha=1}, b_4^{\alpha=1}]$.

Two MFs are hence obtained, as shown in Figure 1, where the α-cut at level $\alpha = 0.3$ is also reported, providing the confidence interval at the level of confidence $p = 1 - \alpha = 0.7$.

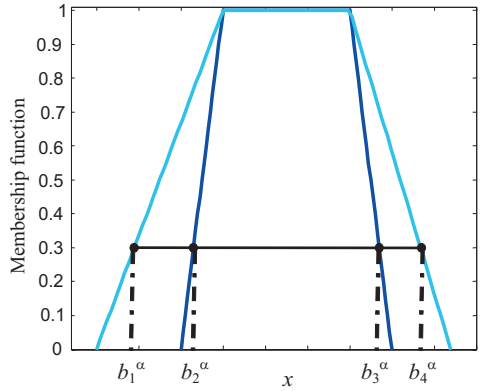


Figure 1. *Example of RFV and determination of the confidence interval at confidence level 0.7*

It has been proven [SAL 07, FER 05] that RFVs can be effectively employed to represent measurement results and that the impact of the different contributions to uncertainty can be retrieved from the analysis of the three sub-intervals defined by the two membership functions. Interval $[b_2^{\alpha}, b_3^{\alpha}]$ is related to non-random contributions, while intervals $[b_1^{\alpha}, b_2^{\alpha}]$ and $[b_3^{\alpha}, b_4^{\alpha}]$ are related to random contributions [SAL 07, FER 05]. Therefore, RFVs can be considered as a generalization of FVs also under the metrology point of view.

Mathematical tools

The first step in the comparison procedure of two measurement results, when they are represented in terms of RFVs, is to define an ordering rule for stating if an RFV is lower, equal or greater than another one, and assess how credible this statement is.

In the most general case, the ordering rule should consider both the internal and external MFs of the RFV, so that the impact of the different contributions to uncertainty on the decision can be analyzed. However, when this analysis is not required, as in most cases, when only the overall confidence interval within which the measurand value is supposed to lie is considered, it is sufficient to consider only the external MF of the RFV. In the following, only this MF will be considered, for the sake of clarity, though all mathematical derivations can be straightforwardly applied also to the internal one, if necessary.

Different methods are available, in the literature, for the comparison of two FVs [KLI 95, FOR 96], though they are not totally satisfying when metrological issues have to be solved.

Similarly to Kerre and Fortemps and Roubens methods [KLI 95, FOR 96], the proposed method considers suitable areas subtended by the MFs and is based on the Hamming distance and the fuzzy-max operator, whose definitions are recalled here.

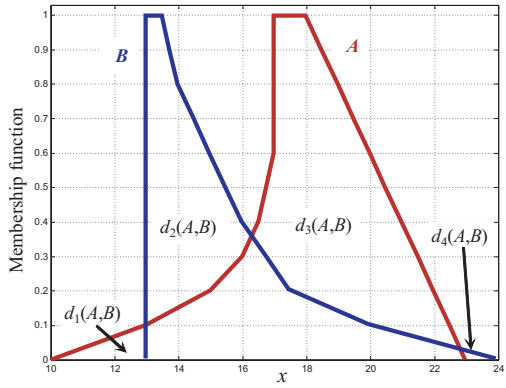


Figure 2. *Fuzzy variables A and B considered in the comparison process. The different areas considered by the method are shown*

Given two FVs A and B, such as the ones in the example of Figure 2, the Hamming distance is defined as:

$$d(A,B) = \int_{-\infty}^{+\infty} \big|\mu_A(x) - \mu_B(x)\big| dx, \tag{3}$$

and the fuzzy-max operator $\mathrm{MAX}(A,B)$ is defined by:

$$\mu_{\mathrm{MAX}(A,B)}(z) = \sup_{z=\max(x,y)} \min\big[\mu_A(x), \mu_B(y)\big], \tag{4}$$

$\forall x, y, z \in X$. Equation (4) provides the MF of the FV shown in Figure 3 where (4) has been applied to the two FVs A and B in Figure 2.

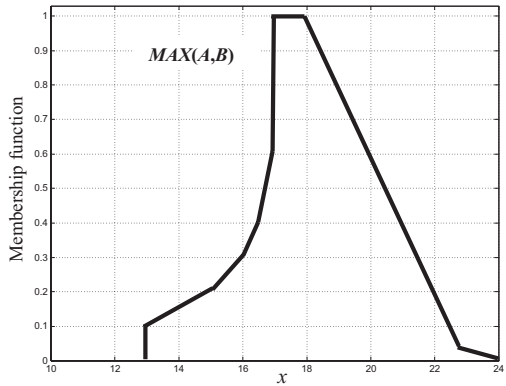


Figure 3. *Fuzzy variable obtained by applying the fuzzy-max operator to variables A and B in Figure 2*

The Hamming distance between two FVs is generally given by the sum of the four areas $d_1(A, B)$, $d_2(A, B)$, $d_3(A, B)$ and $d_4(A, B)$, shown in Figure 2. It can be readily proven [KLI 95] that the greater value of the Hamming distance, the smaller the overlap between the two variables, and the Hamming distance takes its maximum value when the two variables are completely separated. Therefore, $d(A, B)$ is an index of how much the two variables differ from each other.

It is also interesting to evaluate the Hamming distance between each considered FV A and B and the FV provided by $\mathrm{MAX}(A, B)$, that is:

$$d_A(A, B) = d\big(A, \mathrm{MAX}(A, B)\big); \tag{5}$$

$$d_B(A, B) = d\big(B, \mathrm{MAX}(A, B)\big). \tag{6}$$

It can be readily proven [SAL 07] that the above Hamming distances (5) and (6) can be obtained from the four areas shown in Figure 2 as:

$$d_A(A, B) = d_1(A, B) + d_4(A, B); \tag{7}$$

$$d_B(A, B) = d_2(A, B) + d_3(A, B); \tag{8}$$

The proposed approach

The four areas in Figure 2 can be interpreted as follows:

- $d_1(A, B)$ is an indicator of how much $A < B$;
- $d_2(A, B)$ is an indicator of how much $B < A$;
- $d_3(A, B)$ is an indicator of how much $A > B$;
- $d_4(A, B)$ is an indicator of how much $B > A$.

These areas can be pairwise joint, since both $d_1(A, B)$ and $d_4(A, B)$ are related on how much A is lower than B, and both $d_2(A, B)$ and $d_3(A, B)$ are related on how much A is greater than B.

Let us now consider the fuzzy intersection area and the fuzzy union area, defined as [KLI 95]:

$$\mathrm{Int}(A, B) = \int_{-\infty}^{+\infty} \min\big[\mu_A(x), \mu_B(x)\big]\, dx; \tag{9}$$

$$\mathrm{Un}(A, B) = \int_{-\infty}^{+\infty} \max\big[\mu_A(x), \mu_B(x)\big]\, dx. \tag{10}$$

By considering (3), (9) and (10), it can be readily proven that:

$$d(A, B) = \mathrm{Un}(A, B) - \mathrm{Int}(A, B)$$

that leads immediately to:

$$\mathrm{Un}(A,B) = d_1(A,B) + d_2(A,B) + d_3(A,B) + d_4(A,B) + \mathrm{Int}(A,B)$$

and, taking into account (7) and (8):

$$\mathrm{Un}(A,B) = d_A(A,B) + d_B(A,B) + \mathrm{Int}(A,B).$$

The following relationship is obtained, by referring all terms to $\mathrm{Un}(A,B)$:

$$\frac{d_A(A,B)}{\mathrm{Un}(A,B)} + \frac{d_B(A,B)}{\mathrm{Un}(A,B)} + \frac{\mathrm{Int}(A,B)}{\mathrm{Un}(A,B)} = 1 \tag{11}$$

It is now possible to define the following coefficients, all ranging in the $[0,1]$ interval:

$$C_{\mathrm{lo}}(A,B) = \frac{d_A(A,B)}{\mathrm{Un}(A,B)} = \frac{d_1(A,B) + d_4(A,B)}{\mathrm{Un}(A,B)} \tag{12}$$

$$C_{\mathrm{gr}}(A,B) = \frac{d_B(A,B)}{\mathrm{Un}(A,B)} = \frac{d_2(A,B) + d_3(A,B)}{\mathrm{Un}(A,B)} \tag{13}$$

$$C_{\mathrm{comp}}(A,B) = \frac{\mathrm{Int}(A,B)}{\mathrm{Un}(A,B)} \tag{14}$$

The following considerations apply:

• Coefficient (12) represents the credibility coefficient of statement: "A is lower than B". In fact, it is equal to one only when A is located on the left side of B and does not overlap with B. In this situation it is $A < B$ with full certainty. Moreover, (12) is equal to zero when A is located on the right side of B and does not overlap with B. In this situation it is surely $A > B$ and therefore statement "A is lower than B" is wrong with full certainty. For intermediate positions of A and B, (12) takes values between 0 and 1.

• Coefficient (13) represents the credibility coefficient of statement: "A is greater than B". In fact, it is equal to one only when A is located on the right side of B and does not overlap with B. In this situation it is $A > B$ with full certainty. Moreover, (13) is equal to zero when A is located on the left side of B and does not overlap with B. In this situation it is surely $A < B$ and therefore statement "A is greater than B" is wrong with full certainty. For intermediate positions of A and B, (13) takes values between 0 and 1.

• Coefficient (14) is not zero if and only if A and B do somehow overlap. It can be immediately recognized that the more they overlap, the greater $\mathrm{Int}(A,B)$ and the greater the value taken by (14). Moreover, (14) is equal to one only if $\mathrm{Int}(A,B) = \mathrm{Un}(A,B)$, which occurs only if $\mu_A(x) = \mu_B(x), \forall x$, that is, if the two considered variables are perfectly equal. Therefore, (14) represents the credibility coefficient of statement "A is compatible with B", or "A is equal to B". Of course, this last formulation should be used only when the considered FVs have equal MFs.

The above points lead to conclude that coefficients (12), (13) and (14) quantify the credibility of statements: "$A < B$", "$A > B$" and "$A = B$", respectively. It is worth noting that, due to (11), the three coefficients always add up to one, that represents, correctly, full certainty that either $A < B$, or $A > B$, or $A = B$. It can be concluded that the method fully exploits the available evidence and does not leave any ambiguity.

Once the credibility of these statements has been quantified, it is possible to define suitable ordering rules, supporting the decision that "$A < B$", or "$A > B$", or "$A = B$".

Despite the obtained credibility coefficients are unambiguous, different rules can be defined, according to the importance assigned to the credibility of one of the three statements. In other words, a suitable decision rule can be chosen on the basis of the amount of credibility required to take a decision by the considered application.

In many practical technical applications, it is often required to assess whether a quantity is greater or lower than another one, while there is little concern to state whether a quantity is equal to another one. In general, it is stated that two quantities are "equal", or, better, "compatible", when there is no available evidence that they differ for more than a given quantity.

These considerations lead to the following decision rules.

• If $C_{\mathrm{gr}}(A, B) > C_{\mathrm{lo}}(A, B)$ statement: "A is greater than B" holds. In this case, the value of $C_{\mathrm{gr}}(A, B)$ is the credibility factor with which the decision is taken.

• If $C_{\mathrm{lo}}(A, B) > C_{\mathrm{gr}}(A, B)$ statement: "A is lower than B" holds. In this case, the value of $C_{\mathrm{lo}}(A, B)$ is the credibility factor with which the decision is taken.

• If $C_{\mathrm{lo}}(A, B) = C_{\mathrm{gr}}(A, B)$ statement: "A is compatible with B" holds. In this case, the value of $C_{\mathrm{comp}}(A, B)$ is the credibility factor with which the decision is taken. Under practical situations, the above mathematical condition is never met. Therefore, condition: $|C_{gr}(A, B) - C_{lo}(A, B)| < \epsilon$ should be considered, where ϵ is a suitable positive constant, depending on the particular considered application. For instance, ϵ could be a suitable fraction of $\mathrm{Un}(A, B)$. Moreover, when this condition is met and $C_{\mathrm{comp}}(A, B) = 1$, it is possible to state that $A = B$.

If the above rules are applied to the FVs shown in Figure 2, coefficients (12), (13) and (14) take the following values:

- $C_{\mathrm{gr}}(A, B) = 0.725$;
- $C_{\mathrm{lo}}(A, B) = 0.032$;
- $C_{\mathrm{comp}}(A, B) = 0.243$.

Therefore, it can be stated that $A > B$, with a credibility factor $\lambda = C_{\mathrm{gr}}(A, B) = 0.725$.

Discussion

The method proposed in the above section provides unambiguous comparison of measurement results when they are expressed in terms of FVs or RFVs. A decision can always be taken, based on the comparison, and a credibility factor λ in the range $[0, 1]$ is associated to the decision.

The value of the credibility factor represents the amount of evidence supporting the decision. It is worth noting that the capability of taking a decision, with the given rules, does not depend on the absolute value of the credibility factor, but on the relative values of the three coefficients (12), (13) and (14). In the limit case, a decision can be taken also with a zero credibility factor, if no direct evidence supports the decision.

To fully perceive this statement, let us consider the two FVs A and B shown in Figure 4.

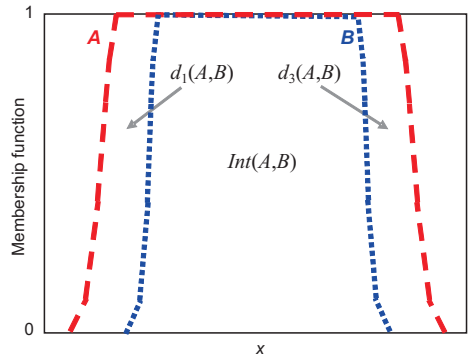


Figure 4. *Example of two compatible FVs*

It can be immediately recognized that FV B is perfectly centered in FV A. This is mathematically confirmed by the fact that $C_{\mathrm{gr}}(A, B) = C_{\mathrm{lo}}(A, B) = 0.15$. Therefore, it is possible to state that A is compatible with B, and the credibility of this statement is given by the credibility factor $\lambda = C_{\mathrm{comp}}(A, B) = 0.7$.

In this situation, the credibility factor with which the decision is taken is quite high. However, let us assume that the width is decreased of FV B, while keeping unchanged the location of its mean value. Areas $d_1(A, B)$ and $d_3(A, B)$ in Figure 4 increase, still remaining equal to each other, while area $\mathrm{Int}(A, B)$ decreases.

This means that $C_{\mathrm{gr}}(A, B)$ and $C_{\mathrm{lo}}(A, B)$ increase, still remaining equal to each other. Therefore, according to the above defined rules, we can still state that "A is compatible with B", although the credibility factor $C_{\mathrm{comp}}(A, B)$ decreases, being directly related to area $\mathrm{Int}(A, B)$.

In the limit case when FV B degenerates into a crisp value, which is the typical case when a single value threshold is considered in the comparison, and this value is located exactly in the middle of FV A, the given rules provide $C_{\mathrm{gr}}(A, B) = C_{\mathrm{lo}}(A, B) = 0.5$ and $C_{\mathrm{comp}}(A, B) = 0$.

In this case, the only possible decision is still "A is compatible with B", though the credibility of such a statement is nil. From a metrological point of view, this is absolutely acceptable, since there is no evidence supporting the statement that the result of a measurement is equal to a single crisp value. However, in this example, there is no evidence supporting the statement that FV A is greater or lower than the given threshold B, and therefore no other decision can be taken than stating that A is compatible with B.

Conclusions

An original method has been proposed to take into account measurement uncertainty while comparing measurement results.

The method is based on the representation of measurement results in terms of FVs or RFVs, since this representation has proved to be very effective in representing a measurement result together with its associated uncertainty, whatever the contributions (random or non-random) to uncertainty.

The major advantage of the proposed comparison method, over other available methods, and especially the one encompassed in [ISO 98], is that it always provides unambiguous results and associates a credibility factor, in the range $[0, 1]$, with the taken decision.

These credibility factors quantify the available evidence that supports a decision and could be used to analyze the implications of the taken decision further on. This can be extremely useful, for instance, to estimate the credibility of the final measurement result when it is provided by an algorithm with embedded *if* ... *then* ... *else* ... structures, or to estimate the risk of a decision taken on the basis of a measurement result.

Bibliography

[FER 04] FERRERO A. and SALICONE S., "The random-fuzzy variables: a new approach for the expression of uncertainty in measurement", *IEEE Trans. Instr. Meas.*, vol. 53, no. 5, p. 1370–1377, 2004.

[FER 05] FERRERO A. and SALICONE S., "The theory of evidence for the expression of uncertainty in measurement", *12 International Metrology Congress*, Lyon, France, June 20-23, 2005.

[FER 06] FERRERO A. and SALICONE S., "A fully-comprehensive mathematical approach to the expression of uncertainty in measurement", *IEEE Trans. Instr. Meas.*, vol. 55, no. 3, p. 706–712, 2006.

[FOR 96] FORTEMPS P. and ROUBENS M., "Ranking and defuzzification methods based on area compensation", *Fuzzy Sets and Systems*, vol. 82, no. 3, p. 319–330, 1996.

[IEC 92] IEC-ISO, Guide to the Expression of Uncertainty in Measurement, 1992.

[ISO 98] ISO EN 14253-1, Geometrical Product Specification (GPS) - Inspection by measurement of workpieces and measurement instruments - Part 1: Decision rules for proving conformance or non-conformance with specification, 1998.

[KLI 95] KLIR G. J. and YUAN B., *Fuzzy sets and fuzzy logic. Theory and applications*, Prentice Hall PTR, NJ, USA, 1995.

[MAU 00] MAURIS G., BERRAH L., FOULLOY L. and HAURAT A., "Fuzzy handling of measurement errors in instrumentation", *IEEE Trans. Instr. Meas.*, vol. 49, no. 1, p. 89–93, 2000.

[MAU 01] MAURIS G., LASSERRE V. and FOULLOY L., "A fuzzy approach for the expression of uncertainty in measurement", *Measurement*, vol. 29, p. 165–177, 2001.

[MAU 06] MAURIS G., "Possibility Expression of Measurement Uncertainty In a Very Limited Knowledge Context", *AMUEM 2006*, Sardagna, Trento, Italy, April 20-21, 2006, p. 19–22, 2006.

[SAL 07] SALICONE S., *Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence*, Springer series in reliability engineering, Springer, New York, NY, USA, 2007.

[URB 03] URBANSKY M. and WASOWSKI J., "Fuzzy Approach to the Theory of Measurement Inexactness", *Measurement*, vol. 34, p. 67–74, 2003.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Calibration and recalibration of measuring instruments: a Bayesian perspective

E. Mathioulakis and V. Belessiotis

Solar & other Energy Systems Laboratory, NCSR "DEMOKRITOS"
15310, Agia Paraskevi Attikis - Greece

ABSTRACT: The scope of this communication is to present an analysis of problems related to the calibration of the measuring instruments, based on the concept of conditional probabilities and the Bayes formula. Existing knowledge on the metrological behavior of the instrument is treated as prior information which will be combined with probabilistic information drawn from the calibration itself, in order to produce an updated state of knowledge concerning the response of the instrument. Thus, a probabilistic representation of the parameters defining the calibration curve is produced which considers the history of the instrument, as well as the uncertainties associated with the measurements taken during the calibration.

Introduction

Calibration is a key concept in metrology because it enables the qualitative and quantitative characterization of measurement results. It is also a vital activity in the field of tests, as it provides reliability required for any quality control of products and services.

The simple idea behind calibration is so that the response from a measuring instrument can be compared to the value of a physical quantity realized by a standard as this value is known with a given and small enough uncertainty. The objective is to ensure indication traceability of the device under calibration, that is, to establish a relationship with the corresponding international standard through an uninterrupted chain of comparisons, each with a given uncertainty.

Measuring instruments are very often calibrated on several points equally distributed over the expected range of use, pointing out the fact that calibration and testing laboratories generally have to provide an evaluation of the uncertainties over an interval of values instead of over a single measurement point. Calibration is then accomplished on a certain number of points and the objective is to find a relationship between the calibrated instrument response and its respective reference value. In order for this relation to be used efficiently in the future for the determination of the value of the measured, it is important to estimate the uncertainty that is associated with the corrected response of the instrument.

In practice, the calibration result is very often described in terms of calibration ratio, or in a more general way, a calibration curve, which is, in almost all cases, a linear function between reference value R and indication V of the instrument in almost all cases:

$$R=a+bV \tag{1}$$

The objective is to determinate offset a and slope b, as well as to estimate the uncertainty characterizing these values. The result of the calibration can be subsequently used to determine the value of the measured physical size from the instrument indication as well as to estimate the uncertainty of this value (reverse evaluation). This methodology is implicitly applied even in the case where relation (1) is a simple equality (a=0 and b=1), insofar as an estimation of uncertainties must still be provided. In any case, the approach adopted must be applied in a way that it will lead to valid uncertainty estimations, in accordance with commonly accepted guidelines [1].

The method generally used to determine the calibration curve is the least square regression because of the wide availability of software making its application easy. In its simplest form, it has the disadvantage of being really efficient only when the uncertainty relative to the reference values (standard) is insignificant compared to the uncertainty relative to the response of the calibrated instrument.

There is also another conceptual disadvantage: the least squares method is not parametric, thus it cannot lead to probabilistic estimations of the gradient and slope of the calibration curve, since it does not use the probabilistic character of information obtained during calibration [2]. Because of this, information which could be available before calibration from previous calibrations or manufacturer specifications cannot be considered in the current calibration procedure.

The scope of this paper is to present an alternative calibration methodology involving information from previous calibrations, based on a Bayesian approach. This approach takes into consideration uncertainties linked to indications of the calibrated instrument as well as to standard reference values, leading to estimations compatible with the probabilistic character of the instrument response. It also makes it possible to consider the state of available knowledge from the instrument history and to produce a coherent image of its behavior.

The Bayesian approach

According to traditional statistics, probabilities are assigned to random events appearing under well controlled experimental conditions. Probability is then defined as the relative appearance frequency of an event after an unlimited number of repetitions of the experiment. There are cases however where this approach is not

suitable, especially with non-observable variables or variables which, by definition, do not generate random values. In addition, traditional statistics can only process information from repeatable observations, such as those from scientific decisions, information characterizing the metrologic quality of measures (calibration certificates or specifications), from personal experience or physical restrictions.

The Bayesian statistic considers measuring data as constants and the measurand as a random variable. The state of knowledge about a physical quantity is represented by a distribution of probabilities describing the degree of plausibility that we can associate with all values reasonably attributable to this quantity. The Bayesian approach can be understood as a quantitative description of the learning process and, more precisely, of the process of updating knowledge relative to a variable on the basis of new observations.

Take for example possible values x of a variable X, with a state of available knowledge before observation that is represented by distribution $f(x)$. This state of knowledge will be modified in the light of a new observation during which data d will be obtained. A new state of knowledge will be established, then represented by a distribution $f(x|d)$, by combining previously available information and information contained in observed data. The incorporation of measured data in the existing state of knowledge, requires knowledge of distribution $f(d|x)$ which represents the probability of observing data d given a particular value x of X. However, once the measurement is done, we can no longer speak of data d as if they were variables. It is then more suitable to think about distribution $f(d|x)$ as a model of observation occurrence expressing the knowledge of the random mechanism as if the unknown cause X was identified. For this reason $f(d|x)$ is expressed as a function of the unknown X, because of observations d, called function of likelihood and denoted as $l(x|d)$.

The Bayes theorem is only one way of formalizing the update knowledge mechanism about X, by combining previous distribution $f(x)$ and information from the observation in order to produce a final distribution [2]:

$$f(x|d) \propto l(x|d)\, f(x) \tag{2}$$

The proportionality symbol "$\propto$" here is used instead of the equality symbol to emphasize that there is a constant which can be calculated according to the normalization condition, i.e. by integrating the left part of equation (2) with fixed d.

Even though the Bayes theorem has been known for at least two centuries, its application in metrology is much more recent and coincides with a more general renewal of Bayesian ideas within several sciences. Since measure is mainly the science of learning by observation, it constitutes a convenient field for the application of Bayesian reasoning. It is not by chance that ISO GUM, the most commonly accepted reference text in terms of uncertainty estimation, introduces an evaluation of Type B uncertainties based on the concept of prior distribution as a model for the state of

knowledge relative to a variable [1]. A large number of studies have appeared these last few years concerning different possible applications of the Bayes theorem in metrology, including a complement to ISO GUM which is in preparation [3, 4, 5, 6, 7].

Calibrating in a Bayesian perspective

Calibrating by comparison

Generally, the procedure of calibration by comparison procedure consists of detecting the relation between response V of the instrument under calibration and the "real" standard value R of the quantity measured as this one is approximated by the indication of the reference instrument. Calibration is done in N points so that the whole measurement range is properly covered. The reference value must remain as constant as possible during calibration at each point i, ($i=1...N$).

Determination of uncertainties associated with V and R values at each calibration point is vital for the estimation of the uncertainty associated with measurements that will be taken with the calibrated instrument. This uncertainty constitutes a probabilistic estimation representing the state of knowledge about V and R values and must be determined on the basis of experimental data for each specific calibration. We should note that the quality of measurements, therefore the uncertainty, is not necessarily the same at each calibration point, as it depends on the behavior of the instruments used (precision, stability and repeatability), on the quality of work performed by personnel involved, ambient conditions, etc.

In general the instrument is calibrated by comparing N readings v_i, each characterized by its own uncertainty $u_{v,i}$, with reference value r_i characterized by its own uncertainty $u_{r,i}$. The general mathematical problem consists of finding a model with M parameters a_m to represent the series N observations (r_i, v_i) with the maximum of precision.

$$R(V)=R(V; a_1...a_M) \tag{3}$$

The model that we use most often in calibration is a linear relation between the independent variable, which is the indication V of the instrument, and the dependent variable R which is the reference value. The linear model seems quite reasonable if we consider the fact that corrections which are usually required involve the offset (zero shifting) and slope (sensitivity).

In the case of the least squares method, we search for a and b coefficients minimizing the sum S of squares of residuals, i.e. of the differences between the reference values and model predictions:

$$S = \sum_{i=1}^{N} [r_i - (a + bv_i)]^2 \tag{4}$$

This approach does not allow for statistical inferences, since we cannot use the probabilistic character of data and coefficients *a* and *b* are considered as constants with no probability distributions. However, the objective is to build a model for producing probabilistic conclusions on coefficients, and the most important is to be able to determine the predictive distribution of values which will be estimated with the help of the calibrated instrument.

We can only draw probabilistic conclusions concerning probable *a* and *b* values if we put forward certain hypotheses which, in the case of the linear model (equation (1)) come down to the following propositions:

At each calibration point, given the indication *v*, the "real" value of reference *r* follows a distribution with a most probable value being a linear function of *v*

$$r = a + bv \tag{5}$$

or, following an alternative parameterization:

$$r = a_o + b(v - \bar{v}) \tag{6}$$

where $\bar{v}$ and $\bar{r}$ are the weighted averages of observed indications for the complete calibration range.

$$\bar{v} = \frac{\sum_{i=1}^{N} v_i \sigma_i^{-2}}{\sum_{i=1}^{N} \sigma_i^{-2}}, \quad \bar{r} = \frac{\sum_{i=1}^{N} r_i \sigma_i^{-2}}{\sum_{i=1}^{N} \sigma_i^{-2}} \tag{7}$$

The advantage of this parameterization is that estimations $\hat{a}$ and $\hat{b}$ *of* a_0 and *b* respectively can be considered as independent, enabling factoring of the likelihood function.

The values observed during calibration are characterized by uncertainties. The state of knowledge concerning the observations made can be represented by a normal distribution with zero average and a known variance σ^2. In the case of least squares methodology, we usually consider that reference values r_i are known with insignificant uncertainties and that indications v_i are the only ones characterized by errors. Here, however, we consider the general case in which all observations are uncertain, which better corresponds to calibration reality. In addition, we accept that

errors are independent from one calibration point to another. We will see later how to determine this variance.

Information from calibration

During calibration, we observe the response from the calibrated instrument by setting the measured quantity to a presumed "real" reference value. The joint likelihood of observations for the i^{st} calibration point is therefore a function of their joint probability density, which is a function of two coefficients a_0 and b. These coefficients present fixed given but unknown values. The likelihood corresponds to the probability of observing values r_i and v_i, given the specific values of a_0 and b. In reality it represents relative weights associated with all possible coefficient values based on observations:

$$L(a_0,b;r_i,v_i) = f(r_i,v_i / a_0,b) \propto e^{-\frac{1}{2\sigma_i^2}[r_i-(a_0+b(v_i-\bar{v}))]^2} \tag{7}$$

From the moment when all observations are independent from each other, the likelihood of all observations is equal to the product of individual likelihoods:

$$L(a_0,b;r_1,v_1,...,r_N,v_N) \propto \prod_{i=1}^{N} e^{-\frac{1}{2}\left[\frac{r_i-(a_0+b(v_i-\bar{v}))}{\sigma_i}\right]^2} \tag{8}$$

or even:

$$L(a_0,b;r_1,v_1,...,r_N,v_N) \propto e^{-\frac{1}{2}\sum_{i=1}^{N}\left[\frac{r_i-(a_0+b(v_i-\bar{v}))}{\sigma_i}\right]^2} \tag{9}$$

The part in brackets of the exponent can be written in the following alternative form after simplification:

$$S_r - 2bS_{rv} + b^2 S_v + \sigma_0^{-2}(a_o - \bar{r})^2$$

with

$$S_r = \sum_{i=1}^{N}\left(\frac{r-\bar{r}}{\sigma_i}\right)^2,\ S_v = \sum_{i=1}^{N}\left(\frac{v-\bar{v}}{\sigma_i}\right)^2,$$

$$S_{rv} = \sum_{i=1}^{N}\frac{(r-\bar{r})(v-\bar{v})}{\sigma_i^2},\ \sigma_0^{-2} = \sum_{i=1}^{N}\sigma_i^{-2}.$$

This formulation of the exponent makes it possible to factorize the likelihood into three parts:

$$L(a_0,b;r_1,v_1,...,r_N,v_N) \propto e^{\frac{1}{2}\frac{S_{rv}^2}{S_v}-S_r} e^{-\frac{1}{2S_v^{-1}}\left(b-\frac{S_{rv}}{S_v}\right)^2} e^{-\frac{1}{2\sigma_o^2}(a_o-\bar{r})^2} \quad (10)$$

Since the first part of this expression is not a function of coefficients a_0 and b, it can be absorbed as a constant during normalization of their subsequent distribution. The joint likelihood can then be written as the product of two individual likelihoods:

$$L(a_0,b;r_1,v_1,...r_N,v_N) \propto L(a_0;r_1,v_1,...r_N,v_N)\, L(b;r_1,v_1,...r_N,v_N) \quad (11)$$

with:

$$L(a_0;r_1,v_1,...,r_N,v_N) \propto e^{-\frac{1}{2\sigma_o^2}(a_o-\bar{r})^2} \quad (11a)$$

$$L(b;r_1,v_1,...,r_N,v_N) \propto e^{-\frac{1}{2S_v^{-1}}\left(b-\frac{S_{rv}}{S_v}\right)^2} \quad (11b)$$

Since the joint likelihood is factored into two individual likelihoods, it is presumed that they are independent of one another. The likelihood of slope b is a normal distribution with best estimation (average) of S_{rv}/S_v and variance of S_v^{-1}, whereas the likelihood of offset a_o also takes a normal form with a best estimation of $\bar{r}$ and variance of σ_o^2.

Determination of variance σ^2

A particularly important point is the determination of variance σ_i^2 which characterizes observations (r_i, v_i) at each calibration point. Even if we consider that errors relative to reference values are insignificant, the determination of variance σ_i^2 is not easy, unless we write the calibration curve in reverse, i.e. by presuming that the "real" measured value is the independent variable. This approach is not quite compliant with the future use of the calibrating curve, as this curve will enable us to calculate the "real" corrected value (dependent variable) from the observed indication (independent variable).

In a more general and more realistic case, conditions, we must consider errors linked to reference value observations as well as those linked to the instrument response. The former are caused by the uncertainty characterizing the state of

knowledge of the "real" value (uncertainty and stability of the standard during measurement), whereas the latter result from the dispersion of values indicated by the instrument at each calibration point. In reality, variance σ_i^2 is the quantity variance $r_i - (a_o - b(v_i - \bar{v}))$ and can be determined with the help of the error propagation law, based on variances $\sigma_{r,i}^2$ and $\sigma_{v,i}^2$ characterizing observations r_i and v_i respectively [1]:

$$\sigma_i^2 = \sigma_{r,i}^2 + b^2 \sigma_{v,i}^2 \tag{12}$$

The fact that coefficient b is not known in advance, as its determination is the desired goal, requires that the work has to be done iteratively: presuming first an initial value for b, then use it to calculate variance σ_i^2, calculate a new value for b, recalculate the variance, and so on until convergence. The value determined by the least squares method can be used as initial value. Concerning the calculation of variances, we can also consider that this initial value is a very good approximation and that the iterative process is not really necessary.

Prior information

One of the advantages of the Bayesian approach is that it makes it possible to take into consideration the state of knowledge relative to the behavior of the instrument, available before the initiation of the calibration process. In the case of a new instrument we can consider, for example, manufacturer specifications (offset or sensitivity errors, accuracy), or information from scientific literature concerning the implemented measuring method. We can also consider the results from previous calibrations if they exist, with the logic of functional continuity of the instrument. Finally, for the case of an instrument for which no prior specific information is available, the state of knowledge about this instrument is only represented by a non-informative probability distribution.

In the general case we consider a prior joint distribution which, by accepting the validity of the independence hypothesis, is proportional to individual distributions:

$$f_o(a_o, b) \propto f_o(a_o)\, f_o(b) \tag{13}$$

In most cases, prior information from a previous calibration would be represented by a normal distribution, whereas information based on the knowledge of error margins (accuracy) would be represented by a rectangular distribution. However, a state of total ignorance, corresponding to an instrument void of any calibration history, would be formulated by an informative distribution, i.e. by a multiplying constant.

Joint a posteriori distribution

According to the Bayes formula, joint *a posteriori* distribution after the fact is proportional to the likelihood multiplied by the joint *a priori* distribution:

$$f(a_o,b \mid r_1,v_1,...,r_N,v_N) \propto L(a_0,b;r_1,v_1,...,r_N,v_N)\, f_o(a_o,b) \tag{14}$$

The combination of equations (11), (13) and (14) leads to joint *a posteriori* distribution which can be factorized into two distributions:

$$f(a_o,b \mid r_1,v_1,...r_N,v_N) \propto f(a_o \mid r_1,v_1,...,r_N,v_N)\, f(b \mid r_1,v_1,...,r_N,v_N) \tag{15}$$

with:

$$f(a_o \mid r_1,v_1,...,r_N,v_N) \propto L(a_0;r_1,v_1,...,r_N,v_N)\, f_o(a_o) \tag{15a}$$

$$f(b \mid r_1,v_1,...,r_N,v_N) \propto L(b;r_1,v_1,...,r_N,v_N)\, f_o(b) \tag{15b}$$

The determination of marginal distributions can prove to be calculation intensive for the general case of distributions of any form, even if the techniques necessary are widely spread today, e.g. Monte-Carlo simulations. In most practical cases the use of normal, rectangular or constant distributions allows for the analytical evaluation of final distributions.

If we consider for example normal *a priori* distributions $\left(\tilde{a}_o,\sigma^2_{\tilde{a}_o}\right)$ and $\left(\tilde{b},\sigma^2_{\tilde{b}}\right)$ for a_o and b respectively, we easily obtain normal *a posteriori* distributions $\left(\hat{a}_o,\sigma^2_{\hat{a}_o}\right)$ and $\left(\hat{b},\sigma^2_{\hat{b}}\right)$, by combining equations (11) and (15):

$$\sigma_{\hat{a}_o} = \left(\sigma^{-2}_{\tilde{a}_o} + \sigma^{-2}_{o}\right)^{-0.5},\ \sigma_{\hat{b}} = \left(\sigma^{-2}_{\tilde{b}} + S_v\right)^{-0.5} \tag{16a}$$

$$\hat{a}_o = \tilde{a}_o \frac{\sigma^2_{\hat{a}_o}}{\sigma^2_{\tilde{a}_o}} + \bar{r}\frac{\sigma^2_{\hat{a}_o}}{\sigma^2_{o}},\ \hat{b} = \tilde{b}\frac{\sigma^2_{\hat{b}}}{\sigma^2_{\tilde{b}}} + \bar{r}S_v\sigma^2_{\hat{b}} \tag{16b}$$

It should noted that in the case where a non-informative *a priori* distribution is used, the results illustrated above are quite comparable to those obtained by traditional statistical methods and the weighted least squares method in particular [8].

Measurements performed by the calibrated instrument

The main objective of calibration is to be able to predict the "real" value r' of the measured quantity starting from indication v' of the calibrated instrument. The best estimation is explicitly given by the calibration curve. The remaining question is to determine the quality of this estimation, knowing that a_0 and b are random variables with a known distribution.

The predictive distribution of r', noted as $f(r' \mid v', r_1, v_1, ..., r_N, v_N)$ can be determined using the Bayes theorem, by integrating the *a posteriori* distribution of r' given the uncorrected indication of the instrument and the calibration results, with the parameters a_0 and b considered as nuisance parameters. Since observation (r',v') is independent from observations taken during calibration and a_0 and b are independent from r', we can conclude that r' will follow a distribution with average $\hat{a}_o + \hat{b}\left(v' - \bar{v}\right)$. The variance of this distribution, which is determined by the error propagation law, is equal to $\sigma_{r'}^2 = \sigma_{\hat{a}_o}^2 + (v' - \bar{v})^2 \sigma_{\hat{b}}^2$ [1]. The same law also enables us to determine the distribution of parameter a of equation (1), starting from the relation $\hat{a} = \hat{a}_o - \hat{b}\bar{v}$.

Conclusions

The Bayesian approach provides a coherent and efficient framework for treating the issue of measuring instruments calibration. It makes it possible to consider experimental data from the calibration itself as well as information available from previous calibrations or technical characteristics of the instrument. It also allows us, in the case of calibration by comparison, to take into consideration the quality of observations and more precisely of uncertainties associated with reference values and instrument indications. The methodology adopted leads to realistic probabilistic estimations for the parameters of the calibration curve and consequently for measurements implemented with the calibrated instrument.

References

[1] ISO, GUM: Guide to the expression of uncertainty in measurements, Switzerland: ISO ed., 1995.

[2] M. W. Bolstad, Introduction to Bayesian Statistics, New Jersey: Wiley-Interscience, 2004.

[3] I. Lira and G. Kyriazis, "Bayesian inference from measurement information", Metrologia, Vol. 36, pp. 163-169, 1999.

[4] K. Weise and W. Woger, “A Bayesian theory of measurement uncertainty”, Meas. Science Tech., Vol 4, pp. 1-11, 1993.

[5] G. Agostini, Bayesian Reasoning in Data Analysis, New Jersey: World Scientific, 2003.

[6] R. Kacker and A. Jones, “On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent”, Metrologia, Vol. 40, pp. 235–248, 2003.

[7] W. Bich, G. Cox and P. M. Harris, “Evolution of the Guide to the Expression of Uncertainty in Measurements”, Metrologia, Vol. 43, pp. 161–166, 2006.

[8] E. Mathioulakis and V. Belessiotis, “Uncertainty and traceability in calibration by comparison”, Meas. Sci. Technol., Vol. 11, pp. 771–775, 2000.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Using the correlation analysis in electrical measurements

Tatjana Barachkova

Virumaa College of Tallinn University of Technology
35 Järveküla tee 30328 Kohtla-Järve, Estonia
Email: Tatjana.Barashkova@mail.ee

ABSTRACT: Revealing the distinctions between estimations of certain measured values in the different schemes of measurement quite often becomes the purpose of experiments.
Thus it is necessary to check, whether this difference is caused by random deviations of the experiment. The research into correlation at measurement will allow us to execute the above-stated check, as the random deviations reduce researched correlation connection between values.
In a metrological practice there are routine statistical researches corresponding to the simple scheme of the etalons compared in pairs.
KEYWORDS: The collective working standard of variable voltage, method of definitions of dispersion and correlation coefficient

The dependence of the correlation coefficient on random deviations

It is a well-known fact that casual deviations affect the essential influence on the correlation rate between the physical units. The correlation link studied between the initial variables has be weakened:

– if estimates of casual factors are not interrelated;

– if the estimates are not dependable on the measured parameters;

– if they are normally distributed;

– if they have zero mathematical expectations and final dispersions.

In other words, the correlation rate of measured variables under the effect of casual deviations can be lower by the absolute volume than correlation rate of initial signs. On the other hand the measured correlation rate for independent sizes may be different from zero as a consequence of casual dispersion of the results being measured.

Thus, we came to the conclusion that casual deviations affect the significance of the correlation rate. The situation mentioned is typical when measuring too small values and demands the application of an alternative method of determining dispersion estimates of measurands making use of correlation analysis. To evaluate

the effectiveness of the suggested method of determining dispersion estimates the group-working standard of variable voltage (Barashkova, 2002) [1] is under investigation. It consists of N unit–keepers of physical value.

A much more convenient and reliable method of determining dispersion estimates consists of the double measurement of estimates of physical values f_1, δ_1 and f_2, δ_2, kept by group standard with the following statistical processing of the results. For the two arrays received average $\bar{f}_1, \bar{\delta}_1$ and $\bar{f}_2, \bar{\delta}_2$ and standard deviations $\sigma_{f1}, \sigma_{\delta 1}$ and $\sigma_{f2}, \sigma_{\delta 2}$ of relative and angle corrections of keepers must be calculated by collating with a reference standard. Furthermore we define correlation coefficients r_f, r_δ which should be considered as autocorrelation rates of parameters f_1, f_2 and δ_1, δ_2.

The technique of correlation analysis applied in electrical measurements

Performing two cycles of measuring parameters f and δ we obtain the following array of values:

$$\begin{vmatrix} f_{11} \\ f_{21} \\ f_{31} \\ \vdots \\ f_{i1} \end{vmatrix} \quad \begin{vmatrix} \delta_{11} \\ \delta_{21} \\ \delta_{31} \\ \vdots \\ \delta_{i1} \end{vmatrix} - \text{for the first moment of time}$$

$$\left| f_{12}\ f_{22}\ f_{32} \ldots f_{i2} \right| \quad \left| \delta_{12}\ \delta_{22}\ \delta_{32} \ldots \delta_{i2} \right| -$$

for the second moment of time

For the received arrays, average values of $\bar{f}_1, \bar{\delta}_1$ and $\bar{f}_2, \bar{\delta}_2$ and standard deviations $\sigma_{f_1}, \sigma_{\delta_1}$ and $\sigma_{f_2}, \sigma_{\delta_2}$ are calculated under formulae similar to the case with the parameter f_1.

$$\bar{f}_1 = \frac{1}{N} \sum_{i=1}^{N} f_{i1} \tag{1}$$

$$\sigma_{f1} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (f_{i1} - \bar{f}_1)^2}\ , \tag{2}$$

where

f_{i1} –relative correction of i–keeper, obtained at the first collation of the secondary voltage of i– standard and reference standard.

Then correlation rates which should be considered as autocorrelation rates in this case are defined.

$$r_f = \frac{\sum_{i=1}^{N}(f_{i1} - \bar{f}_1)(f_{i2} - \bar{f}_2)}{(N-1)\,\sigma_{f_1}\,\sigma_{f_2}} \tag{3}$$

$$r_\delta = \frac{\sum_{i=1}^{N}(\delta_{i1} - \bar{\delta}_1)(\delta_{i2} - \bar{\delta}_2)}{(N-1)\,\sigma_{\delta_1}\sigma_{\delta_2}} \tag{4}$$

where

δ_{i1} –corresponding angle correction of *i*–standard, set in minutes and obtained at the first collation of *i*–standard and reference standard;

f_{i2} –relative correction of *i*–keeper, obtained under the repetitive collation of the secondary voltage of i–standard and reference standards;

δ_{i2} –corresponding angle correction of *i* –standard, set in minutes and obtained at the repetitive collation of *i* –standard and reference standards.

It is supposed that data of two arrays must not essentially differ as the unit of a physical value, kept by the group–working standard, is stable in time. Let correlation rate r_f characterize the measure of linear dependence between values f_1 and f_2. Let us admit that $f_1^* = a f_2 + b$ is a linear function of the best average quadratic approach to the value f_1. Let us designate the estimate of amendment deviations e.g. by Δf :

$$\Delta f = f_1 - (a f_2 + b)$$

In this case, the mathematical expectation of the estimate Δf is equal to zero and the relation of its dispersion to the dispersion of the estimate f_1 is determined only by value of a correlation rate:

$$\frac{\sigma_{\Delta f}^2}{\sigma_{f1}^2} = 1 - r_f^2 \tag{5}$$

If the estimates of casual factors, causing the diversity of the measured values happen to be independent, then the statistical characteristics mentioned are linked with the following interrelation:

$$\sigma_{f1}^2 = \sigma_{\Delta f}^2 + \sigma_{f1^*}^2 \tag{6}$$

Formula (3) can be presented in this way

$$r_f^2 = \frac{\sigma_{f1^*}^2}{\sigma_{\Delta f}^2 + \sigma_{f1^*}^2} = \frac{\sigma_{f1^*}^2}{\sigma_{f1}^2} \tag{7}$$

It is not hard to see that (5) provides an alternative chance to define the estimates of parameter dispersions, kept by the given group standard.

Analysis of calculations of dispersion estimations of the measured value

Definition of relative correction $\gamma_1(f_1^*)$ using a simple ratio of dispersion analysis (6)

It should be noted that under high casual deviations, absolute corrections of estimates of average quadratic deviations σ_{f1} and $\sigma_{\Delta f}$ are equal. Taking the condition of equality σ_{f1} and σ_{f1^*} we obtain

$$\gamma_1(f_1^*) = \frac{2 \cdot \Delta\sigma_{\Delta f}}{\sigma_{f1^*}} \tag{8}$$

where

$\Delta\sigma_{\Delta f}$ −the absolute correction of the deviation estimate Δf

Based on the results of *N*–fold measuring of a parameter f_{i1} for the same object we define the absolute correction $\Delta\sigma_{\Delta f}$. At the normal law of distribution of the possible values of the measured quantity with the confidence probability $p = 0.99$ and the coverage factor $k = 3$ the estimate sought is equal to

$$\Delta\sigma_{\Delta f} = 3\sigma_{\Delta f}$$

Thus

$$\gamma_1(f_1^*) = \frac{6\sigma_{\Delta f}}{\sigma_{f_1^*}} \tag{9}$$

The confidence interval for relative correction of any keeper is equal

$$\Delta\sigma_{f_1^*} = 6\sigma_{\Delta f} \tag{10}$$

Definition of the relative correction $\gamma_2(f_1^*)$ using the autocorrelation coefficient

Under conditions of high casual deviations, taking into account formula (3), it is possible to write that

$$\gamma_2(f_1^*) = \frac{\Delta r_f}{r_f} + \frac{\Delta\sigma_{f_1}}{\sigma_{f_1}} = \frac{\Delta r_f}{r_f} + \frac{\Delta\sigma_{\Delta f}}{\sigma_{f_1}} \tag{11}$$

In this case the 2-fold casual values f_1 and f_2 are made use of. Under conditions of large casual deviations when measuring a normal casual value the correlation rate between f_1 and f_2 is equal to $r_f^{'}$. In the absence of casual deviations the rate of correlation link between initial casual values is defined by the value r_f. In conformity with the main rules of correlation analysis

$$r_f^{'} = \frac{r_f}{1 + \frac{\sigma_{\Delta f}^2}{\sigma_{f_1^*}^2}} = r_f \frac{\sigma_{f_1^*}^2}{\sigma_{f_1}^2}$$

In this case when determining r_f the following value can be taken as an absolute correction:

$$\Delta r_f = r_f - r_f \frac{\sigma^2_{f_1^*}}{\sigma^2_{f_1}} = r_f (1 - r_f^2) \tag{12}$$

Using the last formula, the expression for $\gamma_2(f_1^*)$ obtains the following form:

$$\gamma_2(f_1^*) = 1 - r_f^2 + \frac{3 \cdot \sigma_{\Delta f}}{\sigma_{f\,1}} \tag{13}$$

Confidence interval for relative correction of any keeper i

$$\Delta\sigma_{f_1^*} = \sigma_{f_1}(4 - r_f^2) \tag{14}$$

The last expression is simplified by the conditions of equality

σ_{f1} and $\sigma_{\Delta f}$ σ_{f1} and σ_{f1^*}

Having made an analogous consideration for the correlation rate r_δ between values δ_1 and δ_2, we obtain the following conclusions of dispersion and correlation:

$$\Delta\sigma_{\delta_1^*} = 6\sigma_{\Delta\delta} \tag{15}$$

$$\Delta\sigma_{\delta_1^*} = \sigma_{\delta_1}(4 - r_\delta^2) \tag{16}$$

Example of using correlation analysis in electrical measurements

As a result of experiment the following data arrays were received:

$$\begin{vmatrix} 0.03\ \% \\ 0.03\ \% \\ 0.037\ \% \end{vmatrix} \quad \begin{vmatrix} -3.3\ \text{min} \\ -3.13\ \text{min} \\ -2.88\ \text{min} \end{vmatrix}$$

– for the first moment of time

$$\begin{vmatrix} 0.0293\ \% \\ 0.0265\ \% \\ 0.0258\ \% \end{vmatrix} \begin{vmatrix} -3.2\ \text{min}, \\ -2.48\ \text{min}, \\ -2.87\ \text{min} \end{vmatrix}$$

– for the second moment of time

The average values of the relative and angular corrections of the keepers in accordance with each moment of time are equal:

$$\overline{f}_1 = 0.032\% \qquad \overline{f}_2 = 0.027\%$$
$$\overline{\delta}_1 = -3.1\,\text{min} \qquad \overline{\delta}_2 = -2.85\,\text{min}$$

The average quadratic deviations of the relative and angular corrections of the keepers in accordance with each moment of time:

$$\sigma_{f_1} = 0.0040\% \qquad \sigma_{f_2} = 0.0019\%$$
$$\sigma_{\delta_1} = 0.21\,\text{min} \qquad \sigma_{\delta_2} = 0.36\,\text{min}$$

Based on formulae (3) and (4) the factors of correlation are found:

$$r_f = -0.93$$
$$r_\delta = 0.51$$

Based on formulae (7) and (14) the corrected ratings of dispersion of the relative and angular corrections of the keepers with the appropriate confidence intervals are found.

Finally,

$$\sigma^2_{f_1^*} = (0.0037\%)^2 \quad \Delta\sigma_{f_1^*} = 0.013\%$$
$$\sigma^2_{\delta_1^*} = (0.11\,\text{min})^2 \quad \Delta\sigma_{\delta_1^*} = 0.11\,\text{min}$$

Conclusions

The received results allow us to make the conclusion about efficiency of application of factor of auto correlation for definition of ratings dispersion especially at large casual deviations.

References

[1] T. Barashkova, “Interlaboratory comparison of voltage transformer”, *Proc. of the 3rd International Conference Industrial Engineering – New Challenges to SME*. Tallinn: Tallinn Technical University, 2002.

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Validation of industrial measurement processes

Roman A. Tabisz

Rzeszow University of Technology. Faculty of Electrical and Computer Engineering
Department of Metrology and Measurement Systems
ul. Wincentego Pola 2, 35-959 Rzeszow, Poland rtabisz@prz.rzeszow.pl

ABSTRACT: Metrological characteristics of measurement processes and the aims of their industrial application are described. The need for validation of measurement processes is justified, as well as clarification of the principles of carrying it out. The main assumptions of the project carried out, are presented. The purpose of this project is the development and improvement of the information system – CAMPV (Computer Aided Measurement Processes Validation). This system contains the metrological database, among others, and is intended for regional metrological centers.

Introduction

Industrial Measurement Processes are being utilized in order to establish the numerical value of the critical characteristics of products or of manufacturing processes. On the basis of measurement results, important decisions concerning conformity assessments or correcting actions are being taken in the industry. This action is aimed at an improvement in the quality of products or manufacturing processes. Accuracy of the decisions taken depends on the essential phase of the accuracy and the credibility of the measurement results. The probability of wrong decisions which are caused by the inaccuracy of measurement results is called a measurement decision risk. The level of this risk is optimized by different factors, such as manufacturing costs, costs of the customer complaint or costs bound with the effects of dangerous events.

In every measurement process employed, there are a few factors, such as the measurement method, the measurement procedure, measurement equipment and its calibration, the operator and conditions, from all of which, the measurement process is realized. All mentioned factors influence the final result of measurements in to a lesser or greater degree. Since these factors are characterized by a certain individual variability, the following results make the realization of the measurement processes different from each other. Therefore, the final result of the measurements should be expressed in the form of a range of numerical values, rather than in the form of a single number.

If the length of this range, called the measurement process variability (***MPV***) is well known, and the location reference to specification limits (***SL***), is known, then

assessing the level of the measurement decision risk is possible. It is also possible to carry out the reverse action, if the permissible value of the measurement decision risk of the given intended use will be established first. We can then estimate the permissible value of ***MPV*** for the process which will be applied for the given intended use.

In both situations, a clear-cut result is only achieved when the character of the variability of the critical characteristic of the assessed object is well-known, and when well known metrological characteristics of the measurement process applied for measuring this feature, is outlined.

Every measurement process can be useful for the realization of one intended use of measurements but it may turn out that for the realization of a different intended use this process is not useful. For this reason a separate assessment of the usefulness of the measurement process is always necessary when it is applied for the realization of a different intended use. In industrial metrology, a notion is applied: relative usefulness of the measurement process. This notion was led in the publication [1]. This publication [1] is recognized as an important reference publication in the field of industrial measurement systems. Assessment of the relative usefulness of all measurement processes which provides documented evidence to the fact that the metrological characteristics of these processes are meeting the metrological requirements of the intended uses of measurements, is called validation. The validation is a set of metrological activities that establish the metrological requirements of the intended use of measurements, periodic calibrating, the statistical assessment of the metrological characteristics of the measurement equipment and the issue of certificates of validation. An information system equipped with a metrological database, can be very helpful for leading the validation of industrial measurement processes. Designing and developing such a system is the purpose of the project carried out by the author at the Department of Metrology and Measurement Systems of the Faculty of Electrical and Computer Engineering of the Rzeszów University of Technology.

Measurement processes and their metrological characteristics

In publications from the field of metrology, it is possible to meet different definitions of measuring processes [2],[3],[4]. Making an attempt at describing the problem of the validation of industrial measurement processes, the latest result of works of the international group JCGM/WG-2 was taken into consideration. This result is a draft of the international metrological vocabulary [5] which will be issued in the form of the international standard of the International Organization for Standardization (ISO). After many years of discussions led by a competent organization and people, they established the vocabulary to define the characteristic of measurements, of physics, quantities and the definition of measurement characteristics executed to the needs and problems in chemical and biological

analyzes. In this new vocabulary they also restored the equality of two metrological approaches – "classical approach" and "Uncertainty approach". In the draft of this new vocabulary [5] the measurement process is defined under the notion of "measurement", in the following way:

> ***measurement*** *(2.1.VIM-draft-2006)*
> ***process of experimentally obtaining one or more quantity values*** *that can reasonably be attributed to a quality.*
>
> *NOTES:*
>
> ***1-*** *Measurement implies comparison quantities or counting of entities.*
>
> ***2-*** *Measurement presupposes description of the quantity commensurate with the intended use of the measurement result, a measurement procedure and calibrated measuring system operating according to a specified measurement procedure.*

The graphical interpretation of this definition, also taking into consideration the valuable interpretation of the measuring process given in the publication [2], is depicted in Figure 1.

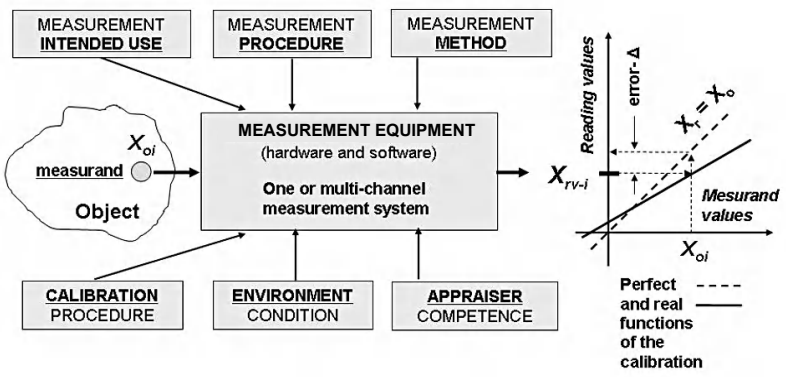


Figure 1. *The cause and effect diagram of inaccuracy of the meaurement process which was prepared according to the definition proposed in the draft [5], and taking interpretation into consideration are given in [2]*

The conclusion from this drawing is that readings numerical values $-X_{rv-i}$ which are obtained in every repeated realization of the measurement process, will be equal to true values $-X_{o-i}$ of the measurand only in the situation in which the applied, measurement process is working according to an ideal calibration curve. An ideal

calibration curve is only assured with the ideal transformation of the true value of the measurand on the read (or calculated) value in the final phase of the measurement process. For such an ideal calibration curve, equation (1) is true only.

$$X_{rv-i} = X_{o-i} \tag{1}$$

Such a situation can become known only as a very rare coincidence. In fact, the function of the calibration isn't perfect and readied values $-\boldsymbol{X_{rv\text{-}i}}$, differ from true values of measurand $-\boldsymbol{X_{o\text{-}i}}$ Differences between these values are errors of measurements $-\boldsymbol{\Delta}$.

It is possible to express an accurate value of the error included in the measurement result which was obtained in one realization of the measurement process with equation (2)

$$\Delta_i = X_{rv-i} - X_{o-i} \tag{2}$$

Industrial measurement processes taken for conformity assessment of products to the requirements of technical specifications, are made in the situation in which the true value of the measurand $-\boldsymbol{X_{o\text{-}i}}$ is not known. Products or sub-assemblies for manufacturing them are new and unknown objects. Controlled parameters of processes of manufacturing also take turns in certain limits which often are not fully known. In such cases, calculating of the measurement errors and their correction is not possible without having additional knowledge. Collecting of suitable metrological knowledge about the characteristics of measurement systems used in industrial measurement processes, is needed to create the possibilities of pointing out measurement errors and their correction. The additional knowledge about the influence of individual factors involved in the measurement process on the trueness and the precision of measurement results is also needed. It is possible to get the first part of the required knowledge by making periodical experiments in the calibrating of the measurement system carried out in the reference conditions, or in use conditions. Experiments in calibration should be led in regular and sensibly determined time intervals, with the help of ***W**orking **S**tandards (**WS**)* which represent reference values – $\boldsymbol{X_{REF}}$ which are described with the uncertainty – $\boldsymbol{U_{REF}}$. It is possible to get the second part of the knowledge by performing additional experiments of a different type. These repeatability and reproducibility experiments are, in short, called "R&R experiments" and experiments of shared assessment are known as interlaboratory comparisons. It is possible, and we should also take into consideration, knowledge contained in the technical specifications of measuring equipment delivered by its producers. These specifications contain declarations as to the maximum permissible errors – $\boldsymbol{\Delta_{MPE}}$. These declarations most often have the form of mathematical formulas, expressed as in equation (3).

$$\left|\Delta_{MPE}\right| = \left|a\% \cdot X_{berv}\right| + \left|b\% \cdot X_{N}\right| \tag{3}$$

where:

$\boldsymbol{X_{berv}}$ – *best estimation of the readings value (most often arithmetic average);*

$\boldsymbol{X_N}$ – *value of the nominal measuring range of the measurement device;*

a, b – *values declare*d by producers (most often different for different nominal measuring ranges).

The metrological knowledge extracted from data received with the help of earlier mentioned experiments and that are contained in technical specifications of measurement equipment, should be collected, analyzed and verified. Only after verification can this knowledge be used for the correction of measurement errors and the validation of measurement processes. Collecting data needing to be obtained and the verification of metrological knowledge is at present, more comfortable with the help of an appropriate information system containing, among other things, a metrological database. An example one of such metrological database was described in the publication [6]. Accepting the appropriate data model is an essential problem in every database creation process. This data model should enable data collection methods and then methods of extracting it from the database so that it is possible to make the correct statistical analyzes and simulations needed for outlining the values of the metrological characteristics of the measurement process that is under supervision.

The first assumption which was taken by the author with the aim of designing the data model for the creation of the metrological database was the assumption that every measurement process uses simple or complex measurement systems (equipment). The simple measurement system contains one elementary measurement channel. The complex measurement system is composed of many elementary measurement channels. In the created metrological base a set of results for calibrating of the elementary measurement channel will be the most basic particle of data. They also made an assumption that on the basis of historical data of calibrating of elementary measurement channels there will be numerical values of the trueness measure and of the precision measure. Making this assumption, a "classical approach" was used and restored in the final draft [5].

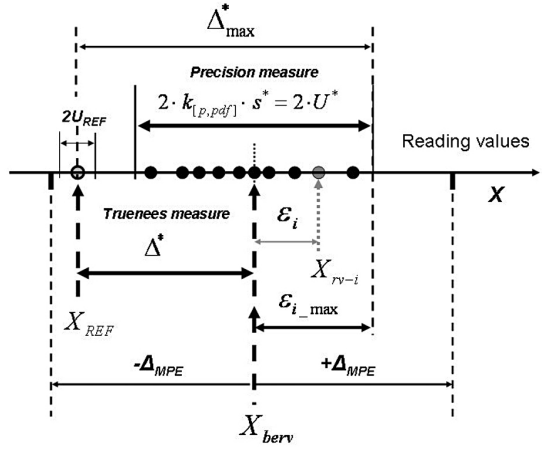


Figure 2. *Graphical interpretation of the "classical approach" which enables the calculation of measures of trueness and precision*

In Figure 2, a graphical interpretation of the "Classical approach" is shown. This approach enables the calculation of the numerical measures of trueness and precision. Such an approach is possible in the situation when the known value is being measured in the measuring process. In this case it is value $-X_{REF}$ which is represented by ***W**orking **S**tandard –**WS**,* determined with the uncertainty $-U_{REF}$.

This interpretation was worked out for the case in which was used one *Working Standard* and repeated measurements of this Standard were repeated in the same conditions of the repeatability. Curiosities of these kinds of experiments in which repeated realizations of measurement processes are used, and in which are used one or several working standards, were analyzed in paper [7]. From this paper it is possible to conclude that in such cases many problems appear in connection with accepting the appropriate model. A good model is one where you should consider all conditions in which they are being made. The interpretation which is introduced in Figure 2 can be accepted to obtain the numerical measure of the trueness and precision for each calibration point of the given measurement range of the elementary measurement channel. The experimental standard deviation $-s^*$ is a numerical measure of precision, of which a value can be accepted as value of the standard uncertainty $-u_s$. The value of the expanded uncertainty $-U^*$ enumerated as the product of the coverage factor $-k_{[p,\ pdf]}$ and experimental standard deviation- s^*. Also a different measure of precision is important for the validation, with which double the value of the expanded uncertainty is calculated with the help of formula (4). The interval of values determined in this way is called the Measurement Process Variability –***MPV***, and is used for designing of the conformity assessment diagrams.

$$MPV^* = 2 \cdot U^* = 2 \cdot k_{[p.pdf]} \cdot s^* \tag{4}$$

where:

$\boldsymbol{k}_{[p,\ pdf]}$ is a coverage factor determined for the established confidence level – ***p*** and for the right kind of probability density function – ***pdf.***

A numerical measure of the trueness is experimental deviation (bias) – Δ*, which it is possible to calculate, using the equation (5)

$$\Delta^* = X_{berv} - X_{REF} \tag{5}$$

If we assume that the amendment P is a consistent value of the systematic error adopted with the opposite sign ($\boldsymbol{-\Delta_s}$), we can fill it by writing:

$$-\Delta_s = P \approx P^* = -\Delta^* \tag{6}$$

If value of this experimental deviation (bias) $\boldsymbol{-\Delta^*}$ is outlined in a similar way for several points of the measurement range, and we assume that this value is approximately equal to the true values of systematic errors $\boldsymbol{-\Delta_s}$ it is then possible to make an amendments table or outline a linear correction function of the type:

$$\Delta^* = A \cdot X_{berv} + B \tag{7}$$

If, however, such a linear corrective function is not appropriate in defining the different non-linear function, a corrective function will be necessary.

For every elementary measurement channel it is possible to outline relations (8) and (9) which will be very useful in practice.

$$s^* = f(X_{berv}) \tag{8}$$

$$MPV^* = f(X_{berv}) \tag{9}$$

Relation (8) is possible to use for the simulation of the uncertainty propagation in the multichannel measurement systems. However, relation (9) can be used for creating diagrams of the conformity assessment for single-channel measurement systems.

"Classical approach" and "uncertainty approach" – applied together

If an appropriate corrective function is accessible to every elementary measurement channel, correcting systematic errors in all channels which can be used in the complex measurement system will be possible. In such a situation, using the "Uncertainty approach" will be possible to express the final results obtained by the multichannel measurement system. In Figure 3, the conception of simultaneously applying two metrological approaches, is presented. This conception takes into consideration the recommendations of the international guide for expressing the uncertainty of measurement results [8] and Supplement-1 to this guide [9]. "Classical approach" is used to define two metrological characteristics of elementary measurement channels: the trueness ***(Δ*)*** and precision ***(s*=u*).*** "Uncertainty approach" is applied for obtaining one metrological characteristic of the multichannel measurement system. This characteristic is the expanded uncertainty – $\boldsymbol{U_Y}$ from the final results obtained from the complex measurement system.

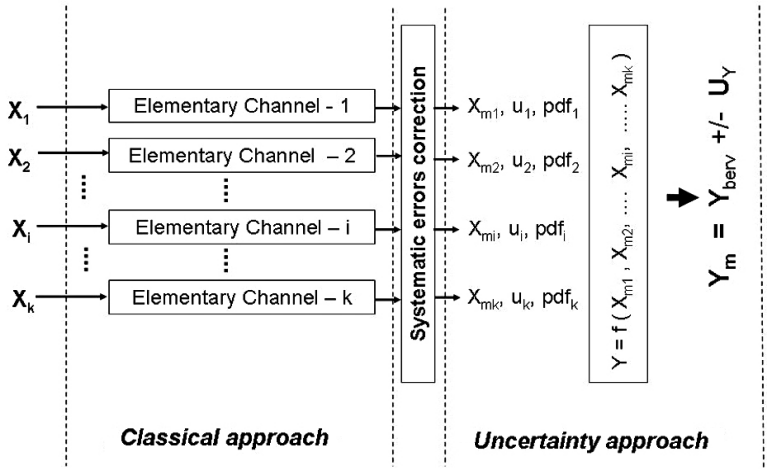


Figure 3. *Conception of simultaneous application of two metrological approaches. The "classical approach" and the "uncertainty approach"*

Possessing an information system for outlining the value of metrological characteristics (4) and (5) and for collecting them in the metrological database is possible in practice for the correction of systematic errors and carrying out the validation. This action will then be based on credible values of metrological characteristics outlined for simple and complex measurement systems. The realization of the proposed conception in the industry should be an important factor of the supervising system of measurement processes.

Principles of the validation of measurement processes

Using the presented conception of simultaneously applying two metrological approaches for holding a measurement decision risk, the lower than accepted level is not enough. This is only one crucial factor. Carrying the validation of every measurement process which will be used to realize the determined purpose, is still needed. The validation should provide documented confirmation that the given measurement process is fulfilling the requirements of the intended use of the measurement.

Intended uses of measurement processes can be different in the industry. It can be a simple conformity assessment to the specification of the critical characteristic of a single product. In such a situation, the criterion of the assessment belongs to the interval values defined with unilateral limits **[-∞, *SL*]**, [***SL*, +∞]** or double-sided limits [***LSL*, *USL***]. However, it can be a more complex intended use, as in the example outlining the numerical value of the processes capability indexes (C_p), (C_{pk}) or creating the value of Shewarth's control charts.

In every situation in which the intended use of a measurement process will be different, the estimation of the measurement decision risk will also be in a different way. It means that Metrological Requirement of the Intended Use of Measurement [***MRIUM***] can be calculated in a different way. The general flow diagrams of the validation of measurement processes in all situations will be the same and will consist of four stages:

1- describing Metrological Requirements of the Intended Use of Measurement – ***MRIUM***;

2- describing Metrological Characteristics of the Measurement Process – ***MCMP***

3- ***MRIUM*** comparison to ***MCMP***;

4- issue of the certificate stating that ***MCMP*** are fulfilling ***MRIUM***.

In Figure 4 the flow diagram of the validation of measurement processes which fulfil requirements [10] can be seen.

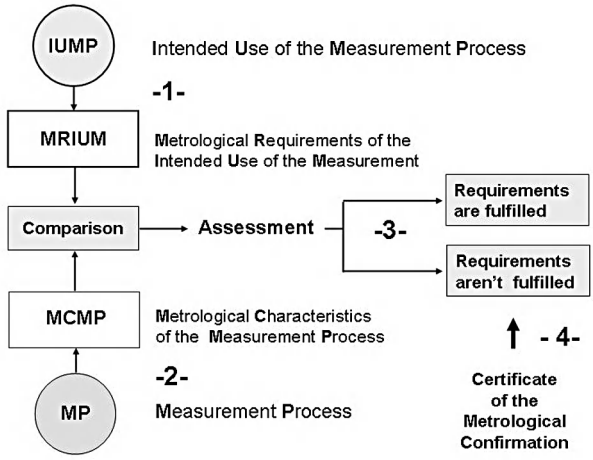


Figure 4. *The flow diagram of the validation of measurement processes*

These four validation stages introduced in Figure 4 can be realized with the help of the metrological information system which was designed and is still being developed by the author. This system is called Computer Aided Measurement Process Validation – CAMPV and contains three main parts:

A – the laboratory stand for calibrating elementary measurement channels;

B – metrological database accessible through the interactive web page;

C – the stand for statistical analyzes and simulation.

The CAMPV system is an improved version of the CAMCP system prototype described in the publication [11].

Conclusions

Based on the presented conception of the simultaneous application of the "classical" and "uncertainty" metrological approaches, the metrological information system called CAMPV – Computer Aided Measurement Process Validation, was designed and is still under development by the author. The main principles of the validation applied in this system conform to the recommendations of the international standard [10]. The knowledge and experiences obtained during creating and implementing the CAMPV system and the ensuing information infrastructure can be, in the future, used by different interested laboratories, in particular by laboratories of the companies which are located in the "AVIATION VALLEY"[12]. Rzeszow University of Technology is also one of the members of this association.

References

[1] WHEELER D., LYDAY R.W., "Evaluating the Measurement Process". Second Edition. SPC Press. Knoxville.1989

[2] Mari L.P., "Model of the measurement process". *Handbook of Measuring System Design.* Edited by Sydenham P.H, Thorn R. John Wiley & Sons.Ltd, 2005. Vol.3. article.104.

[3] FERRIST.L.J., "A new definition of measurement". Measurement. Vol.36, 2004, pp.101-109.

[4] Measurement Process Characterization. Part.2 of the Engineering Statistics Handbook. www.itl.nist.gov/div898/handbook/

[5] International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM) 3rd edition. Final draft 2006-08-01, JCGM/WG-2 Document N318, ISO VIM (DGUIDE 99999.2).

[6] GROSS H., HARTMANN V., JOUSTEN K., LINDNER G., Generic System Design for Measurement Databases – Applied to Calibrations in Vacuum Metrology, Bio-Signals and a Template System. *Advanced Mathematical and Computational Tools for Metrology-VII,* World Scientific Publishing Co. 2006, pp-60-72.

[7] PAVESE F., *Some notes on replicated measurements in metrology and testing: classification into repeated or non-repeated measurements*. (Printed paper before Round Table of the TC7 and T21 IMEKO Committees, during XVIII IMEKO World Congress, Brazil, 2006).

[8] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML. *Guide to the Expression of Uncertainty in Measurement.* 1995, Second edition.

[9] Evaluation of measurement data – Supplement-1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using Monte Carlo method. JCGM 1 Final draft 2006.

[10] ISO-10012:2003 Measurement management systems – Requirements for measurement processes and measuring equipment.

[11] TABISZ R. "Computer aided metrological confirmation processes", in *Proceedings of the XVIII IMEKO World Congress*. September 17-22, 2006, Rio de Janeiro, Brazil.

[12] "Aviation Valley" http://dolinalotnicza.pl.

This work is financed from funds allocated for research in 2007-2009, by the Ministry of Science and Higher Education of the Polish Government. It is known as research project No. **505 015 32/2885.**

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

iMERA and the impact of metrology R&D on society

Gert Rietveld

NMi Van Swinden Laboratorium, P.O. Box 654, 2600 AR Delft, The Netherlands
e-mail: GRietveld@NMi.nl

Introduction

Metrology research and development (R&D) is quite costly and generally takes many years. It is therefore important to develop those standards and reference materials that are actually needed by industry and society. In order to prove that this metrology R&D is really relevant, NMIs and universities should be able to give an indication of the impact of metrology R&D on society.

The actual measurement of the impact of metrology R&D appears to be quite difficult. Most of the impact of metrology research projects will be indirect, by increasing generally available knowledge and by providing better measurement facilities to customers. Even if a direct economic impact of a research project can be identified, it is difficult to quantify this impact since it is often achieved via several, mixed factors.

In this paper we present an overview of the methods used throughout Europe for the impact measurement of metrology R&D, as well as the results of a workshop recently held on this subject.

The inventory and the workshop are both part of the iMERA project [1], which aims to increase the impact from national investment in European metrology R&D. The final aim of the iMERA project is a joint European metrology R&D programme. Such European-wide cooperation is one of the answers to the metrology dilemma: the fact that challenges of existing and new areas in metrology grow at a larger pace than the available budgets.

Impact measurement

If the research and development in metrology aims to serve industry and society, it is important to measure the actual impact of metrology R&D. The metrologist's favourite quote concerning measurement also applies to impact assessment –

"*If you cannot measure it, you cannot improve it*."

Sir William Thompson, Lord Kelvin (1824-1907)

The results of impact assessment for metrology activities are vitally important, since metrology must compete for government funding with other scientific areas. Clear studies that can unequivocally demonstrate that metrology R&D provides excellent value for money will be essential in persuading funders of metrology R&D.

As a working definition of impact assessment for the project the following description was made:

The impact assessment of metrology R&D is an estimation of the "*change*" *(in size, quality, or value) caused or influenced by the metrology R&D activities. This change is usually assessed in areas of scientific progress, the economy (e.g. industrial competitiveness, wealth creation) and wider societal quality of life (e.g. health, security and climate change).*

Showing impact is of special relevance to the iMERA project, since it is one of the explicit aims of this project to enhance the impact of metrology R&D in Europe via increased cooperation. Therefore a specific task in iMERA is devoted to the methods that are available and in use throughout Europe for the measurement of the metrology R&D impact. The following activities were developed within the iMERA project:

- gathering information on the present status of impact measurement in Europe via a questionnaire for iMERA partners;
- literature study on impact studies performed by non-iMERA partners;
- organization of a workshop on impact measurement of metrology R&D.

Questionnaire

At the beginning of the iMERA project (summer 2004), a questionnaire was sent out to all iMERA partners covering the following issues:

- whether formal methods are in use to measure the impact of metrology research, and if so, what these methods are;
- who requires the use of such methods;
- involvement of outside experts;
- the use of the results of the impact studies;
- meaning of impact in the studies that are performed;
- other metrics apart from direct economic impact;
- how often impact is measured and what timescales are considered.

The questionnaire was completed by 14 NMIs.

Relevance of impact measurement

The relevance of the measurement of the impact of metrology R&D is emphasized by the vast majority of the NMIs. As the use of the results of such studies they mention

- justification of money invested in current and past projects;
- prioritization of proposed, new, projects. This is the subject of another task of the iMERA project, where indeed three out of the six suggested criteria for prioritizing projects in a European metrology research programme are related to the estimated economic and social impact of the proposed projects;
- process improvement, e.g. improvement of the prioritization process, of project design, and of KT strategies to increase the future uptake;
- marketing: impact studies give important additional knowledge of the market for metrology R&D and thus helps to perform market analyses. Success stories (industrial take-up and scientific successes) that demonstrate the impact can be used as a marketing tool to attract new customers.

Here, the justification of past or running projects was seen as the major use of impact. Success stories would greatly help to show that the money invested in metrology R&D was well spent, and would furthermore help the policy makers to understand the relevance of metrology R&D.

Audience of impact studies

Due to the apparent relevance, there is a clear need for metrology R&D impact studies. In general, such studies are requested by the government, quite frequently supported by a metrology advisory board, since they provide the main funds for metrology R&D in a country. Also NMI management is interested in such studies since a high impact proves the relevance of the existence of the NMI and is the basis for (increasing) customer income.

Methods of impact assessment

A core finding of the questionnaire is that many NMIs have difficulty in performing serious impact measurements, even with the help of economic experts. In approximately half of the countries no impact measurement is performed at all. Of those that do perform some kind of impact measurement, this is mostly covering the complete measurement infrastructure and not focused on the impact of R&D. Furthermore, impact measurement of metrology R&D is often limited to case studies or to measurement of more *indirect* impact parameters like scientific output, international cooperation, and knowledge transfer. In fact only in one country (UK)

a continuous effort is present in the development and application of models for measuring the economic impact of metrology R&D.

It is noted that a common language does not exist: e.g. some NMIs consider scientific papers are a *direct impact*, others consider this an *output*, but would argue that the *impact* only comes from actions relating to the paper's content.

Frequency of impact studies

If studies are performed, this is generally done every 3 to 4 years. This period also corresponds to the general time scale that is considered in the studies. This coincidence is not surprising since research projects and metrology research programmes typically run for 3 to 4 years.

Literature study

References [2 – 8] contain a series of recent papers on metrology impact measurement activities outside Europe. The report of J. Birch to the OIML [8] is focussed on legal metrology and in that sense not relevant for metrology R&D, but it contains a very useful inventory of metrology impact papers, with short descriptions of the contents of the major papers.

The study of Williams [4] was commissioned by the EU, DG-Research, as part of the FP5 growth programme. It considers the general role of measurement activity in the EU economy, rather than just R&D, even though the discussion of EU funding concentrates on R&D.

The impact measures used contain society (quality of life) benefits as well as economic measures. The quantitative economic methods are used for the costs and benefits of measurement across the whole EU. Here the impact is estimated by contribution to GDP growth, and it uses the measurement share of patents to set the contribution to the total growth.

The non-economic measures of society level impacts is in the form of six case studies at the industry or technology sector level. Here, interviews are used to examine how measurement is used and what the effects are of a lack of measurement provision. The value of measurement is estimated from the amounts spent on equipment and activity.

A paper by Semerjian and Watters [2] gives an excellent overview of impact measurement at NIST (USA) in the past decade. The impact measurement is performed via a series of case studies [3]. Some of them are informal case studies, where NIST activities are related to the total expenditure in a certain area and an

estimate is made of the contribution of the NIST activities to this expenditure. Another series of studies are more formal economic impact studies that are quite detailed and mostly performed by external contractors. Quantitative estimates are provided either as benefit-to-cost ratios or as rates of return to the nation (social rate of return). These studies are retrospective since impact data are usually only available a few years after project initiation.

In general, the NIST studies tend to be purely economic and only consider direct impact. They barely mention quality of life aspects, nor consider economic impact via indirect parameters.

In 2001 a study of the general impact of the NRC-INMS (Canada) on the economy was performed by external consultants as part of a five-year strategic planning process [5]. The most quantitative part of the study focuses on the impact of measurement services (calibrations) via the so-called "ISO proxy model". This model estimates the value of the multiplication factor or calibration fan-out via the willingness of companies to pay for ISO accreditation.

The core benefit of metrology R&D is estimated via document review and some case studies. In the case studies "big winners" were identified and subsequently were asked very detailed questions to reflect exactly how the NMI research served their needs.

In general, the literature study of impact measurement activities confirmed the findings of the questionnaire among the partners in the iMERA project. Only a few impact studies have been performed, and those that are performed generally cover the complete national measurement system, with relatively little attention for the economic impact of metrology R&D.

Workshop

Since the questionnaire mainly revealed the shortcomings of the present situation in Europe, a workshop was held on 15th October 2006 in Turin, just before the midterm meeting of the iMERA project. The aim of this workshop was to make a further step towards better metrology R&D impact measurement in Europe via discussions on:

- aspects of impact measurement,
- the relevance of impact measurement to an eventual European Metrology Research Programme (EMRP),
- how metrology R&D impact measurements should preferably be carried out.

During the workshop, presentations were given on the results of the impact questionnaire, and on actual impact studies performed by DTI (UK), by NIST (US), and by the IRMM (BE).

Relevance of impact assessment

The relevance of impact studies to the EMRP was widely recognised by the workshop participants. It is felt to be an important tool for showing the success of the EMRP: showing impact is one of the best ways of proving that "the EMRP has made a difference" and that the money in this research is well spent. The results of impact studies are therefore especially important for the funding agencies of metrology R&D (government, EU). However, a series of stakeholders were also mentioned that would profit from impact studies like lobby and interest groups, risk managers, industry, universities and the general public.

Methods of impact assessment

The presentations given at the workshop clarified that impact studies resulting in clear "success stories", presented in a widely understandable "socio-economic language", are most convincing in showing the relevance of metrology R&D.

It was felt that it would be useful to compare projects' predicted impacts, with measured impact during the project, and a number of years after the project end. This would validate any impact assessment methods and models used, and improve the prioritization process.

In the afternoon breakout session of the workshop four different practical approaches to metrology R&D impact measurement were discussed:

- case studies;
- industry area or sector approach;
- overall review of metrology R&D impact;
- impact of fundamental metrology research.

Each of these approaches appears to have its own merits and demerits. Case studies are relatively easy to do and the results can be easily presented to a non-specialist audience. The main disadvantage is that only a selected number of case studies can be performed and therefore a representative selection of cases should be made.

The sector approach naturally focuses on industrial needs, and has more flexibility and possibility of being aligned with the interests of the funders of the

metrology R&D in that sector. Again, the selection of sectors requires careful consideration, especially when a European-wide study is performed, since the sector should be relevant to all or at least the majority of the countries concerned.

Making an overview of the impact of the total metrology R&D program in Europe has the advantage that no selection of subject or area has to be made, and that data is easier to collect and felt to be less subjective. On the other hand, this approach is very broad and therefore might have the risk of not telling too much, since specific impact and new services might not always be (only) caused by the research performed. Therefore, this approach will be most effective when the evaluation is limited to a new area of research, or to a new technology (e.g. nanometrology, optical clocks).

Impact assessment for long-term fundamental metrology research naturally has to be limited to scientific indicators like papers, patents, awards, etc. which can be easily collected. On the other hand, such indicators are not a primary measure of impact on society and therefore might not impress the funders of research and the public.

In the concluding session of the workshop, the need for impact measurement was emphasized again by the workshop participants. Each of the four approaches to impact measurement were felt to be useful to some or more extent, and it was proposed to extend the work in this iMERA task to an actual trial of several impact studies based on these approaches.

Conclusions and recommendations

It is one of the explicit aims of iMERA to increase the impact of metrology R&D in Europe by increasing the cooperation among NMIs. It is therefore important to measure the impact of R&D projects, both as a justification of finished projects as well as a tool for prioritization of new project proposals. This relevance of impact measurements, and the need for "success stories" proving the relevance of the metrology R&D, is broadly recognised by the NMIs in Europe.

The inventory of the present status in Europe via the questionnaire shows that the measurement of socio-economic impact of metrology R&D projects is hardly performed. One of the main reasons is the lack of suitable tools for quantitative measurement of economic impact and impact on quality of life. As a consequence, impact measurment is mainly limited to more *indirect* impact parameters like scientific output, international cooperation and knowledge transfer. The main finding of the inventory is that

there is a large gap between the wish for metrology R&D impact measurements and the availability of adequate methods for actually doing this.

In order to make a step forward towards better metrology R&D impact measurement in Europe, a workshop was held with all iMERA partners. In this workshop, the necessity of metrology R&D impact measurement resulting in clear "success stories" was again stressed by the participants. One proposal was to make impact measurement a continuous activity within the EMRP, possibly best performed by a specialist working group with access to independent impact assessment specialists. This working group would expedite progress towards validated impact studies particularly in the sphere of socio-economic impact assessment where current knowledge is limited. A joint approach would share costs, which can be extensive, and pool current knowledge.

Concerning the methods of impact measurements, it was concluded that each of the four approaches discussed in the workshop to some or more extent should be followed within the EMRP. In order to be better prepared for such impact studies, presently a few trial impact assessments are being performed.

References

[1] See www.euromet.org/projects/imera. Partners in the iMERA project are European NMIs and governments with an R&D programme in metrology.

[2] "Impact of measurement and standards infrastructure on the national economy and international trade", H.G. Semerjian and R.L. Watters, Measurement 27 (2000), pp 179 – 196.

[3] The individual case studies of impact measurement NIST has performed are available on the Internet under http://www.nist.gov/director/planning/impact_assessment.htm#recent

[4] "The assessment of the economic role of measurement and testing in modern society", Dr. Geoffrey Williams, 2002.

[5] "*National Metrology Institute – Its Value to Canadian Economy and Society*", J. Lusztyk, 2002 NCSL International Workshop and Symposium

[6] "Evaluation of Vote Research, Science and Technology Output Class 10: National Measurement Standards", Opus International Consultants, 2002. Available on the web at

http://www.morst.govt.nz/?CHANNEL=STATISTICS+AND+EVALUATIONS&PAGE=Statistics+and+evaluations

[7] "*Potential Economic Impact of the CIPM Mutual Recognition Arrangement*", KPMG Consulting, 2002. The report was carried out on behalf of the BIPM and is available on the BIPM web site under http://www.bipm.org/en/cipm-mra/economic.html

[8] "*Benefit of Legal Metrology for the Economy and Society*", J. Birch, 2003. This report is vailable on the OIML website at http://www.oiml.org/public

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Overview

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Automation of testing procedures in accredited metrology laboratories

A. Silva Ribeiro [1], J. Alves de Sousa [2], M. Pimenta de Castro [1] and R. Barreto Baptista [2]

[1] National Laboratory of Civil Engineering, Lisbon, Portugal, asribeiro@lnec.pt, mpcastro@lnec.pt
[2] Madeira Regional Laboratory of Civil Engineering, Funchal, Portugal, jasousa@lrec.pt, rbaptista@lrec.pt

ABSTRACT: The automation of measurement processes has advantages but also has risks, generally related to the integrity of the data. Depending on the nature of the operations to be carried out and the required robustness and reliability, the use of automation in acquisition, processing and control should be adequately assessed by a validation procedure in order to assure the fulfilment of its requirements.
It is also important to have information about metrics that are able to evaluate the performance of these tools, allowing the traceability of errors and the knowledge of the complete set of data necessary to any correction procedure. The main risks involved, the performance evaluation and some notes on future trends are discussed.

Introduction and motivations

As a consequence of the increasing demand for the calibration of instruments, the use of automation to perform different tasks of the measurement process is also increasing. However, there are other reasons behind the present shaping of metrology laboratories, namely, the modern complexity of systems requiring the assurance of minimum system failure and the optimization of human resources (skills, capabilities, expertise) available in order to use them more efficiently.

Documented procedures related to each of these different issues should be developed under a Quality Management System (QMS). The development of QMS in laboratories and in industry increased the need of metrological facilities in order to achieve quality in measurements. Metrology became one of the several branches of QMS, bringing its own requirements related to traceability and measurement uncertainty evaluation.

The modern approach to many measuring systems makes use of complex technological solutions where automation is largely applied. The high performance required to many of those measurement systems depends on the robustness and reliability of hardware and software functionality, assuring quality requirements [1][2].

The use of automation in metrological laboratories has similar requirements to those found in testing laboratories and industry, being necessary to perform risk analysis in order to demonstrate compliance with standard requirements. Considering that the automation process is usually based on software and hardware interfaces, it becomes extremely important to perform validation of these tools and their interaction, assuring the integrity of the information that leads to measurement results, as prescribed in the EN ISO 17025 [3].

This paper describes the type of metrological testing procedures and their requirements in two laboratories as well as the automation developed for the procedures, followed by a discussion concerning the integrity and validation of information. In the following sections the general approach adopted in two accredited metrological laboratories will be discussed, detailing the procedures, the parameters and the techniques used to validate components in automated measurement systems, followed by a discussion of the results obtained and the implications of these in the evaluation of measurement uncertainty.

Benefits and risks of automation in metrological processes

The introduction of automation in metrological processes is mainly due to 3 reasons: the complexity of modern measurement systems; the need to increase performance reducing human occupation with repetitive operations and errors due to human interpretation; and operational security.

The complexity of measurement systems is related to several tasks, such as:

– high frequency in data acquisition;

– heavy data acquisition and memory requirements;

– control response of systems;

– simultaneous observation of multiple quantities;

– heavy processing work;

– permanent observation of critical quantities;

– redundancy.

Regarding the optimization of human resources, the automation can be applied, i.e., substituting human work that does not require special skill or expertise, to repetitive tasks and to processes where it decreases the risks of failure related to human interpretation or behavior.

The operational security reasons more commonly considered are:

– testing systems with risk of human integrity;

– monitoring quantities with risk of human integrity;

– measuring quantities with risk of human integrity.

When considering the use of software as an automation tool in the metrological processes, some of the common risks are related to:

– configuration of systems;

– data acquisition and processing;

– communications;

– synchronizations;

– fulfilment of requirements;

– unstable methods, boundary treatment and compilation;

– numerical stability and statistical processes;

– security and integrity of information;

– failure/recovery behavior.

These risks entail the need to perform "Risk and Integrity Analysis" (RIA) integrated in general validation procedures [4][5].

The level of validation should be defined by the metrologist based on comprehensive perspective of the nature of the problems and their models[1], performing extensive validation where high or critical risks are found and moderate analysis with less effort in other cases. Typically, the effort should be proportional to the level of consequences and to the probability of failures of the analysed system.

Considering that today embedded software with extended functionalities is integrated to almost every measuring instrument and system, its validation became a

1 It should be taken into account that validation is a part of the larger perspective of verification and validation of systems including design, model definition, prototyping, hardware/software interfaces, installation and maintenance.

common requirement [3]. In this context, the definition of validation should be considered [6][7]: "*confirmation and provision of objective evidence that the particular requirements for a specific intended use are fulfilled*".

Developed validation procedure

In order to perform the validation of software used in the two laboratories previously mentioned (LNEC and LREC), a validation procedure has been developed. The following description is based on that procedure.

According to EN ISO 17025 [3] *commercial off-the-shelf software (e.g. word processing, database and statistical programmes) in general use within their designed application range may be considered to be sufficiently validated*". Therefore, the validation procedure applies to custom or modified off-the-shelf software.

The developed validation procedure includes a sequence of 5 main steps [8]:

– specification of the main characteristics of the software under validation process;

– identification of requirements and testing parameters;

– risk analysis combined with the definition of the *software integrity level (SIL)*;

– selection of the validation techniques, identification of the reference, test and validation parameters and definition of the criteria to use;

– test and conclusion about the validation process.

Each of these topics can be described in detail, however, one of them is considered to be crucial to the validation process: the risk analysis and the definition of the *SIL*. This parameter is of major importance since it summarizes the knowledge of prior topic information providing a quantitative value which is the basis for posterior decisions, especially the extension of validation and the selection of appropriate techniques.

The need to quantify the risk related to the use of software, containing qualitative elements, is based on the definition of the *SIL* considering four input parameters, each of them scaled in 4 degrees (from 1 – minimum, to 4 – maximum). The input parameters and the scales are:

– criticity (related to the type of consequences related to the failure of the system under analysis, scaled from level 1 – monitoring of low influence quantities – to level 4 – critical to human integrity);

– complexity of control (from simple input of data, level 1, to complex interaction between the system and the results obtained, classified as level 4);

– complexity of data processing (from level 1, e.g. the linear regression applied to input data, to level 4, e.g. use of complex algorithms);

– legal requirements (conditions that can impose specified levels of quality assurance and, therefore, can influence the *SIL* quantification).

The criteria based on these parameters is to adopt the maximum of the levels stated in the first three parameters and consider legal requirements as an additional input parameter that is able to change the final value of the *SIL*. The sequence including the parameter definition and further selection of techniques is illustrated in Figure 1.

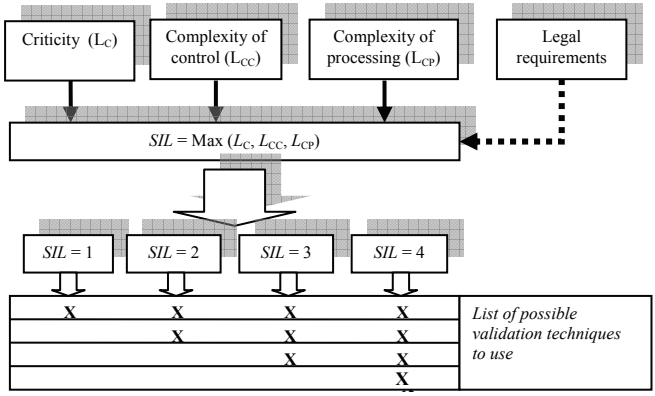


Figure 1. *Definition of the SIL parameter followed by the selection of validation techniques*

The *SIL* parameter takes integer values between 1 and 4, expressing a degree of *criticity* according to the nature of the system under analysis with corresponding level of criticality as described in Table 1.

SIL	*Description*
1	Low-risk (e.g. monitoring of quantities)
2	Moderate risk
3	High risk (e.g., typical commercial risk related to measurement results in testing and in calibration)
4	Critical risk (e.g., risk related with human integrity in testing and calibration – high temperature, dangerous materials, high pressure)

Table 1. *Description of the nature of SIL risks*

The following step is the selection of the appropriate validation techniques (Figure 1) based on the *SIL* result, considering that different types of techniques are adequate to different degrees of integrity (e.g., *low risk* systems can be tested with techniques like consistency analysis or mathematical specifications, while high risk systems should be validated using techniques such as numerical stability test, code revision, Petri nets, boundary behavior evaluation or formal specifications) [8].

The nature of the selected techniques determines the type of reference and test parameters [9] and the validation criteria which usually combines these parameters with a numerical tolerance.

Some practical applications

In metrological laboratories many different tasks can be performed with automated systems, some of which have an important influence on the main calibration activity. These systems can be validated by considering two perspectives:

– the validation of a software component;

– the validation of the system as a whole.

The two examples selected intend to illustrate the approach used in both perspectives, using two examples. The first is related to software embedded in a resistance bridge acting as a “black box”. The second is related to an automated process of measuring metallic moulds using a coordinate measuring machine (CMM).

The first of these examples concerns the evaluation of temperature (ranging from 0ºC to 231ºC, approximately) based on the ratio of electrical resistances measurement using a "black box" software integrated into a resistance bridge. In Figure 2 the three parts of the system that can be distinguished are presented: the input quantity obtained by measurement (electrical resistance R), the programmed parameters obtained by previous calibration (curve coefficients a and b and the electrical resistance at the triple point of water, R_{tpw} – used in the evaluation of electrical resistance ratio $W=R/R_{tpw}$) and the processing of input data using the built-in instrument software that gives the temperature as output quantity. According to the International Temperature Scale of 1990 (ITS90) [10] coefficients a and b, are obtained solving two equations with two unknown variables of the type:

$$W(T_{90}) - W_r(T_{90}) = a\left[W(T_{90}) - 1\right] + b\left[W(T_{90}) - 1\right]^2 \tag{1}$$

where T_{90} represents the temperature (in K) and $W_r(T_{90})$ stands for the reference function, given by,

$$W_r(t_{90}) = C_0 + \sum_{i=1}^{9} C_i \left[\frac{(t_{90}/^oC - 481)}{481}\right]^i \tag{2}$$

where t_{90} is the temperature in Celsius and C_i are constants of this reference function with values given in ITS90 [10].

The risk analysis of the system led to a classification of the software as *SIL* = 3 (commercial risk) and a combination of two validation techniques was adopted as main test: *mathematical specifications* combined with *reference data set testing* (as illustrated in Figure 3).

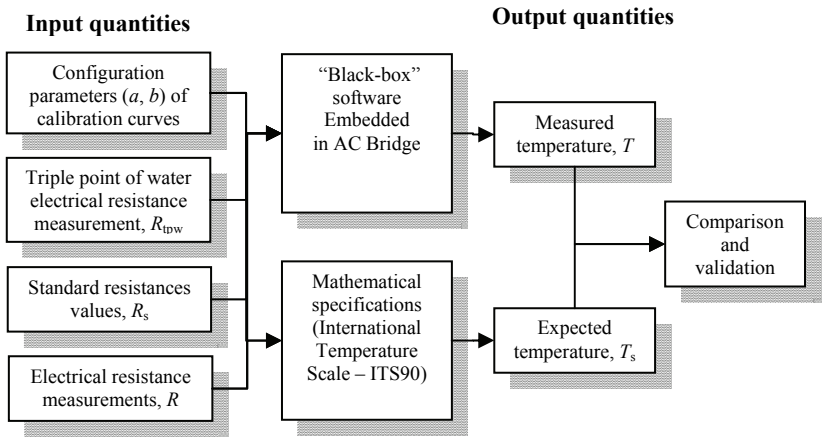


Figure 2. *Input and output quantities in the first practical case*

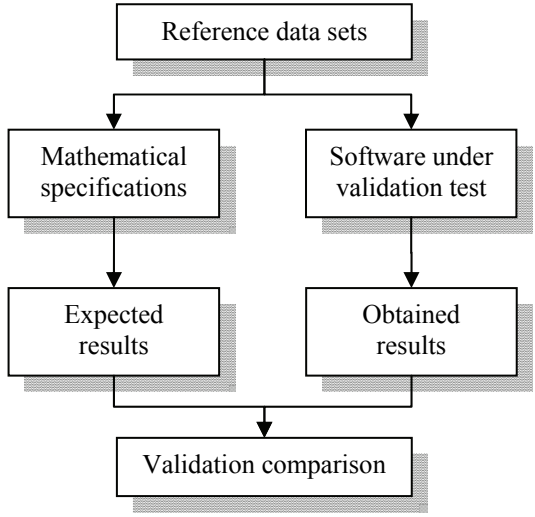


Figure 3. *Validation procedure combining reference data sets testing and mathematical specifications test*

The criteria adopted was that the difference between the reference output and the test output should not be higher than 0.010ºC, a value taken as quantification of the uncertainty due to this expected discrepancy. In the case tested the results obtained confirmed the fulfilment of the stated criteria and the instrument software was accepted, see Figure 4.

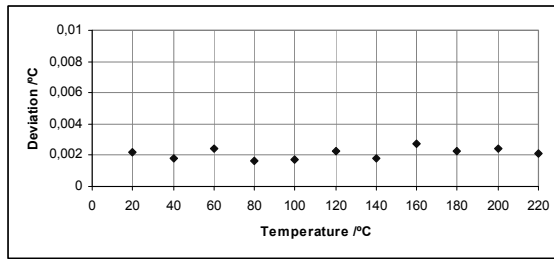


Figure 4. *Deviations in temperature of the observed value to the reference value in the validation test*

The second example concerns a automated system validation, related to geometric quantities verification in cubic, prismatic and triple metallic moulds [11] using a 3D coordinates measuring machine (CMM) as standard (Figure 5). The quantities to be measured are distance, angle and parallelism between faces, planicity of each face and rugosity.

Figure 5. *Testing of metallic triple mould*

This system is composed by different elements as seen in Figure 6, and is able to perform the following tasks: data acquisition; communication; data storage; data processing; comparison of results with tolerances; and certificate issue. All these tasks are grouped in a automated measurement system. The uncertainty evaluation and the acceptance criteria are based on software connected to the CMM own software, using the generated output CMM data.

The results of the automated process are influenced by a set of external inputs, as seen in Figure 6. The system was considered as a whole in order to define the *SIL*, bearing in mind its classification of 3 (commercial criticality type). According to the

system *SIL* classification, validation techniques were selected considering the major parameters to be validated, as summarized in Table 1.

The critical part of this process is the spatial behavior of the measuring system, since the temporal behavior, due to the low frequency involved, does not require an extensive validation process, which explains why the following discussion will be centered on the spatial testing components and results.

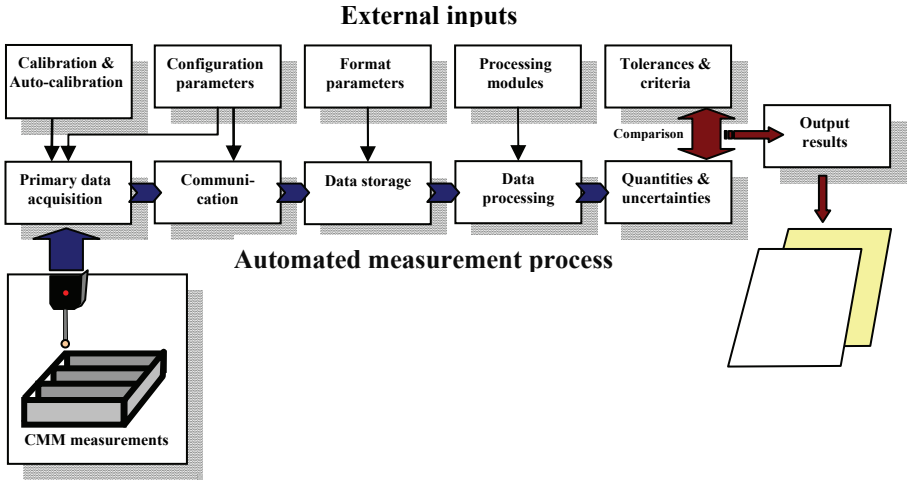


Figure 6. *Triple mould test using a CMM based automated process*

Component to validate	Validation techniques adopted	parameters under test	Reference parameters
Primary data acquisition	Temporal performance Reference test sets	Geometric quantities	2D standard grid
Communication	Interface analysis Critical timing analysis	String structure, buffer sizes, baud rate and other configuration parameters	Pre-defined configuration parameters
Data storage	Temporal performance Integrity test	Stored data	Primary data
Data processing	Reference test sets	Geometric quantities Function variables	Reference standard gauges
Evaluation of quantities estimates	Mathematical specifications	Output quantities	Data stored
Evaluation of uncertainties estimates	Reference test sets Mathematical specifications	Output measurement uncertainties	Input quantity data and mathematical specifications
Tolerances comparison	Numerical stability Boundary test sets	Acceptance criteria	Tolerances and results obtained

Table 2. *Validation techniques and parameters used in the CMM automated system*

It should be pointed out that, with respect to data processing and estimate evaluation, the techniques mentioned can be used separately or grouped together depending on the nature of the variables and algorithms. In the case of data processing, most of the testing was performed using reference test sets and stored data, whereas the evaluation of estimates was tested mainly using the mathematical specifications applied in the algorithms, and the uncertainty estimates were tested combining both techniques using stored data to apply the GUM procedure.

The illustration of the implementation of the validation tests is presented regarding three of the components: primary data acquisition, communication and data processing.

The primary data test was performed using a standard 2D matrix with known coordinates from (10 mm, 10 mm) to (100 mm, 100 mm) applied to the three different planes (OXY, OYZ and OXZ), followed by the measurement of these coordinates by the system in an automated process. The measured coordinates are graphically presented in Figure 7 showing the agreement with the predicted point (cross of gridlines). The comparison used a metric based on the deviation function and the metric acceptance criteria of ± 0.0005 mm, which was achieved.

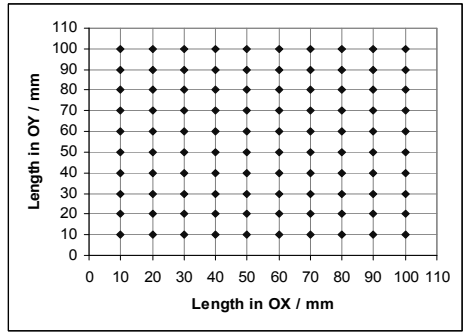


Figure 7. *Output coordinates in data acquisition test using a 2D standard grid*

The communication software was tested programming configuration settings, introducing input data and comparing the observed output with expected output, as presented in the scheme of Figure 8.

The data processing test was performed with standard gauges (block gauges, angular gauges and ring gauges) comparing the reference values with the generated values (using an automated process) of the geometric quantities in different ranges. The tests were performed considering the three planes (OXY, OYZ and OXZ). The validation results are expressed in Table 3 (the tolerances were determined based on the contributions considered with the uncertainty budget).

The examples presented indicate the diversity of solutions that can be adopted in order to validate components of systems.

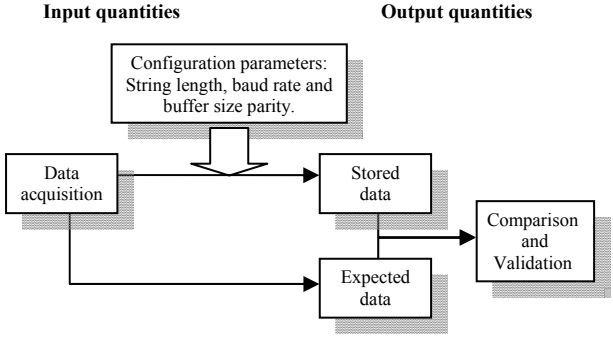


Figure 8. *Input and output quantities involved in the communication analysis*

Quantity	**Nominal values**	**Tolerances**	**Validation result** / mm		
			OXY	OYZ	OXZ
Length	100 mm	± 0.001 mm	< 0.001	< 0.001	< 0.001
Paralelism	100 mm	± 0.002 mm	< 0.001	< 0.002	< 0.002
Planicity	10 mm x 20 mm	± 0.002 mm	< 0.001	< 0.001	< 0.001
Roundness	100 mm/200 mm	± 0.002 mm	< 0.001	< 0.002	< 0.002
Angle	90º	± 0.05´	< 0.02´	< 0.03´	< 0.03´

Table 3. *Validation process tolerances and results*

Quantity	**Standard gauges**	**Validation result**		
		OXY	OYZ	OXZ
Length	Block gauges	Passed	Passed	Passed
Paralelism	Block gauges	Passed	Passed	Passed
Planicity	Block gauges	Passed	Passed	Passed
Roundness	Ring gauges	Passed	Passed	Passed
Angle	Calibrated Squares	Passed	Passed	Passed

Table 4. *Validation acceptance/rejection results*

Conclusions

The complex solutions adopted in modern automated measurement systems require that every influence component should be tested and validated as a part of the quality system assurance. This task is usually time consuming and the right effort is a matter of high importance for the selection of techniques as well as for the development of the tests.

Studies must be carried out in situations where the knowledge of the intrinsic parts of the systems and their functionalities is essential to define appropriate metrics, since these have a major role on the definition of the conditions and on the criteria that is the subject of a validation procedure.

It should be emphasized that the application of validation procedures to systems can also provide benefits in the correction or development of more reliable and robust solutions to the intended use and, therefore, it can be considered as an interesting tool applied in modern Quality Management Systems (QMS).

The software in automated systems is usually an important part of the uncertainty budget. Frequently its contributions are not readily obtained (e.g., the first “black-box” problem) and assumptions are often made. Therefore, the validation of systems has an important role not only in the assurance that these

assumptions are correct but also, in achieving an appropriate degree of confidence based on the application of validation techniques that can be documented and provided to clients, in metrology laboratories and in any other measurement activity.

References

[1] ISO/IEC 9126: 1991 – Information Technology – Software Product Evaluation. Quality Characteristics and Guidelines for their Use, International Organization for Standardization, Geneva (Switzerland).

[2] EN ISO/IEC 12207: 1995 – Software Life-Cycle Processes, International Organization for Standardization, Geneva (Switzerland).

[3] EN ISO 17025: 2005 - General requirements for the competence of testing and calibration laboratories. International Organization for Standardization, Geneva (Switzerland).

[4] Schulmeyer, G. G. and Mackenzie, G. R., Verification & Validation of Modern Software-Intensive Systems, Prentice Hall, Upper Saddle River (USA), 2000. ISBN 0-13-020584-2.

[5] Cox, M. G., Harris, P. M., Johnson, E. G., Kenward, P. D. and Parkin, G. I., Testing the Numerical Correctness of Software, Crown, National Physical Laboratory, Teddington (UK), 2004. ISSN 1471-0005.

[6] CEI/IEC 61508-1: 1998 – Functional Safety of Electric/Electronic/Programmable Electronic Safety-Related Systems. General Requirements, International Electrotechnical Commission, Geneva (Switzerland).

[7] CEI/IEC 61508-3: 1998 – Functional Safety of Electric/Electronic/Programmable Electronic Safety-Related Systems. Software Requirements, International Electrotechnical Commission, Geneva (Switzerland).

[8] Wichmann, B., Parkin, G. and Barker, R., Validation of Software in Measurement Systems. Software Support for Metrology Best Practice Guide No. 1, Crown, National Physical Laboratory, Teddington (UK), 2004. ISSN 1471-4124.

[9] Cox, M. G. and Harris, P. M., "The design and use of reference data sets for testing scientific software", Analytica Chimica Acta, V.380, **2**, 339-351, 1999.

[10] H. Preston-Thomas, "The International Temperature Scale of 1990 (ITS-90)", Metrologia 27, 3–10 (1990).

[11] EN 196-1: 2006. Métodos de Ensaio de cimentos. Parte 1: Determinação das Resistências Mecânicas. Instituto Português da Qualidade, Monte da Caparica (Portugal).

Transverse Disciplines in Metrology:
Proceedings of the 13th International Metrology Congress, 2007 – Lille, France
French College of Metrology

Index of Authors