MATHMET 2016

International Workshop on

„Mathematics and Statistics for Metrology“

295. PTB-Seminar

07 – 09 November 2016, PTB, Berlin, Germany

BOOK OF ABSTRACTS

Organizers

Markus Bär (PTB/EURAMET, Germany), Clemens Elster (PTB, Germany), and Antonio Possolo (NIST/SIM, USA)

Scientific Committee

Markus Bär (PTB/EURAMET, Germany), Maurice Cox (NPL, UK), Kensei Ehara (NMIJ, Japan), Stephen Ellison (LGC, UK), Clemens Elster (PTB, Germany), Nicolas Fischer (LNE, France), Alistair Forbes (NPL, UK), Gregory Kyriazis (INMETRO, Brazil), Juris Meija (NRC, Canada), Tony O' Hagan (University of Sheffield, UK), Leslie Pendrill (SP, Sweden), Francesca Pennecchi (INRIM, Italy), Antonio Possolo (NIST/SIM, USA)

Invited Speakers

Olha Bodnar (PTB, Germany), Nicolas Fischer (LNE, France), Alistair Forbes (NPL, UK), Tony O' Hagan (University of Sheffield, UK), Sebastian Heidenreich (PTB, Germany), Leslie Pendrill (SP, Sweden), Adrian Sandu (VirginiaTech/NIST, USA), Blaza Toman (NIST, USA) Adriaan van der Veen (VSL, the Netherlands)

Content

Quantification of the uncertainty due to the geometrical influence factors for dimensional measurements using computed tomography

Evelina Ametova[a], Massimiliano Ferrucci[a,b], Suren Chilingaryan[c], Wim Dewulf[a]

[a] *Department of Mechanical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium*
[b] *Engineering Measurement Division, National Physical Laboratory, Teddington, United Kingdom*
[c] *Institute of Data Processing and Electronics, Karlsruhe Institute of Technology, Karlsruhe, Germany*

e-mail: evelina.ametova@kuleuven.be

X-ray Computed Tomography (CT) is a non-destructive imaging technique that employs penetrating electromagnetic radiation to visualize, analyse and quantify both internal and external structures of a test object. In recent years, increased focus on the application of CT for dimensional quality control has motivated research on acceptance of CT scanners as metrological instruments. A main driving factor is the application of CT for non-destructive inspection of parts with complex inner geometries in relatively short time. Tomographic reconstruction generates a 3D volumetric representation of the measured object from a series of 2D projections acquired at different perspectives. Edge-based segmentation is used to define transitions between materials, which results in a surface model of the object. Surface sampling is then used to extract a point cloud representation of the surfaces, which can be used to perform coordinate measurements.

One of the more common tomographic reconstruction algorithms for cone-beam CT is filtered back-projection (FBP), which maps 2D projections onto 3D space. The accuracy of the reconstructed volume by FBP relies on the accurate knowledge of the geometrical CT acquisition parameters, *i.e.* relative position and orientation of X-ray source, rotation axis and detector. Calibration procedures provide users with estimates of the CT geometry. However, uncertainties in the estimated geometrical parameters propagate to uncertainties in the reconstructed volume and, consequently, to dimensional measurements performed on the volumetric data. Quantification and propagation of such sources of uncertainty must be considered to provide a complete uncertainty analysis and to establish confidence intervals for dimensional measurements obtained from volumetric data. Due to the complex dependence between geometrical parameters and dimensional measurements, analytical expressions linking geometrical parameters to dimensional measurements are currently not available; therefore, the conventional GUM method cannot be applied. The use of Monte Carlo (MC) computer simulation is a valuable alternative to the GUM method. However, the computational load associated with CT metrology chain does not allow a direct application of MC simulation. The required computational resources and time would be prohibitive once all image processing steps are included in the stochastic simulation.

In this study, we propose a simplified CT model for quantification of uncertainty in dimensional measurements due to the uncertainty in the parameters of the data acquisition geometry. Since dimensional measurements are performed on the point cloud corresponding to segmented surface, *i.e.* in coordinate space, we propose to omit the computationally expensive reconstruction of the full CT volume and quantify uncertainty directly in the coordinate space. During the CT reconstruction, each ray recorded by the detector is traced back to the measurement volume. If CT acquisition and reconstruction parameters are consistent, all the rays from different projections corresponding to one point in the measurement volume will intersect in this point. In the presence of deviation of parameters, the point will be back-projected as volumetric distribution of points. This distribution is used to determine a three-dimensional confidence interval for each point in the volume. The efficacy of the proposed approach was confirmed on simulated data in the presence of individual uncertainty contributors.

Magnetic Resonance-based EPT: Noise Propagation in an Ill-Posed Problem

Alessandro Arduino[1,2], Oriano Bottauscio[2], Mario Chiampi[1], Francesca Pennecchi[2], Luca Zilberti[2]

[1]Dipartimento Energia – Politecnico di Torino – Corso Duca degli Abruzzi, 24 – 10129 Torino, Italia

[2]Istituto Nazionale di Ricerca Metrologica (INRIM) – Strada delle Cacce, 91 – 10135 Torino, Italia

Abstract

Since its first proposal in 1991 [1], many attempts have been done to recover the electric properties tomography (EPT) of biological tissues from the measurements of a component of the radiofrequency magnetic field during magnetic resonance imaging [2-3]. The useful additional quantitative information to standard imaging results that would be available with a successful EPT implementation motivates its development. The most direct application of EPT is in clinical oncology, since some kinds of cancerous aggregates show a higher electric conductivity with respect to healthy tissues. In addition, dosimetry would also take advantage of this technology, for example in hyperthermia treatment planning, where the estimate of the specific absorption rate would be specialized for the patient, and in ultrahigh field magnetic resonance imaging safety, in which a real time estimate of the specific absorption rate would localize possible hotspots and ensure patients' safety.

Traditionally, EPT is performed through a local inversion of the electromagnetic equations. Nevertheless, in the last years some promising methods, which deal with the EPT inverse problem by using the optimal control theory, have been proposed [4-6]. Differently from local methods, the latter approaches avoid the estimation of the first and second order spatial derivatives of the measured field, which is a critical operation in presence of noise. Amongst the optimal control based methods, the specialization of the contrast source inversion (CSI) technique to electric properties tomography is of particular interest, because, in addition to the dielectric properties, it estimates the electric field induced by the measured radiofrequency magnetic field, becoming a candidate for the real time specific absorption rate estimation.

Since EPT is an ill-posed problem, regularization techniques have to be taken into account when working in the optimal control framework. For the implementation of the CSI method, a naïve regularization consists in stopping the procedure at a certain iteration, in order to keep the noise contribution to the recovery error lower than the error due to the numerical approximation. More sophisticated multiplicative regularization strategies, based on the bounding of the total variation of the electric properties distribution, can be employed. In any case, a metrological analysis on the propagation of the noise from the input measurements to the output data, *i.e.* the estimates of the electric properties distribution and the induced electric field, is required in order to make the method reliable for *in vivo* applications.

In the talk, the application of the CSI method to the magnetic resonance-based EPT and its regularization strategies will be described. Then, the talk will focus on the study of noise propagation in a two dimensional model problem where virtual measurements, obtained by the numerical resolution of a direct electromagnetic problem on a realistic domain, are used as input data to the CSI method. Estimates of the corresponding reconstructed electric properties distribution will be compared to the original values, in order to assess the reliability of the method.

References

[1] Haacke EM, *et al.* "Extraction of conductivity and permittivity using magnetic resonance imaging." Phys in Med and Biol 1991;36:723-34

[2] Katscher U, *et al.* "Recent Progress and Future Challenges in MR Electric Properties Tomography." Comp and Math Meth in Med 2013;2013:546-62.

[3] Zhang X, *et al.* "Magnetic-Resonance-Based Electrical Properties Tomography: A Review." IEEE Rev in Biomed Eng 2014;7:87-96.

[4] Ammari H, *et al.* "Magnetic resonance-based reconstruction method of conductivity and permittivity distributions at the Larmor frequency." Inverse Problems 2015;31:105001.

[5] Balidemaj E, *et al.* "CSI-EPT: A Constrast Source Inversion Approach for Improved MRI-Based Electric Properties Tomography." IEEE Trans on Med Imaging 2015;34:1788-96.

[6] Borsic A, *et al.* "An Inverse Problems Approach to MR-EPT Image Reconstruction." IEEE Trans on Med Imaging 2016;35:244-56.

MATHMET - the European Centre for Mathematics and Statistics in Metrology

M. Bär (Chair, MATHMET)

The European Centre MATHMET has been established considering the need for collaborative research on mathematical and statistical methods for metrology, since modern metrology requires expert knowledge of and relates on further advances in applied mathematics, statistics, numerical computation and related state-of-the-art computational tools. Therefore, the aim of MATHMET is to undertake collaborative research, publish guidelines, case studies and best-practice guides and to provide a website as a unique reference point of information. The talk will provide an overview of MATHMET´s goals and activities.

MATHMET emerged from a collaborative research project funded by the European Metrology research program (EMRP) of EURAMET, the association of European metrology research institutes. The founding members of MATHMET are four European national metrology institutes (NMIs) from France (LNE), the United Kingdom (NPL), Germany (PTB) and Sweden (SP). MATHMET membership is open to further European NMIs and other interested institutions and organisations.

MATHMET activities are arranged around selected research areas covering many aspects of modelling and simulation and statistical data analysis in metrology: uncertainty evaluation, key comparisons, regression and inverse problems, uncertainty quantification in complex models, conformity assessment and dynamic measurements. Joint work will be carried out in dedicated MATHMET projects funded by the member institutes and through externally funded research projects.

For more information visit the official MATHMET homepage: http://mathmet.org

Bayesian inference for measurements of ionizing radiation under partial information

Olha Bodnar, Rolf Behrens and Clemens Elster

Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin

Decision threshold and detection limit (i.e., minimum detectable activity) are perhaps the most important characteristics in radiation measurements, especially in radioactivity measurements. The calculation of these two characteristics is usually based on the seminal paper of Currie (1968) who applied test theory of conventional statistics. This methodology is widely used today and provides the basis of the international standard ISO 11929 (ISO (2010)). The physical models involved in radiation measurements contain additional, not exactly known parameters such as the detection efficiency of the employed measuring system. Bayesian inference is well suited to take into account available prior knowledge about these parameters. Bayesian approaches utilizing the posterior predictive distribution have already been suggested in the physical literature (e.g., Kirkpatrick et al. (2015), Michel (2016)). We propose a Bayesian treatment which is based on the posterior for the background effect, obtained in a preceding background measurement, and the posterior for the gross effect. The proposed treatment leads to an alternative way of calculating the decision threshold. The employed prior distribution utilizes knowledge from previous calibration measurements as well as the posterior distribution for the background effect. The proposed approach is compared with present methods in terms of results for simulated and real radiation activity measurements.

References

Currie, L. A. (1968). Limits for qualitative detection and quantitative determination. application to radiochemistry. Analytical Chemistry, 40:586–593.

ISO (2010). ISO 11929, Determination of the characteristic limits (decision threshold, detection limit and limits of the confidence interval) for measurements of ionizing radiation – Fundamentals and application. Geneva: ISO.

Kirkpatrick, J., Russ, W., Venkataraman, R., and Young, B. (2015). Calculation of the detection limit in radiation measurements with systematic uncertainties. Nuclear Instruments and Methods in Physics Research A, 784:306–310.

Michel, R. (2016). Measuring, estimating, and deciding under uncertainty. Applied Radiation and Isotopes, 109:6–11.

INFLUENCE OF AN OUTLIER ON CONFIRMATION OF THE UNCERTAINTIES DECLARED BY KC PARTICIPANTS

A. Chunovkina, A. Stepanov

The D.I. Mendeleyev Institute for Metrology (VNIIM), Russia

A.G.Chunovkina@vniim.ru

The paper deals with evaluation of key comparison (KC) data in the presence of an outlier. Quite often the outlier is excluded from the data in order to avoid its influence on KCRV and, finally, to a statement of degree of equivalence for all KC participants. But on the one hand the practice of outlier exclusion is questionable and on the other hand the outliers are not always detected by the procedures used with a high probability. We should take into account that KC results are the basis for confirmation of measurement uncertainties declared by participants. The aim of the paper is to show on the basis of a quantitative analysis how the outlier influences decisions relating to a confirmation of its own measurement uncertainties as well as to confirmations of measurement uncertainties reported by other laboratories.

Conventional approach for KC data evaluation is considered. Key comparison reference value (KCRV) is calculated as a weighted mean and criterion E_n is used for confirmation of measurement uncertainties declared by the participants. Influence of the outlier is quantitatively expressed by following probabilities. $P_1 = P\{E_{n,i} \le 2, for\ all\ 1 \le i \le N\}$ – is the probability at which all the participants including the outlier confirm the reported measurement uncertainty, $P_2 = P\{E_{n,i} \le 2, for\ all\ 1 \le i \le N\ i \ne i_0\}$ – is the probability at which all the participants except for the outlier confirm the reported measurement uncertainty and $P_3 = P\{E_{n,i} > 2, for\ i = i_0\}$ – is the probability at which the outlier is detected as an outlier and consequently the associated uncertainty isn't confirmed. The above probabilities are calculated and analyzed as functions of the outlier's weight, ω_{i_0}, and its bias, λu_{i_0}, which is taken proportional to the associated measurement uncertainty u_{i_0}. Normal distributions as well as two-side power (TSP) distribution family are considered for the measurement data modeling. As an example the data from the report of CMC-P-K.12 are considered.

The analysis of the results obtained shows that the outlier with small weight does not significantly influence probability P_2 within a wide range of λ values. The analysis of the probability P_3 for different weights shows that for $\lambda > 4$ the outlier is recognized by the E_n criterion with a high probability for all the distributions considered. It should be stressed that for $\lambda < 4$ the P_2 probability values obtained for normal and TSP distributions differ greatly. Moreover, even for comparably small values of λ a decrease of probability P_2 is significant in case of the outlier with a heavy weight. The results of analysis can be used for some recommendations concerning reasonable restrictions of the range of weights of measurement data used for KCRV evaluation.

The solution of large structured non-linear least squares problems with an application to earth observation

Maurice Cox Arta Dilo
Peter Harris
National Physical Laboratory, Teddington, UK

In earth observation where it is required to obtain knowledge of various climate variables, a sensor is calibrated given observations of spectral radiance provided by a reference sensor. For the first sensor observations are available of various quantities such as radiance, counts and thermodynamic temperature that can be converted to spectral radiances using a calibration function. The total number of observations available may be $O(10^7)$, whereas the number of calibration parameters is very small, 3 or 4, say.

The problem of calibrating the first sensor can be expressed as a generalized distance problem, also known as an errors-in-variables problem. In this formulation, the data consist of an observation vector $\boldsymbol{r}$ provided by the reference sensor and several observation vectors $\boldsymbol{x}^{\text{data}}$, $\boldsymbol{y}^{\text{data}}$, etc. related to the sensor to be calibrated. Unknown quantities to be estimated are calibration parameters $\boldsymbol{a}$ and vectors $\boldsymbol{x}$, $\boldsymbol{y}$, etc., for which $\boldsymbol{x}^{\text{data}}$, $\boldsymbol{y}^{\text{data}}$, etc. are realizations. Stated standard uncertainties and covariances are to be taken into consideration. The calibration model has the form $r = f(\boldsymbol{a}; x, y, \ldots)$. The formulation as a non-linear least squares (NLS) problem involves a design matrix of dimension $m \times n$ and a covariance matrix of dimension $m \times m$, where, for instance, $m \approx 2 \times 10^7$ and $n \approx 1.5 \times 10^7$.

In the absence of covariances, when applying an NLS algorithm the user typically specifies a function that returns the vector $\boldsymbol{f}$ of weighted residuals, the sum of squares of which form the objective function, and the corresponding Jacobian matrix $\boldsymbol{J}$ given values for $\boldsymbol{a}$, $\boldsymbol{x}$, $\boldsymbol{y}$, ... In the presence of covariances, a transformation using the Cholesky factorization of the covariance matrix is used to convert the resulting objective function into a 'sum of squares'.

A solution method is considered for the above problem. Advantage is taken of the structure of the design and covariance matrices to ensure a computationally viable solution algorithm, and to permit the required covariance matrix for $\boldsymbol{a}$ to be determined by uncertainty propagation. Some thoughts are given on the extension to a more difficult class of problems where a set of sensors are to be harmonized in terms of spectral radiance.

3D Finite Element Modelling for Transient Behaviour Study in Electrical Energy Metrology

Marija Cundeva-Blajer
Ss. Cyril and Methodius University in Skopje,
Faculty of Electrical Engineering and Information Technologies-Skopje, R. Macedonia

The instrument transformers, beside the electricity meters, represent the most applied devices in electrical energy metrology and are included in the legal metrology infrastructure in the area of power and energy. Their metrological characteristics in steady-state as well as in transient regimes are specified by the international standards [1], so for their proper design and further measurement use, these characteristics must be known with low uncertainty. The transient behaviour of the instrument transformers is also of great significance for electrical protection purposes.

In this contribution a complex transient analysis of a 20 kV combined current-voltage instrument transformer will be presented. The transient mathematical modelling is conducted for different designs of the combined instrument transformer. The designs are derived from previous genetic algorithm stochastic optimization process. The objective in the optimization process is the minimisation of the metrological errors (voltage, current and phase displacement errors) of the instrument transformer in steady-state regime. This is a multivariable objective function with constraints defined by the international standards, [1], in steady state regime as well as in transient regime. The transient study is for the purposes of further improved metrological design, realization and usage of the combined instrument transformer novel prototype.

The combined instrument transformer is a very complex non-linear electromagnetic system with two magnetic cores (voltage and current measurement core) and four windings (high and low voltage windings, and high and low current windings) in one chassis. The electromagnetic phenomenon in this compound system is described by the Maxwell equations system, [2], which after introduction of the auxiliary quantity-magnetic vector potential, is transformed into a system of non-linear partial differential equations with variable coefficients. The transient analysis is performed through an originally developed complex Matlab-Simulink model, with embedded database of results acquired from three-dimensional finite element iterative calculation, [2]. The finite element calculations are derived by an original program package FEM-3D, developed at the Ss. Cyril and Methodius University in Skopje, [2]. This methodology of 3D finite element modelling for transient behaviour analysis of the combined instrument transformer with its complex geometry and construction of different non-linear materials represents a computationally intensive case study, which implies usage of optimised algorithms for modelling. This is an inverse modelling problem, which is a frequent case in metrology, [3 and 4].

In the full version of the paper different metrological characteristics derived at various transient regimes of the particular transformer designs will be presented, compared and discussed. A novel metrologically improved prototype of the combined instrument transformer will be realized. The characteristics derived through mathematical modelling will be experimentally verified by laboratory testing.

References:

[1] IEC 61869-4, Edition 1.0, 2013-11: Instrument transformers, Part 4: Additional requirements for combined transformers, Geneva, 2013

[2] M. Cundeva-Blajer "Application of Numerical Methods in Metrology of Electromagnetic Quantities", Journal of Accreditation and Quality Science Vol. 20, Issue 6 (Series on Advances in Mathematics for Applied Sciences-Vol.86, Advanced Mathematical and Computational Tools in Metrology and Testing X, World Scientific Publishing Co. Pte. Ltd), Springer Berlin Heidelberg, 2015 pp. 140-148

[3] K. K. Semenov, G. N. Solopchenko "Inverse Problems in Theory and Practice of Measurements and Metrology", Journal of Accreditation and Quality Science Vol. 20, Issue 6 (Series on Advances in Mathematics for Applied Sciences-Vol.86, Advanced Mathematical and Computational Tools in Metrology and Testing X, World Scientific Publishing Co. Pte. Ltd), Springer Berlin Heidelberg, 2015, pp. 330-339.

[4] M. Baer, S. Bauer, K John, R. Model and R. W. dos Santos, "Modelling measurement processes in complex systems with partial differential equations: from heat conduction to the heart", Adv. Math. & Comp. Tools in Metrology, 7, Ser. on Advances in Mathematics for App. Sciences 72, World Scientific, Singapore, 2006, pp.1-12

Uncertainty evaluation of nanoparticle diameter measurements with AFM by modeling experimental data

B. De Boeck, N. Sebaihi, M. Coenegrachts, M. Dobre, J. Pétry
SMD, FPS Economy, Koning Albert II-laan 16, 1000 Brussels, Belgium

Dimensional characterization of nanoparticles is getting increasingly important in nanotechnology research. Unfortunately, dimensional measurements are often not comparable because they lack traceability and a reliable uncertainty evaluation. Nevertheless, this is indispensable to make any statement about the measurand. In this paper, we describe a methodology to calibrate AFM diameter measurements and to evaluate the corresponding uncertainty by performing a series of experiments.

The measurand we want to quantify in a traceable way is defined as the height of a set of nanoparticles, deposited on a flat mica surface. The AFM output parameter which is a natural measurement for this measurand is the maximal measured z-value of a particle in the coordinate system of the AFM. Since the commercial AFM we use is not directly traceable to the meter, we need to use certified reference grids of step heights to perform measurement calibrations. Preferably, we perform a multiple point calibration where the different points correspond to grids with nicely spread step heights covering the relevant range of nanoparticle heights.

In our laboratory we have 5 reference grids (R1-R5) to our disposal with nominal step heights between 8nm and 180nm. Suppose that the knowledge about every quantity is described by a pdf and that the pdf's corresponding to the measured step heights $h_{Ri;m}$ and the certified step heights $h_{Ri;a}$ are known ($i = 1,..,5$). To evaluate the uncertainty compliantly with the Bayesian GUM approach, we introduce the linear regression model through the 5 calibration points ($h_{Ri;m}$,$h_{Ri;a}$) as the measurement model

$$h_a = \alpha\, h_m + \beta + \varepsilon,$$

where h_m is the measured height of the nanoparticles, h_a is the measurand, and ε is a normal deviation from linearity. The pdf's for the quantities α, β and ε display the variability present in the calibration points and the model uncertainty. They are approximated by taking N samples ($h^j_{Ri;m}$,$h^j_{Ri;a}$) ($j = 1,..,N$) from the 5 calibration points, calculating the estimated intercept β^j and slope α^j, and the mean squared error $s^{2,j}$ by performing a frequentist linear regression, and merging the data together. The distribution of α and β is approximated by the N values α^j and β^j. The distribution of ε is approximated by taking a random value from $N(0, s^{2,j})$ for each $j = 1,..,N$.

The pdf's for the quantities $h_{Ri;m}$ and h_m are yet to be determined. Since an AFM measurement is physically very complex, a complete measurement equation is not available. Therefore, an observation equation is used as measurement model. Firstly, we have minimized the effect of factors that may influence the measurements by fixing ambient and sample conditions, and by following well-defined procedures under intermediate precision conditions.

Subsequently, the variation caused by remaining influencing effects under these conditions is assessed by designing a series of experiments and analyzing the resulting measurement data. In our experiments all remaining fixed effects are varied in an optimal way using a DOE approach, while the other effects are randomized. Finally, a Bayesian mixed effects model is fitted to the data to determine the pdf's for $h_{Ri;m}$ and h_m. The resulting pdf's capture the intra-laboratory precision uncertainties.

Keywords: AFM, measurement uncertainty, GUM, nanoparticles, mixed effects model.

Abbreviations:

- AFM: Atomic Force Microscopy
- pdf: probability density function
- GUM: Guide to the expression of Uncertainty in Measurement
- DOE: Design of Experiment

Title: Uncertainty Quantification, Molecular Dynamics, and the Glass-Transition Temperature of Aerospace Polymers

Andrew Dienstfrey
Paul Patrone
National Institute of Standards & Technology

Abstract:
Computer simulations are increasingly being used to inform decision making in both industrial and public policy settings. While this strategy saves time, money, and allows for efficient exploration of alternative scenarios, such use of models and simulations requires that their outputs be carefully assessed for their assumptions and accuracy so as to promote stability and confidence in the decision making process. Uncertainty quantification (UQ) is the field of study providing the foundation for such assessment techniques in scientific computing. In this talk I will present recent results in development of uncertainty quantification applied to estimates of the glass-transition temperature derived from molecular dynamics simulations of aerospace polymers. Elements of this uncertainty analysis---non-linear regression, and analysis of mixed effects---will be familiar to metrologists. However, the application of these tools to molecular dynamics "measurements" represents a new avenue for their use. Time permitting I will discuss prospects, challenges, and a view of the role for metrology institutes with respect to uncertainty quantification in scientific computing more generally.

Bayesian Approach for Improved Estimation of Reflectance Spectra

M. Dierl [1], T. Eckhard [2], B. Frei [2], M. Klammer [2], S. Eichstädt [1] and C. Elster [1]
[1] Physikalisch-Technische Bundesanstalt, Berlin, Germany
[2] Chromasens GmbH, Konstanz, Germany

The reflectance behavior of a surface can be characterized by its spectral reflectance function whose accurate knowledge is important for many applications in industry as well as color and imaging science [1]. Spectral reflectance can be estimated from responses of multispectral imaging systems. This estimation, however, involves the solution of an ill-posed inverse problem which requires the application of regularization techniques. Several approaches have been proposed to this end, where the means of regularization is typically chosen in an ad hoc way [2]. On the other hand, often valuable prior knowledge is available which could be utilized. We develop a Bayesian approach that makes use of such prior information, leading to a well-posed reconstruction problem without the need for the incorporation of any further regularization.

The proposed Bayesian approach consists of two stages. In the calibration stage, spectral sensitivity curves are determined for each camera channel of the multispectral imaging system. In the subsequent stage additional measurements are taken. By using Bayes' theorem, these measurements are combined with the complete results of the calibration step to estimate the reflectance spectrum. One important part of this Bayesian analysis is the elicitation of available prior knowledge and its transformation into a prior distribution. In our case the considered prior knowledge results from consulting experimenters and experts as well as previous monochromator and spectrophotometer measurements. After deriving the prior distributions for the calibration and the measurement stage, we show that these priors provide enough information to turn the ill-posed inverse problem into a well-posed one.

The precise calculation of the posterior requires numerical means, and we present results of a Markov chain Monte Carlo procedure. However, for industrial applications often a more efficient calculation scheme is a prerequisite, and we derive an analytical approximation by introducing several simplifications. The approximate solution can be calculated very efficiently and is well suited even for real-time applications. The proposed Bayesian approach is demonstrated for the reconstruction of spatially resolved color spectra from continuous line scan camera measurements recorded at the company Chromasens. Results for a series of responses from a current multispectral imaging system are presented and discussed. A comparison to estimates obtained by an established method based on linear and Gaussian kernels [3] suggests that the Bayesian approach performs advantageous and is well suited for applications in this industrial area.

[1] J. Y. Hardeberg, Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches, Parkland, FL: Dissertation.com, 2001.

[2] P. A. Jansson, Deconvolution of Images and Spectra, San Diego: Academic Press, 1997.

[3] T. Eckhard, E. M. Valero, J. Hernández-Andrés and V. Heikkinen, "Evaluating logarithmic kernel for spectral reflectance estimation - effects on model parametrization, training set size, and number of sensor spectral channels," *J. Opt. Soc. Am. A,* vol. 31, pp. 541-549, 2014.

Study design and regression models for reference material commutability assessment

Stephen L R Ellison

LGC Limited, Teddington, Middlesex, TW11 0LY, UK

Abstract

ISO Guide 34:2009 [1] requires that reference material producers have procedures that include the assessment of commutability "(where appropriate)", and additionally provides an Informative Annex (Annex B) that describes the concept of commutability and gives a brief indication of the general principle of experimental assessment of commutability. A recent ISO REMCO position paper [2] provides additional guidance on where commutability assessment is appropriate, stressing the particular importance of commutability in the field of clinical analysis.

Recent international guidance on the production of commutable reference materials, [3] together with related guidance for QC and PT materials [4], describes the basic principle of commutability assessment for a typical clinical application. The experimental design uses 20 or more different 'routine' materials (excluding the RM) and is appropriate for a within-laboratory study using single observations or averaged replicate observations. An error-in-variables regression analysis is recommended as the basis for assessment of the candidate reference material's commutability with the routine samples.

Such a study is straightforward in a laboratory experienced in the use of all of the methods in question and where all (including any proposed reference method used for certification) are comparatively economical to perform. It is less amenable to cases where the reference method is particularly expensive and is very different in principle to that used by working clinical laboratories. We have therefore applied a methodology that uses measurements from a carefully chosen set of working laboratories, combined with measurements from a reference method in a different, reference, laboratory, to assess the commutability of clinical reference materials.

The methodology is applied to the evaluation of commutability of two materials; one certified for a macrocyclic immunosuppressant and one for several metals at levels of clinical interest. The practical implications for study design are discussed, including both within- and between-laboratory factors. The statistical analysis of the resulting data is considered in detail. It is shown that the methods applicable to a simpler design cannot necessarily be applied directly to an inter-laboratory study over multiple methods and materials, and approaches to adapting the statistical analysis for this purpose are discussed.

References

1. ISO Guide 34:2009 - General requirements for the competence of reference material producers. ISO, Geneva, 2009
2. ISO REMCO position paper: The need for assessment of commutability of reference materials, http://isotc.iso.org/livelink/livelink?func=ll&objId=16787892&objAction=Open, Accessed 30 May 2015
3. CLSI C53-A, *Characterization and qualification of commutable reference materials for laboratory medicine; Approved guideline*, Clinical Laboratory Standards Institute (2010), ISBN 1-56238-726-X
4. CLSI EP14-A3, *Evaluation of commutability of processed samples; Approved Guideline*, Clinical Laboratory Standards Institute (2014), ISBN 1-56238-971-8

Authors: A. Fiebach and S. Schmelter

Title: Multiphase flow simulations in large pipes

Abstract:

Currently, well over half of the world’s energy demand is satisfied from oil and gas.When oil is extracted from a well it typically exists as a multiphase flow, compromising time-varying ratios of oil, water and gas. Typical multiphase measurement systems can have an uncertainty of 20% or greater under field conditions. One of the aims of the European EMRP project ENG58 "Multiphase Flow Metrology in the Oil and Gas Sector" (MultiFlowMet) is to understand where this measurement uncertainty derives from. To gain insight, an experimental intercomparison between several multiphase test facilities is complemented by CFD (Computational Fluid Dynamics).

This contribution deals with numerical simulations of two- and three-phase flows in horizontal pipes as well as in a vertically installed Venturi meter. For both geometries, the generation of different patterns (stratified, wavy, slug, etc.) is investigated. Therefore, different volume flow rates leading to different superficial velocities are considered. The results are validated by comparison with experimental and numerical data from the literature. Furthermore, the influence of inflow fluctuations as well as of different numerical discretization schemes on the resulting pattern is studied.

Dr. André Fiebach andre.fiebach@ptb.de

Physikalisch-Technische Bundesanstalt (PTB)
Arbeitsgruppe 8.41 Modellierung und Simulation

Bayesian approach to evaluate and use measurement devices' calibration function

Nicolas Fischer[(1)], Catherine Yardin[(1)], Laurent Leblond[(2)], Benoit Savanier [(3)], Marc Sancandi[(4)], Merlin Keller[(5)],

[(1)] Laboratoire National de Métrologie et d'essais (LNE)

[(2)] PSA Peugeot-Citroen

[(3)] Centre Technique des Industries Aérauliques et Thermiques (CETIAT)

[(4)] Commissariat à l'énergie atomique et aux énergies alternatives (CEA)

[(5)] Elecrticité de France (EDF)

Calibration of measurement devices is a daily task in any metrological field. The new definition of calibration of the International Vocabulary of Metrology (VIM) specified a two-step procedure for calibration. In a first step a calibration function is estimated and in a second step the previous relation is used to predict a new measurement. According to this definition metrologists are to provide the measurement calibration function and associated uncertainties. To address this issue, Bayesian approaches and case studies have been recently proposed [1] [2] [3] [4] [5] to the practitioners for doing so. They are based on Bayesian inference, authors have showed the benefit that allows the user to take into account its prior information in the estimation process. They rely on classic Bayesian regression model and are not specific to metrologists' calibration issue in the sense that they did not consider uncertainty on the dependent variable (usually denoted X). In the process of calibration of a measurement device, both the standard and the resulting signal are subject to uncertainties and errors-in-variables regression models have been proposed accordingly. However, these models are not commonly used in a Bayesian framework in metrology. We propose then a generic Bayesian calibration model that may be adapted to suit different calibration situations and some guidance for a simplified inference.

[1] C. Elster and B. Toman. Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate. Metrologia, 48:233{240, 2011.

[2] G. J. P. Kok, A. M. H. van der Veen, P. M. Harris, I. M. Smith, and C. Elster. Bayesian analysis of a ow meter calibration problem. Metrologia, 52:392{399, 2015.

[3] A. Allard, N. Fischer, G. Ebrard, B. Hay, P. Harris, L. Wright, D. Rochais, and J. Mattout. A multi-thermogram-based Bayesian model for the determination of the thermal di_usivity of a material. Metrologia, 53:S1{S9, 2016.

[4] C. Elster, K. Klauenberg, M. Walzel, G. W□ubbeler, P. Harris, M. Cox, C. Matthews, I. Smith, L. Wright, A. Allard, N. Fischer, S. Cowen, S. Ellison, P. Wilson, F. Pennecchi, G. Kok, A. van der Veen, and L. Pendrill. A Guide to Bayesian Inference for Regression Problems. Deliverable of EMRP project NEW04 Novel mathematical and statistical approaches to uncertainty evaluation, 2015.

[5] K. Klauenberg, G. W□ubbeler, B. Mickan, P. Harris, and C. Elster. A tutorial on Bayesian Normal linear regression. Metrologia, 52:878{892, 2015.

Metrology for time-varying quantities:

A first look at some basic issues via time series

Michel FLIESS & Cédric JOIN

Ecole polytechnique & Université de Lorraine

Michel.Fliess@polytechnique.edu & cedric.join@univ-lorraine.fr

For the purpose of many modern scientific and technological topics and according to several authors (see, *e.g.*, S. Eichstädt, Analysis of Dynamic Measurements, PhD, TU Berlin, 2012), there is an urgent need for dynamic metrology: measurements of time-varying quantities need to be investigated and clarified. The basic questions about measurement uncertainty and confidence intervals, which are still discussed in static metrology (see, *e.g.,* R. Willink, Measurement Uncertainty and Probability, Cambridge University Press, 2013.), become of course much more involved. It is moreover not at all obvious that the probabilistic and statistical tools which have been developed for the static case still make sense in this time-varying context.

A new setting on time series should permit to overcome most of the above difficulties. It is based on a rather recent theorem by P. Cartier and Y. Perrin (Integration over finite sets, in F. & M. Diener (Eds): Nonstandard Analysis in Practice, Springer, 1995, pp. 195-204), which is expressed in the language of *nonstandard analysis*. Successful applications to financial engineering, and to renewable energy and road traffic management have already been developed.

When taking time into account, repeated measurements may be viewed as a time series $\mathbf{X}(t)$. The Cartier-Perrin theorem, which does not necessitate any probabilistic modeling, tells us that, under a very weak technical condition, $\mathbf{X}(t)$ may be written as follows:

$$\mathbf{X}(t) = E(\mathbf{X})(t) + QFluct(\mathbf{X})(t)$$

$E(\mathbf{X})(t)$ is the *mean* (or *average*, or *trend*); $QFluct(\mathbf{X})$ denotes the *quick*

fluctuations around 0 (the integral of QFluct(**X**) over any finite time interval is *infinitesimal*, *i.e.*, “very small”)).

The mean E(**X**)(t) yields, on one hand, a new understanding of the *target values*. If it is “approximately” constant, this value is said to be constant: we are back to static metrology. On the other hand the quick fluctuations QFluct(**X**)(t) = **X**(t) – E(**X**)(t) around the mean E(**X**)(t) permit to study *uncertainty* and/or *errors*. Although a standard deviation may be easily defined as well as other higher order statistical quantities, the familiar *confidence intervals* should be replaced by new *confidence bands* (they are to some extent reminiscent of the familiar *Bollinger bands* in financial engineering). This is due to the fact that the distribution of the quick fluctuations is not Gaussian in many concrete situations. It is crucial to emphasize that our bands may be introduced independently of any probabilistic and/or statistical assumption. When E(**X**)(t) is approximately constant, *i.e.*, in the case of static metrology, note that the above approach of uncertainty might still be interesting.

Computer illustrations from mechanical engineering and signal processing will be displayed. They are inspired by concrete case-studies (see, *e.g.*, R. Morales *et al.*, Online signal filtering based on the algebraic method and its experimental validation, Mechanical Systems and Signal Processing, 66-67, 2016, pp. 374-387). The nature of the mathematical tools, which are also useful for applying the Cartier-Perrin theorem, is mainly algebraic.

Uncertainty analysis associated with repeatability and reproducibility studies

Alistair Forbes*and Kavya Jagan
National Physical Laboratory, UK

One of the basic steps in characterising the behaviour of a system that is subject to random effects is to observe a number of outputs of the system and use these observations to make inferences about the system. The inferences are usually made on the basis of parametric model in which the samples are modelled as being drawn from a probability distribution that depends on a finite number of parameters. An important example of this approach is in determining the repeatability of a measurement system in the construction of its uncertainty budget. Repeatability and reproducibility studies in measurement system analysis and analysis of variance studies also follow this approach.

The Guide to the Expression of Uncertainty in Measurement considers the case of repeated sampling from a Gaussian distribution under the assumption that the samples are drawn independently. In practice, we cannot be sure that the samples are drawn from a Gaussian distribution, nor can we be sure that the samples are independent. The objectives of this paper are firstly to examine the consequences of these assumptions in evaluating measurement uncertainty and secondly to provide alternative models that are both plausible and computationally tractable.

The motivation for the alternative models comes from the fact that in an uncertainty budget it is possible to associate an approximate time scale with each component of the budget. Repeatability is associated with short time scales (seconds), environmental effects a range of times scales (hourly, daily, seasonally), bought-in components from calibration, longer time scales (yearly), and material parameters extracted from reference data may be regarded as constants but still have associated uncertainties. This suggests

*alistair.forbes@npl.co.uk

a model in which we associate to the jth effect, a variance parameter and a timescale parameter, the latter representing the average rate at which the effect is resampled. The variances and timescales jointly specify a statistical model of the time series associated with the measurements. We consider the cases where the two sets of parameters are specified *a priori* or are drawn from pre-specified distributions. We also discuss experiments where the underlying distribution is assumed to be a rectangular or arcsine distribution or, more generally, a symmetric beta distribution.

Type A uncertainty evaluation and propagation in a Bayesian framework

Nicola Giaquinto, Giuseppe Maria D'Aucelli

Dept. of Information and Electrical Engineering, Politecnico di Bari
Via E. Orabona 4, I-70125 Bari, Italy
Email: nicola.giaquinto@poliba.it, giuseppemaria.daucelli@poliba.it

Estimating the mean X of a Gaussian distribution with unknown variance, by averaging a sample $\boldsymbol{x} = [x_1, \dots, x_n]$, is an ubiquitous operation in metrology. The standard uncertainty of the mean $\hat{x}$ has, currently, two different definitions, the first given by the GUM, the second by the Supplement 1 (S1) to the GUM. It is well-known that, although both internationally recognized, GUM and S1 are not perfectly compatible [1].

The definition of the GUM (Clause 4.2.3) is $u(\hat{x}) = s/\sqrt{n}$, being s the Bessel-corrected standard deviation of the sample $\boldsymbol{x}$; the definition of S1 (Clause 6.4.9.4) is instead $u_B(\hat{x}) = u(\hat{x}) \cdot \sqrt{n-1}/\sqrt{n-3}$. The S1 definition is "Bayesian", since u_B is the standard deviation of the posterior pdf of X (scaled and shifted t-distribution), derived by a suitable prior distribution (reference prior). This definition is being proposed for a revised version of the GUM [2], [3].

The S1 uncertainty u_B has raised some criticism [4], and first of all the following ones:

- u_B needs at least 4 measurements to be evaluated (contrary to the GUM uncertainty u, which needs only 2);
- u_B is considerably higher than u for small values of n (+73% for $n = 4$)

For these reasons, u_B creates a problematic discrepancy with previous practice, e.g. as regards calibration certificates. On the other hand, it is maintained that using u_B is more correct, and practically advantageous. First of all, the higher values of u_B with respect to u express the fact that u_B takes into account the "uncertainty on uncertainty", i.e. the uncertainty of u^2 as estimation of the unknown variance. Indeed, the GUM uncertainty u must be complemented by the degrees of freedom $\nu = n - 1$, a measure of "reliability" of u, that needs to be propagated via the Welch-Satterthwaite formula. S1, instead, is designed to compute coverage intervals by direct propagation of distributions (PoD) [5], without using the Welch-Satterthwaite formula [3], and in this framework it is correct to compute the propagation of u_B, and not of u.

The paper tries to answer the following questions related to the use of u_B and, in general, to the S1 theoretical framework applied to "Type A" uncertainties:

1) Is it a real problem, for the Bayesian S1 framework, the need of at least 4 measurements in the $u_B(\hat{x})$ formula?
2) Is $u_B(\hat{x})$ sufficient to compute satisfactory coverage intervals for X, without knowledge of the degrees of freedom $\nu = n - 1$?
3) In a simple linear measurement model like $Y = \sum a_i X_i$, which Bayesian distribution should be assigned to Y?

As regards question 1), it is readily pointed out that Bayesian credible intervals are easily computable for any $n \geq 2$ (just like the GUM uncertainty $u(\hat{x})$) and therefore the constraint $n \geq 4$ is simply overcome. As regards question 2), for the purpose of interval evaluation the degrees of freedom ν must be known, and

therefore, using u_B instead of u does not bring particular disadvantages or advantages. As regards question 3), it is already known that PoD yields coverage intervals with long-run success rate (LRSR) considerably higher that the stipulated coverage probability [4], [6]. It is suggested that, instead of the PoD method described in S1, Y is assigned a scaled and shifted t-distribution with proper parameters. In particular, the degrees of freedom are determined by the Welch-Satterthwaite formula.

The conclusion is that, while the Bayesian approach to uncertainty evaluation is perfectly viable, PoD should not be mechanically applied to Bayesian pdfs, both for linear and nonlinear problems. In propagation problems, the pdf of the output quantity should always be computed according to well-established Bayesian procedures, and validated in terms of LRSR. Since the Bayesian standard uncertainty u_B is designed for a simplified computation of the PoD, it is of doubtful practical usefulness.

References

[1] C. Elster, "Bayesian uncertainty analysis compared with the application of the GUM and its supplements," Metrologia, vol. 51, no. 4, p. S159, 2014.

[2] I. Lira, "The GUM revision: the Bayesian view toward the expression of measurement uncertainty," Eur. J. Phys., vol. 37, no. 2, p. 025803, 2016.

[3] W. Bich, "Revision of the 'Guide to the Expression of Uncertainty in Measurement'. Why and how," Metrologia, vol. 51, no. 4, p. S155, 2014.

[4] R. Willink and R. White, "Disentangling classical and Bayesian approaches to uncertainty analysis," Technical Report No. CCT/12-08, 2012.

[5] P. M. Harris and M. G. Cox, "On a Monte Carlo method for measurement uncertainty evaluation and its implementation," Metrologia, vol. 51, no. 4, p. S176, 2014.

[6] H Tanaka and K Ehara, "Validity of expanded uncertainties evaluated using the Monte Carlo Method", in Advanced Mathematical and Computational Tools in Metrology and Testing (World Scientific, 2009) pp 326-329.

[7] C. Elster and B. Toman, "Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using the Supplement 1 to the Guide: a comparison," Metrologia, vol. 46, no. 3, p. 261, 2009.

[8] G. A. Walker and J. G. Saw, "The Distribution of Linear Combinations of t-Variables," Journal of the American Statistical Association, vol. 73, no. 364, pp. 876–878, 1978.

[9] B. D. Hall and R. Willink, "Does 'Welch-Satterthwaite' make a good uncertainty estimate?" Metrologia, vol. 38, no. 1, p. 9, 2001.

[10] M. Ballico, "Limitations of the Welch-Satterthwaite approximation for measurement uncertainty calculations," Metrologia, vol. 37, no. 1, p. 61, 2000.

A compressive sensing approach to estimating ocean sound from shipping measured using an acoustic sensor network

Peter Harris[1], Rachel Philip[2], Jared Tanner[2], Andrew Thompson[2], Stephen Robinson[1] and Lian Wang[1]

[1]National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, UK
[2]Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK

Increasingly there is concern over the effects of anthropogenic acoustic noise on marine life, and the influence this noise poses for the sustainability of key marine species and the health of our seas. The European Union's Marine Strategy Framework Directive (2008/56/EC) requires member states to monitor ocean acoustic noise in order to demonstrate Good Environmental Status, and this has led to interest in improved understanding of the noise radiated by anthropogenic sources, and in acoustic noise maps, which describe the spatial variation of acoustic noise levels within a region of interest.

There are different approaches to estimating the acoustic noise map. One is based on modelling alone. It uses knowledge of the sources of sound that are known to be present and makes use of models for the acoustic output of those sources and for sound propagation in the ocean. Another is based on measurement alone. It uses a network of acoustic sensors at discrete locations within the region of interest to record the sound produced by all sources (whether their locations are known or not). Recently, however, there has been interest in an approach that uses a combination of modelling and measurement, and consideration is given to how such an approach might be implemented [1]. The approach uses measurements of ocean sound provided by an acoustic sensor network, partial information about the locations of the sources of the sound, estimates of the sound levels of those sources, and models of the acoustic output of the sources and for sound propagation in the ocean, to estimate the sound field produced by the sources anywhere within a region of interest. The approach involves a source reconstruction step, which uses a compressive sensing algorithm to exploit the sparsity of the representation of the sound field in terms of the sources, and a sound prediction step.

A set of numerical simulation experiments is used to compare the results obtained with the different approaches, as well as to investigate the influence of the design of the sensor network on the results. Although the spatial resolution to which the source reconstruction problem can be solved is generally limited, reliable estimates of aggregated measures of the sound field can be obtained compared to those provided by other approaches.

In practice, measurements of ocean sound are available for a sequence of time instants, and there is also partial information about how the locations and sound levels of the sources evolve in time. Consideration is also given to how this time dependence may be exploited to improve the accuracy to which (a) the source reconstruction problem (which becomes a source tracking problem) can be solved and, consequently, (b) the estimate of the sound field (and its evolution in time) can be determined.

[1] Peter Harris, Rachel Philip, Stephen Robinson and Lian Wang. Monitoring anthropogenic ocean sound from shipping using an acoustic sensor network and a compressive sensing approach. *Sensors* 2016, **16**, 415; doi: 10.3390/s16030415.

Refractive Index Determination from Spectral Extinction Cross Sections of Ensembles of Spheres

Jonas Gienger, Sebastian Heidenreich and Jörg Neukammer
PTB Berlin

May 2, 2016

Abstract

We present a method to infer the refractive index (RI) of a substance from spectral measurements of the extinction cross section of an ensemble of spheres made of this substance and embedded in a non-absorbing medium with known optical properties. Examples are suspensions of polystyrene spheres or sphered red blood cells (RBCs) in water. For a given ensemble, the extinction cross section can be computed analytically using the Mie solution for the scattering problem. The spheres in these suspensions are polydisperse - the particle size follows a certain unknown probability distribution. In the RBC ensemble the intracellular hemoglobin concentration, directly determines the RI, is not constant either. Thus, both, size and concentration distributions have to be taken into account in the inference problem, which we solve by nonlinear optimization of ensemble extinction cross sections. The experimental data are extinction measurements, given at discrete wavelengths $\lambda_1,..., \lambda_N$. Because one needs to infer ensemble properties like mean diameter or hemoglobin concentration and the corresponding variances as well, the problem to find the refractive index $n(\lambda_i)$ for all i = 1,..., N is under-determined. Using generic properties of the function $n(\lambda)$ and a suitable parametrization, the analysis can be restricted to a function-subspace of lower dimensionality. The problem thus turns into an overdetermined least squares problem, which is to be solved for all wavelengths at once.

Uncertainty quantification for optical metrology

S. Heidenreich, H. Gross and M. Bär

The estimation of reliable uncertainties of measurands represents one of the core tasks in metrology. According to the "Guide to the expression of uncertainties in measurements" (GUM) and its supplements uncertain input variables of nonlinear problems are considered as random variables and output fluctuations are obtained by Monte Carlo (MC) samplings. However, many applications in metrology, physics and engineering use partial differential equations (pde) which are often solved by finite element methods (FEM). Such methods are computationally expensive and the determination of uncertainties according to the GUM is often unfeasible. As a representative example for a computationally expensive problem in optical metrology, we present the statistical inverse problem of scatterometry. Scatterometry is an indirect optical method for the determination of profile parameters of periodic nano-structures on surfaces. Thereby, geometry parameters describing the profile are reconstructed from measured refraction signals by solving an inverse problem. For the uncertainty quantification of these geometry parameters we use the Bayesian approach. Here uncertainties are typically obtained from posterior distributions by using expensive Markov-Chain Monte Carlo (MCMC) methods. For more efficiency we will apply surrogate models and integration techniques to reduce time demanding computations and to accelerate Bayesian inference. We discuss the approximation error and the convergence of the different surrogate models. Furthermore, we use MCMC samplings to determine geometry parameter, its uncertainties and to reconstruct the posterior distribution. The methods presented can be employed in general and applications beyond scatterometry are desirable.

A customized SQP Method for Multiple-Geometrical Element Fitting with large Data Sets from optical and CT Measurements[*]

Daniel Hutzschenreuter[**]

Abstract

Optical and Computer Tomography (CT) measurement systems have gained an increasing importance for 3D Coordinate metrology in the past years. The rapid development of new applications and measurement capabilities is a great challenge for developers of Coordinate measuring machine (CMM) software as well as for metrological institutes to provide traceability for different measurands. At present PTB together with a German CMM manufacturer are investigating methods for inspection of flange rings with diameters of up to 5 meters and more than 50 bore holes that are typically used when mounting the components of wind turbines. In order to access the mountability a virtual counter part of the flange with multiple-geometrical elements is fitted to all bore holes at once. A Sequential Quadratic Programming method (SQP) for Minimax-programs that frequently need to be solved at these applications is presented. By very simple heuristic modifications and computational parallelization it is capable of highly accurate and efficient calculations with large data sets as it is the case for wind turbine inspection as well as optical and CT measurements. Test data sets with an á-priori known optimal solution are used to quantify the performance of the software. Results are discussed according to experimental data for calculations with medium and large scale applications. The work is concluded by an outline of further beneficial applications of the SQP method in different areas of Coordinate metrology like geometrical element calculations and the development of software test capabilities.

Key words: Sequential Quadratic Programming, Coordinate metrology, fitting software, geometrical elements, computer tomography

---

[*] The work has been undertaken as part of the national German technology transfer project “3D-Lochbildeinpassung für Koordinatenmesssysteme mit taktilen und optischen Sensoren sowie CT-Systeme” funded by the Federal Ministry for Economic Affairs and Energy (MNPQ-Transfer 05/14).

[**] Physikalisch-Technische Bundesanstalt Braunschweig und Berlin, Division 1: Mechanics and Acoustics (Daniel.Hutzschenreuter@ptb.de)

Adjustment procedures that minimise relative entropy

Kayva Jagan*, Peter Harris, Alistair Forbes and Maurice Cox
National Physical Laboratory, UK

One of the difficult problems in analysing interlaboratory comparison (ILC) data is what to do when the data is judged to be inconsistent. The starting point for the analysis of ILC data is usually a model [2] of the form

$$x_j \in \mathrm{N}(\alpha, u_j^2), \quad j = 1, \ldots, n, \tag{1}$$

where α is the value of the measurand, x_j is the estimate of α produced by the jth laboratory and u_j its associated standard uncertainty. Given the model (1), the distribution for the consensus value for α is derived from the weighted least squares estimate. The observed sum of squares of the weighted residual errors, $R(\mathbf{u})$, which we regard as a function of the uncertainties $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$, is used to judge the consistency of the ILC data in terms of how far, in probabilitic terms, $R(\mathbf{u})$ is from its expected value of $n - 1$.

If the input data is judged to be inconsistent, there are a number of adjustment procedures, such as the Birge and Mandel-Paule procedures [1, 6], that inflate some or all of the input uncertainties to derive a new set $\tilde{\mathbf{u}}$ for which $R(\tilde{\mathbf{u}})$ achieves a desired level of consistency, $R(\tilde{\mathbf{u}}) = c^2$. In this paper, we look for an adjusted $\tilde{\mathbf{u}}$ that meets the desired consistency constraint and minimises how far $\tilde{\mathbf{u}}$ is from $\mathbf{u}$ in terms of the relative entropy or Kullback-Leibler divergence [4, 5]:

$$\min_{\tilde{\mathbf{u}}} D_{\mathrm{KL}}(\tilde{\mathbf{u}}||\mathbf{u}) = \frac{1}{2}\left[\sum_{j=1}^{n} \log\left(\frac{u_j^2}{\tilde{u}_j^2}\right) + \sum_{j=1}^{n} \frac{\tilde{u}_j^2}{u_j^2} - n\right],$$

subject to $R(\tilde{\mathbf{u}}) = c^2$ and $\tilde{u}_j \geq u_j$, $j = 1, \ldots, n$. We also consider a formulation that adds an L_1 penalty term $K\|\tilde{\mathbf{u}} - \mathbf{u}\|_1$ to the objective function in

*kavya.jagan@npl.co.uk

order to achieve consistency with the least number of uncertainties adjusted, echoing the 'largest consistent subset' approach considered in [3].

References

[1] R. T. Birge. Probable values of the general physical constants. *Rev. Mod. Phys*, 1:1–73, 1929.

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

[3] M.G. Cox. The evaluation of key comparison data: determining the largest consistent subset. *Metrologia*, 44(3):187–200, 2007.

[4] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 55:79–86, 1951.

[5] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[6] R. C. Paule and J. Mandel. Consensus values and weighting factors. *J Res. Natl. Bur. Stand.*, 87:377–385, 1982.

Sampling to assure future reliability

Katy Klauenberg and Clemens Elster

Physikalisch-Technische Bundesanstalt (PTB)
Abbestr. 2-12, 10587 Berlin, Germany

Ensuring measurement trueness, compliance with regulations and conformity with standards are key tasks in metrology which are often considered at the time of inspection. Current practice usually does not verify quality after or between inspections, calibrations, laboratory comparisons, conformity assessments, etc. Statistical models describing the behavior over time may assure reliability, i.e. they may give the probability of functioning, compliance or survival until some future point in time.

It may not always be possible or economic to inspect a whole population of measuring devices or other units. Selecting a subset of the population according to statistical sampling plans and inspecting only these, allows to conclude on the quality of the whole population with a certain confidence. Combining these issues of sampling and aging, raises questions such as 'How many devices need to be inspected, and at least how many of them must conform, such that one can be sure, that more than $100p\%$ of the population will comply until the next inspection?'

Methods from reliability demonstration, such as the prevailing Weibull binomial model, quantify the confidence in future reliability on the basis of a sample. We adapt the binomial model to be applicable to sampling without replacement and simplify the Weibull model such that sampling plans may be determined on the basis of existing ISO standards. No additional information and no software is needed. We establish new sampling plans for utility meter surveillance, which are required by a recent modification of German law.

However, such a simple, table-based approach is not capable of including all available information. A Bayesian framework may offer a more efficient and also more interpretable solution to ensure that $100p\%$ of a population functions until some future point in time. Such an approach is described for the exponential hypergeometric model and the resulting sampling plans are compared with the frequentist ones.

BAYESIAN AND LEAST-SQUARES APPROACHES FOR ESTIMATING PARAMETERS OF DECAY PROCESSES

Gregory Kyriazis, INMETRO, Brasil

The Bayesian and least squares approaches to regression problems were summarized recently in [1], where the authors compared numerically the results of both approaches in a simple, yet relevant, nonlinear example, namely, the estimation of the decay time of an exponential decay process with background. This model has been applied for instance in the context of nuclear magnetic resonance measurements [2][3].
The goal of this paper is to investigate further that example. A Bayesian approach using approximations based on the posterior mode is applied here. This method has been applied to spectrum analysis [4] and more recently to waveform metrology [5].
Probability distributions for the decay time are calculated for two rectangular prior distributions with null lower limit whose upper limits differ by two orders of magnitude. Noninformative priors are assumed for the other model parameters as was done in [1]. The analysis is performed for different signal-to-noise ratios (SNR). Experiment designs which differ on the sampling interval adopted are also analyzed. The mode and the mean of the distributions are calculated and the pros and cons of adopting such summaries as estimates of the value of the decay time are discussed. Credible intervals are also calculated.
A simple least-squares method similar to that used in [6] is also employed here. It was confirmed that generalized inverses need to be used for low SNR to ensure convergence of the algorithm. The least-squares estimate and associated uncertainty are calculated for several SNR and sampling intervals. The uncertainty, as pointed out in [1], has a different interpretation from that evaluated using the Bayesian approach. Confidence intervals are also reported.
The estimates, associated uncertainties and intervals obtained from the Bayesian and least-squares approaches are finally compared numerically. The estimates in general coincide if one chooses the mode as the Bayesian estimate. The cases where they do not are those where the prior support does not cover the least-squares estimate location. This conclusion is valid for moderate and high SNR and for all sampling intervals selected. For low SNR and large sampling intervals, the estimates also do not coincide when the posterior increases monotonically no matter how large the upper limit of the rectangular prior distribution is or when the posterior is bimodal. Also, for low SNR, the posterior mean tends to infinity as the upper limit of the rectangular prior distribution tends to infinity. So if one chooses the mean as the Bayesian estimate, the difference between the Bayesian and least-squares estimates for low SNR will increase progressively with the upper limit of the prior distribution. The estimate is not robust under changes in the prior. As expected, the uncertainty associated with the estimates increases with decreasing SNR. However, even for low SNR, one can always reach the low uncertainties achieved for high SNR by reducing appropriately the sampling interval thus ensuring that the data are sampled in the region where the variation of the signal is large. Merely increasing the number of samples in other regions of the signal does not produce the same effect.

References

[1] C. Elster and G. Wübbeler, "Bayesian regression versus application of least squares - an example," *Metrologia* **53** Issue 1 (2016) S10
[2] G. L. Bretthorst, "How accurately can parameters from exponential models be estimated? A Bayesian review, Concepts in Magnetic Resonance Part A **27A**(2) (2005) 73-83

[3] G. L. Bretthorst, W. C. Hutton, J. R. Garbow and J. J. H, Ackerman, Exponential parameter estimation (in NMR) using Bayesian probability theory, Concepts in Magnetic resonance Part A **27A**(2) (2005) 55-63
[4] G. L. Bretthorst, "Bayesian Spectrum Analysis and Parameter Estimation", in *Lecture Notes in Statistics* **48**, J. Berger, S. Fienberg, J. Gani, K. Krickeberg, and B. Singer, Eds. New York: Springer-Verlag (1988)
[5] G. A. Kyriazis, "A simple algorithm for Bayesian signal analysis with applications to metrology", *IEEE Trans. Instrum. Meas.* **60** no. 7 (2011) 2314-2319
[6] G. A. Kyriazis, P. M. Ramos and A. Cruz Serra, "Bayesian and least-squares algorithms for estimating signal harmonics: A comparative study", *Measurement* **45** (2012) 2203-2212

Error-in-variables models in calibration

I. Lira[1] and D. Grientschnig[2]

[1]Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Metrolytik, 8600 Bruck an der Mur, Austria

Calibration curves are routinely used in metrology. They are constructed by obtaining estimates $\boldsymbol{y} = (y_1, \ldots, y_n)$ of the responses of a measuring system or instrument when subject to corresponding stimuli $\boldsymbol{x} = (x_1, \ldots, x_n)$. This information is used to calculate the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ of the curve

$$y = f_{\boldsymbol{\theta}}(x).$$

The conventional ways of estimating these parameters are those of least-squares and maximum likelihood, about which abundant literature is available. Alternatively, Bayesian methods have been gaining increasing popularity of late, in part because of the relatively recent discovery that Markov Chain Monte Carlo (MCMC) allows computing the probability distributions that result from the analysis. Useful guidance on the Bayesian approach can be found in [1].

In calibration, it is normally assumed that the stimuli $\boldsymbol{x}$ are perfectly known and that the data $\boldsymbol{y}$ can be taken as independent realizations of Gaussian distributions centred at $f_{\boldsymbol{\theta}}(x_i)$, $i = 1, \ldots, n$. However, in practice, the former assumption is never strictly fulfilled and in some instances it may be grossly inadequate. Therefore, in general, not only the uncertainty of the responses has to be taken into account, but also that of the stimuli.

'Error-in-variables' regression models have been developed to address this problem. Our intent with this presentation is to provide some guidance on the Bayesian formulation of these models and on the applicable MCMC computational techniques. For comparison, the least-squares approach used in [2] is also presented.

References

[1] C. Elster et al. *A Guide to Bayesian Inference for Regression Problems.* Deliverable of EMRP project NEW04 Novel mathematical and statistical approaches to uncertainty evaluation, 2015. Available at www.ptb.de/emrp/new04-publications.html.

[2] T. Bartel, S. Stoudt, and A. Possolo. Force calibration using errors-in-variables regression and Monte Carlo uncertainty evaluation. *Metrologia*, 53:965–980, 2016.

The use of morphological methods for the extraction of topographical features from complex additive manufactured surfaces

S. Lou, X. Jiang, W. Zeng, H. Abdul-Rahman, P. J. Scott

Email: s.lou@hud.ac.uk

EPSRC Centre for Innovative Manufacturing in Advanced Metrology, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH

Additive manufacturing (AM) processes have the potential to produce highly complex, customisable and multifunctional parts at lower material and energy costs. However the AM processes, particularly using metal powders, are far from perfection. The complex nature of powder AM processes tends to produce component surfaces that are very rough, showing significant defect features, including large isolated “bumps” due to partially melted particles attached to the surface, repeating steps generated by successively adding layers, surface pores and re-entrant features. These defect features are often superimposed on the complex form of the AM products. Following the existing evaluation methods for traditional machined surfaces and using extant standards, current industrial practices and research work cannot achieve credible results due to insufficient understanding of the non-Euclidean nature of AM surfaces. It is proposed in this work to use morphological methods to suppress the impact of surface form and extract surface topographical features.

The morphological operations with circular structuring elements are applied to approximate the general form of the complex surface. Closing operation will suppress deep valleys and opening operation will remove sharp peaks. These operations or their combination effects will generate a smooth reference surface such that surface topographical features can be excluded. Instead of following the traditional route which takes the assumption that the surface is planar and normally subtracts the “form” from the primary surface to obtain the residual surface, topographical features are defined as the height function over the reference surface, i.e. each sample position is attached with a height value. The morphological watershed segmentation based on Maxwell’s theory and Pfaltz graph is then performed on the “heights” aiming to result the reasonable boundary of these topographical features. A primary example of using the proposed method is the extraction of bump features from the complex AM surface. Characterising the bump features can be useful for detecting process malfunction.

Frequentist behaviour of Bayesian inferences, and vice versa

A O'Hagan

Department of Probability and Statistics, University of Sheffield, United Kingdom

Abstract

From the frequentist perspective, the Bayesian posterior mean is a biased estimator and a Bayesian 95% credible interval does not have 95% coverage. But it is easy to show that the expected bias of the posterior mean is zero and the expected coverage of a 95% posterior credible interval is 95%. This talk will try to make sense of results like these, and analogous findings for Bayesian behaviour of frequentist inferences.

MATHMET 2016 – International Workshop on Mathematics and Statistics for Metrology

L R Pendrill, SP Metrology, Box 857, S-501 15 Borås, Sweden

The Metrology of Decision Risks and Qualitative Measurements

In response to meeting societal challenges (education, healthcare, climate change, environmental control), where human perception, uncertainty and decision making are key, one is increasingly seeking the same kind of quality-assurance for qualitative measurements, as is established for more objective measurements in physics and engineering [Fisher *et al* 2016, EMPIR NeuroMet].

Metrological treatment of quantities in these and other similarly qualitative contexts, such as decisional capability or relative concentration, is challenging since data typically lie on an ordinal scale, where distances on different portions of scales are not linear or even not fully known and the usual tools of statistics will in many cases not work reliably. Generalized linear modelling (GLM) with logistic regression is an established approach, providing a link between the 'output' (response) of a system and the 'input' (explanatory) value, z, of the item being assessed. There is an extensive family of GLM models, which can be derived generally in terms of minimising the Kullback-Leibner distance between posterior and prior entropy using Lagrange multipliers subject to restraints on the sum of probabilities/fractions to 100% and on a mean value, for instance.

One particular version of GLM – the Rasch invariant measure psychometric model – by resolving the link function z into a difference, $\vartheta - \delta$, in probe and object attribute estimates, is of particular interest since it opens up opportunities of establishing metrological references for qualitative measurements in terms of quantities such as ability, ϑ, and task challenge, δ [Pendrill & Petersson 2016]. An approach, apparently equally applicable to both physical, social and other qualitative measurement, models inferences in terms of how well a system performs an "assessment", be it traditional "instruments" of social science (questionnaires, ability tests, etc.), compositional data (e.g. Alzheimer biomarkers), or when assessing compliance to a specification limit as covered since 2012 by JCGM 106. Decisional capability, i.e. the probability, $P_{success} = 1 - \alpha$, of correct decisions, is linked to measurement uncertainty typically by evaluating the decision risk, α, in terms of cumulative distribution functions beyond a specification limit.

Can one relate the 'qualitative' Rasch metrics to corresponding metrological parameters (measured value, uncertainty)? Can extending the traditional dichotomous (go/no-go) decision risks to more than two categories form a model for a metrological treatment of qualitative measurements more generally? Examples will be given in univariate (e.g. trade-off between expected utility and uncertainty) and multivariate (principal component regression) cases.

References

EMPIR Project 15HLT04 NeuroMet, Innovative measurements for improved diagnosis and management of neurodegenerative diseases, 2016 - 9

W P Fisher, Jr., A J Stenner, L R Pendrill, T Vosk, S J Cano 2016, "Social Measurement on Trial: Are the Physical and Social Sciences Compatible Paradigms?", in *Metrology across the Sciences: Wishful Thinking?*, IMEKO TC1-TC7-TC13 Joint Symposium 2016, Berkeley, 3-5 August 2016

JCGM 2012, "Evaluation of measurement data – The role of measurement uncertainty in conformity assessment", JCGM 106:2016 Document accompanying the Guide to the Expression of Uncertainty in Measurement, http://www.bipm.org/utils/common/documents/jcgm/JCGM_106_2012_E.pdf

L R Pendrill & N Petersson 2016, "Metrology of human-based and other qualitative measurements", *Measurement Science & Technology*, Special issue (in press 2016), http://iopscience.iop.org/0957-0233/page/Forthcoming%20articles#id2

Niclas Petersson, SP Metrology, Box 5607, S-114 86 Stockholm, Sweden

Invariant Maximum Likelihood Difference Scaling

In 2003 Laurence T. Maloney and Joong Nam Yang presented a stochastic model of suprathreshold perceptual differences based on difference measurement in *J Vis* (Journal of Vision). Their model is called *Maximum Likelihood Difference Scaling* (MLDS). They use the MLDS method on a sample of 11 linearly spaced color patches with blends of red and green that ranges from 100% red to 100% green (in LMS space). The *perceived* differences between the color patches are on the other hand far from linearly spaced for an observer! In the experiment, the observer is subject to a quadruple of color patches presented in pairs. She compares the difference between the two samples in the first pair with the difference between the two samples in the second pair. The observer then selects the pair she thinks has the largest difference. The process is repeated with several different quadruple combinations in a systematic way where the selected pair scores a 1 and the other pair scores a 0.

The MLDS method assumes *normal distributed errors* of the decision variable and determines the most probable interval lengths between the different color patches with maximum likelihood estimation. We suggest a slightly different approach to address this type of problem. Instead of assuming normal distributed errors we introduce *Invariant Maximum Likelihood Difference Scaling* (IMLDS) based solely on the Thurstone invariance criteria and Rasch objectivity. That is, our only model assumption is that the perceived difference actually can be measured on an objective interval difference scale. We also use maximum likelihood estimation to derive our result based on the observer's selections. IMLDS improves the MLDS method in a small but significant way by the choice of model assumptions.

TITLE Chemical and Mineralogical Compositions
— Data Analysis and Uncertainty Evaluation

AUTHORS Adriaan M. H. van der Veen (a), Paul Stutzman (b), Ewelina T. Zalewska (a), Antonio Possolo (b, PRESENTER)

AFFILIATION (a) Van Swinden Laboratorium (VSL, The Netherlands), (b) National Institute of Standards and Technology (NIST, USA)

ABSTRACT The measurement of complete chemical and mineralogical compositions, for example of biogas mixtures emanating from landfills, of Portland cement clinkers, or of holocrystalline rocks like granites, produces amount-of-substance fractions, or mass or volume fractions, that add up to 100 %. This implies that, in the case of biogas, say, an increase in the amount fraction of methane will entail corresponding decreases in the amount fractions of carbon dioxide, nitrogen, or other components. Or, in the case of a cement clinker, an increase in the mass fraction of alite will be accompanied by a decrease in the mass fraction of belite.

That is, the constraint imposed by a constant sum (the so-called "closure effect") induces spurious correlations that do not reflect, and indeed contaminate the expression of any underlying associations and interactions between the components of the composition, which typically exist, for example in the case of a clinker, where the cooling regimen may induce conversion of alite into belite.

We review the key advances in the development of statistical models and methods intended to address the peculiarities of compositional data, starting with the pioneering contributions of Karl Pearson (in biometry) and Felix Chayes (in petrography), and culminating in John Aitchison's discovery of suitable transformations for data constrained to lie in the simplex, and the further advances that have been made since.

And we illustrate how compositional data should be analyzed, for example to compute consensus values in key comparisons like CCQM-K23, or to assign values to certified reference materials like NIST Standard Reference Material 2686a, and to produce corresponding uncertainty evaluations. In particular, we will show that the evaluations produced neglecting the closure effect tend to be too small.

TITLE The NIST Consensus Builder

AUTHORS Amanda Koepke, Thomas Lafarge, Blaza Toman,
Antonio Possolo (PRESENTER)

AFFILIATION National Institute of Standards and Technology (NIST, USA)

ABSTRACT The NIST Consensus Builder (NICOB) is a web-based application that serves to combine measurement results obtained by different laboratories or by the application of different measurement methods, into a consensus estimate of the value of a scalar measurand, and to qualify this estimate with an evaluation of measurement uncertainty that captures not only the stated uncertainties associated with the individual measured values, but also any additional component of uncertainty that manifests itself only when these measured values are intercompared.

In addition, the NICOB can also report the differences between the individual measured values and the consensus value, and the differences between pairs of values measured by different laboratories or methods, and will qualify these differences with evaluations of associated uncertainty. In the context of Key Comparisons, these differences and associated uncertainties are called (unilateral, and bilateral, respectively) Degrees of Equivalence (DoE).

The NICOB offers three different procedures for consensus building, catering to different needs and situations: the DerSimonian-Laird procedure, a Bayesian procedure with mildly informative prior distributions, and the Linear Opinion Pool. The DoEs are computed consistently with each of these methods, and Monte Carlo methods (either the parametric bootstrap or Markov Chain Monte Carlo) are used for uncertainty evaluation and propagation.

We will review the principles that drive the design of the NICOB, outline the methods implemented in it, and illustrate their application to data from several interlaboratory studies, collaborative trials, and multiple method comparisons, in concrete examples from chemistry, medicine, dimensional metrology, thermometry, and radio-nuclide metrology.

Assessment of CT image quality using Bayesian classification

M. Anton, M. Reginatto and C. Elster

Physikalisch-Technische Bundesanstalt Braunschweig and Berlin

X-ray tomography (CT) is one of the most frequently applied imaging techniques in clinical routine today. The larger the employed dose, the higher the quality of the recorded image and hence the reliability of a diagnosis drawn from the image. However, ionising radiation constitutes a potential health risk for the patient and there is a trade-off between image quality and the dose received by the patient. Users of medical CT devices strive to use the minimum dose that still guarantees sufficient image quality. However, currently there is no agreed quantitative measure of CT image quality and corresponding research has received growing interest (e.g. [1, 2]).

One of the currently favoured approaches to quantify CT image quality is the so-called task-specific quality assessment (cf. ICRU report 54 [3]). Usually, the considered task is a simplified version of a radiologic diagnosis such as detecting / verifying a lesion. A "classifier" is applied to a number of test images and its ability to detect an image containing a lesion is estimated in terms of the Receiver Operating Characteristic (ROC). The area under the ROC (AUC) is then taken as the figure of merit for image quality of the considered device. AUC can take values between 0.5 and unity, where AUC=0.5 means the outcome is totally random and AUC=1 corresponds to an ideal detector.

Initially, the classification was performed by human observers, which is costly and time consuming. Therefore, mathematical model observers have been developed to perform this task, with the aim of predicting human performance. Meanhwile there is a vast literature on different observer models: ideal, ideal linear or anthropomorphic observers including either eye filters or so-called channels mimicking properties of the human visual system (for an overview see [4] and references therein).

In this contribution we give an overview on the task-specific quality assessment in medical imaging (with focus on CT). One of the current challenges, particularly from a metrological point of view, is the reliable evaluation of the uncertainty associated with the derived image quality. Another challenge is the ability to incorporate prior knowledge (e.g. about characteristics of lesions etc.). So far, employed model observers are based on classical statistics and we introduce a Bayesian classifier [5]. We illustrate this classifier and compare its performance to that of current model observers. The Bayesian classifier lends itself in a natural way to characterise the uncertainty associated with the image quality, and we contrast this uncertainty with that proposed for the employed model observers based on classical statistics.

References

[1] L. M. Popescu and K. J. Myers. CT image assessment by low contrast signal detectability evaluation with unknown signal location. *Medical Physics*, 40(11):111908, 2013.

[2] A. Wunderlich, F. Noo, B. D. Gallas, and M. E. Heilbrun. Exact confidence intervals for channelized hotelling observer performance in image quality studies. *Medical Imaging, IEEE Transactions on*, 34(2):453–464, 2015.

[3] ICRU. ICRU Report No. 54: Medical imaging – the assessment of image quality. *Journal of the ICRU*, 6(2):NP, 2006.

[4] J. Y. Vaishnav, W. C. Jung, L. M. Popescu, R. Zeng, and K. J. Myers. Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Medical Physics*, 41(7):071904, 2014.

[5] S J Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. Dover, New York, 2nd edition, 2005. Reprint, originally published 2nd ed. by Krieger 1982.

Robust inference with nonlinear dynamical models using L1 and Huber norms

Adrian Sandu

Virginia Tech and National Institute for Standards and Technology, USA

Presence of large errors in some observational data, e.g., data collected from a faulty instrument, negatively affect the quality of the data assimilation results. We discuss a systematic framework for robust data assimilation. The approach is based on replacing the traditional L2 norm formulation of data assimilation problems with formulations based on L1 and Huber norms. Both variational and ensemble-based inference methods are cast in the new robust framework. Numerical experiments using the Lorenz-96 and the shallow water on the sphere models illustrate how the new algorithms outperform traditional EnKF approaches in the presence of data outliers.

The challenge of implementing the GUM and its Supplements

The view from inside an industrial laboratory

Steve Sidney HN Dip EE
Director: National Laboratory Association – South Africa
P.O. Box 298, Persequor Park, Pretoria, South Africa, 0020.

Tel +2712 349 1500, Fax +2712 349 1501, email: steves@nla.org.za

Abstract

Viewed from the pinnacle of the traceability chain it may well be that understanding and implementing a technical standard/guide such as the *Guide to the expression of uncertainty in measurement* and its more recently released Supplements, constitutes scientific endeavour. However, when viewed from the bottom of the same chain it is a totally different matter entirely.

For so-called 'industrial' laboratories the task can best be considered challenging and in some cases impossible.

For those who spend their days in national metrology institutes in the pursuit of excellence and 'shining the marble', understanding these documents just forms part of the day-to-day job.

However, in a commercial industrial laboratory the main concern is usually figuring out how quickly a sample or a device can be calibrated or tested and issues such as uncertainty of measurement, getting the lab's standards calibrated and of course accreditation are all considered impediments to getting the job done.

This presentation will seek to highlight the problems and challenges that this group of laboratories face when looking at how to establish a reasonable and valid uncertainty.

Having spent the past 16 years in particular trying to bridge the gap between the top and the bottom of the traceability chain, the author will provide the raison d'être for those who are responsible for reviewing the GUM to keep in mind whilst they undertake their task.

Contribution to MATHMET2016 (Mathematics in Metrology)

7-9 November 2016

Compositional Calibration

K. Gerald van den Boogaart, R. Tolosana Delgado, Axel D. Renno

An emerging field in the statistics of (geo)chemical data is compositional data analysis. In this field the ratio of concentrations between elements is seen as the essential information, while the absolute concentration is considered irrelevant. E.g. for physical age determination the ratio of concentration of various nuclides is very important, while the absolute concentrations change subject to dilution and depletion processes. In many applications chemical data is rescaled to 100%, or transformed with a log ratio transformation. In both cases the information about the absolute concentration is lost.

The aim of the contribution is to initiate a discussion on how the compositional approach might modify the tasks and methods of metrology. Asking a sligthly different question might lead to different answers. We will consider the calibration of an LA-ICP-MS (laser ablation inductive coupled plasma mass spectrometer), which measures multiple, but not all elements (quasi) simultaneously.

If log ratios are considered the relevant quantity, we should be able to measure them unbiasedly and provide a precision assessment of their measurement. We show that both aspects are problematic, when the log ratio is applied to measurements independently calibrated with respect to absolute standards. The same individual absolute errors can correspond to very different relative errors due to inherent correlation between the counts.

A supplementary concept for a traceable calibration of relative concentrations is introduced and discussed. In the example we use a method based on multivariate Poisson regression for a compositional calibration of LA-ICP-MS measurements. A compositionally calibrated measurement does not carry absolute information, but only relative information about a subset of elements. The concept of a compositional error compatible with the variance concepts in compositional data analysis is introduced. This error can not be computed from classical absolute calibration information of single elements and provides essential information for compositional methods.

In the case of the LA-ICP-MS the absolute amount of material turned into plasma is strongly matrix dependent, while the counting ratios for different elements are much less so. In such cases a relative calibration can guarantee better compositional precision and allows the useage of much loosely matrix-matched standards in a well-defined range of matrix compositions.

Quantifying uncertainty of chemical measurements via Observation Equations

Blaza Toman, Michael A Nelson, Katrice A Lippa, Mary Bedner

National Institute of Standards and Technology

USA

Abstract

Complex chemical measurements often take a form of a designed experiment which allows quantification of variability in the data due to environmental and operator factors, and also employ a measurement equation to quantify the effects of other factors such as internal or external calibration standards and other known input variables. Multiple samples are usually analyzed. As such, the uncertainty evaluation is not purely bottom up (based on the measurement equation) or top down (based on the experimental design), but inherently contains elements of both. This type of uncertainty analysis is suitable for Observation Equations, specifically by an application of Bayesian hierarchical statistical analysis. In this talk we describe such a procedure in detail, and illustrate it using two examples, one is the evaluation of chemical purity of a folic acid material via quantitative ^{1}H-nuclear magnetic resonance spectroscopy, and the other is a measurement of mass fraction of 25-hydroxyvitamin D3by LC-MS using internal standards.

Contribution to MATHMET2016 (Mathematics in Metrology)
7-9 November 2016

TITLE: Filtering lack of microhomogeneity in reference materials for microanalytical methods

Authors: van den Boogaart, K.G.; Renno, A.D.; Tolosana-Delgado, R.T.

Abstract:

Homogeneity is a relative property of a sample in relation to the measurement (analytical method), the analyte, and the intended purpose, like the usage as a reference material (RM). The verification of homogeneity is essential to define a RM as fit for purpose. In this context, there have been recent efforts to check the possible superiority of synthetic RMs over natural ones. The assessment of homogeneity is an integral part of these synthesis tests and of the following certification for use as RM. With regard to their spatial variability, five types of microheterogeneity of RMs can be found in the literature, depending on which is the source of heterogeneity that it presents: random, systematic, periodic, nugget and island.

This contribution presents a first attempt towards such tests of microhomogeneity for discussion. In a first step, we define a stochastic random function model that will describe each of the types of microheterogeneity mentioned before. Then, in a quite natural manner a particular sampling strategy for each of them is derived in the second step, with the goal to filter out the undesired source of variability. In the third step, we derive a strategy of characterization of the material, namely a strategy of estimation of the heterogeneity properties of the RM that should be used to certify the reference nature of the material. These the adequacy of these strategies is shown in this contribution by using simulations of the several heterogeneity structures and of the proposed sampling and characterization strategies.

For instance, for the case of a random heterogeneity we may assume that the concentration of the target element is described by a random function (RF). If the covariance function of this RF would be known, the sampling strategy would be to repeat measurements on random positions of a very fine regular grid in such a way that the variance of their average decreases as fast as possible, using as many locations as necessary to ensure that it falls below the method specifications. Finally, the way to characterise the RM would require calibrating the concentration of the target element on a coarser grid, on as many locations as necessary to appropriately set the covariance function, using classical concepts and models of Geostatistics. Similar strategies can be derived for the rest of the heterogeneity structures, like robust methods for nugget heterogeneity or geostatistical concepts related to intrinsic functions of order *k* for systematic heterogeneity.

Evaluation of measurement uncertainty in fluid phase equilibrium calculations

Adriaan M.H. van der Veen

Chemistry Group, Van Swinden Laboratorium (VSL), Thijsseweg 11, 2629 JA Delft, the Netherlands

Abstract

Reference data of fluid properties such as (saturated) vapour pressure, liquid and gas density, viscosity, among many others are widely used in science, engineering, trade and safety applications. These reference properties are often obtained by processing the data with appropriate models to enable the interpolation of the measured data to other conditions than those used in the measurement, or to calculate such properties from measurement data. Most compilations of reference properties do not state uncertainties at all, which makes the listed data unsuited for use the aforementioned applications.

Although many calculations are quite involved and iterative, the framework of the *Guide to the expression of Uncertainty in Measurement* (GUM) and its supplements can be used to perform a comprehensive evaluation of the uncertainty of reference properties. We show how the GUM and its supplements can be used to compute the compressibility of the liquid and vapour phases using a cubic equations-of-state, the derivation of thermodynamic properties from vapour pressure data, and the calculation of the liquid and vapour phase compositions in a flash calculation. The paper underlines once more the versatility of the GUM methods in propagating measurement uncertainty.

Authors: A. Weissenbrunner and A. Fiebach

Title: Prediction of systematic errors of flow meters behind elbows

Abstract:

The goal of the project "EnEff:Wärme:On-site calibration of flow meters in district heating" is to increase the efficiency of district heat generation and transport. This can be achieved by a reduction of the measurement uncertainties of heat quantities. The idea is to develop a method which permits on-site calibration of installed flow meters. In district heating systems, elbow alignments are necessary in almost all pipe assemblies. Double elbows out-of-plane are of special interest, since they introduce strong disturbances in the flow field.

In this contribution the reaction of ultrasonic and electromagnetic flow meters to disturbed flow profile is investigated. Additionally a method is presented which combines laser optical measurements (LDV) with Computational fluid dynamics (CFD) to get an more accurate estimation of the flow rate by LDV measurement along a single path, under disturbed flow conditions.

Dr. André Fiebach andre.fiebach@ptb.de

Physikalisch-Technische Bundesanstalt (PTB)
Arbeitsgruppe 8.41 Modellierung und Simulation

Title: Modelling uncertainty in digital PCR measurement.

Digital PCR (dPCR) is a technique for estimating the concentration of a target nucleic acid. It is the main technique used for value assignment of copy number and copy number ratio for DNA reference materials. It works by loading a sample into a panel comprising a large number of partitions, amplifying the target and using a fluorescent marker to identify which partitions contain the target. The standard analysis uses only the proportion of partitions containing target to estimate the concentration. It depends on the assumption that the initial distribution of molecules in partitions is Poisson. This assumption has not previously been tested.

We describe a way to extend such analysis using the cycle quantification (C_q) data that may also be available. The model uses a generalisation of the Poisson distribution. Previously we reported on the results from a small number of data sets. The model has now been applied to C_q data from a larger number of dPCR experiments involving 3 types of target (attenuated, virulent and plasmid) at 3 concentrations. Results show statistically significant deviation from the Poisson distribution, depending on the target type, but the resulting biases are small and can be in either direction. Theoretical calculations indicate that the size of the relative bias in estimates of concentration increases with concentration. For routine usage the bias may reasonably be ignored, but for highly accurate quantitation, e.g. of reference materials, it should be taken into account. The model also includes a separate efficiency for the first amplification cycle, which the results show varies substantially between the samples. We discuss the possible factors behind these differences.

Philip Wilson | Statistician | Statistics | Science and Innovation |
LGC | Queens Road | Teddington | Middlesex | TW11 0LY | UK |
Tel: +44 208 943 7372 |

Quantitative Magnetic Resonance Imaging: A Regression Approach

Gerd Wübbeler, Selma Metzner and Clemens Elster
Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin

Abstract

Magnetic resonance imaging (MRI) has become a standard diagnostic tool in clinical routine. The technique is based on the magnetic resonance of proton spins in human tissue. The resonance condition varies for different tissues which is utilized to form an image. Basically, MRI data sample the spatial Fourier transform of the position-dependent magnetization caused by the proton spins. By suitably varying the applied magnetic fields the k-space is scanned and an inverse Fourier transform then produces the image (see, e.g., Nishimura (1996)). While standard MRI provides physicians with valuable insights about abnormalities of the tissue (e.g. for tumor detection), the absolute proton density and the spin relaxation times of the tissue cannot be determined. Estimation of these parameters, together with an uncertainty quantification, poses an interesting challenge from a metrological point of view.

Recently, attempts have been made to obtain quantitative information from MRI about the underlying physical parameters within short acquisition times suitable for clinical routine. Ma et al. (2013) have suggested to use a large number of randomly chosen scans each of which providing only a partial sample of the k-space. Combined with a novel way of analyzing these data this so-called fingerprinting procedure avoids severe aliasing artefacts caused by the undersampling (cf. Fig. 1) and is basically capable of retrieving the physical parameters at each voxel. The procedure applies template matching based on a huge database of potential responses to the scans at each single voxel generated by a physical model (i.e. the Bloch equations). Davies et al. (2014) argued that this template matching approach is closely related to compressed sensing and they suggested a related, iterative procedure which is provably convergent under suitable assumptions.

In this contribution we briefly review quantitative MRI and propose an alternative, novel approach in terms of the solution to a related regression problem. A large-scale nonlinear least-squares optimization algorithm is developed based on an appropriate second-order approximation of the Hessian that can be efficiently calculated. The algorithm is capable of identifying the underlying physical model within reasonable calculation times, cf. Fig. 1. Simulation results suggest that the proposed approach yields improved estimates of the physical parameters when compared with the two methods mentioned above. The proposed approach also yields uncertainties for the estimates which appear to provide a reasonable characterization of their accuracy.

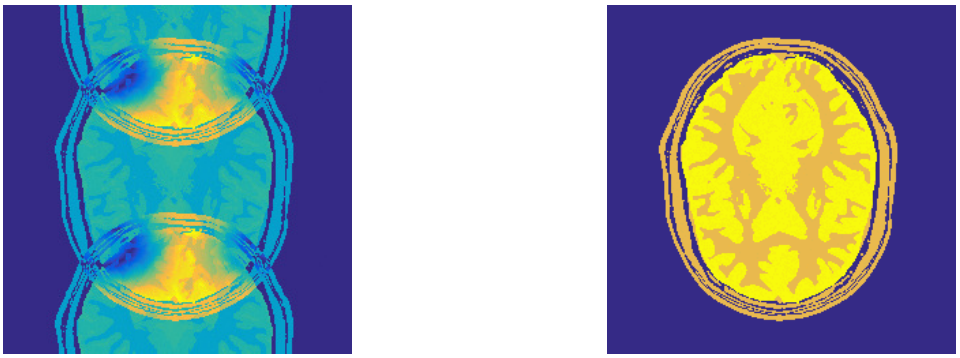

Figure 1: Inverse Fourier transform of a single undersampled scan in k-space (left) and reconstruction of proton density for human brain tissue from several undersampled scans (right).

References

Davies, M., Puy, G., Vandergheynst, P., and Wiaux, Y. (2014). A compressed sensing framework for magnetic resonance fingerprinting. SIAM Journal on Imaging Sciences, 7(4):2623–2656.

Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., and Griswold, M. A. (2013). Magnetic resonance fingerprinting. Nature, 495(7440):187–192.

Nishimura, D. G. (1996). Principles of magnetic resonance imaging. Stanford University.

Complex Structured Surface Characterisation by using PDE based Adaptive Nonlinear Diffusion Filter

Wenhan Zeng*, Xiangqian Jiang, Shan Lou, Hussein Abdul-Rahman, Paul Scott

EPSRC Centre for Innovative Manufacturing in Advanced Metrology, School of Computing and Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK

*Corresponding Author Email: z.wenhan@hud.ac.uk

Manufactured parts with *complex structured* surfaces have been widely used in automobile, bio-engineering, medical and consumer electronics etc. Compared with traditional 'stochastic surface', the complex structured surfaces have two significant characteristics: one is that they have *complex base surface* (reference surface/mean surface) which has complex shape, the other is that they have *deterministic features* with high aspect ratio on the base surface. The scale of complex structured surfaces ranges from the macro scale down to the nano scale. Specific examples include Flexible/wearable electronics, MEMS/NEMS devices, micro moulding, micro fluidic systems, defined geometry abrasives, and bio integration coatings. The ability to adequately characterise these complex structured surface geometry features is crucial in the optimisation and control of such functional device/components. From the function point of view, the accuracy of the geometrical features (line width, step height etc.) of these complex structured are generally more important than the roughness. It is therefore the basic requirements of the filtration techniques are that they should not only filter out the reference surface/mean surface and reduce the measurement noise, but do so without blurring or changing the location of the features. However, traditional single filtration techniques are lack of the ability to preserve the feature boundary with high accuracy. For example, when using the Gaussian filter for the MEMs surface, the boundary of the lines and steps are smoothed. As a consequence, it is very difficult to exactly evaluate the width of lines and heights of steps. In this paper, a Partial Differential Equation (PDE) based adaptive nonlinear diffusion filter combined with the robust filtration to process complex structured surface is proposed. In this method, robust filtration acting as pre-processing step to remove the complex reference surface and the diffusion filter is used to extract the structured features. The proposed diffusion filter is based on the PDE method and can be seen as a nonlinear heat equation, which describes the distribution of heat (or variation in temperature) in a given region over time. The diffusivity function based on the gradient of the surface can help to separate the internal region area of boundary area of the measured feature. In this way the diffusion process will take place mainly in the interior regions (line, step, etc.) of the surface, and it will not affect the region boundaries where the magnitude of gradient is large. The model proposed here has been coded with Matlab. Experimental work shows that the proposed filter can separate the geometrical feature (especially the line and step) from roughness and measurement noise and outliers with ideal edge preserving property.

Abstract for MATHMET 2016 - International Workshop on Mathematics and Statistics for Metrology

Titel: Modelling the Variance: The Variance Function R-Package VFP

Author: Andrea Geistanger, Roche Diagnostics GmbH, Penzberg, Germany

Modelling of the variance function, i.e. the relationship between the variance and the mean, is necessary for the construction of precision profiles for immunoassays or other measurement methods, going over a wide measurement range. Sadler et al. [1] developed an applicable estimation method based on the maximum approximate conditional likelihood (MACL) method, which is implemented in the stand-alone software Variance Function Program 12. 0[2]. In order to use this method conveniently we re-implemented the estimation approach in the VFP R-package, which will be soon available on the open access repository for R-packages CRAN[3].

The VFP package has the same features as the Variance Function Program 12.0:

- Implementation of the MACL estimation algorithm for the variance function
- Estimation of confidence intervals for the precision profile
- Easy graphical representation of the precision profile, either in terms of Variance against mean, SD against mean or CV (Coefficient of variation) against mean.

The VFP package has some additional features:

- The user can specify any kind of linear or non-linear relationship between variance and mean
- The estimation of confidence and prediction intervals of the precision profile can be done based on a first- or second order Taylor approximation.

Within the talk the main concept of the MACL estimation approach is presented. Afterwards the estimation and graphical functions of the VFP package are discussed. Also examples of the application of the first-or second order Taylor approximation for the estimation of confidence intervals are shown. Finally an overview of precision profiles for immunoassays and different non-linear variance functions is given.

[1] Sadler WA, Smith MH. A reliable method of estimating the variance function in immunoassays. Comput Stat Data Anal 1986; 3:227-239

[2] Sadler WA. http://www.aacb.asn.au/professionaldevelopment/useful-tools/variance-function-program-version-110. Accessed November 16, 2015

[3] The Comprehensive R Archive Network . https://cran.r-project.org/ . Accessed April 2016

Linking the Results of Key Comparisons and Supplementary Comparisons of Regional Metrology Organizations

Oleh Velychko[1], Stanislav Karpenko[2]

State Enterprise «Ukrmetrteststandard»
Kyiv, Ukraine
velychko@ukrcsm.kiev.ua[1], s.r.karpenko86@gmail.com[2]

ABSTRACT

The Regional Metrology Organization (RMO) key comparison (KC) is linked to the corresponding International Committee for Weights and Measures (CIPM) KC by using the measurements of joint participating laboratories known as linking National Metrology Institutes (NMIs). The RMO KC is used to demonstrate the global equivalence of the measurement results and it includes the procedure how the results of RMO KC are linked to the corresponding CIPM KC. A supplementary comparison (SC) is a comparison, usually carried out by an RMO to meet specific (regional) needs which are not covered by KC, for instance, measurements of specific artifacts or measurements of parameters not within the "normal" scope of the Consultative Committees (CC) of CIPM [1].

We present a procedure for linking the results of a Regional Metrology Organization (RMO) key comparison (KC) and supplementary comparison (SC) to those institutes which have taken part in the RMO KC or in the SC and also in the corresponding CIPM KC. The RMO results of KC and SC are linked by a correction factor which is determined as the difference of degrees of equivalence from the results of the participating laboratories in both comparisons. Furthermore, we determine the uncertainty of the correction factor which is then used in the calculation of the corrected degrees of equivalence and their uncertainties. The proposed linking procedure is illustrated for two real data sets where the results of EURAMET KC and COOMET SC for Electrical Power [2] and results of COOMET KC and COOMET SC for Capacitance [3] are linked.

Keywords: key comparison, supplementary comparison, linking, regional metrological organization, degrees of equivalence.

References:

[1] CIPM MRA-D-05 «Measurement comparisons in the context of the CIPM MRA».

[2] O. Velychko, S. Karpenko, V. Gachok, O. Akhmadov «COOMET.EM-S2 Supplementary Comparison of 50/60 Hz Power», Final Report, Febr. 2013, SE «Ukrmetrteststandard».
[3] O. Velychko, S. Shevkun. Evaluation results of comparison of the national standards of capacitance unit // *XXI IMEKO World Congress* "Measurement in Research and Industry" August 30 - September 4, 2015, Prague, Czech Republic.

Numerical investigation on partial flow sampling for a unique soot aerosol dilution setup

Authors: G. Lindner, A. Fiebach, S. Schmelter, A. Nowak (PTB, Berlin, Germany)

Abstract:

In order to develop a primary soot aerosol standard for the calibration of soot aerosol instrumentation for automotive industry, a high-accuracy soot generator as well as a well-defined aerosol conditioning, dilution, and homogenization unit is needed. One goal of this development is to establish an independent calibration service for soot particle number concentration measurements in engine exhaust emissions following the new ISO standard 2789.

Hence, in order to optimize the spatial aerosol distribution of soot generator output, to ensure a stable and spatially homogeneous mixing and dilution configuration, an accompanying CFD studies were performed.

In this configuration the simulated mass fraction of dilution air and soot aerosol based on spatially - dependent mixing of gases as carrier fluid is investigated. A L-shaped sampling pipe is inserted in the aerosol flow to simulate the behaviour of the sampling line. In the simulations is showed that the mean field variables like velocity magnitude are in the expected range at different locations. Then the mixing process is observed on different downstream positions of the dilution system. To simulate a possible non-centrally position of sub-tube system an approach to shift the sample face in equidistant increments were applied.

We found that the centered position of sampling tube is able to collect that fraction of mass spectrum which is very well homogenized after a downstream distance of approximately 0.5 m. The relative error on mass fraction in relation of the main cross section to shifted sub-cross section respectively has the smallest value near the center position.

Uncertainty calculation in the characterization study of reference material for spirit beverages – A practical approach

Igor Renato B. Olivares; Danieli Cristina Marcki; Guilherme Teruaki Kuwahara de Toledo

University of São Paulo, Chemistry Institute (IQSC-USP), São Carlos/SP, Brazil

Reliability and traceability of a measurement are vital for any test laboratory. Thus, it's possible to note a rise in the number of laboratories requesting accreditation on the ISO/IEC 17025:2005, according to the "International Laboratory Accreditation Cooperation" (www.ilac.org).

Within this context, an essential tool used in quality control are the certified reference material (CRM), identified by their metrological traceability which brings in their certificates all the values of their properties and associated uncertainty.

The uncertainty of a specified property from a CRM indicates the quality of the measurements and the procedures used in their production (characterization study, homogeneity and stability study), the total value is given by the sum of their individual uncertainties. Spirits beverages are known for being stable and homogeneous, so the uncertainty related to the characterization usually tends to be the more significant.

During the characterization stages, when working with chromatography its common to use the internal standard method to reduce the variability that occurs due the replicates of analyses, consequently reducing the uncertainty.

The three most important sources of uncertainty when characterizing a CRM candidate of spirit beverages using gas chromatography with flame ionization detector (GC-FID) and the internal standard method are: Uncertainty related to the calibration curve, uncertainty of the concentration of the internal standard and the uncertainty related to the precision of the analysis. Between this three sources, the calibration curve is the more expressive contributor to the total value.

This work demystifies the method to calculate the uncertainty during the characterization stage of a CRM candidate of spirit beverages, utilizing as its base a simplified approach of the Eurachem guide. As a result, a practical application utilizing the Excel® and a rearrangement of the equation used to determine the uncertainty of the calibration curve is shown. This approach was used to determine the concentration of 9 analytes contained in a sample of a candidate of certified reference material of whisky in a practical, quickly and simple way.

A polynomial estimation of measurand and parameters for non-Gaussian symmetrically distributed data samples based on cumulants

Zygmunt Lech Warsza[1] Serhii W. Zabolotnii[2],

[1] Industrial Research Institute for Automation and Measurements PIAP,
Al. Jerozolimskie 202, 02-486 Warszawa, Poland,
email: zlw1936@gmail.com
[2] Cherkasy State Technological University, Ukraine,
email: zabolotni@ukr.net

Abstract. The non-standard method for evaluating of the average and standard deviation of non-Gaussian-distributed symmetrically sampling with a priori partial description (unknown PDF) is proposed. This method of statistical estimation is based on the apparatus of stochastic polynomials and uses the higher-order statistics (moment & cumulant description) of random variables. The analytical expressions for finding estimates and analyzing their accuracy to the degree of the polynomial $s = 3$ are given. It is shown that the uncertainty estimates for received polynomial are generally less than the uncertainty estimates obtained based on the mean (arithmetic average). Reduction factor depends on the MSE values of higher order cumulant coefficients, characterizing the degree of the sampling distribution differences from the Gaussian model. The results of statistical modelling, based on the Monte Carlo method, confirmed the effectiveness of the proposed approach are presented.

Keywords:, non-Gaussian model, stochastic polynomial estimator, cumulant coefficients, mean value, variance, uncertainty.

Z. L. Warsza[1] and **M. J. Korczyński**[2]
[1] Industrial Research Institute of Automation and Measurement (PIAP)
Warszawa, e-mail: zlw@op.pl
[2] Technical University of Lodz e-mail: jerzykor@p.lodz.pl

Comparison of Statistical Properties of Skewness and Kurtosis of Small Samples of the Normal and Uniform Populations

Abstract. In many cases it occurred in practice, that a small number of measurement observations can be collect only for various reasons. Then also such samples from a normally distributed population and other symmetric distributions including uniform can be asymmetric. A reason of small measurement sample might be:
- the lack of a larger number of objects to be tested, e.g. for the validation of the method used in only a few accredited laboratories, - high cost of measurements or limited time of their execution, - the investigated object might be destroyed or its properties after test are changing irreversibly, - the inability to re-perform the measurements, e.g. in the study of a distant terrain and in medicine.

According to GUM recommendations, the best estimator is regarded, as a mean value of collected data, and the uncertainty of type A as a parameter, which characterize data scattering, basing on calculated of the variance of the sample. It may be useful to examine how other statistical parameters of small and very small samples, where the type of distribution is a priori not known. In this paper, the as the first step of this research will point the skewness and kurtosis statistics of not autocorrelated data taken at random from a normally distributed population, and for comparison from a uniform distribution. Monte Carlo simulation method is used to analyze statistics of small samples parameters. We hope that knowledge about statistics of skewness and kurtosis should allow to obtain a more reliable estimate of the standard deviation and the uncertainty of the measurand value estimator from samples of a small number of measurement observations, when the range of their distribution is known. Few examples of skewness coefficient distributions are given below.

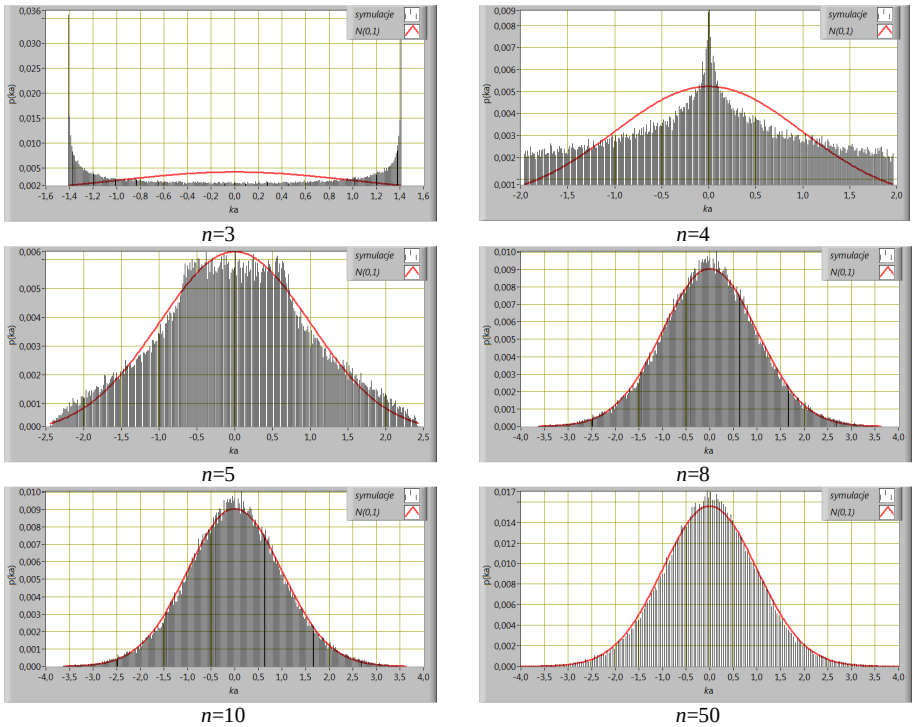

Fig. 1. Some distributions of skewness coefficient $k_a \equiv g$ for samples of low number elements n

Improved Method for the Calculation of Expanded Uncertainty

Igor Zakharov, Olesya Botsyura

National scientific center “Institute of metrology”, Kharkov, Ukraine
Kharkov National University of Radioelectronics, Ukraine
e-mail: newzip@ukr.net

We suggest a new method for the calculation of the expanded uncertainty and the coverage factor that improves the approach considered in Committee draft JCGM 100 201X (GUM-CD). The new technique takes into account the presence of type B uncertainty and can be used when the measurement data are non-normally distributed. The main idea behind the approach is the application of the kurtosis method in the calculation of the expanded uncertainty which shows a good performance when the distribution of data belongs to a symmetrical family of distributions. An improvement of the kurtosis method is proposed and it is compared to the conventional approach as suggested in GUM-CD as well as to the method based on the Monte Carlo simulation following Supplement 1 to GUM. We find that the new approach results in the values of the expanded uncertainty which are very close to the one obtained from GUM-S1, while the GUM-CD method produced larger values of the expanded uncertainty in all of the considered cases. Finally, the theoretical findings were illustrated by using several measurement data.

Keywords: extended uncertainty, coverage factor, kurtosis method.

Andrzej Zięba
AGH University of Science and Metallurgy
Kraków, Poland
andrzej.zieba@fis.agh.edu.pl

Robust alternative for the uncertainty of repeated measurement based on Rousseeuw's Q_n estimator

The use of median $\tilde{x}$ as the robust equivalence of arithmetic mean $\bar{x}$ is generally accepted. Due to high Gaussian efficiency $2/\pi \cong 0.64$ of median there is not much point of using still more efficient estimators of scale. One should bear in mind that variance adapted as a measure of effectiveness exaggerates differences between estimators. Was the standard deviation used as such a measure, median is merely 20% less efficient that the mean.

Median absolute deviation MAD is at present the most common robust equivalent of standard deviation. Its disadvantage is, however, a relatively low efficiency of 37%. Robust estimator Q_n with much higher efficiency was introduced by Rousseeuw and Croux [1]. It is defined by the formula: $Q_n = c_n$ first quartile of $\{ |x_i - x_j| : i > j \}$. Thus, Q_n is calculated from the set of all $n(n-1)/2$ differences of sample elements. Factor c_n dependent on sample size n is used to achieve consistency with standard deviation for Gaussian distribution, its asymptotic value is $c_\infty = 2.2219$. As commented by Wilrich [3]: "with a breakdown point of 50% and an efficiency of 82.3% Q_n seems to be a robust estimate that cannot be outperformed by others". One can name two additional advantages of Q_n: it is independent of the tuning constant and its definition does not involve location estimator, on a contrary to both median and MAD.

Robust formula for the uncertainty of median based on MAD was introduced by Müller [4]. The expression

$$u(\tilde{x}) = \sqrt{\frac{\pi}{2}\frac{n}{n-1}}\ \frac{Q_n}{\sqrt{n}} = 1.253\ \frac{Q_n}{\sqrt{n-1}}$$

is its adaptation utilising the use of Q_n instead of MAD. The factor under the square root represents the not widely known closed formula for the efficiency of median for the finite sample [4]. A higher efficiency with the respect to MAD means that one can use $0.82 / 0.37 = 2.2$ times smaller sample to obtain the same accuracy in determination of the measurement uncertainty.

[1] Rousseeuw P. J. and Croux C.: Alternatives to the median absolute deviation. *Journal of the American Statistical Association* **88** (1993) 1273–1283.

[2] Wilrich P. T.: Robust estimates of the theoretical standard deviation to be used in interlaboratory precision experiments. Accred Qual Assur (2007) 12:231–240.

[3] Müller J. W.. Possible advantages of a robust evaluation of comparisons, *J. Res. Natl. Inst. Stand. Technol*. **105**(2000) 551–554.

[4] Kenney, J. F. and Keeping, E. S., in *Mathematics of Statistics, Pt. 1, 3rd ed.* Princeton, NJ: Van Nostrand, pp. 32-35, 52-54, 211-212, 1962.

List of Participants

Aerts, Maarten (Norwegian Accreditation, Norway)
Ametova, Evelina (KU Leuven, Belgium)
Anton, Mathias (PTB, Germany)
Arduino, Alessandro (INRIM, Italy)
Bär, Markus (PTB, Germany)
Bodnar, Olha (PTB,Germany)
Buran, Pavel (PTB, Germany)
Chunovkina, Anna (Mendeleyev Inst.of Metrology, Russia)
Cox, Maurice (NPL, United Kingdom)
Crößmann, Felix (Hottinger Baldwin Messtechnik GmbH, Germany)
Cundeva-Blajer, Marija (University in Skopje, Republic Macedonia)
D'Aucelli, Giuseppe Maria (Politecnico di Bari, Italy)
De Boeck, Bert (National Metrology Institute SMD, Belgium)
Dienstfrey, Andrew (NIST, USA)
Dierl, Marcel (PTB, Germany)
Du Toit, Pieter (National Metrology Institute, South Africa)
Eichstädt, Sascha (PTB, Germany)
Ellison, Stephen (LGC Ltd., United Kingdom)
Elster, Clemens (PTB, Germany)
Fiebach, André (PTB, Germany)
Fischer, Nicolas (LNE, France)
Fliess, Michel (LIX- Ecole polytechnique, France)
Forbes, Alistair (NPL, United Kingdom)
Frei, Bernhard (Chromasens GmbH, Germany)
Geistanger, Andrea (Roche Diagnostics GmbH, Germany)
Giaquinto, Nicola (Politecnico di Bari, Italy)
Groß, Hermann (PTB, Germany)
Heidenreich, Sebastian (PTB, Germany)
Hutzschenreuter, Daniel (PTB, Germany)
Jagan, Kavya (NPL, United Kingdom)
Jasiński, Robert (Central Office of Measures, Poland)
Kamugasa, Solomon William (CERN, Switzerland)
Karpenko, Stanislav (State Enterprise „Ukrmetertеststandard", Ukraine)
Kessel, Rüdiger (METRODATA, Germany)
Klauenberg, Katy (PTB, Germany)
Knotek, Stanislav (Czech Metrology Institute)
Koepke, Amanda (NIST, USA)
Kyriazis, Gregory (INMETRO, Brazil)
Lehnert, Judith (PTB, Germany)
Lindner, Gert (PTB, Germany)
Lira, Ignacio (Universidad Catolica de Chile, Chile)
Lou, Shan (University of Huddersfield, United Kingdom)
Makarava, Natallia (PTB, Germany)
Martins, Luis (LNEC, Portugal)

Metzner, Selma (PTB, Germany)
Mieke, Stephan (PTB, Germany)
Niedermayer, Thomas (PTB, Germany)
O'Hagan, Tony (Unviersity of Sheffield, United Kingdom)
Orzepowski, Marcin (Central Office of Measures, Poland)
Pendrill, Leslie (SP, Sweden)
Petersson, Niclas (SP, Sweden)
Possolo, Antonio (NIST, USA)
Pythoud, Frédéric (METAS, Switzerland)
Rathsfeld, Andreas (WIAS, Gemany)
Reginatto, Marcel (PTB, Germany)
Renato Bertoni Olivares, Igor (São Paulo University, Brazil)
Renno, Axel (Helmholtz-Institut Freiberg, Germany)
Ruggaber, Benjamin (TechnoTeam Bildverarbeitung GmbH, Germany
Rynn, James (University of Manchester, United Kingdom)
Sandu, Adrian (Virginia Tech/NIST, USA)
Sayanca, Leopoldo (KIT, Germany)
Schachtschneider, Reyko (PTB, Germany)
Schäfer, André (Hottinger Baldwin Messtechnik GmbH, Germany)
Schmähling, Franko (PTB, Germany)
Schmelter, Sonja (PTB, Germany)
Schmeyer, Ellen (PTB, Germany)
Sidney, Steve (National Laboratory Association, South Africa)
Sousa, João (IPQ, Portugal)
Stavridis, Manuel (PTB, Germany)
Tolosano Delgado, Raimon (Helmholtz-Institut Freiberg, Germany)
Toman, Blaza (NIST, Germany)
Van den Boogart, Gerald (Helmholtz-Institut Freiberg, Germany)
Van der Veen, Adrian (VSL, The Netherlands)
Vlachakis, Vasileios (CERN, Switzerland)
Warsza, Zygmunt (Central Office of Measures, Poland)
Wöger, Wolfgang (PTB, Germany)
Weissenbrunner, Andreas (PTB, Germany)
Wilson, Philip (LGC Ltd., United Kingdom)
Witkovskỳ, Viktor (Slovak Academy of Sciences, Slovakia)
Wübbeler, Gerd (PTB, Germany)
Zakharov, Igor (National Scientific Centre, Ukraine)
Zeng, Wenhan (Univesity of Huddersfield, United Kingdom)
Zięba, Andrzej (AGH University of Science and Technology, Poland)