

itENBIS and INRIM Joint Workshop

Mathematical and Statistical Methods for Metrology

30-31 May 2019

INRIM

Istituto Nazionale di Ricerca Metrologica
Strada delle Cacce, 91 – Torino, Italy

<http://www.msmm2019.polito.it/>



**POLITECNICO
DI TORINO**



Scientific Committee

João Alves e Sousa (IPQ)
Rossella Berni (Università di Firenze)
Walter Bich (INRIM)
Alen Bosnjakovic (IMBIH)
Oriano Bottauscio (INRIM)
Maurice Cox (NPL)
Severine Demeyer (LNE)
Stephen L. R. Ellison (LGC)
Clemens Elster (PTB)
Nicolas Fischer (LNE)
Alistair Forbes (NPL)
Fiorenzo Franceschini (Politecnico di Torino)
Rainer Göb (University of Wuerzburg)
Katy Klauenberg (PTB)
Alessandra Manzin (INRIM)
Francesca Pennechi, Co-chair (INRIM)
Jasmine Pétry (FOD Economie)
Antonio Pievatolo (IMATI – CNR)
Jacek Puchalski (GUM)
Jörg F. Unger (BAM)
Amalia Vanacore (Università di Napoli)
Adriaan van der Veen (VSL)
Grazia Vicario, Co-chair (Politecnico di Torino)
Luca Zilberti (INRIM)

Local Organizing Committee

Antonella Bianchi (PoliTo)
Enrico Bibbona (PoliTo)
Silvia Cavallero (INRIM)
Alessandra Durio (UniTo)
Roberto Fontana, Chair (PoliTo)
Gianluca Mastrantonio (PoliTo)
Elisabetta Melli (INRIM)
Francesca Pennechi (INRIM)

Mathematical and Statistical Methods for Metrology MSMM 2019

FINAL PROGRAMME

Thursday 30 May 2019

08:00-09:00	Registration
09:00-09:30	<p>Welcome Room: Conference Hall</p> <p>Maria Luisa Rastello (Scientific Director of the Istituto Nazionale di Ricerca Metrologica - INRIM), Grazia Vicario (<i>itENBIS</i>, Politecnico di Torino) and Francesca Pennechi (Istituto Nazionale di Ricerca Metrologica - INRIM)</p>
09:30-10:20	<p>Keynote Lecture 1 Room: Conference Hall Session Chair: Maurice Cox</p> <p>Anthony O'Hagan (Emeritus Professor of Statistics, School of Mathematics and Statistics, The University of Sheffield, UK), Measurement or Estimation? <i>page 1</i></p>
10:20-10:50	<p><i>Coffee break</i> Room: Expo room</p>
10:50-11:50	<p>Plenary Session JCGM Working Group on the Expression of Uncertainty in Measurement (GUM) Session Chair: Anthony O'Hagan Room: Conference Hall</p> <p>10:50-11:10 The new perspective of the Guide to the expression of uncertainty in measurement, Walter Bich (JCGM-WG 1 chair, Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 2</i></p> <p>11:10-11:30 Informative Bayesian Type A uncertainty evaluation, especially for very few observations, Maurice Cox (JCGM-WG 1, National Physical Laboratory - NPL) <i>page 4</i></p> <p>11:30-11:50 Applying knowledge in routine measurement in type A evaluations of standard uncertainty, Adriaan van der Veen (JCGM-WG 1, Nederlands Metrologisch Instituut - VSL) <i>page 6</i></p>
11:50-12:50	<p>Plenary session Machine Learning and Predictive models Session Chair: Luca Zilberti Room: Conference Hall</p> <p>11:50-12:10 Laplacian Mode Clustering - Leveraging manifold learning for mode finding, Kavya Jagan (National Physical Laboratory - NPL) <i>page 8</i></p> <p>12:10-12:30 Segmentation of magnetic resonance brain images using deep learning, Bozen Jovanovski (Faculty of Computer and Information Science, University of Ljubljana) <i>page 10</i></p> <p>12:30-12:50 The problem of “calibration” of predictive models, Giampaolo E. D'Errico (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 12</i></p>
12:50-14:00	<p><i>Lunch break</i> Room: Expo room</p>

<p>14:00-15:00</p>	<p>Parallel session Uncertainty evaluation 1 Session Chair: João Alves e Sousa Room: Conference Hall</p> <p>14:00-14:20 Monte Carlo estimation of calibration uncertainty for instrumented indentation testing machines, Giacomo Maculotti (DIGEP, Politecnico di Torino) <i>page 14</i></p> <p>14:20-14:40 Modelling of correlation and estimation of coverage region in multivariate measurements with correlated or uncorrelated sources of uncertainty, Jacek Puchalski (Central Office of Measures - GUM) <i>page 16</i></p> <p>14:40-15:00 Uncertainty propagation in quantitative imaging of electric properties using Monte Carlo method, Alessandro Arduino (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 18</i></p> <hr/> <p>Parallel session Conformity assessment and Process/Quality control 1 Session Chair: Alistair Forbes Room: Seminar room</p> <p>14:00-14:20 The role of the measurement uncertainty in the decision making process of conformity assessment, Alen Bosnjakovic (Institute of Metrology of Bosnia and Herzegovina - IMBiH) <i>page 20</i></p> <p>14:20-14:40 Conformity assessment of multicomponent materials or objects using compositional data, Francesca Pennechi (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 23</i></p> <p>14:40-15:00 Assessment of inspection strategies in assembly manufacturing processes, Elisa Verna (DIGEP, Politecnico di Torino) <i>page 25</i></p>
<p>15:00-16:00</p>	<p>Parallel session Uncertainty evaluation 2 Session Chair: Adriaan van der Veen Room: Conference Hall</p> <p>15:00-15:20 An attempt to introduce the use of covariances to evaluate the uncertainty of the detection efficiency in k_0-NAA, Marco Di Luzio (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 27</i></p> <p>15:20-15:40 Measurement uncertainty evaluation and correlation analysis of gas flow measurements in a system of three parallel pipes, Zijad Dzemic (Institute of Metrology of Bosnia and Herzegovina - IMBiH) <i>page 29</i></p> <p>15:40-16:00 A statistical method for the evaluation of near-surface air measurement uncertainty due to nearby roads, Graziano Coppa (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 31</i></p> <hr/> <p>Parallel session Conformity assessment and Process/Quality control 2 Session Chair: Maria Sole Pellegrino Room: Seminar room</p> <p>15:00-15:20 On the detection of anomalous values of Radioxenon in IMS stations of CTBTO, Michele Scagliarini (Dipartimento di Scienze Statistiche, Università di Bologna) <i>page 33</i></p> <p>15:20-15:40 The evaluation of chronic alcohol abuse biomarkers in hair samples: the interpretation of cut-off values, Marco Vincenti (Dipartimento di Chimica, Università degli Studi di Torino) <i>page 35</i></p> <p>15:40-16:00 Assessment and management of occupational exposure to airborne dust in a quality approach, Rebecca Nebbia (Politecnico di Torino) <i>page 37</i></p>

16:00-16:30	<p><i>Coffee break</i> Room: Expo room</p>
16:30-17:30	<p>Parallel session Coordinate metrology Session Chair: Grazia Vicario Room: Conference Hall</p> <p>16:30-16:50 Statistical analysis of coordinate metrology data with Q-DAS, Suela Ruffa e Marco Zeno (Hexagon Manufacturing Intelligence) <i>page 39</i></p> <p>16:50-17:10 EUCoM: Evaluating the uncertainty in coordinate metrology, Alessandro Balsamo (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 41</i></p> <p>17:10-17:30 Uncertainty evaluation in coordinate metrology based on approximate models of CMM behaviour, Alistair Forbes (National Physical Laboratory - NPL) <i>page 43</i></p> <hr/> <p>Parallel session Time and frequency metrology Session Chair: Alen Bosnjakovic Room: Seminar room</p> <p>16:30-16:50 Allan variance and novel signals from optical metrology: how to avoid aliasing, Claudio Calosso (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 45</i></p> <p>16:50-17:10 Role of dead times and correlations in the measurements of optical clocks, Marco Pizzocaro (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 47</i></p> <p>17:10-17:30 Two-Sample covariance and multichannel tracking DDS for measuring the frequency stability of cryogenic sapphire oscillators, Claudio Calosso (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 49</i></p>
17:30-18:10	<p>Parallel session Chemometric methods Session Chair: Stephen L. R. Ellison Room: Conference Hall</p> <p>17:30-17:50 Effective validation of chromatographic analytical methods, Eleonora Amante (Dipartimento di Chimica, Università degli Studi di Torino) <i>page 51</i></p> <p>17:50-18:10 Chemometric handling of Raman spectra for live systems monitoring and susceptibility tests, Luisa Mandrile (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 53</i></p> <hr/> <p>Parallel session Optimization algorithms Session Chair: Oriano Bottauscio Room: Seminar room</p> <p>17:30-17:50 Calibration Curve Computing (CCC) software v2.0: a new version of INRIM's software for regression analysis, Maricarmen Lecuna (Politecnico di Torino) <i>page 55</i></p> <p>17:50-18:10 Advanced statistical techniques to assess nonclassicality in ensembles of single photon emitters, Ivo Pietro Degiovanni (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 57</i></p>
18:10-20:30	<p><i>Welcome party at the INRIM</i> Venue: INRIM gardens (or Expo room in case of bad weather)</p>

Friday 31 May 2019

08:30-09:00	Registration
09:00-09:50	<p>Keynote Lecture 2 Room: Conference Hall Session Chair: Walter Bich</p> <p>Giulio D'Agostini (Professor of the Physics Department of the University of Rome "La Sapienza", Roma, Italy, and INFN Roma 1), p-values: meaning, misconceptions and dangers, but also their practical utility if used <i>cum grano salis</i> page 59</p>
09:50-10:30	<p>Plenary Session Statistical indices Session Chair: Grazia Vicario Room: Conference Hall</p> <p>09:50-10:10 The median scaled difference: an outlier-resistant and model-independent indicator of anomalies for Key Comparison data, Stephen L R Ellison (Laboratory of the Government Chemist - LGC) <i>page 60</i></p> <p>10:10-10:30 Consistency of subjective evaluations: an investigation into the paradoxical behaviour of inter-rater agreement coefficients, Maria Sole Pellegrino (Dipartimento di Ingegneria Industriale, Università di Napoli "Federico II") <i>page 61</i></p>
10:30-11:00	<p><i>Coffee break</i> Room: Expo room</p>
11:00-12:20	<p>Plenary Session Simulated experiments and numerical modelling Session Chair: Alessandra Manzin Room: Conference Hall</p> <p>11:00-11:20 Virtual measurements on patients with orthopaedic prostheses during MRI, Luca Zilberti (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 63</i></p> <p>11:20-11:40 Advanced modelling of stochastic distributions of magnetic nanostructures, Riccardo Ferrero (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 65</i></p> <p>11:40-12:00 Modelling of magnetic nanobead transport in a microvascular network, Marta Vicentini (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 67</i></p> <p>12:00-12:20 Accuracy in Electrical Resistance Tomography: from measurements to maps, Alessandro Cultrera (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 69</i></p>
12:20-13:30	<p><i>Lunch break</i> Room: Expo room</p>
13:30-15:00	<p>Plenary Session EMPIR EMUE Project "Advancing measurement uncertainty - comprehensive examples for key international standards" Session Chair: Francesca Pennecchi Room: Conference Hall</p> <p>13:30-13:40 EMUE: Introduction to the Project, Maurice Cox (National Physical Laboratory - NPL)</p> <p>13:40-14:00 Two-point and multi-point interpolation of calibration data, Maurice Cox (National Physical Laboratory - NPL) <i>page 71</i></p>

	<p>14:00-14:20 Getting started with uncertainty evaluation using the Monte Carlo method in R, Adriaan van der Veen (Nederlands Metrologisch Instituut - VSL) <i>page 73</i></p> <p>14:20-14:40 Examples of measurement uncertainty evaluation applied to microflow, flow and thermal comfort, João Alves e Sousa (Instituto Português da Qualidade - IPQ) <i>page 75</i></p> <p>14:40-15:00 Uncertainty estimation in multi-channel data-acquisition systems, Alessio Carullo (ACCREDIA-DT Torino - Politecnico di Torino) <i>page 77</i></p>
15:00 -16:00	<p>Plenary Session Bayesian methods Session Chair: Giulio D'Agostini Room: Conference Hall</p> <p>15:00-15:20 The posterior mean-power of normal variables, Giovanni Mana (Istituto Nazionale di Ricerca Metrologica - INRIM) <i>page 79</i></p> <p>15:20-15:40 Bayesian inference on the parameters of the truncated normal distribution and application to reverberation chamber measurement data, Carlo Carobbi (Dipartimento dell'Ingegneria dell'informazione, Università degli Studi di Firenze) <i>page 81</i></p> <p>15:40-16:00 An optimal Bayesian design for a reliability study, Rossella Berni (Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Università degli Studi di Firenze) <i>page 83</i></p>
16:00-16:15	<p>Closure Room: Conference Hall</p>

Measurement or Estimation?

Anthony O'Hagan

Invited Speaker - Emeritus Professor of Statistics, School of Mathematics and Statistics, The University of Sheffield (UK)

In any structure or enterprise, it is important to ensure that the foundations are solid. We can go on adding more fancy elements but if the foundations are weak then at some point cracks will appear. Metrology is a fascinating field, the GUM is a remarkable document and some truly outstanding technical methods have been built on this foundation, but it is a mess. So this talk is about rebuilding the foundations and making them solid and fit for purpose. Any such rebuilding must ask fundamental questions. What do we mean by (the) measurement? What is the meaning of standard uncertainty, and is it fit for purpose?

My answers may be a little radical, so it is important to ask: are they realistic? The metrology community has already rejected change once because they didn't like, and didn't accept the need for, the practical consequences. This is a much harder question for me because I'm a statistician, not a metrologist, but I will do my best to address the reality of everyday metrology.

The new perspective of the Guide to the expression of uncertainty in measurement

Walter Bich¹

Key words: measurement uncertainty, Joint Committee for Guides in Metrology

1 General

The Guide to the expression of uncertainty in measurement, hereafter the *Guide*, is the undisputed reference as concerns uncertainty evaluation in metrology. It has served now for more than 25 years, and its great merit has been the world-wide harmonization of the previously disparate procedures for uncertainty evaluation.

The *Guide* is now published under the joint copyright of the eight International Organizations whose core business is measurement, and which constitute the Joint Committee for Guides in Metrology (JCGM) [1].

In the Scope of the *Guide*, it is declared that "This *Guide* establishes general rules for evaluating and expressing uncertainty in measurement that can be followed at various levels of accuracy and in many fields — from the shop floor to fundamental research".

Despite its broad scope and the international influence of the Organizations holding its copyright, the GUM is barely known, let alone adopted, beyond the comparatively restricted circle of the National metrology institutes, calibration and industrial laboratories, and legal metrology. No wonder for this: the GUM is essentially based on Gauss' law of propagation of errors. This law has enormous merits and still is fit for purpose in innumerable applications. But, as it always happens, no single rule can accommodate the totality of problems, and the GUM method serves a range of problems which is narrower than the claimed scope.

If the GUM is to be more widely known and adopted within the JCGM Member Organizations and possibly beyond them, it is necessary that a range of techniques broader than the existing ones be included. The JCGM-WG1 already developed in the past a number of documents ancillary to the GUM (see [2]). However, the

¹ Walter Bich
INRiM, strada delle Cacce 91, 10135 Torino, Italy, w.bich@inrim.it

JCGM-WG1 subsequently realized that a radical change of perspective was needed. This GUM New Perspective was elaborated and submitted to the JCGM, which endorsed it in 2017.

In this GUM New Perspective, the interrelationships among the various documents is highlighted in a more user-oriented way. An introductory document sets the scene for the expression, use and evaluation of measurement uncertainty, and gives a general outline of what is involved and directs the reader to the document of interest. All the other documents focus on a specific purpose. This new organisational structure emphasizes that users with different measurement problems require different guidance, which will now be available in largely stand-alone documents. With this shift in perspective, a sustainable structure is created which underlines that different methods for uncertainty evaluation exist next to one another, alongside specific application documents. The structure enables the JCGM to respond in a flexible manner to emerging needs from the measurement community and will furthermore enable the community to benefit from the improved accessibility of the documents. An essential aspect is that the preservation of the inheritance built up since 1993 with the current GUM is also accommodated in this new perspective, thus meeting requests for its grandfathering raised by some influential users. Finally, to underline the relationship between the documents and their commonalities, the entire suite of documents will be published under the common title *Guide to the expression of uncertainty in measurement*. All documents become part of this suite, with the introductory document as part 1. In conclusion, the “new GUM”, Parts 0, 1, 4, 5 and 6 currently being available (with existing numbering in parentheses), will be the whole collection of documents:

Guide to the expression of uncertainty in measurement (GUM)
Part 0 Current "grandfathered" GUM (ex JCGM 100)
Part 1 Introduction (JCGM 104)
Part 2 Concepts (JCGM 105)
Part 3 Developing and using measurement models (JCGM 103)
Part 4 Propagation of distributions using a Monte Carlo method (JCGM 101)
Part 5 Extension to any number of output quantities (JCGM 102)
Part 6 Role of measurement uncertainty in conformity assessment (JCGM 106)
Part 7 Least squares methods (JCGM 107)
Part 8 Bayesian methods (JCGM 108)
Part 9 Statistical models and data analysis for interlaboratory studies (JCGM 109)
Part 10 Examples of uncertainty evaluation
Part 11 Basic method for uncertainty propagation (JCGM 111)
Part 12 Nominal properties

References

1. <https://www.bipm.org/en/committees/jc/jcgm/>
2. <https://www.bipm.org/en/publications/guides/>

Informative Bayesian Type A uncertainty evaluation, especially for very few observations

Maurice Cox and Katsuhiko Shirono

Key words: GUM, measurement uncertainty, Type A evaluation, few observations

A criticism of the *Guide to the expression of uncertainty in measurement* (GUM) [2] is that it is based on both frequentist and Bayesian thinking. Type A (statistical) uncertainty evaluations are frequentist, whereas Type B evaluations, using state-of-knowledge distributions, are Bayesian. In contrast, making the GUM fully Bayesian implies that a conventional Bayesian approach to Type A uncertainty evaluation for n repeated observations leads to the unpalatable consequence that n must be at least equal to 4, presenting a difficulty for calibration and testing laboratories.

This presentation summarizes a Bayesian analysis of Type A uncertainty evaluation that applies for all $n \geq 2$ [4], as in the frequentist analysis in the current GUM. The analysis is based on assuming that the observations are drawn from a normal distribution (as in the Bayesian analysis in GUM Supplement 1 [3]), but uses an informative prior based on lower and upper bounds for the standard deviation of the sampling distribution for the quantity under consideration.

The main outcome of the analysis is a closed-form mathematical expression for the factor by which the standard deviation of the mean observation should be multiplied to obtain the required standard uncertainty. The method is straightforward to apply in that the expression can be implemented in a few lines of MATLAB and with similar effort in Excel and other environments.

Comparison is made with the Kacker-Jones method [7], which although *ad hoc* has been adopted in ANSI and IEC Standards [1, 5]. Conditions are given in which the presented method outperforms the Kacker-Jones method.

National Physical Laboratory
Teddington TW11 0LW, UK e-mail: maurice.cox@npl.co.uk

National Metrology Institute of Japan, National Institute of Advanced Industrial Science and Technology, Ibaraki 305-8565, Japan e-mail: k.shirono@aist.go.jp

1

Three metrological examples with data provided by domain experts are used to illustrate the approach. In each case only two or three observations are made under repeatability conditions of measurement.

1. *Mass measurement.* Secondary calibration laboratories calibrate sets of weights by comparing each unknown weight with the corresponding standard in a calibrated reference set. Several data sets are considered, each consisting of two or three observations.
2. *Coal-property characterization.* Mass fraction volatile matter is a characterization of coal used for industrial purposes. The relevant standard, ISO 562:2010 [6], stipulates that only two observations shall be obtained of mass fraction volatile matter.
3. *Vacuum-gauge calibration.* To calibrate manometers, the S-factor, the ratio of a reference pressure and an indicated pressure, is employed. Only two observations are commonly made at each reference pressure.

In each example, the Bayesian prior is based on metrological knowledge provided by a domain expert. Validation of the method through sensitivity studies is considered. Possible deficiencies in the approach in some circumstances are identified. Scope for further work is given.

Acknowledgements Valuable information was provided by Dr Walter Bich (INRIM, Italy), Dr Adriaan van der Veen (VSL, The Netherlands) and Dr John Greenwood (United Kingdom Accreditation Service).

References

1. ANSI. *ANSI C63.23: American National Standard Guide for Electromagnetic Compatibility-Computations and Treatment of Measurement Uncertainty*. ANSI, Washington, 2012.
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, AND OIML. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008.
3. 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. Joint Committee for Guides in Metrology, JCGM 101:2008, 2008.
4. COX, M., AND SHIRONO, K. Informative Bayesian Type A uncertainty evaluation, especially applicable to a small number of observations. *Metrologia* 54, 5 (2017), 642.
5. IEC. *IEC/TR 61000-1-6: Electromagnetic compatibility (EMC). General. Guide to the assessment of measurement uncertainty*. IEC, Geneva, 2012.
6. ISO. *ISO 562: Hard coal and coke – Determination of volatile matter*. ISO, Geneva, 2010.
7. KACKER, R., AND JONES, A. On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent. *Metrologia* 40 (2003), 235–248.

Applying knowledge in routine measurement in type A evaluations of standard uncertainty

Adriaan M.H. van der Veen

Key words: Measurement uncertainty, Bayesian inference, type A evaluation, ANOVA, regression

Abstract

In the Guide to the expression of Uncertainty in Measurement (GUM) [1], the evaluation of standard uncertainty from series of observations is based on the normal distribution. This approach is simple and straightforward, but not address the dispersion in precision data in full. Supplement 1 to the GUM (GUM-S1) has attempted to address this issue by introducing a very specific kind of Bayesian evaluation of series of data, which results in the use of Student's t distribution [2]. Not only has this recommendation created an inconsistency between the GUM and GUM-S1 [3], it also left those experimenters with a problem that rely on measurements in duplicate or triplicate with a serious problem. The Monte Carlo method cannot be used as described in GUM-S1, as it requires using the t distribution and this distribution has a defined mean and standard deviation only when the number of observations $n \geq 4$.

The response to this problem has been unhelpful to practioners, as arguments were given such as “one should not use only two or three observations to calculate a standard deviation” and advice was given in the spirit of “use the standard deviation from validation” or “just take more replicates”. None of these responses addresses the concern of most experimenters. They consider a standard deviation obtained for a particular sample or artefact to be more representative for the measurement in question than a standard deviation obtained on a different day, a different sample (or series of samples), and a different measured value.

Adriaan M.H. van der Veen
VSL, Thijsseweg 11, 2629 JA Delft, The Netherlands e-mail: avdveen@vsl.nl

Nevertheless, laboratories possess a wealth of information about their measurements and procedures. For example, a mass laboratory will know the nominal mass of the weight(s) offered for calibration. It will also know the repeatability and reproducibility of its balance, among many other information obtained when verifying or validating its method. Environmental laboratories determining the concentration pollutants in water will usually have little information about the nominal value, but they have a similar information position with regard to the methods they are using. They also know whether their results are meaningfully biased, for example from proficiency testing or using certified reference materials [4]. In this paper, we revisit first briefly Bayesian models for performing type A evaluation of standard uncertainty (e.g., mean and standard deviation, analysis of variance, regression). Then we explain why these models are useful in addressing the above issue with the t distribution. We show that especially for $n \leq 4$ Bayesian methods are very suitable to address the concern that such a pool of data is too small to base a standard deviation on as well as that the standard deviation obtained during the measurement would be more “representative” than a standard deviation from shelf [5]. As a final step, we show how the complicated Markov Chain Monte Carlo method, often required in Bayesian inference, can be replaced in many instances by something simpler that can readily be implemented in any kind of data evaluation. A short excursion is made to other type A methods, such as analysis of variance and regression to demonstrate that the approach can also be used in more complex evaluations.

Acknowledgements The research described in this paper received funding from the Ministry of Economic Affairs and Climate of the Netherlands.

References

- [1] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections*. BIPM, 2008.
- [2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Supplement 1 to the ‘Guide to the Expression of Uncertainty in Measurement’ – Propagation of distributions using a Monte Carlo method, JCGM 101:2008*. BIPM, 2008.
- [3] R Kacker and A Jones. On use of bayesian statistics to make the guide to the expression of uncertainty in measurement consistent. *Metrologia*, 40(5):235, 2003. URL <http://stacks.iop.org/0026-1394/40/i=5/a=305>.
- [4] Adriaan M. H. van der Veen. Revision of ISO Guide 33: good practice in using reference materials. *Accreditation and Quality Assurance*, 20(6):529–532, aug 2015. doi: 10.1007/s00769-015-1167-7. URL <https://doi.org/10.1007/s00769-015-1167-7>.
- [5] Adriaan M.H. van der Veen. Bayesian methods for type A evaluation of standard uncertainty. *Metrologia*, 55(5):670–684, jul 2018. doi: 10.1088/1681-7575/aad103. URL <https://doi.org/10.1088/1681-7575/aad103>.

Laplacian Mode Clustering

Leveraging manifold learning for mode finding

Kavya Jagan, Stéphane Chrétien

Key words: Clustering, Laplacian Eigenmaps, Mode finding

The purpose of clustering is to identify groups in a dataset in the hope of revealing some unforeseen latent discrete variable [4]. A natural way of clustering data is to partition it based on a measure of distance – points close to each other are considered to be members of the same cluster while those far apart are regarded as belonging to different clusters.

The clustering algorithm described in this paper is based on affinities, quantities that are inversely related to distance. We use a Gaussian kernel of the form

$$A(i, j) = \exp\left(-\beta \|x_i - x_j\|_2^2\right),$$

where $\|\cdot\|_2$ denotes the L2 distance between 2 data points. The choice of the scale factor β is critical for the performance of the clustering algorithm and a method of setting β is described. This is however, an open problem and there is scope for further work in this area.

The method incorporates embedding of the data using Laplacian Eigenmaps [3] to identify a change of coordinates that better suits the clustering purpose. It is assumed that the data come from a multivariate multimodal distribution where the modes represent different features. In practice, it has been observed that Laplacian Eigenmaps preserve the modes of this distribution even when the dimension of the data is reduced.

Kavya Jagan
National Physical Laboratory, Teddington, UK, e-mail: kavya.jagan@npl.co.uk

Stéphane Chrétien
National Physical Laboratory, Teddington, UK, e-mail: stephane.chretien@npl.co.uk

The algorithm does not depend on the initialisation of the procedure, unlike iterative schemes like K-means clustering [4] or model-based methods like the EM algorithm [1]. Therefore, the result is always the same for each running of the algorithm. Another advantage of the algorithm is that the number of clusters need not be known in advance. The main drawback of this approach is that it is not valid when the transformed data are of a very high dimension, as mode detection is then well known to suffer from the curse of dimensionality. Although low dimensionality is often obtained by applying the Laplacian eigenmap embedding, some rare intrinsically high dimensional instances may not be amenable to this approach.

The paper also discusses methods to assess the performance of clustering algorithms where there is no ground truth. In some low-dimensional cases it is possible to gauge by eye how good a clustering scheme is. In most cases, however, such assessment is not feasible. In order to provide a more principled approach to assessing clusterings, methods based on sensitivity derived from the theory of coresets [2], silhouettes [6] and model-based methods [5] are discussed.

The methods are illustrated on satellite data relating to subsidence monitoring. The clustering algorithm is used to find clusters of areas showing similar ground motion behaviour.

References

1. Murray Aitkin and Granville Tunnicliffe Wilson. Mixture models, outliers, and the EM algorithm. *Technometrics*, 22(3):325–331, 1980.
2. Olivier Bachem, Lucic Mario, and Krause Andreas. Practical coreset constructions for machine learning.
3. Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.
4. Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Springer series in statistics New York, 2001.
5. Geoffrey McLachlan and David Peel. *Finite mixture models*. John Wiley & Sons, 2004.
6. Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

Segmentation of magnetic resonance brain images using deep learning

Božen Jovanovski, Dejan Georgiev, Jernej Avsenik, and Aleksander Sadikov

Key words: deep learning, brain MRI, image segmentation, machine learning

1 Introduction

Image segmentation is an important task in today's bio-medical research. Magnetic resonance imaging (MRI) of the brain can be troublesome to segment when looking for tumours or other lesions. Manually this task takes a lot of time and automatising it can help medical practitioners and researchers by automating the discovery of a condition, making it comparable, and potentially treatable in time.

Brain tumours can be in a variety of shapes and sizes and their location can be almost anywhere in the brain tissue. They can also have fuzzy borders and can be hard to distinguish from healthy tissue. Because of the nature of this problem we believe that deep neural networks (DNN) are most suited for this task. Our choice of

Božen Jovanovski
University of Ljubljana, Faculty of Computer and Information Science, Večna Pot 113, Ljubljana, Slovenia, e-mail: bj9914@student.uni-lj.si

Dejan Georgiev
Ljubljana University Medical Centre, Department of Neurology, Zaloška cesta 2, Ljubljana, Slovenia
Faculty of Computer and Information Science, University of Ljubljana, Večna Pot 113, Ljubljana, e-mail: dejan.georgiev@kclj.si

Jernej Avsenik
Ljubljana University Medical Centre, Institute of Radiology, Zaloška cesta 7, Ljubljana, Slovenia
e-mail: jernej.avsenik@kclj.si

Aleksander Sadikov
University of Ljubljana, Faculty of Computer and Information Science, Večna Pot 113, Ljubljana, Slovenia e-mail: aleksander.sadikov@fri.uni-lj.si

the network is the generative adversarial network (GAN) [Goodfellow et al(2013)] which has grown in popularity in the past few years. It can learn to mimic any distribution of data and learn a loss function while training to minimise that same loss, making it efficient in image analysis tasks.

2 Methodology

We will train and test our GAN network using different publicly available datasets, comprised not only of images of brains with tumors, but also of brain images with other brain lesions. The data comes from the BraTS Challenge [Menze et al(2015), Bakas et al(2017)], the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, and the Ischemic Stroke Lesion Segmentation (ISLES) Challenge [Maier et al(2017), Kistler et al(2013)]. The BraTS Challenge provides us with MRI images of abnormal brains with tumors, while the ADNI database and the ISLES Challenge provide us with MRI images of abnormal brains with lesions.

Final testing of the learned model will be on MRI brain images from the Ljubljana University Medical Centre, Institute of Radiology. We will evaluate our algorithm using the "Dice score" and "Hausdorff distance" methods. We expect the use of various datasets with different brain abnormalities to give better results than using a single data source.

GAN networks have been used to perform a segmentation of MRI images of different organs. Our plan, however, is to do a segmentation of the different abnormalities that occur in the brain tissues, caused by various diseases, automating and improving the process of discovering and diagnosing patients with these diseases.

References

- [Bakas et al(2017)] Bakas S, et al (2017) Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* 4(170117):n.pag.
- [Goodfellow et al(2013)] Goodfellow I, et al (2013) Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 27 pp 2672–2680
- [Kistler et al(2013)] Kistler M, et al (2013) The Virtual Skeleton Database: An Open Access Repository for Biomedical Research and Collaboration. *Journal of Medical Internet Research* 15(11):e245
- [Maier et al(2017)] Maier O, et al (2017) ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis* 35:250–269
- [Menze et al(2015)] Menze B, et al (2015) The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34(10):1993 – 2024

The problem of “calibration” of predictive models

Giampaolo E. D’Errico¹

Key words: predictive models, probabilistic forecasts, calibration tests, manipulability

1 Introduction

The performance of prediction models can be evaluated according to a variety of criteria: “calibration” is one of great importance. Calibration tests are used to assess probabilistic models. In this area, the meaning of calibration differs from the definition given in the international vocabulary of metrology [3].

In a probability forecast over a sequence of periods, a forecaster declares the probability that an event will be realized in the next period. Forecasts are correct if they coincide with the conditional probabilities of the underlying data generating process. Correct forecasts are “calibrated”: calibration is a property that can be verified by means of calibration tests aimed at comparing the realized frequency of the event with the forecast average, over the same sequence. However – diversely from deterministic models where only forecasts that exactly predict the future events will pass the calibration test – forecasts of probability model that are calibrated are not necessarily correct [1, 2, 7].

The complexity of making a decision on the base of probabilistically predicted events is a valid reason to consult experts (humans or expert systems). A major problem is that not only skilled experts’ forecasts may be unrealized; but also forecasts given at random by a strategic forecaster may seem to be calibrated.

This problem is described in terms of “manipulability” of calibration tests. Application fields of probabilistic predictive models include meteorology, econometrics, clinics, and epidemiology.

2 “Manipulability” of calibration tests

¹ INRIM, e-mail: g.derrico@inrim.it

A forecasting model is defined correct if $f = p(x|f)$, where $p(x)$ is the probability distribution of forecast event x , whose realized frequency results to be $f(x)$ by empirical observation over the same period [4]. A calibration test such that a correct model passes the test with probability 1 is said to control for type I error. The plot of $p(x)$ vs f is called the “empirical calibration curve”: if the curve is diagonal, the expert is said “well calibrated” [2]. Thus, correctness implies calibration. However, calibration does not imply correctness. For example, let a forecast be given with application to a future event x (e.g., $x = 1$ for rain, otherwise $x = 0$) over a sequence of a pair number of periods. Let the observations be constantly $x = 1$ (rain) in odd periods and $x = 0$ in even periods. If rain is uniformly predicted with probability $p = 0.5$, forecasts are calibrated: $f = p = 0.5$ over the whole sequence of periods; but the same forecast is not correct: $f = 1$ ($f = 0$) over each odd (even, respectively) period [7]: this would be a futile prediction, though.

Moreover, calibration tests that control for type I error are unable to reject any futile but strategic models that produce randomized forecasts purposely aimed at passing the test [5, 8]. This kind of tests is said “manipulable”: the problem is that manipulability may introduce a bias when assessing effectiveness of predictive models and reliability of probability forecasts. If the expert’s model is being tested for calibration, a reputed expert might consider declaring futile forecasts to preserve reputation: undergoing the test with probability 1 of success might be preferred to pursuing correctness with related uncertainty [1, 5]. In fact non-manipulable tests [6] can be devised to control for type I error, but they are not of the calibration kind.

3 Conclusion

Calibration of a prediction model is not a valid criterion to distinguish truly expert’s forecasts from those of any futile models designed to pass a calibration test. This is the problem of possible manipulation, by means of strategically randomized forecast, of calibration tests. An outline of this problem has been proposed here.

References

1. Brier G W 1950 Verification of Forecasts Expressed in Terms of Probability *Monthly Weather Review* **78** 1
2. Dawid A P 1982 The Well-Calibrated Bayesian *Journal of the American Statistical Association* **77**(379) 605
3. ISO/IEC Guide 99-12:2007. *International Vocabulary of Metrology*
4. Murphy A H and Winkler 1987 A General Framework for Forecast Verification *Monthly Weather Review* **115** 1330
5. Olszewski W and Sandroni A 2008 Strategic Manipulation of Empirical Tests *Mathematics of Operations Research* **34**(1) 57
6. Olszewski W and Sandroni A 2009 A Nonmanipulable Test *The Annals of Statistics* **37**(2) 1013
7. Sandroni A 2003 The Reproducible Properties of Correct Forecasts *Int J Game Theory* **32** 151
8. Shmaya E 2008 Many Inspections are Manipulable *Theoretical Economics* **3** 367

Monte Carlo estimation of calibration uncertainty for Instrumented Indentation testing machines

Giacomo Maculotti¹, Gianfranco Genta² and Maurizio Galetto³

Key words: Instrumented Indentation Test, Calibration, Uncertainty, Monte Carlo

Extended Abstract

Instrumented Indentation Test (IIT) is a depth sensing technique that provides thorough mechanical characterisation of material properties. It is particularly relevant for enabling hardness characterisation at nano-scale level, where traditional optical instruments are ineffective [2]. IIT requires to perform a loading-unloading cycle, throughout which the applied force, F , and the indenter displacement, h , are measured by devoted transducers. Mechanical properties can be evaluated, once it is established a relationship, $A(h)$, between the area of the contact surface of the indenter in the material, A , and the indenter displacement, h . The functional form of $A(h)$ depends on the indenter shape, e.g. for a Berkovich indenter it is a second order polynomial [2]. However, some errors, e.g. the estimate of first contact point of the indenter on the sample, “zero error” h_0 , the non-null compliance of the testing machine, C_f , and the material elastic recovery, are introduced in the measurement of h and compel a correction. In general, the system stiffness, S_m , is modelled as a series of springs representing for the contribution of the frame compliance, C_f , and of the sample

¹ Giacomo Maculotti
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, e-mail: giacomo.maculotti@polito.it

² Gianfranco Genta
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, e-mail: gianfranco.genta@polito.it

³ Maurizio Galetto
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, e-mail: maurizio.galetto@polito.it

material stiffness, S . With a Berkovich indenter, C_f and the parameters of the polynomial form of $A(h)$ are the major contributions to measurement uncertainty in IIT applications [1]. Thus, they require careful estimation. ISO 14577-2 [3] suggests six methods, two of which, i.e. no.2 and no.4 according to Annex D, enable the parameters evaluation by an iterative procedure, which is cost effective as it does not need the use of an AFM for $A(h)$ parameters evaluation. These two methods only differ for exploiting, respectively, one or two materials for calibration. However, neither the standard nor the literature address how the uncertainty of the calibrated parameters shall be evaluated. Referring to method no.2, in this work, a Monte Carlo approach is adopted to give a solution to this normative gap [4]. Influencing factors are shown in Fig.1, with indication of the assumed distribution and the source of information. Parameter distributions, when not provided from calibration certificate (as for material mechanical properties), are estimated from a set of measurements on a uniform quartz sample.

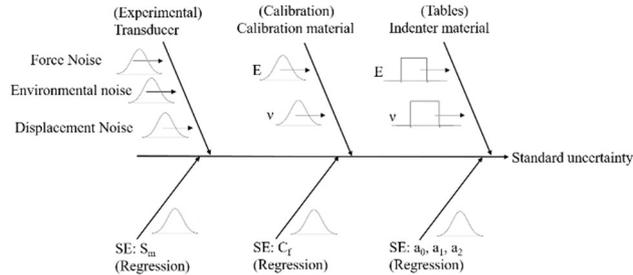


Figure 1: Influencing factors of standard uncertainty $u(x)$ of C_f and $A(h)$ parameters.

Specifically, one hundred Monte Carlo simulations were performed to estimate the standard uncertainty, $u(x)$, of each parameter, which resulted from the combination of the variance of the mean parameter estimates, \bar{x}_i , and the mean of the squared regression standard errors, $SE(\bar{x})_i$, associated to each simulation: $u(x) = \sqrt{\text{var}[\bar{x}] + \mathbb{E}[SE^2(\bar{x})]}$. The obtained results show that the variability sourced from the regression standard error is the dominant contribution to the standard uncertainty of the calibrated parameters. The approach provides a powerful tool to address a metrological comparison of performances of different calibration methods.

References

1. Barbato G., Genta G., Cagliero R., Galetto M., Klopstein M.J., Lucca D.A., Levi R.: Uncertainty evaluation of indentation modulus in the nano-range: contact stiffness contribution. CIRP Ann. 66, 495–498 (2017)
2. Galetto M., Maculotti G., Genta G. Barbato G., Levi R.: Instrumented Indentation Test in the Nano-range: Performances Comparison of Testing Machines Calibration Methods. Nanomanufacturing and Metrology, in press
3. ISO 14577-2:2015 Metallic materials-instrumented indentation test for hardness and materials parameters-part 2: verification and calibration of testing machines. ISO, Genève, Switzerland
4. JCGM 101:2008 Evaluation of measurement data-supplement 1 to the “Guide to the expression of uncertainty in measurements”—propagation of distributions using a Monte Carlo method. JCGM, Sèvres, France.

Modelling of correlation and estimation of coverage region in multivariate measurements with correlated or uncorrelated sources of uncertainty

Jacek Puchalski¹

Key words: type A and type B uncertainties, correlation coefficient, matrix law of propagation of variance, coverage region, Monte Carlo numerical method.

Abstract

In this paper, a Monte Carlo Method (MCM) [1,2] is applied for determining coverage regions for a multivariate output when this depends on two multivariate quantities, whose uncertainty is calculated according to type A procedure, the former, and type B, the latter, respectively. Different values of correlation are considered between components of the first and the second input quantity, but the two are taken as independent. Zero correlation could arise when measurements are performed by different instruments, for example, whereas using the same instrument could lead to a correlation close to 1.

The model

For a number n of measurements of the same quantities which are carried out by the same instrument on different points of the measurement range or by different measuring instruments, the measurands and corresponding uncertainties are given by:

$$y_i = x_{iA} + x_{iB} \quad u_{yi} = \sqrt{u_{iA}^2 + u_{iB}^2} \quad \text{and} \quad i = 1, \dots, n \quad (1a, b)$$

where x_{iA}, x_{iB} are the input quantities to which type A – u_{iA} and type B – u_{iB} uncertainties correspond, respectively. Values y_i and u_{yi} are the measurands estimates and associated uncertainties. Using a vector notation, the joint density distribution function of \mathbf{Y} is numerically simulated as the convolution of two statistically independent distributions of the input multivariate quantities \mathbf{X}_A and \mathbf{X}_B . We assume correlation between components of vector \mathbf{X}_A and vector \mathbf{X}_B , but not between \mathbf{X}_A and \mathbf{X}_B . For case $n = 2$, e. g., the output covariance matrix \mathbf{U}_Y associated with $\mathbf{Y} = [y_1, y_2]^T$ is obtained by the matrix equation of propagation of variances, that is, $\mathbf{U}_Y = \mathbf{S} \mathbf{U}_X \mathbf{S}^T$ [2], where

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad (2)$$

and

$$\mathbf{U}_X = \begin{bmatrix} u_{1A}^2 & \rho_A u_{1A} u_{2A} & 0 & 0 \\ \rho_A u_{1A} u_{2A} & u_{2A}^2 & 0 & 0 \\ 0 & 0 & u_{1B}^2 & \rho_B u_{1B} u_{2B} \\ 0 & 0 & \rho_B u_{1B} u_{2B} & u_{2B}^2 \end{bmatrix}, \quad (3)$$

where $\mathbf{X} = [x_{1A}, x_{2A}, x_{1B}, x_{2B}]^T$.

¹ Jacek Puchalski, Central Office of Measures (GUM), Warszawa Pl, email: jacek.puchalski@gum.gov.pl

Coverage regions

Coverage regions for Y have been simulated for different values of input correlation coefficient and uncertainties [3,4]. Corresponding coverage regions for a quantity $Z = G(Y)$, which is function of Y , have been estimated using MCM. As application examples, the model for transforming the resistances of star circuits and that for determining resistances by balancing a Wheatstone bridge have been studied [5]. Simulations proved that taking into account correlation can significantly change the output coverage region.

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections (2008)
2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, 2011 Supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement - Extension to any number of output quantities'. JCGM 102:2011
3. Dorozhovets M., Warsza Z. L., Propozycje rozszerzenia metod wyznaczania niepewności wyniku pomiarów wg Przewodnika GUM (2), Uściślenie metod obliczeń niepewności typu B, Pomiary Automatyka Robotyka (PAR) 2 2007, s. 6-12
4. J. Puchalski, P. Fotowicz, Coverage region for three - dimensional vector measurand calculated with the use of uncertainty propagation. Ppt: in CD Proceedings of conference: Problems and Progress of Metrology ppm'18 Szczyrk 04-06. June 2018,
5. Warsza, Z. L., Puchalski, J.: Estimation of vector uncertainties of multivariable indirect instrumental measurement systems on the star circuit example. XXII World Congress IMEKO 2018 Belfast. CD Proceedings PO-062 and IOP Conf. Series: Journal of Physics: Conf. Series 1065 (2018) 052026, doi:10.1088/1742-6596/1065/5/052026

Uncertainty propagation in quantitative imaging of electric properties using Monte Carlo method

Alessandro Arduino, Oriano Bottauscio, Mario Chiampi, Francesca Pennecchi and Luca Zilberti

Key words: Uncertainty propagation, Monte Carlo method, magnetic resonance imaging, electric properties tomography, contrast source inversion

1 Introduction

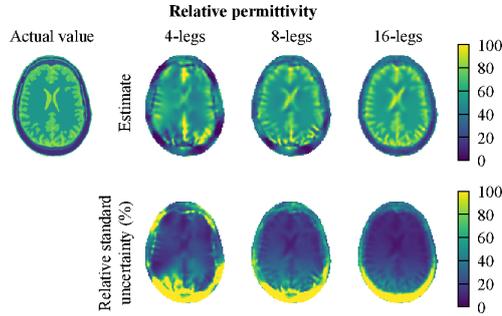
Quantitative imaging of electric properties (EPs) for biomedical applications is a topic of great interest, as proven by the fast development of magnetic resonance-based techniques for radiofrequency EPs tomography (MR-EPT) [4]. In order to move these techniques towards clinical applications, they need to be metrologically characterised. At present, only the original Helmholtz-based MR-EPT technique has been deeply studied [3], and just some preliminary analyses have been produced for more complicated techniques [1]. The same analysis carried out in [1] is here repeated for the method denoted as contrast source inversion global Maxwell tomography (CSI-GMT) [2], whose peculiarity is the elaboration, by means of optimisation theory, of the magnitude of the transmit sensitivities, $|B_1^+|$, only.

2 Materials and methods

The CSI-GMT technique is implemented as described in [2] and applied to a 2D model problem: a section of a human head illuminated by a TEM coil with as many channels as legs. In first approximation, all the uncertainty sources are attributed to

Alessandro Arduino, Oriano Bottauscio, Mario Chiampi, Francesca Pennecchi and Luca Zilberti
Istituto Nazionale di Ricerca Metrologica (INRiM), Torino, ITALY
e-mail: a.arduino@inrim.it, o.bottauscio@inrim.it, m.chiampi@inrim.it, f.pennecchi@inrim.it, l.zilberti@inrim.it

Fig. 1 Estimates and relative standard uncertainties of the relative permittivity distribution computed by the Monte Carlo method. The input transmit sensitivity magnitude is corrupted by spatially uncorrelated random error with SNR 20.



$|B_1^+|$, whilst the incident fields generated in vacuum are assumed perfectly known—a complete uncertainty budget should consider the errors related to them, too.

The Monte Carlo method is applied with $M = 1000$ extractions for each choice of input uncertainty u and number of legs. The exact distribution of $|B_1^+|$ is obtained by simulation, whereas the random error is introduced by extracting, independently for each pixel, $n \sim \mathcal{N}(0, u^2)$ and defining $|B_1^+|_{\text{meas}} = ||B_1^+| + n|$. The external absolute value operation takes into account naively the practical estimation of $|B_1^+|_{\text{meas}}$, proportional to the so-called flip angle, with application of the arccosine function. It is assumed that u is proportional to the spatially averaged sensitivity, $\text{mean}(|B_1^+|)$.

3 Results and discussion

Fig. 1 shows the results obtained when the input uncertainty is $u = 0.05 \text{ mean}(|B_1^+|)$, corresponding to a signal-to-noise ratio (SNR) of 20, and the random error is evaluated independently for each leg of the TEM coil driven at 128 MHz. As expected, better and more accurate estimates are obtained when the number of transmit channels is higher, since the random errors are suppressed by averaging multiple inputs. It is worth noting that higher uncertainties are present in the posterior region. This happens because the head is not centred in the coil and a lower electric field is generated in the nape [1]. The adoption of a weighted cost could overcome this issue.

References

1. Arduino A, *et al* (2017) Monte carlo method for uncertainty propagation in magnetic resonance-based electric properties tomography. *IEEE Trans Magn* 53:2220–2232
2. Arduino A, *et al* (2018) Magnetic resonance-based imaging of human electric properties with phaseless contrast source inversion. *Inverse Problems* 34:084,002
3. Lee SK, *et al* (2015) Theoretical investigation of random noise-limited signal-to-noise ratio in MR-based electrical properties tomography. *IEEE Trans Med Imag* 34:5100,304
4. Liu J, *et al* (2017) Electrical properties tomography based on B1 maps in MRI: principles, applications, and challenges. *IEEE Trans Biomed Eng* 64:2515–2530

The Role of the Measurement Uncertainty in the Decision Making Process of Conformity Assessment

Alen Bosnjakovic, Vedran Karahodzic, Ehlimana Jugo and Zijad Dzemic

Key words: conformity assessment, decision making, measurement uncertainty, distribution function, error function, software, ISO/ IEC 17025

1 Introduction

Measurement uncertainty plays a very important role in most industries related to the conformity assessment of products. Therefore, the measurement uncertainty has an impact on the decision making process of the conformity assessment of the product. Without adequate estimates of measurement uncertainty or its neglecting in the process of making a decision on conformity assessment, the cost of production increases, and production mistakes are guaranteed. Therefore, measurement uncertainty is an indispensable factor for proper making of decisions and risk assessment in the conformity assessment procedure [1],[2],[3].

Most of the industries make a decision on product conformity on the basis of a report that contains measurement results issued by certain laboratory. If, for some reason, the measurement results are not appropriate, the risk of the wrong decision is increased. Also, the selection of the unincorporated laboratories, or laboratories without proven competencies, also increases the risk of the wrong decision. Therefore EN ISO / IEC 17025 [4] standard sets requirements for the qualification of test and calibration laboratories. The first edition of the mentioned standard dates from 1999, and from this period a large number of accredited laboratories, as well as laboratories of the national metrology institutes, base their quality management system

Alen Bosnjakovic, IMBiH, Augusta Brauna 2, e-mail: alen.bosnjakovic@met.gov.ba
Vedran Karahodzic, IMBiH, Augusta Brauna 2, e-mail: vedran.karahodzic@met.gov.ba
Ehlimana Jugo, IMBiH, Augusta Brauna 2, e-mail: ehlimana.jugo@met.gov.ba
Zijad Dzemic, IMBiH, Augusta Brauna 2, e-mail: zijad.dzemic@met.gov.ba

precisely on this standard.

The new edition of the standard, besides the structural changes, also introduces new requirements for the training of the laboratories. Major changes in the new standard edition include risk-based thinking, flexibility in process-related requirements, documented information and organizational abilities, and more. This work does not intend to give instructions for interpreting certain standard points, nor how to implement them in the laboratory. Taking into account the importance of the documented decision making rules, this work aims to demonstrate, through practice examples and through the application of mathematical models, that it facilitates implementation of decision making requirements for the laboratory. The rule for making the decisions brings a new requirement that appears in the new edition of the standard, and it occurs when there is the need for laboratories to evaluate the conformity of results with pre-set criteria taking into account the measurement uncertainty, "...the laboratory shall document the decision rule employed, taking into account the level of risk associated with the decision rule and apply the decision rule" [4].

Throughout the paper, models related with the process of documented decisions, decision-making rules, and conformity assessment based on two scenarios (with and without the use of guard bands) will be presented [5]. All these analysis will be accompanied by the mathematical model which will facilitate their application in practice. Conformity assessment will be determined by using knowledge of distribution functions (for calculating probabilities), and displayed by using the error functions (Q and erfc function) whose values are taken from the tables. The results mentioned above will be confirmed using appropriate software tools. The aim of this paper is to make the code for conformity assessment publicly available to reduce the risk of making wrong decisions in practice. It is planned to implement code in MATLAB/ OCTAVE, using functions and m-files. Output of every code execution will be appropriate probability of acceptance what can serve as acceptance or rejection criteria. In this way, the laboratories that plan the transition from an old edition to the new one, or the laboratories that have already made that transition, will have open access to the program code. Laboratories will have the ability to modify the code for their needs later, or if necessary, to compare them with their existing solutions. Also, the impact of measurement uncertainty on the acceptance or rejection of product conformity with product specification will be analyzed in the work. Practical example on which results of code will be tested is regarding pressure gauges [6].

References

1. Alistair B Forbes. Measurement uncertainty and optimized conformance assessment. *Measurement*, 39(9):808–814, 2006.
2. LR Pendrill. Using measurement uncertainty in decision-making and conformity assessment. *Metrologia*, 51(4):S206, 2014.
3. Francesca R Pennecchi, Ilya Kuselman, Ricardo JNB da Silva, and D Brynn Hibbert. Risk of a false decision on conformity of an environmental compartment due to measurement uncertainty of concentrations of two or more pollutants. *Chemosphere*, 202:165–176, 2018.
4. General requirements for the competence of testing and calibration laboratories - ISO/IEC 17025. 2017.
5. JCGM 106:2012. Evaluation of measurement data – the role of measurement uncertainty in conformity assessment. 2012.
6. Guidelines on the calibration of electromechanical and mechanical manometers - TC-M - version 3.0. 2017.

Conformity assessment of multicomponent materials or objects using compositional data

Francesca Pennechi¹, Aglaia Di Rocco² and Ilya Kuselman³

Key words: conformity assessment, compositional data, Bayesian framework, Monte Carlo method

1 Conformity assessment of multicomponent materials or objects

The JCGM 106:2012 guidance explains the role of measurement uncertainty in conformity assessment (CA) of a single entity, object or system with respect to specified requirements, using the Bayesian framework. According to this framework, an item conforms to a specified requirement if the true value of its property lies in the tolerance interval, where the knowledge about such true property values (prior) and the knowledge about the measurement results (likelihood) are conveyed to a relevant (posterior) probability density function (PDF).

In the recent IUPAC Project 2016-007-1-500, a more general Bayesian approach was elaborated for evaluation of risks of false decisions in CA of multicomponent materials or objects, taking also into account possible correlations between the measured property values of the item components. To recover the conformance probability of the item, this ‘conventional’ approach applies integration of the relevant posterior multivariate PDF on the tolerance multi-domain.

Compositional data (CD) [1,3], such as mass fractions, molar fractions or any other positive quantity ratios summing up to a certain value (typically 1 or 100 %), require special attention due to the “spurious” correlations intrinsically arising between components’ contents because of the summation constraint. A well-known problem in the CD analysis is bumping into singular covariance matrices. In this work, we studied how to deal with CD in conformity assessment.

2 Monte Carlo method for evaluation of conformance probability

Let us consider the simplest CD model: $Y = 100 - X$, where the two variables X and Y are defined on the [0 %, 100 %] interval, for example. The proposed approach consists in resorting to a Monte Carlo (MC) method taking into account the compositional constraint. It means to randomly draw from the univariate PDF of X and then reconstructing Y values as complementary (up to 100 %, in the considered case) to the X ones. The frequency of (X, Y) values generated within the tolerance

¹ Francesca Pennechi, INRIM, Torino, Italy, f.pennechi@inrim.it

² Aglaia Di Rocco, Università di Torino, Torino, Italy, aglaia.dirocco@edu.unito.it

³ Ilya Kuselman, Independent Consultant, Modiin, Israel, ilya.kuselman@gmail.com

multi-domain is the conformance probability of the multicomponent item. For example, at a truncated normal distribution on [0 %, 100 %] for X , with parameters $\mu_X = 30$ and $\sigma_X = 10$, the conformance probability of (X, Y) on the domain [20, 40] x [60, 80] is 68.37 % (at 10^7 MC generations). For comparison, using correlation coefficient equal to 0.9999 and resorting to classical integration of a bivariate truncated normal PDF, leads to a conformance probability of 68.18 %, the difference being due to the impossibility of using correlation 1 in the conventional approach.

The proposed approach was applied to a subset of test results of a PtRh alloy [2], where contents (c , %) of Pt, Rh (mass fractions) and of the eight impurities (sum of mass fractions), are constrained by the following balance: $c_{Pt} = 100 \% - c_{Rh} - c_{imp}$. The conformance probabilities were calculated for 1) the simulated case of zero correlations, 2) the simulated case when only the correlations due to the compositional constraint are applied, and 3) the observed (experimental) correlation coefficients are taken into account, induced as by the compositional constraint, as by the nature of the raw materials of the alloy and technological factors. The calculated conformance probabilities in the required tolerance multi-domain of the alloy compositions were practically the same as obtained by the conventional approach.

Conclusions

- The conventional approach to the evaluation of conformance probability in CA of multicomponent materials or objects is able to automatically take into account the compositional constraint by means of the corresponding covariance matrix, still positive definite.
- A Monte Carlo method including the compositional constraint is proposed as a possible alternative, especially when degenerate PDFs arise. It will be also applied to the evaluation of risks of false decisions.

References

1. Aitchison, J.: The statistical analysis of compositional data. Chapman and Hall, London (1986)
2. Kuselman, I., Pennechi, F.R., da Silva, R.J.N.B., Hibbert, D.B., Anchutina, E.: Total risk of false decision on conformity of an alloy due to measurement uncertainty and correlation of test results. *Talanta* **189**, 666-674 (2018)
3. Pawlowsky-Glahn, V., Buccianti, A.: Compositional data analysis: theory and applications. Wiley, Chichester (2011)

Assessment of inspection strategies in assembly manufacturing processes

Elisa Verna¹, Gianfranco Genta², Fiorenzo Franceschini³ and Maurizio Galetto⁴

Key words: quality control, inspection effectiveness, inspection cost

Extended abstract

In general, assembly manufacturing processes may be decomposed into a number (m) of workstations, i.e., process steps, each one potentially critical in generating defects [2,4]. In each workstation, different quality controls may be performed. Each i -th workstation can be described through three parameters [4]: (i) p_i : probability of occurrence of a defective-workstation-output; (ii) α_i : probability of erroneously signalling a defective-workstation-output after the control (i.e., type-I inspection error); (iii) β_i : probability of erroneously not signalling a defective-workstation-output after the control (i.e., type-II inspection error), where $i=1,\dots,m$. The first parameter (p_i) is a physiological characteristic of the process in normal working conditions, and can be *a priori* estimated using defect-generation models or empirical/simulation methods [3,5]. The estimates of α_i and β_i depend on the characteristics of the inspection procedure and the technical skills and/or experience of the inspector [1]. According to this model, two indicators which depict the overall effectiveness and economic convenience of an inspection strategy may be obtained

¹ Elisa Verna
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy,
e-mail: elisa.verna@polito.it

² Gianfranco Genta
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy,
e-mail: gianfranco.genta@polito.it

³ Fiorenzo Franceschini
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy,
e-mail: fiorenzo.franceschini@polito.it

⁴ Maurizio Galetto
DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy,
e-mail: maurizio.galetto@polito.it

[2]. The first one, D , is the mean total number of defective-workstation-outputs which are erroneously not signaled in all the inspections, defined as [4]:

$$D = \sum_{i=1}^m p_i \cdot \beta_i \quad (1)$$

The second one is the total inspection cost, C_{tot} , which can be expressed as [2,3]:

$$C_{tot} = \sum_{i=1}^m [c_i + NRC_i \cdot p_i \cdot (1 - \beta_i) + URC_i \cdot (1 - p_i) \cdot \alpha_i + NDC_i \cdot p_i \cdot \beta_i] \quad (2)$$

where c_i represents the cost of the inspection activity; NRC_i and URC_i are respectively the necessary- and unnecessary-repair cost; and NDC_i is the cost of undetected defects.

The proposed model and the two inspection indicators, reported in Eqs. (1) and (2), are supposed to have both an analytical and predictive connotation. According to a cost-benefit logic, the combined use of the inspection indicators allows the comparison of alternative inspection strategies in terms of effectiveness and cost, and the selection of the most appropriate according to the manufacturer requirements. This may represent a powerful and practical approach to assist inspection designers in early design phases of new assembly manufacturing processes. For instance, it may be adopted to choose between a complete inspection, i.e., when quality controls are performed after each workstation, and a partial inspection in selected workstations. Once the optimal strategy has been identified through the proposed method, it will be used each time an inspection activity is carried out. Since the proposed indicators enable the identification of the most critical workstations, more effective control procedures may be designed. Furthermore, as the assessment of inspection strategies is closely related to inspection errors α_i and β_i , a study on the impact that inspection errors have on inspection strategies selection is proposed.

The proposed approach may be exploited for a wide range of industrial processes, and is particularly useful in the case of short-run productions, for which most of the statistical process control techniques are unsuitable. In this work, a case study concerning the comparison of three alternative inspection procedures for a production of hardness testing machines is presented.

References

1. Duffuaa, S.O., Khan, M.: Impact of Inspection Errors on the Performance Measures of a General Repeat Inspection Plan. *Int. J. Prod. Res.*, 43(23), 4945-4967 (2005)
2. Franceschini F., Galetto M., Genta G., Maisano D.A.: Selection of quality-inspection procedures for short-run productions. *Int. J. Adv. Manuf. Technol.*, 99(9-12), 2537-2547 (2018)
3. Galetto M., Verna E., Genta G., Franceschini F.: Robustness analysis of inspection design parameters for assembly of short-run manufacturing processes. In: *Proceedings book of the 3rd International Conference on Quality Engineering and Management*, pp. 255-274 (2018)
4. Genta G., Galetto M., Franceschini F.: Product complexity and design of inspection strategies for assembly manufacturing processes. *Int. J. Prod. Res.*, 1-11 (2018)
5. Krugh, M., Antani, K., Mears, L., Schulte, J.: Prediction of Defect Propensity for the Manual Assembly of Automotive Electrical Connectors. *Procedia Manuf.*, 5, 144-157 (2016)

An attempt to introduce the use of covariances to evaluate the uncertainty of the detection efficiency in k_0 -NAA

Di Luzio M and D'Agostino G and Oddone M

Key words: k_0 -NAA, detection efficiency, uncertainty evaluation

Neutron Activation Analysis (NAA) is an elemental technique based on activation of target isotopes by neutron capture reactions and counting of the emitted gamma photons [Alfassi(1990)]. The k_0 -standardisation protocol makes straightforward the application of NAA as a multi-elemental technique by using a single element (monitor) in a standard to quantify target elements (analytes) in samples [De Corte(1987)]. The knowledge of curves modeling the efficiency versus the gamma energy at the counting distance of samples, $\varepsilon_d(E)$, is mandatory. Specifically, the ratio

$$k_E = \frac{\varepsilon_{d_{\text{std}}}(E_m)}{\varepsilon_{d_{\text{sm}}}(E_a)}, \quad (1)$$

is required, where d_{std} and d_{sm} are the counting distances of standard and sample, E_m and E_a are the gamma photon energies selected for the monitor and analyte, respectively. A common used fitting function is

$$\ln[\varepsilon_d(E)] = \sum_{i=1}^6 a_i E^{2-i}, \quad (2)$$

where a_i are constant parameters depending on d [Gilmore(2008)].

Gamma reference sources are counted at d_{std} and d_{sm} to measure $\varepsilon_{d_{\text{std}}}(E_{\text{ref}})$ and $\varepsilon_{d_{\text{sm}}}(E_{\text{ref}})$ at single reference energies, E_{ref} . The natural logarithm in eq. (2) allows to fit the data with a polynomial regression if errors of $\varepsilon_d(E)$ are suitably small.

Di Luzio M
Istituto Nazionale di Ricerca Metrologica (INRIM) e-mail: m.diluzio@inrim.it

D'Agostino G
Istituto Nazionale di Ricerca Metrologica (INRIM) e-mail: g.dagostino@inrim.it

Oddone M
University of Pavia, department of chemistry e-mail: oddone@unipv.it

The most advanced method currently available to obtain k_ϵ is based on a numerical computation starting from dimensions of detector crystal and data collected with reference sources [De Corte(1987)]. Although this method is suitable also for counting emissions of extended samples close to the detector, it omits possible correlations among fitting parameters.

The aim of this study is to develop an alternative method that simplifies the characterization of the detection system and takes into account correlations to correctly evaluate the uncertainty.

The proposed approach considers k_ϵ as the product $k_{\epsilon\Delta E} k_{\epsilon\Delta d}$, where

$$k_{\epsilon\Delta E} = \frac{\epsilon_{d_{\text{std}}}(E_m)}{\epsilon_{d_{\text{std}}}(E_a)} = e^{\left[\sum_{i=1}^6 a_i (E_m - E_a)^{2-i}\right]}, \quad (3)$$

and

$$k_{\epsilon\Delta d} = \frac{\epsilon_{d_{\text{std}}}(E_a)}{\epsilon_{d_{\text{sm}}}(E_a)} = e^{\left[b_1 E_a + b_2 + b_3 E_a^{-2} + b_4 E_a^{-3} + b_5 E_a^{-4}\right]}, \quad (4)$$

In eq. (3), the parameters a_i are obtained by fitting the model $\sum_{i=1}^6 a_i E^{2-i}$ to the $\ln[\epsilon_{d_{\text{std}}}(E_{\text{ref}})]$ values, whereas in eq. (4) the parameters b_i are obtained by fitting the model $b_1 E + b_2 + b_3 E^{-2} + b_4 E^{-3} + b_5 E^{-4}$ to the $\ln\left[\frac{\epsilon_{d_{\text{std}}}(E_{\text{ref}})}{\epsilon_{d_{\text{sm}}}(E_{\text{ref}})}\right]$ values.

We collected experimental data by counting a multi-gamma reference disk source with a 35% relative efficiency Ge detector. As a preliminary test, the gamma source was located at $d_{\text{std}} = 18$ cm and d_{sm} between 2 cm and 16 cm.

The equation models were fitted in a linear regression analysis to obtain the a_i and b_i with covariance matrices.

Since the $\frac{\epsilon_{d_{\text{std}}}(E_{\text{ref}})}{\epsilon_{d_{\text{sm}}}(E_{\text{ref}})}$ values are independent on the gamma emission rate, the resulting correlation between a_i and b_i can be reasonably neglected, thus, the uncertainty evaluation of k_ϵ requires only the knowledge of a_i , b_i and their covariance matrices.

As expected, the k_ϵ uncertainty converges to zero when E_a gets close to E_m if $d_{\text{std}} = d_{\text{sm}}$ (i.e. $k_{\epsilon\Delta d} = 1$). In this case, only the a_i correlation matrix is required to correctly evaluate the uncertainty, as reported in a recent paper focused on k_0 -NAA uncertainty [D'Agostino(2018)].

The method proposed in this study is suitable in the quite common case when samples are counted at different detector distances. In fact, the k_ϵ uncertainty combines the effects due to variability in energies and distances. However, its application becomes cumbersome in the case of extended samples counted close to the detector end-cap.

References

- [Alfassi(1990)] Alfassi (1990) Activation Analysis, vol 2. CRC Press
 [D'Agostino(2018)] D'Agostino (2018) J Radioanal Nucl Chem 318
 [De Corte(1987)] De Corte (1987) The k0-standardization thesis. Rijksuniversiteit Gent
 [Gilmore(2008)] Gilmore (2008) Practical Gamma-ray Spectrometry. Wiley

Measurement Uncertainty Evaluation and Correlation Analysis of Gas Flow Measurements in a System of Three Parallel Pipes

Zijad Dzemic, Vedran Karahodzic and Alen Bosnjakovic

Key words: measurement uncertainty, flow meter, measurements, correlations, calibration

1 Introduction

The objective of the present work is to test the validity of the hypothesis of equal flow for three flow meters in parallel, and to expand the mathematical model for two flow meters into three. The theoretical starting uncertainty model is developed under the assumption that the flow rate is the same in three parallel connected pipes. Within this work, a new calibration model of the flow meters 1:1 and the calibration model 3:1 are developed, and the absolute and relative measurement uncertainty for three flow meters are analyzed.

Following the research done on calibration installation described in this work, an expensive and comprehensive experiment with 135 gas flow measurements through three parallel pipelines has been carried out. Unsolved traceability problems related to flow conditions are studied in this work. Through the experiment it was confirmed that the assumption of the same flow through three flow meters in parallel has measurement standard deviation of 20 %, which is explained in details through the work.

The quantity of fluid measured in the field by the flow meter, which is calibrated in laboratory, is an issue if the measuring quantity is important and/or expensive [1, 2, 3]. A main goal, in that case, is to reach reliable measurements and to apply appropriate verification and calibration techniques to meet the requirements estab-

Zijad Dzemic, IMBiH, Augusta Brauna 2, e-mail: zijad.dzemic@met.gov.ba
Vedran Karahodzic, IMBiH, Augusta Brauna 2, e-mail: vedran.karahodzic@met.gov.ba
Alen Bosnjakovic, IMBiH, Augusta Brauna 2, e-mail: alen.bosnjakovic@met.gov.ba

lished by normative documents. As an upgrade to the procedures described in standards [4, 5, 6], which apply to work in the laboratory and assumes 0 % or 100 % correlation between measurements of flow meters running in parallel, a new model for evaluation of the measurement uncertainty for the three flow meters in parallel is developed, based on the presumptions given in these standards. It is expected that the analysis and the obtained results will provide improved assumptions and measurement results, which will lead to more precise premises for the relations of the flow for (three or even more) flow meters in parallel [7]. Additionally, this analysis could contribute to the improvement of the existing normative documents in this field.

References

1. The European Parliament and Council Directive 2012/27/EU of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC.
2. The European Parliament and Council Directive 2006/32/EC of 5 April 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EC.
3. Gas meters - turbine gas meters (EN 12261:2002/A1:2006, DT). 2002, 2006.
4. Collaborative study- Accuracy (trueness and precision, parts 2,3, ISO 5725:1994). 1994.
5. Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation (ISO 21748:2018). 2018.
6. Measurement of fluid flow - Procedures for the evaluation of uncertainties. 2005.
7. Zijad Dzemic, Brane Sirok, and Janko Drnovšek. Traceability of gas flow measurements in complex distribution systems - uncertainty approach vs. error approach. *Metrology and Measurement Systems*, 2/2019, 2019.

A statistical method for the evaluation of near-surface air measurement uncertainty due to nearby roads

Coppa, Graziano¹, Quarello, Annarosa², and Merlone, Andrea³

Key words: Generalized Additive Models, uncertainty, near-surface air temperature, meteorology, climatology

1 Introduction

The accuracy of near-surface measurements of meteorological variables is influenced by the environmental characteristics of the site where the instruments are placed ([1], [2]). Measurement results and recorded data can be subject to errors and uncertainty if the influence of the measuring site is not fully understood and evaluated. WMO guide #8[3] establishes a qualitative/quantitative classification, by itemizing different site conditions, in terms of obstacles proximity, ground slopes, projected shades etc.

2 Data collection and statistical analysis

In the framework of the EMPIR project MeteoMet 2, in order to deliver a validated analysis aiming at possibly improving the WMO siting classification, a one-year lasting data collection scheme has been devised for evaluating the effect of roads on near-surface air temperature measurements. It consists in a 100-m-long array of identical INRiM-calibrated thermometers equipped with aspirated solar shields, placed on a flat grass field at increasing distances from a road, such that the farthest station fulfils current WMO requirements for a Class 1 site. Quantities of influence are also measured: humidity, solar radiation, wind speed and direction. This data set is too rich to be fully modelled with a parametric model; furthermore, a linear model

¹ Coppa, Graziano
Istituto Nazionale di Ricerca Metrologica (INRiM), strada delle Cacce 91, Torino, e-mail:
g.coppa@inrim.it

² Quarello, Annarosa
IPGP, IGN, ENSG, Université Paris Diderot, Sorbonne Paris Cité, UME 7154 CNRS, Paris, France, e-mail : annarosa.quarello@edu.unito.it

³ Merlone, Andrea
Istituto Nazionale di Ricerca Metrologica (INRiM), strada delle Cacce 91, Torino, e-mail:
a.merlone@inrim.it

is too simple to accurately describe the data. Generalized Additive Models (GAMs) allow the extension of the methods, dropping out the linearity assumption, to deal with multiple predictors, for which we used smoothing splines. This analysis has been used to understand the effect of each quantity of influence on the temperature measurements (Figure 1). Since the analysis was focused on the identification of the largest possible temperature bias due to the influence of the obstacle, the model was instrumental in understanding the best combination of environmental factor that would boost the effect.

3 Results and further work

Results show that the roads influence temperature readings more during nights, when solar radiation stops feeding the road heat reservoir, and when wind is absent or very weak. Other relations (especially with wind direction) are more complicated.

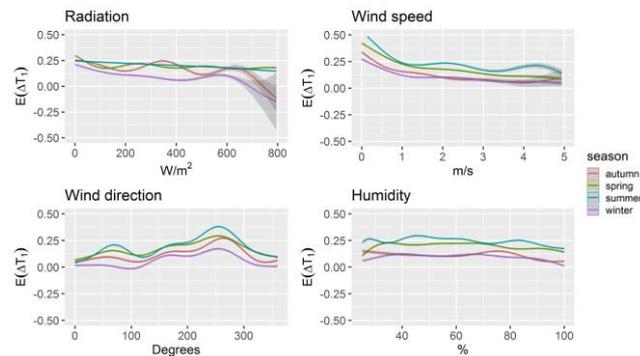


Figure 1: Results of the GAM analysis as applied to the temperature difference between the air nearby the road and the reference sensor at 100 m.

Next steps are to extend the GAM to temperature differences between other measurement points, in order to understand better the dynamics among all the environmental parameters, and introduce other influencing parameters, like temperature itself and its derivative over time. The modelling features could be further investigated by considering complex statistical models.

4 References

- [1] Mahmood, R., Foster, S. A. and Logan D., 2006: The GeoProfile metadata, exposure of instruments, and measurement bias in climatic record revisited. *Int. J. Climatol.*, 26, 1091–1124.
- [2] Pielke, R. A., C. A. Davey, D. Niyogi, S. Fall, J. Steinweg-Woods, K. Hubbard, X. Lin, M. Cai, Y.-K. Lim, H. Li, J. Nielsen-Gammon, K. Gallo, R. Hale, R. Mahmood, S. Foster, R. T. McNider, and P. Blanken, 2007: Unresolved issues with the assessment of multidecadal global land surface temperature trends. *J. Geophys. Res.*, 112, D24S08, doi:10.1029/2006JD008229.
- [3] WMO, https://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf

On the detection of anomalous values of Radioxenon in IMS stations of CTBTO

M. Scagliarini¹, S. Guernelli², G. Ottaviano³, A. Rizzo⁴ and F. Padoani⁵

Key words: abnormal concentration, control charts, interquartile filter, radioactivity

1 Background and some Results

The “Comprehensive Nuclear-Test-Ban Treaty Organization” (CTBTO) supports and promotes the “Entry Into Force” (EIF) of the “Comprehensive Nuclear Test-Ban Treaty”, a treaty that outlaws nuclear test explosions. The core of the CTBTO is a verification regime based on three pillars, the International Monitoring System (IMS), the International Data Centre (IDC) and the On-Site Inspections (OSI). The IMS, when completed, will consist of 337 stations worldwide to monitor the planet for signs of nuclear explosions. Around 90 percent of the facilities are already up and running. The activity concentrations of ^{133}Xe (radioxenon), a fission product of the nuclear reactions, are considered particularly relevant by CTBTO since allow to detect possible underground nuclear tests. This signature, however, can be difficult to recognise due to the anthropogenic releases of radioxenon into the atmosphere from nuclear power plants (NPPs) and from medical isotope production facilities (MIPFs). Making conclusions about a suspected underground nuclear explosion or a release into atmosphere of any kind, considering measurements of radioxenon, may

¹ Michele Scagliarini
Department of Statistical Sciences, University of Bologna, e-mail:
michele.scagliarini@unibo.it

² Sofia Guernelli
Department of Statistical Sciences, University of Bologna, e-mail:
sofia.guernelli@studio.unibo.it

³ Giuseppe Ottaviano
ENEA, Bologna Research Centre, e-mail: giuseppe.ottaviano@enea.it

⁴ Antonietta Rizzo
ENEA, Bologna Research Centre, e-mail: antonietta.rizzo@enea.it

⁵ Franca Padoani
ENEA, Bologna Research Centre, e-mail: franca.padoani@enea.it

prove to be a very complex task due to several factors [1]. One of the objectives of the IDC of the CTBTO is to analyse the radioxenon releases and therefore a classification system has been defined in which a measured activity concentration is classified as anomalous if it is above the “abnormal concentration” value. The abnormal concentration value, or “abnormal limit”, used by the IDC, is defined as: $L_A = Q_2 + \lambda_A [Q_3 - Q_1]$ where Q_1 is the first quartile, Q_2 is the median, Q_3 the third quartile and $\lambda_A=3$. The radioxenon concentrations are measured daily and the L_A value is constantly updated after each observation since the last 365 observations are used for computing the quartiles. Summarising, a concentration is classified as anomalous if it is above the “typical background” of the station established by applying an inter-quartile filter, updated after each observation, to the data.

In this work we aim to compare the IDC method for detecting anomalous values with the results obtained by using an approach based on statistical process control. We used the Shewhart and EWMA control charts [2] for individual measurements for discriminating between natural background and anomalous measurements on selected six IMS stations.: AUX04 Melbourne-AUS; DEX33 Freiburg-D; SEX63 Stockholm-S; NZX46 Chatham Island- NZ; USX74 Ashland-USA; JPX38 Takasaki-J. For each station we performed a Phase I, using the Shewhart control charts, in order to establish the in-control state, or the natural background, for the radioxenon observations. In the subsequent monitoring phase (Phase II) we used both Shewhart and EWMA control charts for detecting concentrations values above the upper control limits. The Shewhart and EWMA charts were set up in such a way to have the same in-control performance, more precisely with an in-control average run length of about 370. The obtained results were compared with the inter-quartile filter used by IDC. As expected we obtained different results. As an example, for the JPX38 station (Japan) over the period 22/02/2015-25/07/2018, the percentages of anomalous concentrations were 5.68%, 2.12% and 20.82% for the inter-quartile filter, Shewhart and EWMA-charts respectively. Currently we are deepening our study to further substantiate the results in order to explore the possibility to integrate the three different approaches. This work is the result of a collaboration between the Department of Statistical Sciences, University of Bologna and the National Data Centre - Radionuclides (NDC-RN) of the ENEA, Bologna Research Centre. The data used in this work were obtained within the framework of a two years’ contract called “virtual Data Exploitation Centre” (vDEC) signed with the Preparatory Commission of the CTBTO. The views expressed herein are those of the authors and not necessarily reflect the views of the CTBTO Preparatory Commission.

References

1. Kalinowski, M, Axelsson, A., Bean, M., Blanchard, X, Bowyer, T., Brachet, G., Hebel, S. McIntyre, J., Peters, J., Pistner, C., Raith, M., Ringbom, A., Saey, P., Schlosser, C., Stocki, T., Taffary, T., and Ungiar, K.: Discrimination of Nuclear Explosions against Civilian Sources Based on Atmospheric Xenon Isotopic Activity Ratios. *Pure Appl. Geophys.* **167**, 517-539, (2010).
2. Montgomery, D.C: *Introduction to Statistical Quality Control*, 6th ed. Hoboken, John Wiley & Sons (2009).

The evaluation of chronic alcohol abuse biomarkers in hair samples: the interpretation of cut-off values

Eugenio Alladio^{1,2}, Giulia Biosa^{1,2}, Alberto Salomone^{1,2}, Tereza Neocleous³, Marco Vincenti^{1,2}

Key words: ethyl glucuronide (EtG); fatty acid ethyl esters (FAEEs); cut-off values; hair samples; chronic alcohol drinkers

1 Introduction

The quantitative determination of chronic alcohol abuse biomarkers is consistently used throughout the world to assess chronic excessive alcohol consumption. Aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyltransferase (GGT), mean corpuscular volume (MCV) and carbohydrate-deficient transferrin (CDT) are the indirect biomarkers most commonly used, whereas direct ethanol metabolites, including ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEE), are highly specific and sensitive, in particular when they are quantified in the keratin matrix, typically scalp hair. The conjoined determination of direct biomarkers currently represents the most accredited strategy for proving chronic alcohol abuse, but a debate is still ongoing, in the scientific community, on the feasible interpretation approaches that can be used to recognize chronic alcohol

^{1,2} Eugenio Alladio, PhD, Giulia Biosa, MSc, Alberto Salomone, Prof. Marco Vincenti, Prof.

Dipartimento di Chimica, Università degli Studi di Torino, via P. Giuria 7, Torino, Italy - Centro Regionale "A. Bertinaria", Regione Gonzole, 10/1, Orbassano (Torino), Italy - email: eugenio.alladio@unito.it; giulia.biosa@edu.unito.it; alberto.salomone@unito.it; marco.vincenti@unito.it

^{1,2} Tereza Neocleous, PhD

School of Mathematics and Statistics, University of Glasgow, University Place, Glasgow G12 8QQ, United Kingdom - email: tereza.neocleous@glasgow.ac.uk

2] Eugenio Alladio, Giulia Biosa, Alberto Salomone, Tereza Neocleous, Marco Vincenti drinkers. Current practice involves comparing the analytical results with cut-off values recognized by regulatory authorities and scientific societies, for both administrative and legal purposes. However, the involvement of cut-off values might result in an arbitrary interpretation by the experts, especially when the measured EtG and FAEEs values are close to the cut-off values and/or contradictory. Furthermore, their detected values may depend on a variety of factors, partly unpredictable, such as the use of cosmetic treatments and seasonality.

2 Materials, methods and aim of the study

Several Multivariate Data Analysis (MDA) techniques and Likelihood Ratio (LR) approaches combining the results from direct and indirect biomarkers, as well as the evaluation of external variability factors, have been performed and tested. MDA approaches (e.g. Logistic Regression, Linear Discriminant Analysis, etc.) and LR models (i.e. uni- and multivariate) were developed on data from clinical and toxicological analyses used as training (reference) set. All MDA and LR models were validated by employing repeated double-cross validation strategies (typically involving venetian blind design and 5 cancellation groups). Furthermore, a R Shiny application have been developed in the open-source R Studio environment [1,2]. Analyses were commissioned by Local Committees for Driving Licences and Alcohol Abuse Treatment Services located in Piedmont (northern Italy), and performed at the Regional Antidoping and Toxicology Center “A. Bertinaria” (Orbassano, Italy). The main goal of the study was to decrease the number of misleading conclusions caused by biased data interpretation procedures. Furthermore, a discussion on the correct/misleading way of establishing and interpreting cut-off values will be performed. Moreover, the development of a R Shiny application might represent a useful and open-source tool to perform a robust and not arbitrary interpretation of chronic alcohol abuse biomarkers. Furthermore, it allows analysts and practitioners to independently test the several MDA and LR models that are implemented in the application on their own data. At the current stage, our studies represent a consistent proof-of-concept for the development of robust and significant interpretation protocols of chronic alcohol abuse biomarkers in hair samples. Nevertheless, our approaches will be necessarily tested on larger number of samples and caseworks.

References

1. RStudio, Inc. Easy web applications in R, 2013. <http://www.rstudio.com/shiny/>
2. RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

Assessment and management of occupational exposure to airborne dust in a Quality approach

Rebecca Nebbia¹

Key words: occupational safety quantification, workplaces, measurement quality evaluation, airborne dust

Foreword

Work related exposure to airborne dust -a common pollutant in many parts of the world [7]- is still a critical problem, and the methods for its quantification are often subjective. In fact, a possible cause of this criticality can be identified in determining the real conditions of pollution in the workplaces, attributable to the delicacy of the various phases of acquisition of the pollution measures. This leads in many situations to poor risk assessment because of incomplete or slapdash data collection. Thus, in this context, it becomes really important to discuss the causes of uncertainty attributable to laboratory and sampling activities, leading to a correct results' representativeness assessment. This is the objective of the PhD research project "Brume" -tutors proff Galetto and Pira- in which a method for worker's exposure assessment is under development, together with a computer assisted technique [2] to help companies and external audit technicians to collect and analyse data rigorously.

The method

Similar to what suggested by the UNI Standard [6], the approach consists of the following steps, manageable through the computer assisted technique [2]:

Step 0-preliminary survey: a basic characterization of the workplace.

¹ Rebecca Nebbia

Politecnico di Torino, Corso Duca degli Abruzzi, 24 10129 Torino, Italy; Università di Torino, piazza Polonia 94 10126 Torino, Italy e-mail: rebecca.nebbia@polito.it

Step 1-investigational survey: it is the exposure assessment phase and entails five sub-steps:

Step 1.1-laboratory activities preliminary to the in-field samplings: sampling lines are set, and the instruments are calibrated.

Step 1.2-in-field sampling: the activities carried out by each individual worker and any other significant parameter are registered, to clear up, in the following step 1.5, the possible outliers.

Step 1.3-laboratory activities following the in-field samplings: the concentration of the pollutant is determined and the causes of possible errors in the transport and manipulation of samples are avoided, and, in case they occurred, the resulting data discarded on the base of practical evidence (known inadequate handling) or suspect of unknown handling accident (using a statistical exclusion principle).

Step 1.4-calculation of the expanded uncertainty and check of quality objectives: The expanded uncertainty (in this case the measurand is the concentration of the pollutant) is calculated following the BIPM Guide GUM [3] and is performed with the uncertainty budget table [1]. This lead a clear and complete result, that considers what is suggested by Unichim methods [4,5], and puts into evidence the sensitivity coefficients and variables such as e. g. pressure and temperature.

Step 1.5-Analysis and interpretation of result: the outliers are pointed out and - where appropriate- excluded from the log-transformed concentration data set. Then the hypothesis of normality of the log-transformed data is corroborated with the Shapiro-Wilk test. Finally, the One-sided Tolerance Limit- OTL Tuggle method or Leidel & Busch method are used to establish compliance/noncompliance of the result [2].

Step 2-combined hygiene and medical survey: the results of the investigational survey should be integrated with medical examinations by Occupational Medicine experts, the accuracy of exposure assessments being an obviously fundamental step for epidemiological studies.

References

1. Barbato, G., Germak, A., Genta, G.: Measurements for Decision Making. Ed. Esc., Bologna (2013)
 2. Bisio, P., Fargione, P., Luzzi, R., Maida, L., Nebbia, R., Patrucco, M.: Computer-assisted technique for airborne dust sampling data representativeness and worker's exposure assessment - CAT-ReADS. GEAM (2018) ISSN 1121-9041
 3. JCGM 100:2008 (GUM) "Evaluation of measurement data - Guide to the expression of uncertainty in measurement"
 4. Metodo Unichim n. 1998:2013 "Ambienti di lavoro - Determinazione della frazione inalabile delle particelle aerodisperse – Metodo gravimetrico"
 5. Metodo Unichim n. 2010:2011 "Ambienti di lavoro - Determinazione della frazione respirabile delle particelle aerodisperse – Metodo gravimetrico"
 6. UNI EN 689: 2018 "Atmosfera nell'ambiente di lavoro – Misura dell'esposizione per inalazione agli agenti chimici – Strategia per la verifica della conformità coi valori limite di esposizione occupazionale"
 7. Yew, W.W., Leung, C.C., Chang, K.C., Zhang, Y., Chan, D.P.: Can treatment outcomes of latent TB infection and TB in silicosis be improved? J.Thorac.Dis. (2019) doi: 10.21037/jtd.2018.12.113
- Sincere thanks to prof. Patrucco for his supervising.

Statistical Analysis of Coordinate Metrology Data with Q-DAS

Marco Massaia¹, Bruno Rolle¹, Suela Ruffa¹ and Marco Zeno¹

Key words: CMM, Statistical Process Control, Measuring System Analysis, QDAS

1 Introduction

Coordinate measuring machines are one of the most common measuring instruments used in different industrial fields to control tolerances of workpieces with high precision. The present work deals with a description of how common statistical tools are applied in the coordinate metrology field. Two different possible data analyses are considered and presented by means of real industrial case studies. The first one refers to the use of measuring system analysis (MSA) to periodically control a Coordinate Measuring Machine (CMM) once it is installed at the customer site, the second is the analysis of the production process capability through the measurement of workpiece samples.

Q-DAS software, part of Hexagon Manufacturing Intelligence, is built to convert characteristic values and process parameters reliably into statistics and to present them appropriately. The main goal of such a software is to pave the way for structured, customised evaluation and control of industrial and metrological processes. In the following the capability of Q-DAS software in managing and analysing CMM data is presented by means of two case studies that make use of two different QDAS modules: *Solara* (MSA module) and *qs-STAT* (Process/Machine Capability module).

1.1 Q-DAS Analysis of CMM data

The procedure for the acceptance of a CMM is regulated by the family of ISO 10360 standards and the machine must be supplied with a certification that it

¹ Hexagon Manufacturing Intelligence, Strada del Portone 113, Grugliasco, Torino, Italy, e-mail: marco.massaia@hexagon.com, bruno.rolle@hexagon.com, suela.ruffa@hexagon.com, marco.zeno@hexagon.com

satisfies the specification. However, standards require only measurements on simple artefacts (calibrates spheres, step gauges or rings), while the customers are often interested into the control of more complex tolerances, thus they ask for extra acceptance tests that are more significant for the specific measurements they need.

With the Q-DAS Solara [1] module it is possible to automatically and periodically perform an MSA of the machine producing Accuracy and Repeatability and Reproducibility Test. It encompasses many guidelines of the automotive industry, even the procedures given in the MSA manual of AIAG's Core Tools [2]. The performance of the CMM is evaluated using a calibrated master measured a proper amount of times, then the software calculates the capability indices (C_g and C_{gk}) and compares them with the applied threshold.



Figure 1: Typical MSA and Process Capability analysis output in Q-DAS software

Once the CMM performance is ensured with an MSA study, the focus can be moved a step forward: on the Statistical Process Control (SPC) whose principal elements are control chart and process capability study [3]. With qs-STAT module the user can engage in the evaluation and continuous improvement of processes based on recognized statistical evaluation methods and procedures. User can view graphs, can choose the period of time to study the evolution of some parameters as he desires. Moreover, the user can trigger measures for process optimization directly from the graphs. The first evaluation can be made via graphics, such as value chart and histograms, analysing the distribution of the process, then capability indexes (C_p and C_{pk}) are computed and compared with the thresholds taking in consideration the best-fitting statistical distribution model.

One of the main pros in using Q-DAS on CMM data is that data are collected automatically from the CMM and loaded into an internal database. The information is centralized, and it is possible to trace any kind of additional data: it is possible to filter measurement and split the statistic evaluation based on (e.g.) date/time, batch number, CMM ID, production machine ID, etc.

The application of Q-DAS software to real industrial case study will be presented and deeply discussed in the paper.

References

1. <https://www.q-das.de/en/>
2. Measurement System Analysis (MSA); Automotive Industry Action Group, 4th edition, 2010
3. Montgomery, Douglas C. Introduction to Statistical Quality Control. Wiley, 2019

EUCoM: Evaluating the Uncertainty in Coordinate Metrology

Alessandro Balsamo¹ and Aline Piccato²

Key words: uncertainty evaluation, coordinate metrology

1 Introduction

Coordinate metrology is a fascinating discipline that bridges a very wide gap between sophisticated mathematics and practical application in industry, and particularly does the evaluation of the uncertainty. On one side, the evaluation is very complex and requires dedicated mathematical and software tools; on the other side, the impact is tremendous in industry, whether the evaluation is carried out properly, or it is not properly, or it is not at all, as unfortunately is often the case.

2 An industrial challenge

Coordinate metrology is widely used for inspection of parts. In most cases, measurements are taken for deciding upon conformance or nonconformance of parts to specifications (tolerances on drawings). This is the field of application of the JCGM 106 [6], which requires that the measurement uncertainty is known and accounted for in the decision. Due to the difficulties in the evaluation, very often this is overlooked, resulting in unreliable decisions. As the turnover of manufacturing is order of € 10¹⁰ in Europe alone, the impact of this poor routine is huge, ranging from economic loss to catastrophic failures (e.g. a faulty blade of an aircraft engine).

¹ Alessandro Balsamo
INRIM, str. delle cacce, 91 – 10135 Torino (IT), e-mail: a.balsamo@inrim.it

² Aline Piccato
INRIM, str. delle cacce, 91 – 10135 Torino (IT), e-mail: a.piccato@inrim.it

3 A mathematical challenge

A major contribution is the CMM geometry error. Models are available [4,1] where the CMM three carriages are subject to 6 roto-translational degrees of freedom each. These 18 d.o.f.'s are functions of a coordinate, requiring 10 to 50 parameters each: several hundred parameters. These parameters are correlated and a full covariance matrix is required, amounting to order of 10^4 input uncertainties to assess.

The derivation of intermediate features does not enjoy a close form solution, and numerical iterations are the only option: the GUM [5] equation (1) is not available.

Features and operators are combined at will. The propagation of the uncertainty is reflected and uncertainty evaluations can only be *task-specific*. The effort risks having not enough return, apart from serialised measurements.

A general solution is based on Monte Carlo simulations (coordinate metrology was first in pioneering this technique [2,7]). However, this requires a significant investment in dedicated software and experimental analysis of individual CMMs. The acceptance of this method is still very limited in practice.

4 The EUCoM project

The EUCoM project [3] aims at tackling this problem in a viable and industry-friendly way. The goal of the project is to develop two methods suitable for as many ISO standards: an *a posteriori* (type A) evaluation based on the dispersion observed in reversal measurements, and an *a priori* (type B) evaluation based on prior information such as standardised CMM performance parameters. The methods will be extensively validated experimentally by all project partners with different CMMs. The project involves 12 partners and 1 permanent collaborator from 10 countries, including 9 NMIs and DIs, 3 universities and 1 research centre.

References

1. Balsamo, A.: Effect of arbitrary coefficients of CMM error maps on probe qualification. Annals of the CIRP **44/1**, 475-478 (1995)
2. Balsamo, A., et al.: Evaluation of CMM uncertainty through Monte Carlo simulations, Annals of the CIRP **48/1**, 425-428 (1999)
3. EMPIR-17NMR03 Standards for the evaluation of the uncertainty of coordinate measurements in industry, 2018-06/2021/05, eucom-empir.eu
4. Hocken, R., et al.: Three-dimensional metrology. Annals of the CIRP **26/2**, 403-408 (1977)
5. JCGM 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in measurement
6. JCGM 106:2012, Evaluation of measurement data – The role of measurement uncertainty in conformity assessment
7. Schwenke, H., et al.: Assessment of uncertainties in dimensional metrology by Monte Carlo simulation: proposal of a modular and visual software. Annals of the CIRP **49/1**, 395-398 (2000)

Uncertainty evaluation in coordinate metrology based on approximate models of CMM behaviour

Alistair B Forbes

Key words: coordinate metrology, uncertainty evaluation

Abstract

The primary task of form and tolerance assessment in precision engineering is to estimate how close a manufactured workpiece is to its ideal geometry, as specified by a technical drawing or CAD specification [ISO(2012)]. Traditionally, such assessment was made using hard gauges, but in recent decades coordinate measuring systems such as coordinate measuring machines (CMMs) are used extensively in industry for this task. However, the evaluation of the uncertainty associated with CMM measurement, for example, is not straightforward. CMM measurement results are subject to a large range of influence factors including kinematic errors, probing effects and environmental effects that are difficult to quantify.

The National Physical Laboratory is participating in a European Metrology Programme for Innovation and Research (EMPIR) led by INRIM, IT, on *Standards for the evaluation of the uncertainty of coordinate measurements in industry*, EUCoM. The overall objective of the project is to develop viable methods for evaluating the measurement uncertainty to support the further development of related standards (in the ISO 15530 series [ISO(2011)]). New viable and standardised methods for evaluating the uncertainty in coordinate measurement would make inspections in manufacturing more reliable and ensure better quality of products.

One uncertainty evaluation approach to be developed by the project is based on using *a priori* information such as an approximate model of CMM behaviour. In this paper, we describe a number of plausible models that capture the main behaviour

Alistair B Forbes
National Physical Laboratory, Teddington, UK, e-mail: alistair.forbes@npl.co.uk

of a CMM and show how they can be used to evaluate uncertainties associated with features or key product characteristics derived from measured coordinate data.

Acknowledgements This work is funded by the European Metrology Programme for Innovation and Research (EMPIR). The EMPIR initiative is co-funded by the European Unions Horizon 2020 research and innovation programme and the EMPIR Participating States.

References

- [ISO(2012)] (2012) The ISO Geometric Product Specifications Handbook. International Organization for Standardization, Geneva
- [ISO(2011)] ISO (2011) ISO 15530: Geometric Product Specifications (GPS) – Coordinate Measuring Machines (CMM): Technique for determining the uncertainty of measurement - Part 3: Use of calibrated workpieces or standards. International Organization for Standardization, Geneva

Allan Variance and Novel Signals from Optical Metrology: How to Avoid Aliasing

Claudio E. Calosso^{1,4}, Cecilia Clivati¹ and Salvatore Micalizio¹

Key words: optical fiber link, atomic clock comparison, Allan variance, modified Allan variance, aliasing.

A basic rule in signal analysis is to use an anti-aliasing filter before data decimation [3]. Surprisingly, this fundamental rule is often missed in time and frequency metrology and this leads to subtle problems when data are analysed. These issues are particularly evident in case of signals coming from the relatively new optical metrology, especially when the Allan variance (AVAR) [1] is calculated.

Time and frequency community is preparing the field for changing the definition of the second from microwave to optical atomic transitions. In this regards, new optical clocks are under development, as well as new fiber links, that are necessary to compare and disseminate them. This revolution is having a strong impact on the characteristics of signals: RF signals, the output of microwave clocks, have the highest spectral purity, since they have not to degrade the phase/frequency information they carry. In the optical domain, RF signals are still present, but here they are the result of beat notes among lasers and carry only part of the information: the difference of laser frequencies/phases. The phase noise now is orders of magnitude higher with respect to the past since the optical carrier frequency is in the terahertz region instead of megahertz. In addition, the novel optical signals have phase spectra that are no longer clean. Bumps and spurs are often present, mainly induced by the sensitivity of lasers and optical fibers to acoustic vibration and seismic noise. They have significant power at relatively high frequencies, mainly in the range from 10 to 10 kHz and these components lead to aliasing if not properly treated.

It is well known that the same signal measured with different types of frequency counters lead to measures that can be very different. In this regard, Λ -counters are preferred to Π -counters, because they are less sensitive to high-frequency noise. It is also well known that, in many cases, the same set of data, analysed with modified

¹ INRIM, Division of Quantum Metrology and Nanotechnology, Torino, Italy.

⁴ CEC is the corresponding author. E-mail: c.calosso@inrim.it

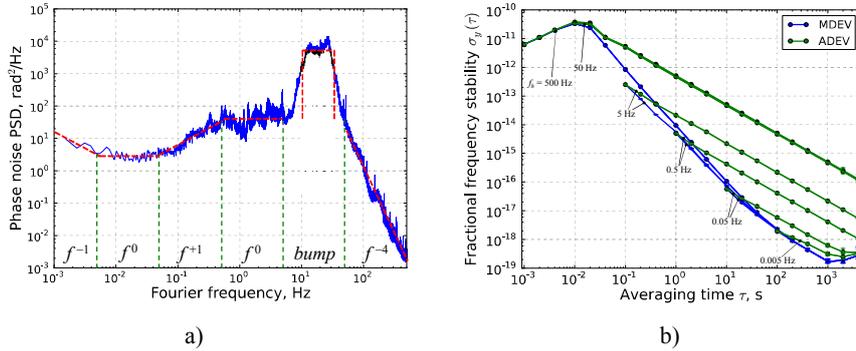


Figure 1: background noise of the Italian fiber link from Torino to Florence (a) and the related Allan and modified Allan deviations as a function of the measurement bandwidth f_h (b). Reprinted from [2].

Allan variance (MVAR), leads to results that are significantly lower with respect to the ones provided by AVAR. The reason is the same: MVAR is less sensitive to high frequency noise than AVAR. However, MVAR is a different estimator with respect to AVAR and, for white frequency noise, the one that we find in primary frequency standards, the difference (-3 dB) is not negligible. In the end, different instruments and different variances applied to the same signal lead to a wide spread of results and it is quite difficult to compare them. Sometimes, for completeness, results are presented with the main four combinations: (Π -counter, Λ -counter) \times (AVAR, MVAR).

In this work, we show that all these issues can be stated in terms of aliasing. We identify two main sources of aliasing: the one induced by the measurement instrument (*instrumental aliasing*), and the one that is hidden in the variance algorithm itself (*variance aliasing*). We show that all these issues can be eliminated by setting properly the anti-aliasing filter that, in metrology community, is well represented by the concept of measurement bandwidth f_h . The latter is always present in tables of variances, but it is seldom used, just because in the past, with RF signals, it was not strictly necessary. Each time f_h is present in the table, it means that aliasing is corrupting the results. At the workshop, we will consider in details all these aspects in the significant case of a frequency dissemination of a clock through the Italian fiber link from Torino to Florence (Figure 1).

References

1. Allan D. W.: Statistics of Atomic Frequency Standards. In: Proc. IEEE V ol. 54, pp. 221-230 (1966).
2. Calosso C. E., Clivati C., Micalizio S.: Avoiding Aliasing in Allan Variance: an Application to Fiber Link Data Analysis. In: IEEE TUFFC, Vol. 63, Issue 4, pp. 1-10 (2016).
3. Nyquist H.: Certain factors affecting telegraph speed. In: Bell Syst. Tech. Jour., vol. 3, pp. 324, (1924).

Role of dead times and correlations in the measurements of optical clocks

Marco Pizzocaro, Filippo Bregolin, Piero Barbieri, Filippo Levi and Davide Calonico

Key words: Optical frequency standards, SI second, Uncertainty evaluation

The second in the International System of Units (SI) is realized by caesium fountains clocks with an uncertainty of few parts in 10^{16} . Optical frequency standards based on optical transitions of several atomic species demonstrated that they can outperform caesium standards both in terms of accuracy and stability, with uncertainties reaching 1×10^{-18} (McGrew et al, 2018). A future redefinition of the SI second based on an optical transition is anticipated (Riehle et al, 2018). The International Committee for Weights and Measures (CIPM) recommended 9 optical transitions as secondary representation of the second.

Absolute frequency measurements of optical transitions and optical frequency ratios are needed to build confidence in optical clocks and assure consistency in view of the redefinition. The values of the secondary representation of the second has been calculated by a least square fit of the available data (Margolis and Gill, 2015). Measurements are usually performed respect to caesium fountains that provides the local realization of the SI second. When a caesium fountain is unavailable optical transitions can be measured respect to the SI second via the International Atomic Time (TAI), as for example in (Hachisu et al, 2017; Baynham et al, 2017).

TAI is a virtual time scale maintained by the International Bureau of Weights and Measures (BIPM) from the comparison of frequency standards in about 80 worldwide laboratories carried out by satellite techniques. BIPM computes TAI in 5 days intervals and gives its deviation from the SI second in one month average published in the bulletin Circular T. Ideally a continuous measurement of the optical frequency over one month should be needed for comparison with the SI second. However optical frequency standards are usually operated only intermittently for few hours at time. Local flywheel oscillators (hydrogen masers) can be used to bridge the gap be-

Marco Pizzocaro · Filippo Bregolin · Piero Barbieri · Filippo Levi · Davide Calonico
Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, 10135 Torino, Italy
e-mail: m.pizzocaro@inrim.it

1

tween the optical measure to one month but the uncertainty introduced by the dead times must be carefully evaluated (Yu et al, 2007).

We have performed the absolute frequency measurements of the ytterbium optical lattice clock at INRIM (Pizzocaro et al, 2017) with a link to TAI taking data for 5 months. At the conference we will present the uncertainty evaluation of the link to TAI for each month, including the contribution of the dead time while using an hydrogen maser as a flywheel.

Moreover a careful evaluation of the correlations in the optical frequency measurements and in the link to TAI is needed. We will discuss possible source of correlations in the available worldwide data.

Our calculations are important in view of the possible redefinition of the SI second but also as a preliminary step to directly incorporate optical clocks into TAI.

Acknowledgements This work has been funded by the European Metrology Program for Innovation and Research (EMPIR) project 15SIB03 OC18, the EMPIR project 18SIB05 ROCIT and the ASI project DTF. This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Unions Horizon 2020 research and innovation programme.

References

- Baynham CFA, Godun RM, Jones JM, King SA, Nisbet-Jones PBR, Baynes F, Roland A, Baird PEG, Bongs K, Gill P, Margolis HS (2017) Absolute frequency measurement of the optical clock transition in with an uncertainty of using a frequency link to international atomic time. *Journal of Modern Optics* 65(5-6):585–591
- Hachisu H, Petit G, Nakagawa F, Hanado Y, Ido T (2017) Si-traceable measurement of an optical frequency at the low 1×10^{-16} level without a local primary standard. *Opt Express* 25(8):8511–8523
- Margolis HS, Gill P (2015) Least-squares analysis of clock frequency comparison data to deduce optimized frequency and frequency ratio values. *Metrologia* 52(5):628
- McGrew WF, Zhang X, Fasano RJ, Schffer SA, Beloy K, Nicolodi D, Brown RC, Hinkley N, Milani G, Schioppo M, Yoon TH, Ludlow AD (2018) Atomic clock performance enabling geodesy below the centimetre level. *Nature* 564(7734):87–90
- Pizzocaro M, Thoumany P, Rauf B, Bregolin F, Milani G, Clivati C, Costanzo GA, Levi F, Calonico D (2017) Absolute frequency measurement of the $^1S_0 - ^3P_0$ transition of ^{171}Yb . *Metrologia* 54(1):102
- Riehle F, Gill P, Arias F, Robertsson L (2018) The CIPM list of recommended frequency standard values: guidelines and procedures. *Metrologia* 55(2):188
- Yu DH, Weiss M, Parker TE (2007) Uncertainty of a frequency comparison with distributed dead time and measurement interval offset. *Metrologia* 44(1):91

Two-Sample Covariance and Multichannel Tracking DDS for Measuring the Frequency Stability of Cryogenic Sapphire Oscillators

Claudio E. Calosso^{1,4}, François Vernotte², Vincent Giordano², Christophe Fluhr³, Benoit Dubois³ and Enrico Rubiola^{2,1}

Key words: ultra-stable oscillators, phasemeter, two-sample covariance, three-cornered hat.

The aim of this contribution is to show how two-sample covariance can be used to improve the short-term resolution of instruments and make them suitable to measure ultra-stable signals. Specifically, this work shows the first measurement of three 100 MHz signals exhibiting fluctuations from 2×10^{-16} to parts in 10^{-15} for integration time τ between 1 s and 1 day. Such stable signals are provided by three Cryogenic Sapphire Oscillators (CSOs) [3] operating at about 10 GHz, also delivering the 100 MHz output via a dedicated synthesizer.

The measurement is made possible by a 6-channel Tracking DDS (TDDS) [1] and the two-sample covariance tool [4], used to estimate the Allan variance (AVAR). The use of two TDDS channels per CSO (see Figure 1.a, upper part) enables high rejection of the instrument background noise, since the residual noises of two different channels are independent. In this regard, the covariance is advantageous with respect to Three-Cornered Hat (TCH) method, because the background converges to zero by theorem, with no need of the hypothesis that the instrument channels are equally noisy, nor of more sophisticated techniques to estimate the background noise of each channel. Thanks to correlation and averaging, the instrument background (AVAR) rolls off with a slope $1/\sqrt{m}$, the number of measurements, at a level below 10^{-15} at $\tau = 1$ s. For consistency check, we compare the results to the traditional TCH method beating the 10 GHz outputs down to the megahertz region (see Figure 1.a, lower part).

¹ INRIM, Division of Quantum Metrology and Nanotechnology, Torino, Italy.

² FEMTO-ST Institute, Dept. of Time and Frequency, Université de Bourgogne and Franche-Comté (UBFC), and CNRS, Besançon, France.

³ FEMTO Engineering, Besançon, France.

⁴ CEC is the corresponding author. E-mail: c.calosso@inrim.it

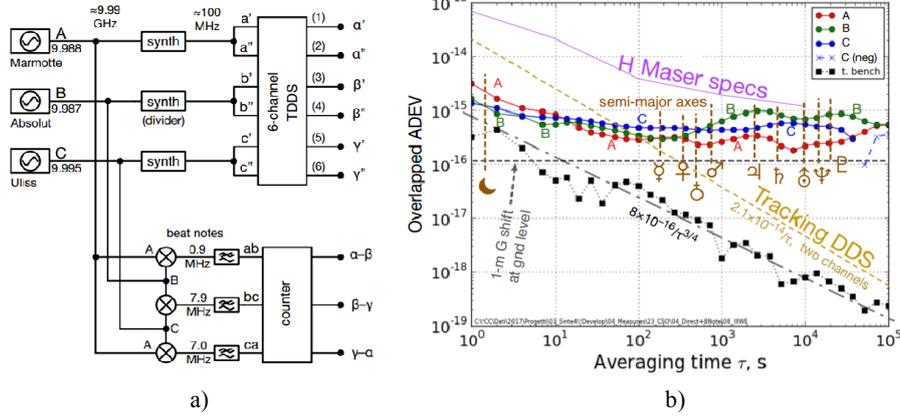


Figure 1: test bench used to measure the frequency stability of three CSOs (a) and the obtained results (b). Reprinted from [2].

Figure 1.b summarizes the relevant results obtained with the TDDS and the covariance at the 100 MHz output of the CSOs. The gold dashed line shows the background noise of the TDDS (AVAR), accounting for two channels. The background noise is $\sigma_y(\tau) = 2.1 \times 10^{-14}/\tau$ and is low enough for the measurement of H masers and the other classical atomic standards. Additionally, the CSOs can be measured when 2-sample covariance is used, provided that each CSO feeds two independent channels. The statistical limit improves and, in this particular case where the duration of the time series is 5.5 days, the instrument background reduces to $\sigma_y(\tau) = 8 \times 10^{-16}/\tau^{3/4}$ (grey dash-dot line). The measurement confirms that the CSO instability is in the low 10^{-15} at 1 s, and below 10^{-15} for $\tau \geq 10$ s up to one day. For reference, the gravitational shift g/c^2 is of 1.09×10^{-16} when a clock is raised by 1 m at the ground level. The best stability is seen for τ between seconds and hours, which is the travel time of the light at interplanetary distances.

This contribution is an extract of [2]. Our goal is to diffuse these concepts to a wider audience, as the one represented by the community of this workshop.

References

1. Calosso C. E.: Tracking DDS in time and frequency metrology. In: Proc. Europ. Freq. Time Forum, Prague, Czech Republic, pp. 747–749 (2013).
2. Calosso C. E., Vernotte F., Giordano V., Fluhr C., Dubois B., Rubiola E.: Frequency Stability Measurement of Cryogenic Sapphire Oscillators with a Multichannel Tracking DDS and the Two-Sample Covariance. In: IEEE TUFFC. DOI 10.1109/TUFFC.2018.2870593 (2018).
3. Fluhr C., Grop S., Dubois B., Kersalé Y., Rubiola E., and Giordano V.: Characterization of the individual short-term frequency stability of cryogenic sapphire oscillators at the 10⁻¹⁶ level. In: IEEE Trans. Ultras. Ferroelec. Freq. Contr., vol. 63, pp. 915–921, (2016).
4. Vernotte F., Calosso C. E., and Rubiola E.: Three-cornered hat versus Allan covariance. In: Proc. Int'l Freq. Control Symp., New Orleans (LA, USA), pp. 282–287 (2016).

Effective validation of chromatographic analytical methods

Eleonora Amante¹, Eugenio Alladio, Cristina Bozzolino, Fabrizio Seganti, Alberto Salomone, Marco Vincenti

Key words: chromatographic analytical methods, validation protocol, calibration.

1 Introduction and purpose of the study

Several organizations and scientists have tried to standardize the validation procedure, according to the purpose of the analysis (*e.g.*, qualitative, quantitative) and the application field, recommending specific parameters to be evaluated and tests to be performed (Shabir, 2004; Taverniers, De Loose, & Van Bockstaele, 2004). For the validation of quantitative methods, a feature of utmost importance is represented by the calibration. Although most instrumental systems should theoretically supply a straight correlation between concentrations and analytical signals, in reality some interfering physical and chemical phenomena may result in a deviation from the expected linear trend and/or heterogeneous distribution of data-point at different concentration levels (Sayago & Asuero, 2004). The occurrence heteroscedasticity is responsible of the erroneous evaluation of other parameters, of which the most relevant is the limit of detection (LOD).

In our daily work with biological matrices, LC-MS/MS and GC-MS-based methods are continuously developed and/or updated to support the ongoing evolution of clinical and toxicological requirements. This changing scenario opens further analytical inquiries not addressed in standardized validation procedures, including practical questions that frequently need careful consideration. To answer these questions, we studied an integrated approach, resulting in the development of an efficient and rigorous validation protocol.

¹ Eleonora Amante

Dipartimento di Chimica, Università degli Studi di Torino, via P. Giuria 7 (Torino)
Centro Regionale “A. Bertinaria”, Regione Gonzole, 10/1, Orbassano (Torino)

email: eamante@unito.it

2 Validation protocol

Our validation protocol integrates the Desharnais' routine (Desharnais, Camirand-lemyre, Mireault, & Skinner, 2017) R procedure of calibration into an inclusive strategy to estimate further validation parameters: intra- and inter-day accuracy and precision, LOD, limit of quantification (LOQ), ion abundance repeatability, selectivity, specificity and carry-over. A scheme of the operating procedure is reported in Figure 1.

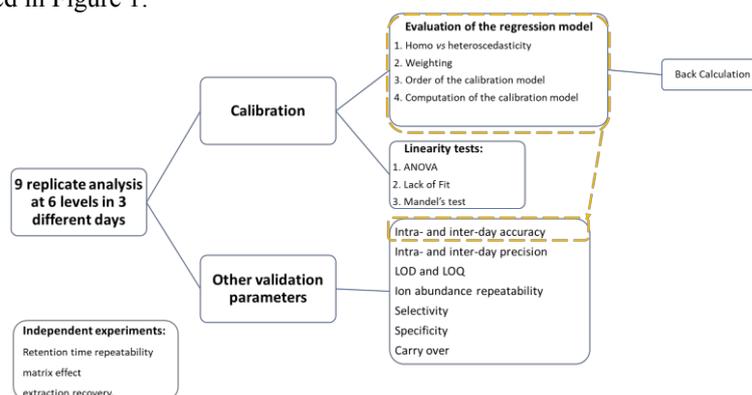


Figure 1: Scheme of the validation protocol.

The new protocol allows to verify the stability of calibration with time and the need to renovate it, measures accuracy and precision at all calibration levels, calculates the best fit parameters, including weighting factors, regression performance and a rigorous evaluation of the LOD and LOQ. Simultaneously, it permits to evaluate selectivity, specificity, ion abundance repeatability and carry over effect without conducting further experiments.

References

- 1 Desharnais, B., Camirand-lemyre, F., Mireault, P., & Skinner, C. D. (2017). Procedure for the Selection and Validation of a Calibration Model I — Description and Application. *Journal of Analytical Toxicology*, 41, 261–268. <https://doi.org/10.1093/jat/bkx001>
- 2 Sayago, A., & Asuero, A. G. (2004). Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of variances. *Critical Reviews in Analytical Chemistry*. <https://doi.org/10.1080/10408340490888599>
- 3 Shabir, G. (2004). Step-by-step analytical methods validation and protocol in the quality system compliance industry. *Journal of Validation Technology*, 10, 210–218. <https://doi.org/10.1038/ncomms11248>
- 4 Taverniers, I., De Loose, M., & Van Bockstaele, E. (2004). Trends in quality in the analytical laboratory. II. Analytical method validation and quality assurance. *TrAC - Trends in Analytical Chemistry*, 23(8), 535–552. <https://doi.org/10.1016/j.trac.2004.04.001>

Chemometric handling of Raman spectra for live systems monitoring and susceptibility tests

Luisa Mandrile¹, Emanuela Noris², Giulia Barzan¹, Alessio Sacco¹, Andrea Giovannozzi¹, Silvia Rotunno², Laura Miozzi², Anna Maria Vaira² and Andrea Rossi¹

Key words: Chemometrics, Raman Spectroscopy, Principal Components Analysis, Partial least squares regression, Partial Least Squares Classification, Antibiotic susceptibility test, Virus susceptibility tests

1 The use of chemometrics for Raman spectroscopy data analysis and applications in life science

Multifaceted and innovative approaches for live systems monitoring are needed at any level in life sciences to improve control strategies in agro-food industry, pharmaceutical and medical fields. Immunological assays, nucleic acid-based techniques (e.g. ELISA and PCR), colony forming unit tests (CFU) and other classical cell viability tests are time-consuming, destructive and expensive.

Raman spectroscopy (RS) combined with chemometric analysis generate a chemical fingerprinting of the sample rapidly, at low operating costs and without sample destruction [1, 3].

Here, we present two examples of application of Raman spectroscopy for live systems monitoring of the acute effects of exogenous agents in a time frame, substantiating the possibility to build predictive models to evaluate i) the susceptibility of plants to virus infection and ii) the outcome of antibiotic treatment in bacterial cells, aiming at susceptibility tests for antibiotic resistance studies [2].

Negative controls and biological replicates were performed to guarantee representative sampling of the natural variability of biological samples and to avoid misleading interpretation of data variability that can occur in small data sets.

¹ Luisa Mandrile, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135, Torino, Italy, e-mail: l.mandrile@inrim.it

² Emanuela Noris, Institute for Sustainable Plant Protection, National Research Council of Italy, Strada delle Cacce 73, 10135, Torino, Italy
e-mail: emanuela.noris@ipsp.cnr.it

Correlations between spectra measured with a laboratory Raman spectrograph (DXR Dispersive Raman by ThermoFisher) on fresh leaflets detached from tomato plants (without impairing plant viability) and virus infection in the first case, and the effect of drugs to *Escherichia coli* cells in the latter were found, efficiently separating the ageing effect that occurs in biological samples. Principal Components Analysis (PCA) was preliminarily applied for easier visualization of systematic spectral differentiation. Then, partial least squares-discriminant analysis (PLS-DA) provided sensitivity and specificity of RS classification for early diagnosis. RS captured the effect of virus infection when the symptoms are still not evident in plants, with accuracy higher than 70 % and 85 % for two viruses, in cross validation. Moreover, RS allowed to detect the immediate action of drugs on bacterial cells (models validated using completely independent sets of data) providing high robustness and generalization capability of the classification model (Figure 1). Raman results were confirmed by comparison with traditional biomolecular tests commonly used for such studies.

RS, assisted by chemometrics, is eligible for in-field analysis using portable RS instruments in plant breeding programs for disease resistance monitoring and in laboratory medicine for antibiotic resistance tests, in which rapid and potentially portable measurement techniques are strongly needed to reduce time and costs of analysis, without impairing reliability of results.

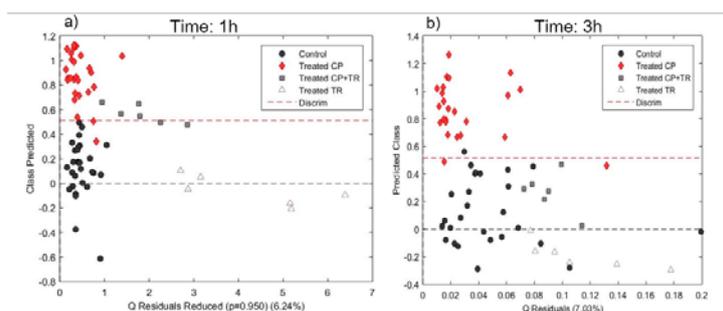


Figure 1: Prediction of the resistance to ciprofloxacin (CP) induced by triclosan (TR) with PLS-DA model of RS (3 LVs). Training (red and black dots) and predicted (white and grey) sets after a) 1 h and b) 3 h of cell growth in presence or absence of antibiotics against Q residuals.

References

1. Danting Y., and Ying Y. (2011) Applications of Raman spectroscopy in agricultural products and food analysis: A review. *Appl Spec Rev* 46: 7, 539-560
2. Schröder U.C., Kirchhoff J., Hübner U., Mayer G., Glaser U., Henkel T., Pfister W., Fritzsche W., Popp J., and Neugebauer U. (2017) On-chip spectroscopic assessment of microorganisms. *J biophotonics* 10: 11, 1547-1557
3. Tineke V., Vercauteren A., Baeyens W., Van der Weken G., Verpoort F., Vergote G., and Remon J. P. (2002) Applications of Raman spectroscopy in pharmaceutical analysis. *TrAC* 21:12 869-877.

Calibration Curve Computing (CCC) software v2.0: A new version of INRIM's software for regression analysis

M. Lecuna¹, F. Pennechi², A. Malengo², P. G. Spazzini²

Key words: regression models, calibration curves, CCC software, data analysis

1 Introduction

The Guide for the Expression of the Uncertainty in Measurement (GUM) defines a set of rules to properly evaluate and report measurement uncertainties. However, many real life cases, such as regression problems, are not explicitly covered by the GUM and related supplements yet. From 2012 to 2015, the EMRP NEW04 project “Novel mathematical and statistical approaches to uncertainty evaluation” was conducted with the aim to provide mathematical and statistical tools for attaining reliable uncertainty evaluation on problems not explicitly covered by the GUM [1]. CCC Software version 1.3 was a software originally developed and tested in the framework of the EMRP NEW04. Its main objective was to provide statistical treatment to flow measurement data by determination of relevant calibration curves [2]. Based on comments from users and recent testing, a set of improvements was proposed and implemented. The new version has addressed problems with the standardisation of data. The program now includes a more complete version of the plot area and the report file, and an improved minimization algorithm that uses the model gradient [4]. Especially the latter novelty improves the quality of the regression results by delivering smaller uncertainties and a chi-squared value, validated against literature analytical results, with a difference below $1e-6$ [3,5].

2 The CCC software: version 2.0 vs. 1.3

CCC Software was developed in MATLAB. The program takes as input a .xls file containing information about experimental values and their covariance matrices, if available, according to a specific format. The output is a plot containing the experimental data against the estimated regression model, the uncertainty of the fitted y values and a report with information on the regression.

¹ Politecnico di Torino, Torino, 10129, Italy, e-mail: maricarmen.lecuna@polito.it

² INRIM, Istituto Nazionale di Ricerca Metrologica, Torino, 10135, Italy, e-mail: f.pennechi@inrim.it, a.malengo@inrim.it, p.spazzini@inrim.it

The interface of version 2.0 respects that of version 1.3, but with different numerated sections to guide the user through the process of construction of the calibration curve (See Figure 1). The first step is the choice of the data file, by typing the path of the file or using the browse button. By loading the data in the version 2.0, the plot area on the right side, shows the experimental points and their uncertainties (if provided).

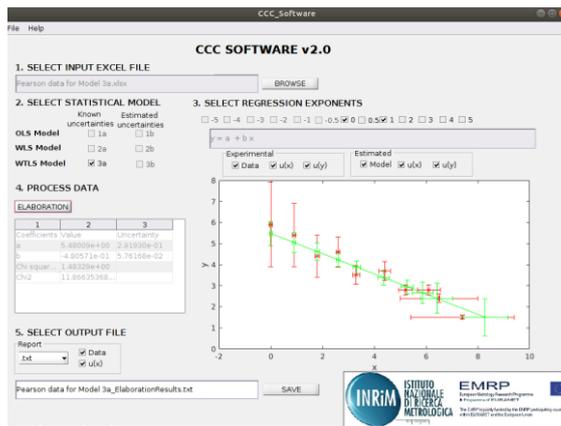


Figure 1. CCC Software version 2.0

The second section concerns the choice of the regression model. As in the previous version, the program automatically unfreezes the possible options based on the type of available data: the possible algorithms include Ordinary least-squares regression (OLS), Weighted least-squares regression (WLS) and Weighted total least-squares regression (WTLS), with and without uncertainty data. The third section allows the user to select the model that will be considered in the regression. The model is built by adding the chosen terms (each one given by its exponent: negative, positive, square root or its inverse). In the box located over the plot area, the chosen model is shown, and it can be changed by ticking the boxes of the desired exponents. In the fourth step, by clicking the 'Elaboration' button, the program performs the regression analysis and shows the resulting regression model in the plot area. The parameters and their uncertainties are displayed in the box under the 'Elaboration' button, along with the chi-squared value as an indicator of the quality of the fit. In the fifth section, a .txt report is generated containing insightful information about the regression analysis, including indicators of the regression quality as well as the fitted x values in the case of the WTLS.

The new version improves the estimation of the regression parameters and their uncertainty evaluation by the inclusion of the Jacobian matrix of the model in the regression process. The outputs, both the report and the plot, are now more readable and customizable, a characteristic that can widen the application of the program beyond the use on calibration of instruments. From June 2019, CCC software version 2.0 will be freely available for public download on the official INRiM website.

References

1. Final Publishable JRP Summary for NEW04 – Uncertainty Novel mathematical & statistical approaches to uncertainty evaluation (2017)
2. INRiM. Calibration Curves Computing – CCC Software User manual (1.3) (2015)
3. Malengo, M. Pennechi, F. Metrologia. 50 654-62 (2013)

Advanced statistical techniques to assess nonclassicality in ensembles of single photon emitters

I. P. Degiovanni¹, P. Traina², E. Moreva², J. Forneris³, S. Ditalia Tchernij⁴, E. Bernardi², F. Picollo⁴, G. Brida², P. Olivero⁴, M. Genovese²

Key words: single-photon sources, subpoissonian statistics

In single-photon metrology [1] the characterization of the emission statistics of the sources in order to quantifying the non-classical properties is of the utmost importance, i. e. the quality of the single photon state produced. This is particularly relevant in quantum cryptography, where uncontrolled fluctuations in the number of photons may open serious security issues. The most used parameter for this characterization is the second order Glauber's correlation function ($g^{(2)}$) [3], which, despite having the advantage of being independent of the quantum efficiency of the detectors, has also several drawbacks, especially when one aims at the characterization of clusters of emitters or single emitters in noisy background. For these systems, new tools based on parameters that are resilient to noise and exploiting multifold coincidence events are being proven to be effective in specific contexts. In this framework we will report on the first experimental demonstration of a recently proposed criterion (Filip's θ function) [4] addressed to detect nonclassical

¹ Ivo Pietro Degiovanni

INRIM, Strada delle cacce 91, 10135 Torino, Italy, e-mail: i.degiovanni@inrim.it

² Paolo Traina, Ekaterina Moreva, Ettore Bernardi, Giorgio Brida, Marco Genovese

INRIM, Strada delle cacce 91, 10135 Torino, Italy

³ Jacopo Forneris

Istituto Nazionale di Fisica Nucleare (INFN) Sez. Torino, 10125 Torino, Italy

⁴ Sviatoslav Ditalia Tchernij, Federico Picollo, Paolo Olivero
Physics Department and "NIS" inter-departmental Centre–University of Torino,
10125 Torino, Italy

behavior in the fluorescence emission of ensembles of single-photon emitters (applied in particular to clusters of Nitrogen-Vacancy centres in diamond) [5]. In a nutshell, the difference between the Glauber's and the Filip's functions is that the former relies on the multi-detection of photons in coincidence, while the latter relies on simultaneous "non-detection" of photons. We will introduce simulation results on the application of a novel technique exploiting higher order Glauber's and Filip's functions ($\theta^{(n)}$ and $g^{(n)}$ with $n > 2$) simultaneously to entirely reconstruct the modes hidden in more complex optical fields such as, e.g., single photon sources in a noise-bath. This mode reconstruction method is based on optimisation algorithms requiring as input data, as said, higher order Glauber's and Filip's functions (that are somehow connected to high order moments of the statistics of the input photons), whose associated uncertainties increase with the order. We show that the use of both $\theta^{(n)}$ and $g^{(n)}$ (rather than using only $g^{(n)}$ as it was done in the past [2]) allows to reduce the functions order necessary to carry on the reconstruction, hence improving its performances.

References

1. C.J. Chunnillall, I.P. Degiovanni, S. Kück, I. Müller, and A. G. Sinclair, *Opt. Eng.* **53**, 081910 (2014)
2. E.A. Goldschmidt, F. Piacentini, I. Ruo Berchera, S.V. Polyakov, S. Peters, S. Kück, G. Brida, I.P. Degiovanni, A. Migdall, and M. Genovese, *Phys. Rev. A* **88**, 013822 (2013)
3. P. Grangier, G. Roger, and A. Aspect, *Europhys. Lett.* **1**, 173 (1986)
4. L. Lachman, I. Slodicka, and R. Filip, *Sci. Rep.* **6**, 19760 (2016)
5. E. Moreva, P. Traina, J. Forneris, I. P. Degiovanni, S. Ditalia Tchernij, F. Picollo, G. Brida, P. Olivero, and M. Genovese, *Phys. Rev. B* **96**, 195209-1 (2017)

**p-values: meaning, misconceptions and dangers, but also their practical utility if used
*cum grano salis***

Giulio D'Agostini

Invited Speaker - Professor of the Physics Department of the University of Rome "La Sapienza", Roma (Italy) and INFN Roma 1

Starting remembering few items of the quite long list of false discoveries announced by physicists in the past decades, I will focus on what **they meant** to report ('the probability of a genuine discovery') by the accompanying "number of sigmas" and why this cannot be achieved by the statistical tool ('p-values') they applied to analyze the data.

Fortunately, there is presently an increasing level of awareness of what p-values are and what they aren't, but there is still a strong resistance by practitioners accustomed to their statistical tool box.

It is then important to understand when p-values can be used as alarm bells for further investigations and when, instead, a more careful probabilistic comparison of the possibly concurring models is required.

The median scaled difference: An outlier-resistant and model-independent indicator of anomalies for Key Comparison data.

Stephen L R Ellison¹

Key words: key comparisons, outlier identification, meta-analysis

A robust pairwise statistic, the pairwise median scaled difference (MSD), has been proposed for the detection of anomalous location/uncertainty pairs in heteroscedastic interlaboratory study data with associated uncertainties [1]. The statistic provides a simple indication of degree of agreement between a particular laboratory's reported result and those of a majority of other participants in the study. The indicator does not depend on a particular choice of key comparison reference value, which is an advantage when there is no independent reference value and where different KCRV estimates can give contradictory indications of the consistency, or otherwise, of a particular laboratory result with the KCRV.

The presentation provides a brief description of the MSD indicator and its distribution in the special case of identically distributed independent results. The resistance of the indicator to secondary extreme values – a common problem even in studies of moderate size – will be demonstrated. Simple rules for inspection will be given, and a rigorous interpretation using a simulation method will be shown. The use of the indicator on some typical key and pilot comparison data sets will be demonstrated.

References

1. Ellison S L R (2018) An outlier-resistant indicator of anomalies among inter-laboratory comparison data with associated uncertainty. *Metrologia*, **55:6** 840 (2018). DOI: 10.1088/1681-7575/aae610. Preprint available as arXiv:1809.01094 (<https://arxiv.org/abs/1809.01094>)

¹ Stephen L R Ellison
Laboratory of the Government Chemist, Queens Road, Teddington, TW11 0LY,
United Kingdom, e-mail: s.ellison@lgcgroup.com

Consistency of subjective evaluations: an investigation into the paradoxical behaviour of inter-rater agreement coefficients

Amalia Vanacore and Maria Sole Pellegrino

In several business and industrial systems as well as in social, behavioural and medical contexts, strategic and/or operational decisions rely on subjective evaluations provided by small groups of human raters who may be either field experts (e.g. visual inspectors) or untrained operators (e.g. consumers).

In such contexts, human raters are asked to rate a set of items according to some technical properties (e.g. classification of faulty material by defect type) and/or perception aspects (e.g. quality, comfort, pain, pleasure, beauty) adopting a dichotomous, nominal or ordinal rating scale. This means that raters act as measurement instruments [3, 11, 12] and they can be a main source of epistemic uncertainty [2], which can be reduced only by selecting *reliable* raters.

Subjective evaluations lack a reference value for checking their trueness, thus rater reliability can be assessed only in terms of their repeatability (i.e. intra-rater agreement) and reproducibility (i.e. inter-rater agreement). The more raters agree on the evaluations they provide, the more comfortable we can be that their evaluations are consistent and exchangeable with those provided by other raters and thus trustworthy [9].

This work focuses on the assessment of rater reproducibility, that is rater ability of providing evaluations that are consistent with those provided by other raters (i.e. inter-rater agreement). The degree of inter-rater agreement over ordinal classifications is commonly measured via weighted κ -type coefficients (e.g. [1, 8]).

Despite their popularity, some κ coefficients have been long criticized because of their *paradoxical behaviour*, that is the dependency on the distribution of ratings over classification categories. This makes κ coefficients misrepresent the level of agreement in the case of unbalanced distributions and it is unclear what the coeffi-

Amalia Vanacore
Dept. of Industrial Engineering, University of Naples "Federico II"
e-mail: amalia.vanacore@unina.it

Maria Sole Pellegrino
Dept. of Industrial Engineering, University of Naples "Federico II"
e-mail: mariasole.pellegrino@unina.it

cients are truly measuring [4, 5].

Our research investigates, via an extensive Monte Carlo simulation study, the paradoxical behaviour of three κ coefficients for inter-rater agreement: Fleiss' K [6], s^* coefficient of Marasini et al. [10] and Gwet's AC_2 [7].

The simulation study has been designed as a multi-factor experimental design with five multi-level factors: rating scale dimension, sample size, number of raters, frequency distributions over rating categories and agreement distributions among raters. Specifically, the frequency distributions mimic the ordinal classification scheme adopted by the i^{th} rater whereas the agreement distributions model the probability that the other raters will provide a certain evaluation given that the rater i 's evaluation is known.

Simulation results suggest that Fleiss's K is strongly influenced by marginal frequency distributions of ratings, whereas s^* and AC_2 are less sensitive. Moreover, all the three analysed coefficients are affected by the sample size and the number of raters.

Key words: Rater reproducibility, κ coefficients, Prevalence paradox, Monte Carlo simulation

References

1. Banerjee, M., Capozzoli, M., McSweeney, L., Sinha, D.: Beyond kappa: A review of interrater agreement measures. *Canadian journal of statistics* **27**(1), 3–23 (1999)
2. Barnhart, H.X., Haber, M.J., Lin, L.I.: An overview on assessing agreement with continuous measurements. *Journal of biopharmaceutical statistics* **17**(4), 529–569 (2007)
3. Bashkansky, E., Gadrich, T., Knani, D.: Some metrological aspects of the comparison between two ordinal measuring systems. *Accreditation and Quality Assurance* **16**(2), 63–72 (2011)
4. Cicchetti, D.V., Feinstein, A.R.: High agreement but low kappa: II. Resolving the paradoxes. *Journal of clinical epidemiology* **43**(6), 551–558 (1990)
5. Feinstein, A.R., Cicchetti, D.V.: High agreement but low kappa: I. the problems of two paradoxes. *Journal of clinical epidemiology* **43**(6), 543–549 (1990)
6. Fleiss, J.L.: Measuring nominal scale agreement among many raters. *Psychological bulletin* **76**(5), 378 (1971)
7. Gwet, K.L.: Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* **61**(1), 29–48 (2008)
8. Gwet, K.L.: *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC (2014)
9. Hayes, A.F.: *Statistical methods for communication science*. Routledge (2009)
10. Marasini, D., Quatto, P., Ripamonti, E.: Assessing the inter-rater agreement for ordinal data through weighted indexes. *Statistical methods in medical research* **25**(6), 2611–2633 (2016)
11. Pendrill, L.: Man as a measurement instrument. *NCSLI Measure* **9**(4), 24–35 (2014)
12. Pendrill, L.: Using measurement uncertainty in decision-making and conformity assessment. *Metrologia* **51**(4), S206–S218 (2014)

Virtual measurements on patients with orthopaedic prostheses during MRI

A. Arduino and O. Bottauscio and M. Chiampi and L. Zilberti¹

Key words: Dosimetry, Magnetic Resonance Imaging, Medical implants, Numerical simulations, Pennes' bioheat equation.

1 Introduction

The use of Magnetic Resonance Imaging (MRI) in patients with orthopaedic implants is more and more increasing. The interaction of the switching magnetic field produced by the gradient coils (GCs) of an MRI scanner with metallic joint arthroplasties can result in a significant heating of the surrounding tissues [2, 3], whose temperature must be therefore monitored. Since invasive in-vivo detections cannot be performed, virtual measurements have to be provided by numerical tools able to solve the multiphysics electromagnetic/thermal problem in human body models. For such a purpose, a computational procedure, devoted to handle the complicated time evolution produced by the MRI sequences, has been developed in the framework of the 17IND01 MIMAS Project. The project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

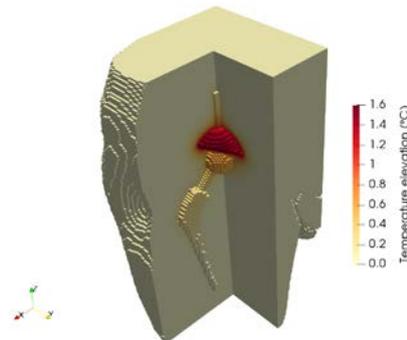
2 Numerical Tool

The numerical procedure has been developed for realistic human models discretized into voxel under two fundamental assumptions. First, the magnetic field produced by

¹ Alessandro Arduino, Oriano Bottauscio, Mario Chiampi and Luca Zilberti
Istituto Nazionale di Ricerca Metrologica, Torino, Italy.
e-mail: a.arduino@inrim.it, o.bottauscio@inrim.it, m.chiampi@inrim.it,
l.zilberti@inrim.it

the currents induced in the tissues is negligible, so that the electromagnetic field problem is solved only inside the metallic prosthesis. Moreover, since the electromagnetic parameters are not affected by the expected temperature variations, the thermal problem can be solved separately using the deposited electromagnetic energy as forcing term. The complicated time evolution of GCs switching suggests an approach alternative to the step-by-step technique, commonly adopted for electromagnetic problems, that would require a very high number of time steps. The proposed procedure is based on a preliminary decomposition of the current signal of each coil into elementary sub-signals, which are then expanded in truncated Fourier series. For each harmonic of the elementary sub-signals, an electromagnetic field problem is solved through an original formulation based on a hybrid finite-boundary element method and the values of the harmonic electric field, \mathbf{E} , computed in the voxels of the prosthesis are stored. The \mathbf{E} -field time waveforms produced by each coil are reconstructed and superposed, allowing the computation of the instantaneous power deposited in each voxel. The thermal problem, described by Pennes' bioheat equation including thermoregulation effects, is developed in the entire human body (or in a suitable portion of it). The solution, in terms of temperature elevation with respect to the temperature before exposure, is provided by a step-by-step procedure based on the finite volume method using Douglas–Gunn time split [1]. The procedure is here used to evaluate the temperature increase in an adult man (Duke of the “Virtual Family”) exposed to the GCs switching of an Echo-Planar Imaging sequence, during a thorax scan in a tubular scanner. The patient has a unilateral hip prosthesis whose metallic parts are made of a CoCrMo alloy with 1.16 MS/m electric conductivity. Fig. 1 shows the temperature elevation after 12 minutes of exposure, which simulates the repeated MR acquisitions to image multiple body slices.

Figure 1: Distribution of the temperature elevation in a body portion including the hip prosthesis. Simulated exposure for 12 minutes to the GCs switching of an Echo-Planar Imaging sequence with readout along z during a thorax scan.



References

1. A. Arduino, O. Bottauscio, M. Chiampi, L. Zilberti: Douglas-Gunn Method Applied to Dosimetric Assessment in Magnetic Resonance Imaging. *IEEE Trans. Magn.* **53**, 5000204 (2017)
2. R. Bruehl, A. Ihlenfeld, B. Ittermann: Gradient heating of bulk metallic implants can be a safety concern in MRI. *Magn. Reson. Med.* **77**, no. 5, 1739-1740 (2017)
3. L. Zilberti et al.: Numerical Prediction of Temperature Elevation Induced Around Metallic Hip Prostheses by Traditional, Split and Uniplanar Gradient Coils. *Magn. Reson. Med.* **74**, 272-279 (2015)

Advanced modelling of stochastic distributions of magnetic nanostructures

Riccardo Ferrero¹ and Alessandra Manzin²

Key words: Magnetic nanostructures, Micromagnetics, Landau-Lifshitz-Gilbert equation, Simulated experiments, Computational methods.

1 Introduction

The need of accurately simulating large numbers of mutually interacting magnetic nanostructures has emerged in many recent applications of nanomagnetism, such as magnetic storage and magneto-logic devices [1], artificial spin-ice structures [5] and magnetic hyperthermia delivery [4]. In the last example, the nanostructures can be stochastically distributed in a 3D medium, a biological fluid or a tissue, at locally high concentrations and with different orientations with respect to the applied field.

In this context, the spatial integration of the Landau-Lifshitz-Gilbert (LLG) equation, typically used in micromagnetics to describe magnetization dynamics at the nanoscale level, can be computationally very demanding. The reason is that we have to simultaneously model the quantum-mechanics exchange interactions, active on a local scale (10 nm) and the magnetostatic interactions, which depend on the global size of the considered system. When introducing a spatial discretization of 5-10 nm, the evaluation of the magnetostatic field on a scale larger than 10 μm is a task that requires the implementation of ad hoc strategies. In micromagnetic tools, one of the most common approaches to speed up magnetostatic field computation is based on Fast Fourier Transform (FFT) techniques. However, FFT techniques, which require structured meshes, are not suitable for the simulation of large numbers of thin-film objects with complex shape and not-aligned orientation. To solve this complex problem, we have developed a 2.5D GPU-parallelized micromagnetic code, able to model a high number of 2D-like nanostructures, randomly distributed in a 3D space.

2 Numerical Model

We consider an ensemble of N equal 2D-approximable magnetic nanostructures distributed in a 3D non-magnetic medium. Each object is discretized with the same mesh of T hexahedra. In each hexahedron the magnetization vector \mathbf{M} is assumed to be uniform and its time evolution is determined by solving the LLG equation. For the generic i -th hexahedron belonging to the n -th object it results in

$$\frac{\partial \mathbf{M}_{i,n}}{\partial t} = -\frac{\gamma}{(1+\alpha^2)} \mathbf{M}_{i,n} \times \left[\mathbf{H}_{eff,i,n} + \frac{\alpha}{M_S} (\mathbf{M}_{i,n} \times \mathbf{H}_{eff,i,n}) \right], \quad 1 \leq i \leq T, 1 \leq n \leq N$$

¹Riccardo Ferrero, INRIM, 10135 Torino, Italy, r.ferrero@inrim.it; Politecnico di Torino, 10129 Torino, Italy.

²Alessandra Manzin, INRIM, 10135 Torino, Italy, a.manzin@inrim.it.

where M_S is the saturation magnetization, γ is the gyromagnetic ratio and α is the damping coefficient. The effective field \mathbf{H}_{eff} is the sum of applied field, magnetocrystalline anisotropy field, thermal field, exchange field and magnetostatic field, in turn decomposed into an internal and an external term. To handle objects with complex shape, the exchange field is calculated with a finite difference technique suitable for non-structured meshes [2]. The internal term of the magnetostatic field is obtained by solving the integral equation derived from Green's theorem application. To speed up the calculation, the external term, which describes the inter-object magnetostatic interactions, is determined by approximating the contribution from each hexahedron as the stray field of a magnetic dipole. Time integration is performed by means of a norm-conserving scheme based on the Cayley transform [3].

3 Results

The numerical model has been applied to calculate the hysteresis behaviour of a large number of stochastically distributed, magnetically interacting, permalloy nanodisks, with variable diameter (from 150 nm to 700 nm) and volume concentration (1-25 %). For certain parameters, the numerical results have been compared to experimental ones, obtained for nanodisks dispersed in an ethanol solution.

As an example, Figure 1(a) shows the hysteresis loop calculated for a set of 50 randomly oriented nanodisks with a diameter of 150 nm, a thickness of 25 nm and a volume concentration of 12%. Figure 1(b) reports the relative remanent magnetization configuration, characterized by out-of-plane magnetic vortex state.

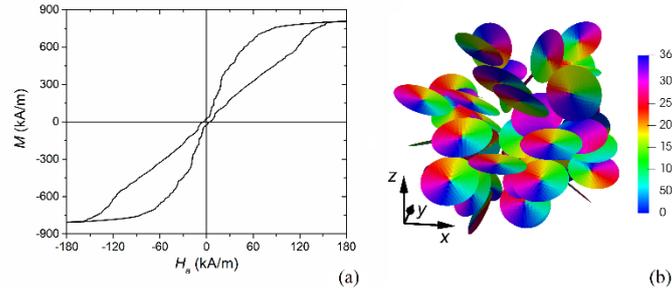


Figure 1: (a) Calculated hysteresis loop for a random distribution of 50 permalloy nanodisks with 150 nm diameter. The external field is applied along x -axis. (b) Magnetization configuration at remanent state. The colour bar represents the angle, in degrees, between magnetization component in xy -plane and x -axis.

References

1. Barrera, G. et al.: Magnetization switching in high-density magnetic nanodots by a fine-tune sputtering process on a large-area diblock copolymer mask. *Nanoscale* 9 (43), 16981-16992 (2017).
2. Bottauscio, O. and Manzin, A.: Spatial reconstruction of exchange field interactions with a Finite Difference scheme based on unstructured meshes, *IEEE Trans. Magn.* 48 (11), 3250-3253 (2012).
3. Manzin, A. and Bottauscio, O.: Connections between numerical behavior and physical parameters in the micromagnetic computation of static hysteresis loops, *J. Appl. Phys.* 108 (9), 093917 (2010).
4. Simeonidis, K. et al.: In-situ particles reorientation during magnetic hyperthermia application: Shape matters twice, *Sci. Rep.* 6, 38382 (2016).
5. Wang, Y. L. et al.: Rewritable artificial magnetic charge ice, *Science* 352 (6288), 962-966 (2016).

Modelling of magnetic nanobead transport in a microvascular network

Marta Vicentini¹, Riccardo Ferrero² and Alessandra Manzin³

Key words: Magnetic biotransport, Magnetic nanoparticles, Blood flow, Complex modelling, Simulated experiments, Computational methods.

1 Introduction

Magnetic nanoparticles can be advantageously used as contrast agents for MRI or mediators for hyperthermia treatment, as well as to label, deliver and separate biomaterials, also thanks to the capability to be magnetically manipulated, independently of microfluidic and biological processes [5]. In *in vivo* applications, one crucial aspect is the control of their transport in the tissue microvasculature and their successive release into target areas, mechanisms that can be driven by external magnetic field gradients [1]. *In silico* models of magnetic biotransport can support the interpretation of complex experimental results as well as the design of novel nanosystems for drug delivery and medical imaging.

In this context, we present a numerical model for the simulation of magnetic nano/microbead transport within a microvascular network, under the action of magnetic and viscous forces. The model, based on the coupling of the Navier-Stokes equations with classical Newtonian dynamics [3], is applied to investigate the role of various factors, like the spatial distribution of the external magnetic field, the viscous drag force and the inter-particle effects (e.g. the magnetostatic dipolar interactions). We also study the influence of bead size and magnetic moment on transport mechanism, tendency to aggregate and accumulation at vessel walls.

2 Physical Model

We have developed a complex numerical model that enables us to calculate the trajectory of an ensemble of magnetic nano/microbeads injected in a blood vessel and magnetically manipulated by an applied magnetic field. The physical phenomena included in the model are: the magnetic force generated by the external field; the viscous drag force due to the interaction with the blood flow; the magnetic dipolar interactions between nearby beads; the steric repulsive force due to the stabilizing effect of the surface coating layers [4]; the surface interactions with the vessel epithelium (collision and adhesion processes) [2]. The perturbation of the blood velocity profile due to the beads and the Brownian motion effects are neglected.

¹Marta Vicentini, INRIM, 10135 Torino, Italy, m.vicentini@inrim.it; Politecnico di Torino, 10129 Torino, Italy.

²Riccardo Ferrero, INRIM, 10135 Torino, Italy, r.ferrero@inrim.it; Politecnico di Torino, 10129 Torino, Italy.

³Alessandra Manzin, INRIM, 10135 Torino, Italy, a.manzin@inrim.it.

Collision and adhesion dynamics for nano/microbeads close to vessel walls are simulated with a probabilistic approach, by calculating an adhesion probability function that takes into account the size of the beads, their instantaneous orientation with respect to the vessel surface, the densities of ligands on bead surface and of receptors on the endothelial substrate [2].

The developed numerical model has been applied to study the influence of bead properties (size and magnetic moment) on transport, aggregation and adhesion processes, focusing on specific spherical nano/microbead with magnetization curve approximated with Langevin function.

3 Results

The developed numerical model has been applied to simulate magnetic nano/microbead transport in a microvascular network, under the assumption of beads with the same chemical and physical properties. Different arrangements for the external field sources have been considered to optimize the accumulation at vessel walls, process driven by magnetic force.

As an example, Figure 1 shows the results obtained for a set of 100 nanobeads with 300 nm size and 0.002 pAm^2 saturation magnetic moment, injected in a vessel with an average cross section radius of $50 \text{ }\mu\text{m}$. The magnetic field is produced by an external cylindrical magnet, in close proximity to the vessel. Specifically, Figure 1(a) illustrates the blood velocity profile in the vessel segment, obtained by solving the Navier-Stokes equations with finite element method, under the hypothesis of laminar blood flow. Figure 1(b) shows the trajectories of the beads after 30 ms from the injection at the vessel entry cross section.

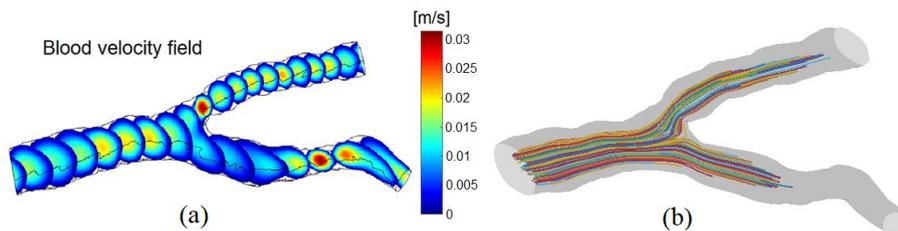


Figure 1: (a) Blood velocity field in a vessel segment extracted from a CT image of the abdomen. The vessel geometry is reconstructed by means of software VMTK (*The Vascular Modeling Toolkit*). (b) Bead trajectories in the vessel segment after 30 ms from the injection.

References

1. Barnsley, L. C. et al.: Understanding the dynamics of superparamagnetic particles under the influence of high field gradient arrays. *Phys. Med. Biol.* 62, 2333-2360 (2017).
2. Decuzzi, P. and Ferrari, M.: The adhesive strength of non-spherical particles mediated by specific interactions. *Biomaterials* 27, 5307-5314 (2006).
3. Furlani, E. P.: *Magnetic Biotransport: Analysis and Applications*. *Materials* 3, 2412-2446 (2010).
4. Satoh, A. and Chantrell, R. W.: Application of the dissipative particle dynamics method to magnetic colloidal dispersions. *Molecular Physics* 104, 3287-3302 (2006).
5. Wilhelm, S. et al.: Analysis of nanoparticle delivery to tumours. *Nature Rev. Mat.* 1, 16014 (2016).

Accuracy in Electrical Resistance Tomography: from measurements to maps

A. Cultrera^{1,*} and L. Callegaro¹

Key words: Electrical Resistance Tomography, Accuracy, Inverse Problems, Image Error

Electrical Resistance Tomography (ERT) is a technique which allows to map the electrical conductivity of the interior of an object by performing electrical measurements only at the object's boundary. The measurements consist in a series of four-terminal resistance measurements performed on a number of contacts. The gathered boundary information, along with numerical methods, allows to retrieve a map of the object's conductivity. From the mathematical point of view ERT is an ill-posed, inverse Laplace problem [Borcea(2002), Calderón(2006)].

ERT has been recently applied to conductive metal oxide thin films with encouraging quantitative results [Cultrera and Callegaro(2016)], and it is currently being extended to graphene characterisation by the same authors.

The assessment of the final ERT image accuracy is an open problem since it depends on a number of factors. One of these factors is the accuracy of the electrical measurements, which are among the input quantities of the problem. We used simulations to get insight of this problem. Figure 1a shows the ERT forward finite element model of a disc (which may represent a thin film sample), with a given conductivity distribution. The model includes 16 electrodes that allow to simulate $n = 208$ true valued transresistance measurements, following a typical ERT *adjacent* measurement pattern [Cultrera and Callegaro(2016)]. Data affected by measurement error, are simulated by adding to each of the n true valued measurements a normally distributed, uncorrelated, random deviation with variance ε set relative to the average of the n true valued measurements. ERT maps corresponding to both true valued and error-affected data are obtained solving the ERT problem with a suitable inverse model. Figure 1b shows a reconstructed image obtained from data affected by a certain amount of error. The relative reconstruction difference D , between the

*e-mail: a.cultrera@inrim.it

1) INRIM — Istituto Nazionale di Ricerca Metrologica, strada delle Cacce, 91, 10135 Torino, IT.

model and a reconstructed image, is defined as :

$$D = \frac{\sum_i (\sigma_{i,rec} - \sigma_i)^2}{\sum_i (\sigma_i)^2}, \quad (1)$$

where $\{\sigma_{i,rec}\}$ are the conductivity values of the elements of the reconstructed image and $\{\sigma_i\}$ the conductivity values of the model. From a series of simulations at increasing ϵ it is observed that D has a lower bound which does not depend on data error. At larger ϵ , D increases with the error. Note that in this example each of the n error-affected measurements is the result of a single extraction of the random deviation so there is only one D value (not a distribution) for a given error level ϵ . A more complete analysis is in progress and will be presented at the conference.

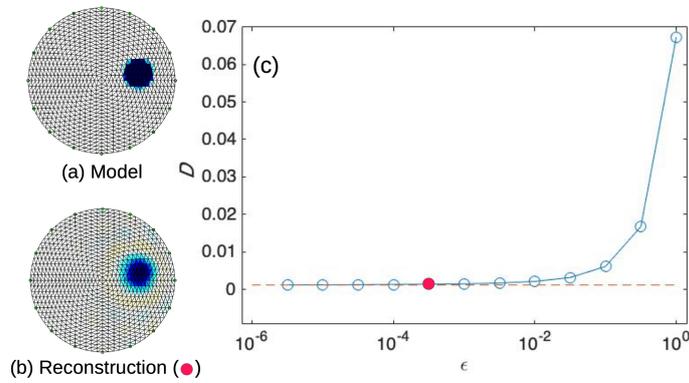


Fig. 1: In (a) the model of a circular sample with a low conductivity spot. In (b) a reconstructed image from data with error. The main plot (c) reports the image error corresponding to increasing error added to the input data. The dashed line marks D of the true value reconstruction. The red circle indicates the value of D corresponding to the reconstruction in (b).

This work has been realized within the Joint Resarch Project 16NRM01 GRACE: Developing electrical characterisation methods for future graphene electronics. This project has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

References

- [Borcea(2002)] Borcea L (2002) Electrical impedance tomography. *Inverse Prob* 18(6):R99–R136
- [Calderón(2006)] Calderón AP (2006) On an inverse boundary value problem. *Comp Appl Math* 25(2-3)
- [Cultrera and Callegaro(2016)] Cultrera A, Callegaro L (2016) Electrical Resistance Tomography of conductive thin films. *IEEE Trans Instrum Meas* 65(9):2101–2107

Two-point and multi-point interpolation of calibration data

Maurice Cox and John Greenwood

Key words: interpolation, calibration, measurement uncertainty, covariance

General

A generic treatment of two-point (straight-line) and multi-point interpolation of calibration data is presented. Uncertainties and covariance (in both variables) are propagated using the law of propagation (LPU) of uncertainty in JCGM 100 (GUM). The approach is applied to measurement of hydrogen ion activity (pH) and mass spectrometer leak detector (MSLD) calibration. There can be covariance associated with values of the stimulus variable because they might correspond to standards produced from a higher-level standard, or be sensitive to errors in common corrections.

Calibration is regarded as a two-step process in which the first stage determines calibration parameters from a calibration data set and the second stage uses those parameters to produce a stimulus (the measurand) corresponding to a response.

Two-point calibration is applied when there is a *bracket* on the measurand given by calibration data. Multi-point calibration is a simple extension of that treatment to an arbitrary number of points. For each interval between pairs of successive points, the single-point treatment can be directly applied. Straight-line or higher-order fitting by least squares is used for validation purposes.

The measurement model is the algorithm used to compute a value of the stimulus variable given the calibration data and a value of the response variable.

National Physical Laboratory
Teddington TW11 0LW, UK e-mail: maurice.cox@npl.co.uk

United Kingdom Accreditation Service
Staines-upon-Thames TW18 3HR, UK e-mail: John.Greenwood@ukas.com

Example 1: pH of a test solution

pH, the negative decadic logarithm (that is, to base 10) of the activity of hydrogen ion in a solution is the most measured kind-of-quantity in chemistry. The electric potential of a suitably constructed cell, for example, a glass electrode and reference electrode, is strongly related to pH and forms the basis of pH measurement.

An approach used to calculate the pH of a test solution is to measure the potential of the cell in the test solution and the solution temperature. Then use is made of certificates for standard solutions that 'bracket' the test solution, and bracketing temperatures from these certificates. The end product is an estimate of the pH of the test solution. Three steps of linear interpolation are involved in the process, which is an algorithmic statement of the measurement model. Uncertainties are propagated through the process to yield the standard uncertainty associated with the pH estimate.

For the data used, by comparison with higher-order interpolation, the error was judged negligible. Moreover, recent publications have indicated that a Monte Carlo method applied to primary pH measurement gave similar results to LPU.

Example 2: calibration of a mass spectrometer leak detector

A leak detector is a commonly used instrument for identifying and quantifying the rate of gaseous material leaving (or entering) an otherwise sealed system. At the heart of such an instrument is a detector that is selectively sensitive to a gas of interest. These detectors are based upon a variety of principles such as solid state chemical sensors and particle counters. A common leak detector is the helium MSLD. For highest accuracy, the calibration of these (and other) detectors is achieved by bracketing the range of interest using two reference leaks.

For the He MSLD a commonly used reference leak is the helium quartz-membrane leak with sealed reservoir. These reference leaks are stable when suitably handled (and stored); however, they have a temperature coefficient that significantly affects the leakage rate when in use and the rate of consumption during periods of storage.

When two bracketing leaks are used there is correlation between the nominal values of the two reference leaks and appreciable correlation between the temperature corrections applied to each reference leak. These correlations influence the uncertainty and covariance associated with the parameters of a linear fit to the MSLD calibration data and so affect the uncertainty associated with use of these parameters to assign a value to an unknown leak. It is demonstrated how this correlation, which is usually neglected, can be incorporated in an LPU type evaluation.

Acknowledgements The work reported is part of the European Metrology Programme for Innovation and Research (EMPIR) project 17NRM05 EMUE. The EMPIR initiative is co-funded by the European Unions Horizon 2020 research and innovation programme and the EMPIR Participating States. Brynn Hibbert provided valuable input to the pH example.

Getting started with uncertainty evaluation using the Monte Carlo method in R

Adriaan M.H. van der Veen and Maurice G. Cox

Key words: Measurement uncertainty, Monte Carlo, R, calibration, testing

Abstract

The evaluation of measurement uncertainty is often perceived by laboratory staff as complex and quite distant from daily practice. Nevertheless, standards describing competence of laboratories such as ISO/IEC 17025 [2] and ISO 15189 [1] require that measurement uncertainty is evaluated and as required reported. In response to this need, a European project entitled “Advancing measurement uncertainty – comprehensive examples for key international standards” started in July 2018 that aims at developing examples that contribute to a better understanding of what is required and aid in implementing such evaluations in calibration, testing and research.

The principle applied in the project is “learning by example”. Past experience with guidance documents such as EA 4/02 [5] and the Eurachem/CITAC guide on measurement uncertainty [6] has shown that practitioners find it often easier to re-work and adapt an existing example than to try to develop something from scratch. In this introductory paper, it is shown how the Monte Carlo method of GUM (Guide to the expression of Uncertainty in Measurement) Supplement 1 [4] can be implemented in R [8], an environment for mathematical and statistical computing. We do so by revisiting the well-known mass calibration example of EA 4/02 and explain how the Monte Carlo method can be used to evaluate the uncertainty, given the description of the input quantities and the assigned probability density functions. It is

Adriaan M.H. van der Veen
VSL, Thijsseweg 11, 2629 JA Delft, the Netherlands, e-mail: avdveen@vsl.nl

Maurice G. Cox
NPL Management Ltd, Hampton Road, Teddington, Middlesex, TW11 0LW, United Kingdom e-mail: maurice.cox@npl.co.uk

also shown how the law of propagation of uncertainty [3, 5] can be implemented in the same environment, taking advantage of the possibility to evaluate the partial derivatives numerically [7], so that these do not need to be derived by analytic differentiation.

The work reported is part of the European Metrology Programme for Innovation and Research (EMPIR) project 17NRM05 EMUE, in which a suite of examples is developed to support, e.g., calibration and testing laboratories, people concerned with conformity assessment and standards developers.

Acknowledgements The project 17NRM05 “Examples of Measurement Uncertainty Evaluation” leading to this application has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme.

References

- [1] ISO 15189 Medical laboratories – Requirements for quality and competence. ISO, International Organization for Standardization, Geneva, Switzerland, Third edition 2012.
- [2] ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories. ISO, International Organization for Standardization, Geneva, Switzerland, 2017. Third edition.
- [3] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008, GUM 1995 with minor corrections*. BIPM, 2008.
- [4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Supplement 1 to the ‘Guide to the Expression of Uncertainty in Measurement’ – Propagation of distributions using a Monte Carlo method, JCGM 101:2008*. BIPM, 2008.
- [5] EA Laboratory Committee. EA 4/02 Evaluation of the uncertainty of measurement in calibration. European Cooperation for Accreditation, September 2013.
- [6] S.L.R. Ellison and A. Williams. Quantifying uncertainty in analytical measurement. Eurachem/CITAC Guide QUAM:2012.P1, 2012. third edition.
- [7] Antonio Possolo. Five examples of assessment and expression of measurement uncertainty. *Applied Stochastic Models in Business and Industry*, 29(1):1–18, oct 2012. doi: 10.1002/asmb.1947. URL <https://doi.org/10.1002/asmb.1947>.
- [8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>.

Examples of measurement uncertainty evaluation applied to microflow, flow and thermal comfort

J. A. Sousa¹, A. S. Ribeiro² and M. G. Cox³

Key words: Uncertainty evaluation, method selection, validation, Monte Carlo method, Bayesian methods

Introduction

A main message in EMPIR project 17NMR05 EMUE – Examples of measurement uncertainty evaluation [1], is that the choice of method in the evaluation of measurement uncertainty should be based on the particular features of the problem in hand, and not on blind recipes. EMUE will contribute to the Joint Committee for Guides in Metrology (JCGM) documents related to measurement uncertainty (such as JCGM 100, JCGM 101, JCGM 102, JCGM 103, JCGM 106), particularly to JCGM 110 which will collect examples to support this view and help users to select the best method for their particular problem. To provide a flavour of that approach, the present study will focus on three case studies related to the fields of microflow, flow and thermal comfort, to exemplify measurement problems where the GUM approach is not satisfactory.

Microflow

Microflow applications are crucial in medical areas such as neonatology and oncology. However, in terms of uncertainty evaluation the very small levels of flow entail problems characterized by a “close to the physical limit” situation. The same happens in nanovolume applications. In these cases the GUM uncertainty framework may not be adequate, since the coverage interval will, in some cases, contain negative values, which is physically difficult to explain. Monte Carlo methods may

¹ J. A. Sousa
IPQ, Rua António Gião, 2, 2829-513 Caparica, Portugal, jasousa@ipq.pt

² A. S. Ribeiro
LNEC, Av do Brasil, 101, 1700-066 Lisboa, Portugal, asribeiro@lnecc.pt

³ M. G. Cox
NPL, Hampton Road, TW11 0LW Teddington, UK, maurice.cox@npl.co.uk

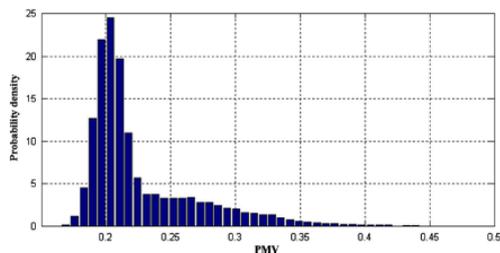
be a good alternative when there are not many negative readings in the application, but may otherwise prove inadequate, since the probability distribution for the measurand may have its mode at zero, which is not very credible to describe satisfactorily the real situation. In the latter circumstances the application of a Bayesian method will be the better approach to deal with the evaluation of measurement uncertainty.

Water supply networks

Water supply networks provide water to consumers, industries, facilities, services and other users. These infrastructures require the use of equipment able to measure many different quantities (flow, volume, level, velocity, pressure, etc.) to assess quality compliance conditions. Management decisions are increasingly supported by information provided by measurements. The lack of knowledge regarding reliable measurement data and associated uncertainty is an issue for the management of water supply networks. For many water providers uncertainty evaluation is difficult. However, the application of the GUM to the mathematical models involved provides simplified equations that can be used in specific circumstances. These more direct approaches apply to common tasks such as constant flow measurement, totalization of volume, and combination of branches of a network.

Thermal comfort

The authors previously carried out work that enabled measurement uncertainties associated with estimates of thermal comfort indices defined in ISO 7730:2005 to be obtained. It was found that the Monte Carlo method was particularly suitable to handle the complexity of the mathematical model that specifies the relation between



the quantities involved. Further work in this area is described that uses raw data from enquiries on thermal comfort, trying to overcome the limitation of the coarse resolution imposed by the standard ISO 7730 on the thermal sensation felt by a particular individual. Consideration is given

to the possibility of using a continuous scale, thus introducing comparability of the scale used in a possibly modified ISO 7730 and an enquiries-based scale.

References

1. EMPIR project 17NRM05 EMUE – Examples of measurement uncertainty evaluation. Program available on www.euramet.org and <http://empir.npl.co.uk/emue/> See also http://empir.npl.co.uk/emue/wp-content/uploads/sites/49/2018/08/17NRM05_Publishable_Summary.pdf

Uncertainty Estimation in Multi-Channel Data-Acquisition Systems

A. Carullo¹, S. Corbellini² and A. Vallan³

Keywords: electrical quantities, data-acquisition systems, uncertainty, correlation

1 Introduction

This paper deals with the uncertainty analysis of multi-channel data-acquisition systems, which are commonly used in the industrial field for the measurement of electrical and non-electrical quantities. According to a modern approach, the result of a measurement is assigned as a coverage interval, i.e. an interval of values that includes the measurand with a stated probability (the coverage probability), which can be estimated using the GUM uncertainty framework [1] or analytical methods [3-4]. Alternatively, numerical methods can be implemented [2], [5-6].

2 The system under analysis

In a multi-channel data-acquisition system, some of the components of the measuring chain are shared among the different channels. This solution, which is based on the presence of a multiplexer (MUX), is very common in industrial applications since it allows a low-cost system to be arranged. However, particular attention has to be paid in the identification of the uncertainty contributions and in the estimation of the correlation among the measured quantities.

¹ Alessio Carullo
ACCREDIA-DT Torino (Italy) - Politecnico di Torino, C. Duca degli Abruzzi, 24 – 10129 Torino (Italy)
e-mail: alessio.carullo@polito.it

² Simone Corbellini
Politecnico di Torino, C. Duca degli Abruzzi, 24 – 10129 Torino (Italy)

³ Alberto Vallan
Politecnico di Torino, C. Duca degli Abruzzi, 24 – 10129 Torino (Italy)

The presence of the MUX introduces uncertainty contributions that are due to the limited insulation between the MUX input channels (*cross-talk*) and to the residual effect of the transient that is fired by the channel switching when the correct settling-time is not complied with. Furthermore, a strong dependence between the measurement of the different channels has to be expected, since the conditioning circuitry (e.g. a differential amplifier) and the Analog to Digital Converter (ADC) are the same for the different channels and are characterized by the same systematic effects. This dependence can be estimated through a Type-B evaluation of the correlation coefficient. The MUX also allows quasi-simultaneous readings to be obtained for the different channels, hence a Type-A evaluation of correlation coefficients r_{ij} between each couple of average estimation \bar{x}_i and \bar{x}_j can be implemented according to the method proposed in [1]. The estimated correlation coefficients r_{ij} are affected by random effects, such as the quantization noise of the ADC, the electronic noise due to the components of the measuring chain and the variation of the influence quantities. When the combined standard uncertainty is evaluated, both Type-A and Type-B correlation coefficients have to be taken into account, also considering that random effects can be minimized if a suitable number of readings in repeatability conditions can be obtained for the measured quantities.

A case study of multi-channel data acquisition system will be described and the uncertainty analysis will be performed highlighting the contributions of each component of the measuring chain.

3 Acknowledgement

The project 17NRM05 “Examples of Measurement Uncertainty Evaluation” leading to this application has received funding from the EMPIR programme co-financed by the Participating States and from the European Union’s Horizon 2020 research and innovation programme.

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. JCGM 100:2008.
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. JCGM 101:2008.
3. Cox M.G., and Harris P.M. - *Best Practice Guide No. 6, Uncertainty evaluation*. Tech. Rep. DEM-ES-011, National Physical Laboratory, Teddington, UK, 2006.
4. Dietrich C.F. - *Uncertainty, Calibration and Probability*. Adam Hilger, Bristol, UK, 1991.
5. Ferber M., Vollaire C., Krähenbühl L., and Vasconcelos J.A. - *Adaptive unscented transform for uncertainty quantification in emc large-scale systems* - COMPEL - The international journal for computation and mathematics in electrical and electronic engineering, 33(3):914–926, 2014.
6. Trincherio R., Manfredi P., Ding T., and Stievano I.S. - *Combined Parametric and Worst-Case Circuit Analysis via Taylor Models*. IEEE Transactions on Circuits and Systems - I: regular papers, Vol. 63, No. 7, pp. 1067–1078, Jul. 2016. DOI: 10.1109/TCSI.2016.2546389

The posterior mean-power of normal variables

Giovanni Mana and Carlo Palmisano

Key words: Probability theory, Statistics, Inference methods, Data analysis

1 Introduction

The posterior probabilities of competing models that differ only by their parametrisation must be the same. This requires priors (named after H. Jeffreys) proportional to the volume-element of the model manifold, where the parameters are the manifold coordinates and whose metric is the Fisher information [Jeffreys(1946)].

However, in the Gaussian case, the uniform Jeffreys' distribution of the mean originates a paradox. It is an inconsistency (named after C. Stein) that occurs when calculating the posterior expectation of the mean power of a large number of variables from single observations of each one [Stein(1956)]. We show that hierarchical modelling and model averaging remove the inconsistency and that, as the number of variables increases, the posterior mean converges to the frequentist estimate.

2 Bayesian model comparison

Let a complete set of mutually exclusive models compete to explain the dataset \mathbf{x} . The posterior probability of the l -th model is

$$P(l|\mathbf{x}) = \frac{Z(\mathbf{x}|l)\Pi(l)}{\sum_m Z(\mathbf{x}|m)\Pi(m)}. \quad (1)$$

where $\Pi(l)$ is the prior probability of the l -model being true,

$$Z(\mathbf{x}|l) = \int_{\mu' \text{space}} L(\mathbf{x}|\mu, l)\pi(\mu|l) d\mu, \quad (2)$$

is the data distribution given the model, and $L(\mathbf{x}|\mu, l)$ is the data distribution given the model and μ parameters. Since, in the absence of prior information, $P(l|\mathbf{x})$ must not depend on the model parametrisation, $\pi(\mu|l)$ must be a Jeffreys' prior.

Giovanni Mana

INRIM – Istituto Nazionale di Ricerca Metrologica, Torino, Italy, e-mail: g.mana@inrim.it

Carlo Palmisano

UNITO – Università di Torino, Dipartimento di Fisica, Torino, Italy

DMA, Torino, Italy, e-mail: palmisan@to.infn.it

3 Stein paradox

Let $x_i \sim N(x_i|\mu_i, \sigma)$ be m realizations of independent normal measurands μ_i having the same, known, variance. The Jeffreys' prior of each μ_i is uniform, resulting in independent normal posteriors, $N(\mu_i|x_i, \sigma)$, of each individual mean.

Suppose we are interested in the mean power, $\theta^2 = |\mu|^2/m$. As m tends to the infinity, a bad situation occurs: the Bayesian analysis predicts that θ^2 is certainly the posterior mean $(\sum x_i^2/m) + 1$, but the frequentist analysis predicts that θ^2 is certainly the maximum likelihood estimate $(\sum x_i^2/m) - 1$ [Giaquinto and Fabbiano(2016)].

4 Proposed solution

The paradox is due to the fact that a uniform variable (i.e., a constant over the whole real line) is certainly larger than any positive value or smaller than any negative one. This is noninformative with respect to the mean, but informative on its square and introduces a bias towards a large value.

We encoded the knowledge of the finitude of the measurands into identical and independent normal priors – hyper-parameterized by the same, unknown, mean and variance – and averaged over this family. In this way, without the use of empirical methods [Efron and Morris(1973)], the posterior mean of a single observation is the maximum likelihood estimate, which is the observation itself.

The posterior expectations of each measurand is between the sample mean and the measured value, shrinking towards one or the other depending on whether the data support the same-measurand or different-measurands hypotheses.

If the data support the different-measurands hypothesis, the model average of the posterior mean-power converges to the frequentist estimate. Contrary, if the data support the same-measurand hypothesis, the model average converges to the power of the sample mean. This result differs from the frequentist estimate, which is blind to this information.

References

- [Efron and Morris(1973)] Efron B, Morris C (1973) Stein's estimation rule and its competitors – an empirical bayes approach. *Journal of the American Statistical Association* 68(341):117–130
- [Giaquinto and Fabbiano(2016)] Giaquinto N, Fabbiano L (2016) Examples of s1 coverage intervals with very good and very bad long-run success rate. *Metrologia* 53(2):S65–S73
- [Jeffreys(1946)] Jeffreys H (1946) An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 186:453–461
- [Stein(1956)] Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, University of California Press, Berkeley, CA, pp 197–206

Bayesian Inference on the Parameters of the Truncated Normal Distribution and Application to Reverberation Chamber Measurement Data

Carlo Carobbi¹ and Ramiro Serra²

Key words: Bayesian inference, truncated normal probability density function, reverberation chamber.

1 Introduction and Statement of the Problem

The in-phase (\parallel) and the quadrature (\perp) components of the field generated inside a lossless and overmoded electromagnetic reverberation chamber (RC) in a particular Cartesian direction follow a normal distribution with zero-mean ($\mu_{\parallel} = \mu_{\perp} = 0$) and equal variances ($\sigma_{\parallel}^2 = \sigma_{\perp}^2$) [1], [2]. The estimation of these two key parameters (mean and variance) is an essential task in order to characterize the statistical properties of the field inside a RC. Physical insight suggests that, due to losses, fields cannot take any arbitrarily large value inside a RC. Hence a more realistic probabilistic model for the field is a truncated normal distribution.

Inference on the parameters of the truncated normal distribution is here carried out on the basis of a set of n observations $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ of the field in a RC and in absence of any prior information. In order to check the validity of the truncated normal probability model, use is made of the two-sample Kolmogorov-Smirnov test by comparing the set of n observations with a set of the same size $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2, \dots, \tilde{\xi}_n)$ generated by using the predictive posterior probability density function (PDF).

¹ Carlo Carobbi

Department of Information Engineering, Università degli Studi di Firenze, e-mail: carlo.carobbi@unifi.it

² Ramiro Serra

Laboratory of Electromagnetic Compatibility, Eindhoven University of Technology Eindhoven, the Netherlands, email: r.serra@tue.nl

2 Mathematical Formulation

We apply Bayesian analysis assuming that the parent distribution of the random variable ξ is a truncated normal distribution. The corresponding PDF is therefore given by $f_{TN}(\xi; \mu, \sigma, a, b)$, where

$$f_{TN}(\xi; \mu, \sigma, a, b) = \frac{f_N(\xi; \mu, \sigma)}{F_N(b; \mu, \sigma) - F_N(a; \mu, \sigma)},$$

$F_N(\xi; \mu, \sigma)$ is the normal cumulative distribution function with mean μ and standard deviation σ , $f_N(\xi; \mu, \sigma)$ is the normal probability density function. a and b are the lower and upper truncation limits, respectively. Note that $f_{TN}(\xi; \mu, \sigma, a, b) = 0$ for $\xi < a$ and $\xi > b$. For the specific physical problem at hand it is further assumed that $\mu > a$ and $\mu < b$ because it is expected that losses do not completely distort the normal distribution which applies to the ideal lossless case.

A vague and improper prior is assigned to the parameters, specifically

$$f_0(\mu, \sigma, a, b) \propto \frac{1}{\sigma(\mu - a)(b - \mu)}.$$

By using Bayes theorem the joint posterior PDF of the parameters is then obtained as

$$f(\mu, \sigma, a, b | \xi) \propto l(\xi; \mu, \sigma, a, b) f_0(\mu, \sigma, a, b),$$

where

$$l(\xi; \mu, \sigma, a, b) = \prod_{i=1}^n f_{TN}(\xi_i; \mu, \sigma, a, b).$$

The predictive posterior PDF is given by

$$f_P(\tilde{\xi} | \xi) = \int_{\mathbf{0}} f_{TN}(\tilde{\xi}; \mathbf{\theta}) f(\mathbf{\theta} | \xi) d\mathbf{\theta},$$

where $\mathbf{\theta} = (\mu, \sigma, a, b)$. Through this formulation confidence is achieved about the appropriateness of the truncated normal distribution particularly under significant loading of the RC, i.e. when losses play a substantial role in the (statistical) field dynamics.

References

1. Hill D.: Electromagnetic Fields in Cavities: Deterministic and Statistical Theories. IEEE, Piscataway (2009).
2. Kostas J. and Boverie B.: Statistical Model for a Mode-Stirred Chamber. IEEE Transactions on Electromagnetic Compatibility, vol. 33, no. 4, pp. 366–370 (1991).

An Optimal Bayesian Design for a reliability study

Rossella Berni

Key words: Design of Experiments, MCMC simulations, Soldering

1 Introduction

Optimal design criteria has recently received growing attention, both at theoretical and computational levels, in part following the increase of computational power. Since the 70s, there is a long history of seminal papers in literature on D and T-optimality, both to estimate model parameters and also to discriminate among models. Furthermore, the building of optimal designs has been improved in a Bayesian framework, by introducing prior distributions on models and parameters and by selecting the optimal design according to the definition of an utility function and its maximization, also by considering a decision analysis framework. Chaloner and Verdinelli (1995) is a well-known and extensive review, which helps to define Bayesian optimal designs through the definition of a specific utility function, depending on the specific purpose of the experiment, by also considering that a decision problem must select the design that maximizes the expected utility. Notwithstanding the generality achieved, in actual applications further flexibility is often needed, for example by defining a utility function in which the cost of each observation depends on the value taken by the independent variable. Moreover, the relevance for costs may be also evaluated by specific weights, which take environmental conditions and technological information into account. Following a decision problem, in Müller et al (2004) the main idea is the definition of an auxiliary distribution $h_J(\mathbf{d})$, proportional to the utility function (nonnegative and bounded) $U(\mathbf{d})^J$, which allows us to simulate from a sequence of augmented probability models.

Rossella Berni
Department of Statistics Computer Science Applications "G.Parenti", University of Florence, e-mail: rossella.berni@unifi.it

2 Bayesian Optimal designs and reliability: our proposal

Three main elements characterized a Bayesian optimal design:

- the definition of decision problem \mathbf{d} relating to a prediction problem in a decision-theoretic model; where the decision rule depends on Y and ξ . We suppose that $Y \in P(\cdot|\theta, \xi; \theta \in \Theta, \xi \in \Xi)$, where Ξ is the design set.
- the definition of a Loss function defined on $D \times \Theta \times \Xi$ (Eaton et al, 1996) or an Utility function, Chaloner and Verdinelli (1995).
- a prior distribution $\pi = p(\theta)$ or a probability model.

We may consider an objective function defined on the basis of a generic utility function $u(d, \theta, y, \xi)$, with probability model $p_d(\theta, y) : p(\theta)p_d(y|\theta)$; $p(\theta)$ is invariant w.r.t decision \mathbf{d} . In general, the optimal decision problem could be defined as in the following:

$$d^* = \arg \max_{d \in D} U(d) \quad (1)$$

with utility defined as:

$$U(\xi) = \int_D \max_{d \in D} \int_{\Theta} U(d, \theta, \xi, y) p(\theta|y, \xi) p(y|\xi) d\theta dy \quad (2)$$

The case study is relating to a soldering process: a comparison of alloy-surface finish combinations considering different component packaging, e.g. four types of components and pin size. Reliability measurements are related to the time to failure for the electrical resistance [Ω] considering a thermal stress cycle. An hierarchical Bayesian model is applied joint with MCMC simulations. Following two utility functions are defined in order to build two optimal Bayesian designs, which also involve costs related to the finishes, component types, and pin. Environmental issues for each unit, by giving a weight (normalized) to each specific cost, are considered.

References

- Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. *Statistical Science* 10, 273-304
- Eaton M, Giovagnoli A, Sebastiani P (1996) A predictive approach to the bayesian design problem with application to normal regression models. *Biometrika* 83, 111-125.
- Müller P, Sansò B, De Iorio M (2004) Optimal bayesian design by inhomogeneous markov chain simulation. *Journal of the American Statistical Association* 99, 788-798.