FINAL PUBLISHABLE REPORT

Grant Agreement number	17NRM05
Project short name	EMUE
Project full title	Advancing measurement uncertainty – comprehensive examples for key international standards

Project start date and duration:	1 July 2018; 36 months	
Coordinator: Dr Maurice Cox, NPL, United Kingdom Tel: 0044 20 8943 6096 E-mail: maurice.cox@npl.co.uk Project website address: http://empir.npl.co.uk/emue/		
Chief Stakeholder Organisation: Joint Committee for Guides in Metrology – Working Group 1 (GUM)	Chief Stakeholder Contact: Dr Walter Bich	
Internal Funded Partners: 1 NPL, United Kingdom 2 BAM, Germany 3 IMBiH, Bosnia and Herzegovina 4 INRIM, Italy 5 IPQ, Portugal 6 LGC, United Kingdom 7 LNE, France 8 NEL, United Kingdom 9 PTB, Germany 10 SMD, Belgium 11 VSL, Netherlands	External Funded Partners: 12 ACCREDIA, Italy 13 LNEC, Portugal 14 UKAS, United Kingdom	Unfunded Partners: 15 AIST, Japan 16 RR, United Kingdom 17 WADA, Switzerland
RMG1: IMBiH, Bosnia and Herzegovina (Employing organisation); NPL, United Kingdom (Guestworking organisation)		

TABLE OF CONTENTS

1. Overview

This project provided a comprehensive set of worked examples illustrating how principles of measurement uncertainty evaluation can support documentary standards and guides. It has promoted uncertainty evaluation according to internationally recognized guides across broad disciplines of measurement. The project has delivered new or improved adaptable examples of and templates for uncertainty evaluation to the Joint Committee for Guides in Metrology (JCGM) as publishers of the internationally acknowledged Guide to the expression of uncertainty in measurement (GUM). It also provided examples to some ten standardisation bodies and other organisations that are specifically related to standards they are developing.

2. Need

Measurement models describe the relationship between (input) quantities we measure or know something about to the (output) quantities of interest (the measurands). In areas such as energy, environment and health care, these models are frequently non-linear and the quantities measured may have substantial uncertainty. In almost all scientific areas, uncertainties associated with the output quantities must be calculated given the input uncertainties. The traditional approach to uncertainty propagation through a model uses the GUM, but because of the above issues the resulting uncertainty so produced may not always be fit for purpose.

Carefully elaborated examples have been developed that are practical and covering many areas of measurement, capable of delivering reliable results, and as far as possible in a form that can be adapted to actual end-users' data and knowledge. Many end-users learn by example rather than from formal guidance material, using the guidance material to support what they have learned when needed.

The examples have been provided to meet end-users' needs, essential in diverse disciplines including traditional metrology (calibration, testing, comparison and conformance) and the sectors environment, energy, quality of life, and industry and society.

3. Objectives

The overall objective is to provide a comprehensive set of examples to illustrate uncertainty evaluation according to the GUM suite of documents. The specific objectives of the project are (note that guidance document JGCM 103 is now known as JCGM GUM-6):

1. To develop examples of measurement uncertainty evaluations capable of acting as template solutions that end users can use for related problems. Examples will include measurement model construction using JCGM 103, application of uncertainty evaluation principles for addressing industrial conformity assessments to support JCGM 106 and taking correlations into account as requested by ISO/REMCO, the ISO committee concerned with reference materials.
2. To derive worked examples of uncertainty analyses using the GUM and other methods to assist users to make informed choices on an appropriate uncertainty evaluation method to use. Examples will include an examination of the extent to which the GUM is appropriate for certain applications or whether the Monte Carlo methods of GUM Supplements 1 and 2, or Bayesian methods, have greater efficacy.
3. To collaborate with JCGM/WG1 (the chief stakeholder), and the standardisation, regulatory and accreditation communities (ISO/REMCO, IEC, CEN, OIML, and ILAC) to ensure that the outputs of the project are aligned with their needs, communicated quickly, and in a form that can readily be incorporated into the JCGM Guides and other documents.

4. Results

We give links to the following JCGM guides so that readers are aware of their contents. We also use convenient abbreviations throughout for these guides.

- JCGM 100:2008 Evaluation of measurement data – Guide to the expression of uncertainty in measurement as 'GUM',
- JCGM 101:2008 Evaluation of measurement data – Supplement 1 to the 'Guide to the expression of uncertainty in measurement' – Propagation of distributions using a Monte Carlo method as 'GUM--S1',
- JCGM 102:2011 Evaluation of measurement data – Supplement 2 to the 'Guide to the expression of uncertainty in measurement' – Extension to any number of output quantities as 'GUM-S2', and
- JCGM 106:2012 Evaluation of measurement data – The role of measurement uncertainty in conformity assessment as JCGM 106.

4.1 Objective 1: Examples of measurement uncertainty evaluations capable of acting as template solutions

ACCREDIA, AIST, BAM, IMBiH, IPQ, LGC, LNE, LNEC, NEL, NPL, PTB, SMD, UKAS and VSL developed 13 worked examples capable of acting as template solutions in the areas of calibration, testing, comparison and conformity assessment. The examples were supported by tutorials on Monte Carlo, Bayesian inference, correlation and reporting measurement results. Many of the examples covered correlation, as specially requested by the international standards committee ISO/REMCO, the correlation tutorial covering the topic in a relatively simple manner since the topic is found difficult by many practitioners.

Calibration, testing and comparison

Two-point and multi-point calibration. The need to interpolate tabular data arises throughout science and technology. A generic treatment of two-point and multi-point interpolation of calibration data was given by UKAS, NPL and IMBiH including the handling of uncertainties and covariances in both variables. The approach was applied to the measurement of hydrogen ion activity (pH) and statements were made that can be carried over to examples in other areas. As a result of collaboration between the NPL mathematicians and the LGC chemists, an anomaly was discerned in the International Union of Pure and Applied Chemistry (IUPAC) recommendations for pH measurement. If the IUPAC approach had been followed, the evaluated uncertainties would be too small because no account is taken of correlation in its recommendations. It was shown that meaningful uncertainties can be obtained by taking correlation into consideration. IUPAC has been informed.

Straight-line calibration. Common practice in calibration is determining a straight-line relationship between two quantities (linear regression), where measured values of the quantities have uncertainties. The intercept and slope of the line can be estimated using appropriate least squares methods. The associated uncertainties can be evaluated by the GUM approach, Monte Carlo or Bayesian methods. These approaches were compared in examples elsewhere in the project illustrating (1) *Calibration of a sonic nozzle*, (2) *Quantifying haemoglobin concentration* and (3) *Calibration of a torque measuring system*.

Mass calibration. In GUM-S1, it is described how the Monte Carlo method (MCM) can be applied to the calibration of a conventional mass of a weight against a reference weight with a nominal mass of 100 g. For this example, a Bayesian evaluation of the measurement was performed. A Bayesian approach differs from MCM in GUM-S1 and the law of propagation of uncertainty (LPU) in the GUM in that it combines prior knowledge about the measurand with the calibration data. From the posterior probability density function so produced, a value, standard uncertainty and a coverage interval for the mass were obtained. For this example, prepared by LNE, IMBiH, IPQ, LNE, PTB and VSL, the Bayesian method provided results close to those given by MCM and to the higher-order variant of LPU, but differed appreciably from the widely adopted first order- LPU. A consequence of the collaboration between all partners involved was the observation that MC and the Bayesian approach take full account of the non-linearity of the measurement model, whereas the basic LPU method and the higher-order variant in general work with approximations to the full model. This point also applied to several other examples.

Fire testing. The objective of the Single Burning Item (SBI) test is to classify construction material regarding its contribution to the propagation of fires in buildings. Material under test is exposed to adverse thermal conditions using a standard combustion item, the SBI, specified within a European standard for the reaction to fire tests for building products. Expressions for the heat release rate (HRR) and the smoke production rate (SPR) given in the standard are complicated functions of measured quantities. The GUM method for estimating HRR and SPR, and evaluating their uncertainties were applied by LNEC, NPL and IPQ to representative data and validated using the Monte Carlo method. The interaction between LNEC and NPL led to improved understanding and simplification of the complex relationships between the quantities involved.

Reassessment of calibration and measurement capabilities. Calibration, testing and other laboratories make a statement of their measurement capabilities in terms of CMCs (calibration and measurement capabilities). In the case of national metrology institutes, those claims are validated by key comparisons (KCs) in which several national laboratories compare their measurements of nominally the same measurand (the quantity to be measured). The CMC claim must be adjusted If the KC does not support it. AIST statisticians and NPL mathematicians furnished a generic analytical approach to the problem of CMC reassessment by making minimal adjustment to the claims. Such an approach, which had previously been lacking, was illustrated by its application to of gauge block measurement, used widely as reference standards in industry, typified by EURAMET-AUV.V-K1.1, a regional comparison in the area of vibration and shock (Figure *1*).

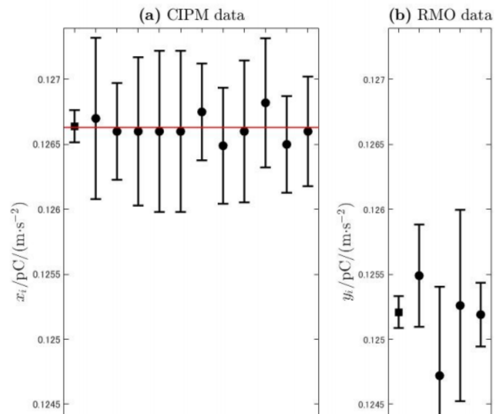

Figure 1 Reported data in (a) CCAUV.V-K1 (the CIPM KC) and (b) EURAMET-AUV.V-K1.1 (the RMO KC), with the horizontal red line showing the consensus value

Analysis of a regional metrology organization key comparison. In metrology, the capabilities of national metrology institutes (NMIs) are compared by means of key comparisons (KCs) conducted by the International Committee of Weights and Measures (CIPM) in which each NMI measures nominally the same artefact. A unilateral degree of equivalence (DoE) in a KC – a measure of the deviation of a laboratory measurement from a consensus value – is used to assess the performance of the laboratory. Comparisons organized by Regional Metrology Organizations (RMOs) must be linked to CIPM KCs. This example gave a method of performing the linking and computing the consequent DoEs for the RMO laboratories. AIST statisticians and NPL mathematicians applied the generalization of the GUM method in GUM-S2 to obtain the associated uncertainties. Conclusions were drawn regarding the efficacy of the approach. An application to key comparison CCL.K-2 relating to gauge block measurements was made (Figure *2*).

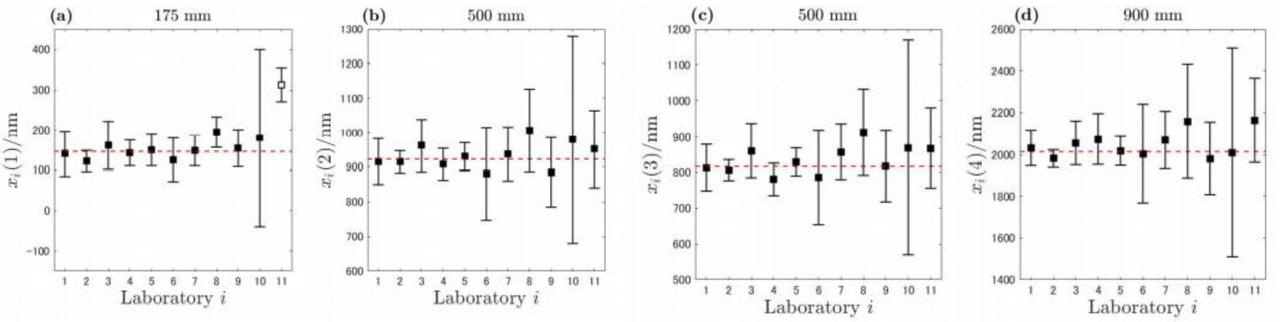

Figure 2 CCL.K-2 gauge block data for four nominal lengths and weighted means as KCRVs (broken horizontal lines) with vertical bars depicting coverage intervals with coverage factor 2

Reference material consumed at a linear rate. There are numerous practical situations in which a quantity of interest changes linearly over time. The mass flow rate from a reference leak is an instance of such a quantity, described in this example in terms of the depletion of helium from quartz membrane reference leaks (Figure 3). UKAS and NPL uncertainty experts developed an example to demonstrate a generally applicable process for modelling the quantity and establishing the uncertainty associated with measured values of the quantity, including those situations where there is covariance within the data. This example included details of the intervening steps that, in published examples, might be omitted in providing the result. The description was given in terms of matrices and vectors (as in GUM-S2) and in the perhaps more familiar notation of subscripted summations (as in the GUM).

Figure 3 Reference leaks externally mounted on a leak detector (photograph courtesy of Vaseco Ltd.)

Factoring effects such as calibration corrections and drift into uncertainty evaluations. Two examples were developed to demonstrate potential danger in the common practice of factoring effects such as calibration corrections and drift into uncertainty evaluations as rectangular distributions. Ways of handling these effects were presented by UKAS, LNE and NPL that are consistent with the GUM suite of documents. These examples illustrated that, despite the availability of appropriate guidance, significant known bias as a result of effects such as calibration corrections, drift or consumption, hysteresis and non-linearity is often not properly handled. This abuse could bias conformity decisions and thereby place the consumer or supplier at an unfair disadvantage. Instances of what can be regarded as ‘good’ and ‘poor’ practice were presented. Using a corrected value, the conformance probability, 0.62, is the shaded area in Figure 4 (left), whereas using the uncorrected value, the conformance probability, 0.90, is the shaded area in Figure 4 (right); the latter (poor practice) approach allows a greater proportion of non-conforming items to be accepted.

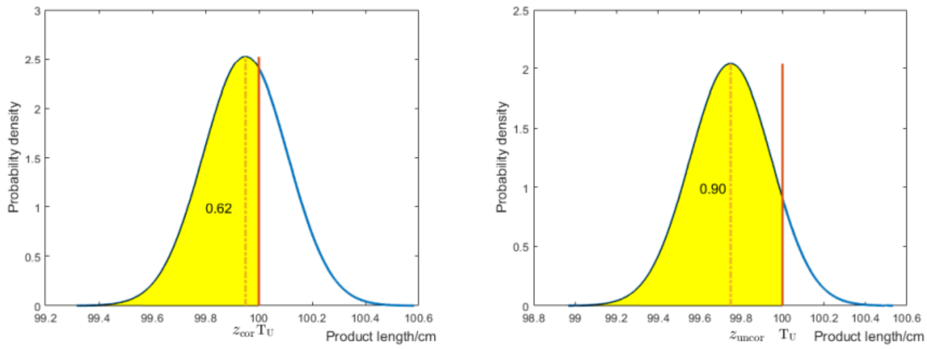

Figure 4 Using a corrected value, the conformance probability p (0.62) is the shaded area in the left figure, whereas using the uncorrected value, p (0.90) is the shaded area in the right figure; the latter (poor practice) approach allows a greater proportion of non-conforming items to be accepted

Conformity to regulation or specification

Multicomponent conformity assessment. Current JCGM guidance JCGM 106 on conformance to regulation or specification deals with a single quantity (shaft diameter, say). That guidance was extended by INRIM, NPL, VSL and LGC to treat multicomponent materials, that is, involving several quantities (shaft diameter and length, say) that jointly should conform. That guidance was illustrated with an example of influenza medication containing four components (pain reliever and fever reducer, cough suppressant, antihistamine, and nasal decongestant) in which each component has tolerances on its content. Risks to the supplier and the customer were assessed. The results were incorporated in a Guide produced by IUPAC/CITAC (the International Union of Pure and Applied Chemistry, and the Co-operation on International Traceability in Analytical Chemistry).

Measurement models involving additive or multiplicative corrections. A common form of presentation for calibration results involves expressing the result as an additive or multiplicative correction. This is the case for vacuum gauges and was illustrated with data using the models described in the international standard ISO/IEC 27893 Vacuum Technology – Vacuum Gauges – Evaluation of the uncertainties of results of calibrations by direct comparison with a reference gauge. The example, prepared by IMBiH metrologists, UKAS and NPL uncertainty experts, demonstrated the effect of model assumptions concerning errors in the reference value. In addition, it was demonstrated how conformance probability can be affected by these assumptions. The example concluded by showing how correlation can be handled for calibration corrections. The examples and conclusions have much wider applicability.

Conformity assessment of mass concentration of total suspended particulate matter in air. This example, developed by INRIM, LNE and LGC, showed how to calculate risks of false decisions in the conformity assessment of test results, according to the framework of JCGM 106, in the case in which a normal distribution is not a valid assumption for modelling prior information on the measurand. As a case study, test results of mass concentration of Total Suspended Particulate Matter (TSPM) in ambient air were considered. Data were collected in the locality of three stone quarries located in Israel and obtained according to methodIO-2.1 of the Environmental Protection Agency (EPA). Some results for the three quarries are given in Figure 5. No action would be taken when a measured value is less than the acceptance limit, that is, when it is conforming with requirements. When a value exceeds the limit, it will be declared as non-conforming and some corrective action will be required. In this case, the producers were provided with a tool for assessing the extent of their responsibility for such failure and making an appropriate response.

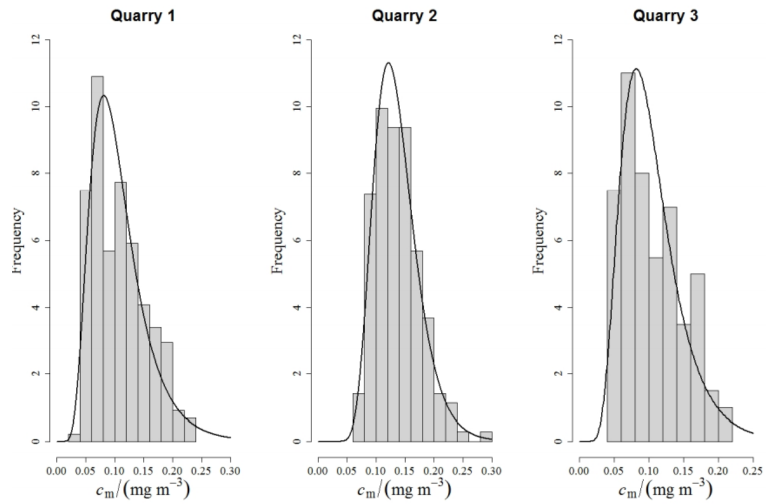

Figure 5 Histograms of the measured TSPM mass concentration values for each quarry and corresponding lognormal probability density functions smoothing the data

Nanoparticle size by atomic force microscopy. A comprehensive framework was presented for uncertainty evaluation for measurement of the mean size of nanoparticles in dispersion samples by Atomic Force Microscopy (AFM). Since no full measurement model existed for this measurement, the joint skills of SMD, NPL, LNE and UKAS were applied to develop a statistical model for the measurement uncertainty evaluation. Random effects and fixed effects were simultaneously considered, taking account of interactions. A Design of Experiment (DoE) was implemented, and a Bayesian approach followed. An optimized DoE was used instead of a full DoE to accelerate the acquisition of data by reducing the number of AFM images. Figure 6 depicts schematically the tip probe at several positions along its path with the broken line showing the measured topography.

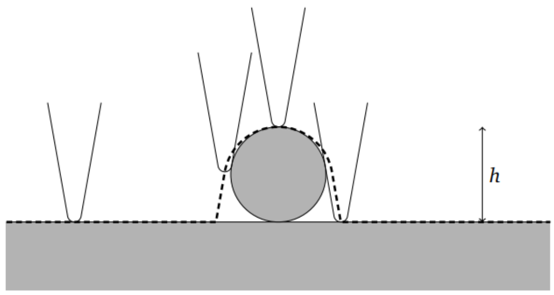

Figure 6 Measurement of nanoparticle size by an AFM with the tip probe represented at several positions along its path, and the broken line showing the measured topography

Measurement traceability according to ISO/IEC 17025:2017. Measurement traceability is commonly obtained from calibration measurements that provide a result in terms of a single value and its associated uncertainty. There are circumstances where, instead, the result may consist of a *range* of possible values. Such situations might arise when a result is provided in the form of the output from a conformity decision- process, for example, as a conformity statement in which a range of acceptable values rather than a specific value is reported. In terms of metrological traceability this style of result provides less information than a specific value but may be sufficient to attain an acceptable target measurement uncertainty for a given application. The standard ISO/IEC 17025 acknowledges the provision of such information in its informative annex A. This example, prepared by UKAS and Accredia with their intimate knowledge of the needs of accredited laboratories, and IMBiH described how such information might be used to propagate traceability. Figure 7 shows the probability density functions for conformance probability p_c = 50 % and various maximum permissible errors and measurement capability indices for an example of geometrical product specification.

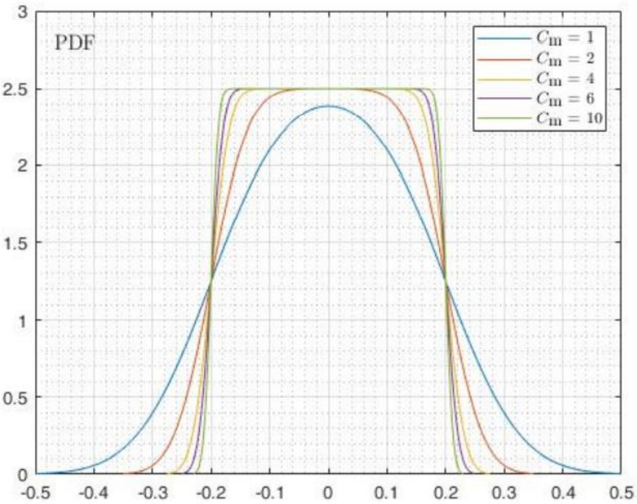

Figure 7 Probability density functions for conformance probability p_c = 50 %, maximum permissible errors MPE = 0.2 and various measurement capability index C_m values

Objective 1: Summary of key outputs and conclusions

Key outputs and conclusions for Objective 1, which was fully achieved, are given below. The outputs added considerable value beyond the state of the art through strong collaboration between the national metrology institutes, designated institutes and accreditation bodies in the project.

Correlation is a recurring theme in many of the examples, which attempt to illustrate its effects in a simple manner and to provide ways of quantifying and accounting for it. For two-point and multi-point calibration, through a real example concerning pH measurement, it was identified that the practice recommended by the International Laboratory Accreditation Cooperation (ILAC) takes no account of correlation. This seemingly innocuous oversight can lead to invalid statements of uncertainty for the pH of a solution under test. Details of how to obtain a valid uncertainty were given.

Linear calibration curves. Many calibration curves in practice are justifiably linear but the method by which straight-line relationships are provided in practice leaves a lot to be desired. To ensure valid uncertainties are provided using the calibration function, it is necessary to take proper account of all available information concerning the data and contextual knowledge on which they are based. Through three examples, constituting activities elsewhere in the project, it is shown how this information can be used to provide reliable uncertainties.

GUM, Monte Carlo or Bayesian? Several examples compare the use of GUM, Monte Carlo and Bayesian methods of uncertainty evaluation, indicating situations in which they are likely to produce comparable results or where one of these methods is preferable. Guidance Is given to assist practitioners in deciding this issue in the case of their own problems.

Competence claims. An example is given of the way claims by laboratories of their competence in terms of CMCs (calibration and measurement capabilities) can be tested against supporting key comparisons and, if necessary, minimally adjusted to ensure compatibility. The example is based on a research paper published by some of the partners just before the start of the project in response to a request by BIPM Director to provide a statistically-based method where previously it had been a matter of judgment by the CIPM committee concerned.

Extending conformance. The current JCGM guidance on conformance to regulation or specification deals with a single quantity but a growing number of situations require the consideration of multicomponent conformance. By way of a practical example, that of influenzas medication, it is shown how the extension from the univariate to the multivariate case can be made, which will provide possible material for a revised JCGM 106.

Calibration corrections. There are many situations in which laboratories do not use a measurement model that enables adequate account to be taken of calibration corrections and drift. Two examples of good practice are given that practitioners can mimic in their own situations. In contrast, poor practice, which is not GUM-compliant and unfortunately often followed, is also illustrated. When used in conformance assessment, it is shown that poor practice can cause a greater proportion of non-conforming items to be accepted.

Traceability provided by a statement of conformance. International standard ISO/IEC 17025:2017 covers traceability provided by a statement of conformance rather than through measured values qualified by uncertainties. According to accreditors, this form of traceability has raised difficulties in calibration and testing laboratories. A detailed worked example was provided to illustrate a way of proceeding in this difficult area, which is hoped will be of considerable assistance to practitioners.

The above outputs added value beyond the state of the art through strong collaboration between national metrology institutes, designated institutes and accreditation bodies.

4.2 Objective 2: Worked examples of uncertainty analyses using the GUM and other methods to assist users to make informed choices on an appropriate uncertainty evaluation method to use

This objective covers four sectors – environment, energy, quality of life, and industry and society. The examples given include three produced by an IMBiH researcher under an EMPIR Research Mobility Grant (RMG) concerning preparation of calibration gas mixtures of ammonia in nitrogen using permeation, flow meter calibration using the master meter method, and pressure drop measurement.

ACCREDIA, BAM, IMBiH, INRIM, IPQ, LGC, LNE, LNEC, NEL, NPL, PTB, RR, SMD, UKAS, VSL and WADA developed 29 worked examples to enable users to make informed choices of uncertainty evaluation methods in the above sectors. The four tutorials mentioned under Objective 1 apply equally to these examples. As in Objective 1, many instances include the treatment of correlated effects, the influence of which is considered in 19 of the 29 examples. Cases are identified in which these effects are minor but in others ignoring correlation yields invalid uncertainty statements. Key outputs included the following. Avoiding the dangers of blind adherence to the GUM approach, as opposed to Monte Carlo or Bayes, with clear guidance on the issue in many examples, including air monitoring and jet engine thrust, is emphasized. Invalid uncertainty statements can also be obtained by not taking account of prior knowledge about the quantities to be measured, which only Bayesian uncertainty evaluation can handle, such as in the examples on greenhouse gas emission inventories, production of reference materials and drug administration in new-borns. Finally, the simpler formulae provided for uncertainty evaluations as in the examples on volume measurements and drinking water supply networks will aid their ready implementation. All examples added considerable value beyond the state of the art through strong collaboration between national metrology institutes, designated institutes, regulators, and industry. The work fully met Objective 2.

<u>Environment</u>

Average areal rainfall – comparison of three methods. Precipitation measurement has diverse applications in contexts such as hydrology, meteorology, and climatology. It is of increasing importance for assessing climate change, both as an indicator and a parameter used in modelling for interpreting climatological phenomena and forecasting. There are several methods for obtaining the quantities of concern, precipitation, and rainfall intensity, for which knowledge of the measurement uncertainty associated with an estimate of a quantity can be valuable for the intended application. For a long-term analysis, several methods are available for calculating the accumulated values of precipitation and the average values observed in given catchment areas and in certain time intervals. However, it is not common to promote information about the impact that the selection of method has on the results. However, it is not common to promote information about the impact that the selection of method has on the results.

This selection was one of the main objectives of the comparative analysis undertaken, that is, the difference that results from this selection regarding the estimate of the quantity and its uncertainty. Examples prepared by LNEC, with its considerable knowledge of the field, and NPL and IPQ metrologists and mathematicians, illustrated the adequacy of the approaches recommended by the GUM and GUM-S1.**Error! Reference source not found. Error! Reference source not found.** shows a schematic of a weighing gauge (left), a weighing gauge in the field (middle) and one of the methods applied, the isohyetal method (right).

Figure 8 Schematic of weighing gauge (left), weighing gauge in the field (middle) and isohyetal method (right)

Quantification of low masses of benzo[a]pyrene. Polycyclic Aromatic Hydrocarbons (PAHs) are toxic contaminants present in all parts of the environment. Among PaHs, benzo[a]pyrene (BaP) is listed in European legislation as a carcinogenic risk marker for all PAHs. Uncertainty evaluation associated with the quantification of such micro-pollutants plays a key role in the reliability of their measurement. INRIM, LGC and VSL compared results obtained by application of the GUM and the Monte Carlo method (MCM) of GUM-S1 to real data sets derived from the quantification of a low mass of BaP spiked on filters commonly used for particulate matter sampling. MCM was shown to give more valid results because of the skewed probability distribution for the measurand (**Error! Reference source not found.**) whereas the GUM assumes a normal distribution.

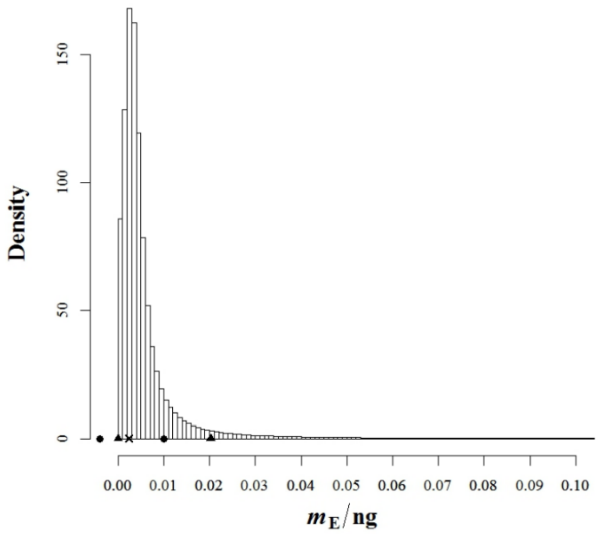

Figure 9 Numerical representation of the PDF for a very small mass of BaP (m_E = 0.003 ng) with circle and triangle symbols indicating the limits of the 95 % coverage interval obtained using the GUM uncertainty framework and MCM, respectively, and a cross indicating the minimum detectable mass of the analytical method

Calibration of an analyser for nitrogen oxides. The European Directive on ambient air quality prescribes the monitoring of nitrogen oxides (NO_x) by means of chemiluminescence as the reference method, which requires the use of proper calibration gas mixtures for instrument calibration. The example, prepared by INRIM and VSL, with their intimate knowledge of uncertainty evaluation and the chemistry involved, gave the uncertainty evaluation for the calibration of a chemiluminescence analyser for NO_x using a multi-point calibration with dynamically prepared calibration gas mixtures obtained by dynamic dilution of standard gas mixtures performed by means of calibrated mass flow controllers. This example addressed the need for a more advanced treatment of correlations arising in such measurements, especially those caused by the use of the same equipment for calibration gas mixture preparation and the use of a single calibration gas mixture from which the dilutions are made.

Routine testing of metals in soil. An ISO standard on soil quality is widely used to determine the levels of toxic metals in soil, important for determining permitted land use, the need for soil remediation, and in some cases for enforcing effluent or disposal regulations. LGC chemists and NPL uncertainty experts developed an example of the evaluation of measurement uncertainty for the routine determination of acid-extractable toxic metals in soil using a combination of acid extraction and atomic emission spectrometry. The example illustrated the general approach taken by international standard ISO 21748, which uses Information on precision and trueness of a routine test procedure to provide an indication of the measurement uncertainty to be expected from the procedure. The example further illustrated the experimental determination of sensitivity coefficients that cannot readily be derived from a mathematical model, examined the evaluation of uncertainties arising from calibration using straight-line regression with zero intercept, and discussed the issues arising in the event of an appreciable bias that is not corrected for when within permitted limits.

Comparison of methods for flow measurement in closed conduits. The growing concern with climate change leading to the need for improvement of water resources management makes the quality of measurement of water service providers a key issue. Uncertainty is a major parameter in the area that allows comparisons to be made and to show the competence of measurement activities. In this context, flow engineers face the difficult task of selecting the technically best and most -cost-effective measuring system for their application. For this purpose, account should be taken of the required accuracy related to the management process including the target uncertainty, the uncertainty of the measuring instruments, the characteristics of the fluid and the conditions of the flow, the installation set up, the metrological management needs, and the data management. In such an analysis, a comparison of the uncertainty achieved by each method can play a major role in the decision making. LNEC and IPQ, with their intimate knowledge of the field, and NPL uncertainty experts prepared an example that compared the use of an electromagnetic flowmeter (shown in Figure 10) and an ultrasonic flowmeter for actual data using the GUM approach to evaluating uncertainty and confirmed using GUM-S1.

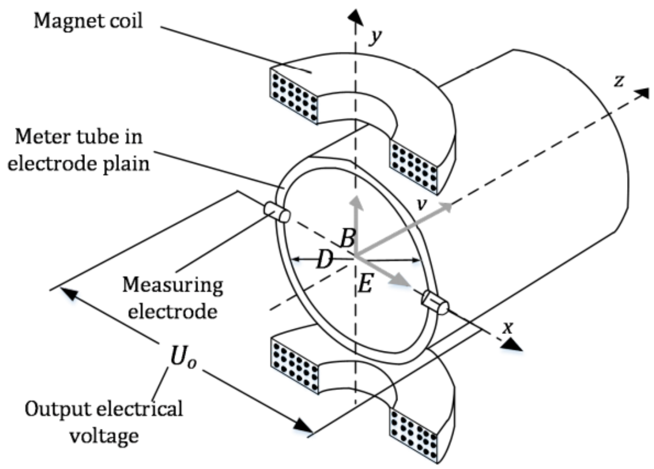

Figure 10 Electromagnetic flowmeter principle of operation, and related quantities

Compiling greenhouse gas emission inventories. The Intergovernmental Panel on Climate Change (IPCC) is the United Nations' body for assessing the science related to climate change. For greenhouse gas (GHG) emissions, IPCC uncertainty guides consider error propagation (essentially the law of propagation of uncertainty in the GUM) and Monte Carlo (MC) methods for uncertainty propagation. Each country submits an annual report to UNFCCC (UN Framework Convention on Climate Change) of its emission estimates for GHGs under the Kyoto Protocol. These estimates are based on contributions for a pollutant from many sources and activity factors for those sources. For agricultural data for a pollutant from the UK contribution, NPL uncertainty experts and chemists at LGC and VSL prepared an example that employed the model to estimate the total emission for that pollutant with GUM and MC methods from the Joint Committee for Guides in Metrology (JCGM) used to obtain the associated uncertainty. The influence of correlations in the data was quantified for the example. Lessons were drawn on the relative merits of the IPCC and JCGM approaches. Agricultural data was chosen since its uncertainty contribution is appreciable compared with that of other sectors (Figure 11).

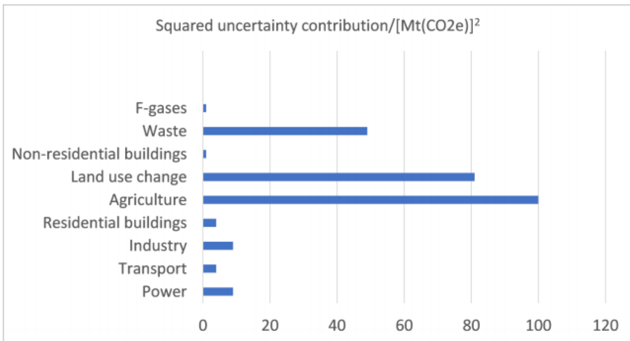

Figure 11 Squared expanded uncertainty contributions across sectors showing the appreciable agriculture contribution

Greenhouse gas emission estimation using inverse modelling. A top-down Bayesian method using a weakly informative- prior was used by NPL to provide greenhouse gas emission estimates. The method was based on the calculation of ambient mixing ratios combined with inverse modelling. It was compared with bottom-up (GUM) methods used for compiling country-scale greenhouse gas emission inventories (see *Compiling greenhouse gas emission inventories*). The first Bayesian-based inversion estimates of $\mathrm{CF_4}$ and $\mathrm{NF_3}$ ($\mathrm{CF_4}$= tetrafluoromethane, $\mathrm{NF_3}$= nitrogen trifluoride) from the East Asian region were obtained. In doing so, a class of successively weaker informative priors, largely removing the influence of available bottom-up information, was used. Consequently, the results of the study were predominantly driven by the data for the period covered but the use of a weak prior had a stabilizing influence on the results. LGC and VSL reviewed the material.

Preparation of calibration gas mixtures of ammonia in nitrogen using permeation. Permeation is one of several techniques for dynamically preparing calibration gas mixtures. It is a dynamic-gravimetric method, which implies that mass flow rates are used, together with information concerning the purity of the materials employed and the molar masses of the components to calculate the composition. One of the mass flow rates originates from the permeation tube ('permeation rate'), the other from a thermal mass flow controller. This example, for the preparation of a calibration gas mixture of ammonia in nitrogen, was developed by an IMBiH metrologist under the RMG scheme with support from chemists at VSL and mathematicians at NPL. The concept of permeation was used, the calculation and the associated uncertainty evaluation being given for the composition of the mixture expressed in amount fractions, as used in many high-end applications. A Bayesian approach for the uncertainty evaluation was used rather than applying the GUM or GUM-S1 since they are incapable of separating the effects of finite resolution and repeatability in the data.

<u>Energy</u>

Totalization of volume measurements in drinking water supply networks. Clean water and sanitation are among the sustainable development goals of the United Nations' 2030 agenda. Water supply networks involve net balances based on data relating to water inflows and outflows provided by metering equipment in many locations. This complexity makes measurement uncertainty a valuable tool to support the analysis of performance and risk related to these utilities. Using volume measurement data, experts in the field at LNEC and IPQ in collaboration with NPL uncertainty experts provided an example giving a straightforward approach for determining the total volume or flow rate at a point in the network. Using the GUM method, simple formulae were derived relating the measurement uncertainty to the number of measurement sites considered. *Figure 12* shows flow measurement experimental data obtained in a water distribution network during 6 days.

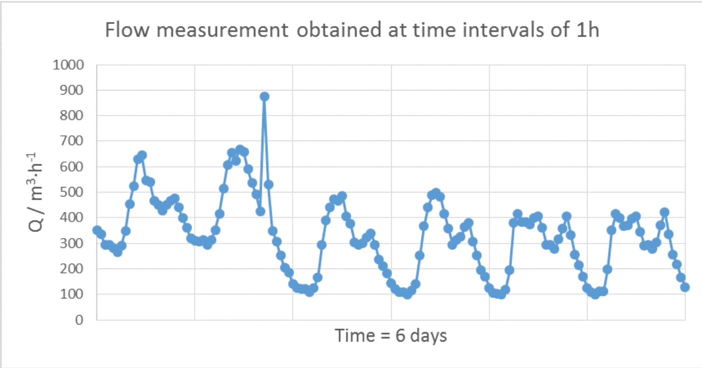

Figure 12 Input flow measurement experimental data obtained in a water distribution network during 6 days with time interval of sampling of 1 h

Flowmeter calibration using an orifice plate. An estimated €200 Bn of natural gas each year in the EU is measured by devices conforming to international standard ISO 5167. An orifice plate (Figure 13) is a key part of the measuring systems used. NEL used a large database, containing some 16 000 measured data values, to evaluate the uncertainty of the orifice-plate discharge coefficient given by the Reader--Harris/Gallagher (1998) equation taking account of the uncertainty in the data and the variability in manufacture permitted by ISO 5167-2. Since the calculated standard uncertainty did not exceed that in ISO 5167-2 by more than 0.01 %, the standard did not need to be changed and confidence in this important standard has hence been increased. The procedure applied here may have benefit in other situations where the uncertainty for an artefact is based on data from other similar artefacts. The example was reviewed by NPL, LNE, IPQ and VSL.

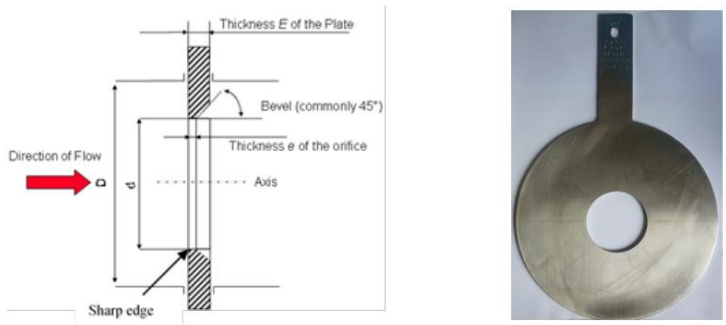

Figure 13 Diagram of an orifice plate and (right) an EEC orifice plate

Calibration of a sonic nozzle. Sonic nozzles are widely used to determine gas flow rates with high precision. They are internationally recognized as a calibration standard for gas flow meters and other flow measurement devices, which in turn facilitate traceable measurements such as in gas and oil pipelines, and in chemical, pharmaceutical and food industries. When calibrating a sonic nozzle, it is recommended to estimate the straight-line relationship between the discharge coefficient of the nozzle and the square root of the inverse Reynolds number for a gas. The slope and intercept of this relation characterise the nozzle, for which reliable estimates and uncertainties are mandatory for the use of the nozzle as a transfer or working standard. This example emphasized the importance of accounting for correlation for a reliable uncertainty evaluation. The use of common reference standards and instruments causes correlation among and between the discharge coefficient and the Reynolds number, and impacts significantly on the uncertainty of the characteristic parameters of the nozzle. It was demonstrated in detail how to evaluate the correlation, uncertainties and estimates for the input quantities of least-squares methods used by applying the Monte Carlo method. As a result, it was confirmed that sonic nozzles can validly be characterized using the GUM. The example was prepared by PTB with support from LNE, NEL and VSL, and constituted one of the *Straight-line calibration* examples in Objective 1.

Load loss and billing associated with fiscal metering. In electrical energy production, transformation, distribution and energy consumption, energy losses generated by transformers form the second largest part of the total losses in the distribution of energy. Measured power transformer losses should be accurate as these are often the object of guarantee and penalty in many contracts and hence play an important role in billing. Costs of losses in power transformers are comparable to the product cost and play a key role in the evaluation of the total costs of obtaining the energy at the point of use. Many European regulations set requirements for reliable loss values. For market surveillance, a fine must be paid if a power transformer facility exceeds the guaranteed limit for losses. International standard IEC 60076-19 recommends that the evaluated uncertainty should be less than the tolerance limit set out in regulations. Measurement uncertainty is often not considered in the agreement between the manufacturer and costumer on who will pay the fine in case the losses exceed the guaranteed value. The standard recommends that guarantee and penalty calculations should refer to the best estimated values of the losses without considering the measurement uncertainties, based on a shared-risk concept, where both parties are aware of and accept the consequences of non-negligible measurement uncertainty. Metrologists and uncertainty experts at IMBiH and VSL showed how the evaluation of load loss uncertainty can be conducted to improve existing practice for power load loss measurements. The GUM method was used for this purpose and validated using GUM--S1.

Energy efficiency and thermal comfort in buildings. Thermal comfort, according to international standard ISO 7730 Ergonomics of the thermal environment, expresses the degree of satisfaction with the thermal environment, a condition varying from person to person. Predicted mean vote (PMV), the average thermal sensation of a large group of persons in the same environment, based on data collected in a controlled climate chamber, was used to quantify thermal comfort. IPQ, LNEC and NPL collaborated to re--express (exactly) the complicated mathematical expression for PMV in the standard in simpler terms. Using the re-expressed model, they carried out the PMV calculation for typical scenarios and evaluated the associated uncertainty using GUM and GUM--S1. The probability distribution obtained using GUM--S1 was far from normal (Figure 14) whereas the GUM assumes normality, indicating the importance of applying a Monte Carlo method in such situations.

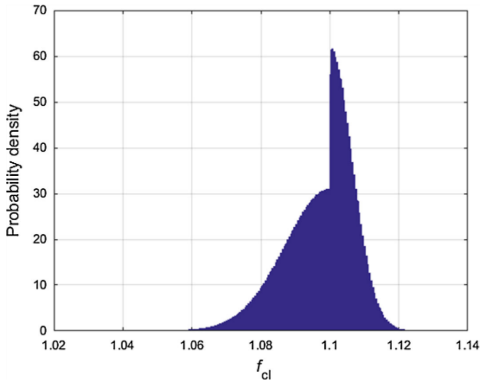

Figure 14 The probability density function for clothing surface area factor f_{cl}, one quantity in the PMV calculation, as produced by a Monte Carlo calculation

Reference material production and certification for gas mixtures. An essential element in the production of certified reference materials and proficiency test materials in batch form is the evaluation of the between bottle- homogeneity. (In)homogeneity accounts for the (small) differences in the property of interest between the bottles and including it in the uncertainty budget for the property value ensures that its estimate and associated uncertainty are valid for each bottle in the batch, rather than for the batch as a whole. The evaluation of between-bottle homogeneity is both a requirement in reference material production as well as in proficiency testing. Chemists and uncertainty experts at VSL and LGC prepared an example that gave the amount fraction of a component in a batch of gas mixtures. A Bayesian approach was used to analyse the data, the main measurand being the between-bottle standard deviation. Traditional analysis of variance failed to quantify the between-bottle homogeneity effect, whereas the Bayesian counterpart provided a probability distribution from which the between-bottle standard deviation could be derived. Open-source software was made available for the calculation.

Flow meter calibration using the master meter method. This example demonstrated the calibration of a gas flow measuring instrument by the so-called ‘master meter’ method, that is, by comparing the measured flow on a reference standard – the master meter (MM) and the measured flow on the meter under test (MUT); see Figure 15 (left). The measurements in this example were obtained using three measurement standards with different measuring ranges and one meter under test in the ‘SARAJEVOGAS’ Laboratory. Measurements of flow, pressure and temperature were obtained according to the provisions of European standards EN 12480 and EN 12261 for turbine and rotary gas meters. The RMG researcher from IMBIH, supported by NPL and VSL, estimated the measurement error of the meter under test for 10 flow rates and deduced the associated standard uncertainties using the GUM. It was confirmed using GUM-S1 that these results were valid – see Figure 15 (right).

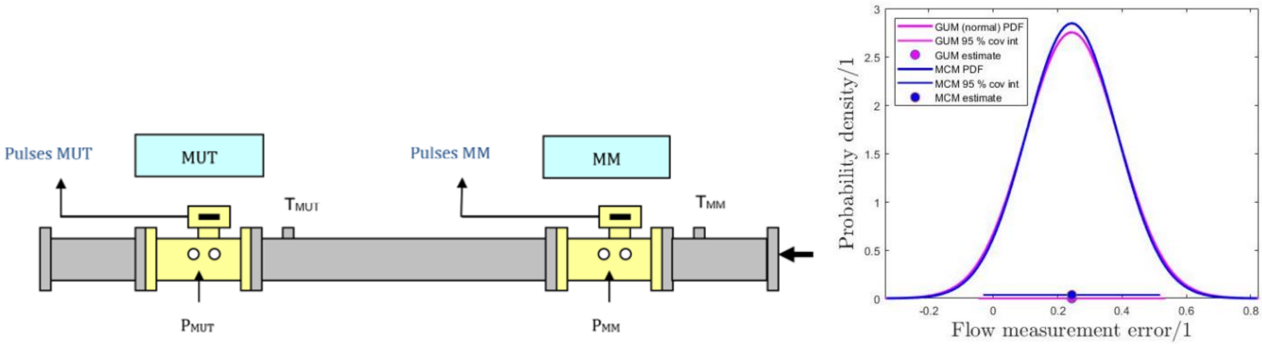

Figure 15 Set-up in the SARAJEVOGAS laboratory, where P denotes pressure and T absolute temperature of the gas, and (right) probability density functions for the GUM and Monte Carlo approaches

Pressure drop measurement. Correlations can sometimes be removed from an uncertainty evaluation by modifying the measurement model. The RMG researcher from IMBIH supported by NPL and UKAS demonstrated such removal by an example considering the pressure drop in a pressurized vessel, in which the effect of correlation between temperature measurements was eliminated. The example was derived from a real-world test involving gas at a pressure up to 50 MPa trapped in a sealed vessel. Probes were used to measure the pressure and temperature inside the vessel. The aim of the test was to show that the uncertainty due to gas leakage (pressure drop in the vessel) can vary for two cases that involved the same measured data but different approaches to evaluation. Correlations within this example arose from the application of the gas equation, a good approximation of the behaviour of many gases under many conditions. In this example, the uncertainty associated with an estimate of pressure drop was appreciably smaller (by over 70 %) when correlation was taken into consideration.

<u>Quality of life</u>

Effect on a computed quantity of considering a 2D or 3D image as a set of pixels or voxels. Two examples were considered. The first example, prepared by NPL with input from SMD, related to molecular radiotherapy where pixel size impacts on the area of a section of an organ or tumour given by the delineation of a 2D image made by an operator or an algorithm. It constituted one step in the estimation of the volume of a 3D region corresponding to a reconstructed 3D image of a region of interest (ROI) such as from a set of parallel planar sections. The area of a section was estimated from knowing the pixels through which the tumour boundary passes. A Fourier series was used to reconstruct the boundary as a smooth curve passing through such a set of pixels. The number of Fourier harmonics (6 in Figure 16) was chosen to balance fidelity and smoothness. The required area was estimated as that of the region enclosed by the Fourier curve and the associated uncertainty evaluated by propagating pixel uncertainties through the calculation using GUM and GUM-S1 approaches.

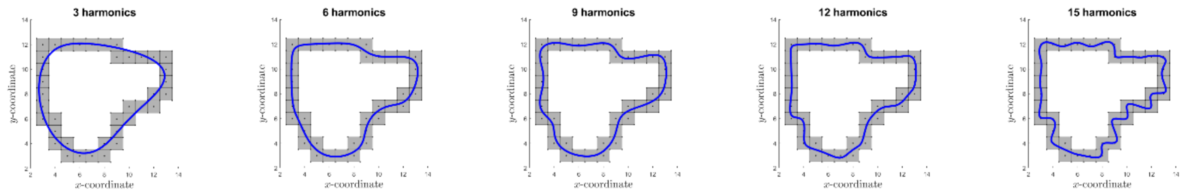

Figure 16 Fourier representations with 3, 6, 9, 12, 15 harmonics of the profile of a section of a tumour

The second example, developed by SMD with support from NPL, related to nanoparticle size in wastewater, daily life products, etc., where atomic force microscopy is used to measure nanoparticles deposited on a flat substrate. The size of a nanoparticle of interest was assessed as the height of the highest pixel in the pixels identified as belonging to the nanoparticle. The example extended example *Nanoparticle size by atomic force microscopy* and used results from it. Focus was placed on the model building and its interpretation by the practitioner for precise uncertainty evaluation or to keep uncertainty from this source under control. A key aspect of the model building related to the extent to which the pixel and nanoparticle 'overlap', Figure 17 showing some configurations. Its application in the classical GUM framework was explained, as well as a possible extension to a Bayesian framework.

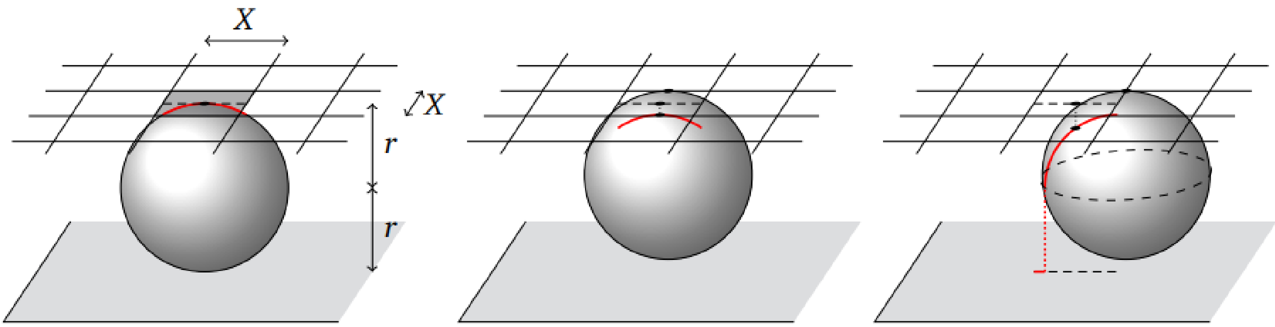

Figure 17 Some different relative positions of sampling grid (pixels) and the nanoparticle with the red line corresponding to the line of measurement: centered (left), off on one side (middle), most disadvantageous (right) relative positions

Cancer characterisation using magnetic resonance-based electrical properties tomography (EPT). An imaging technique – magnetic resonance-based electrical properties tomography (EPT) – monitors properties of biological tissues, to be used as cancer biomarkers. EPT results based on experimental data collected through an MRI scanner on a homogeneous cylindrical phantom provided by collaborator Philips Research Laboratories were analyzed by INRIM with support by NPL. For uncertainty evaluation purposes, repeated MRI scans of the phantom were analysed statistically to evaluate the covariance matrix of the EPT input, and then propagated through the EPT calculation using GUM principles.

Quantifying haemoglobin concentration. The total haemoglobin (Hb) concentration in blood is one of the most frequently measured analytes in clinical medicine because of its significance for evaluating the state of health of a human. Two high-accuracy- methods (the cyanmethaemoglobin and alkaline haematin method) for measuring total Hb concentration in blood were considered. For measurements across a range of values, the comparison of these methods was accomplished by contrasting a straight-line fit of representative data from the methods with the identity line. PTB, with input from LGC, applied the weighted total least-squares method (WTLS), accounting for the uncertainties and the correlation due to common instruments and standards. Uncertainties and correlations were propagated through to the estimate of the slope and intercept according to the GUM. The results were discussed, and recommendations given. Since, in metrology, two methods of measuring the same quantity are often to be assessed whether they agree, the principles of this example can be mimicked in other applications. The example constituted one of the *Straight-line calibration* examples in Objective 1.

Calibration of rheometers. Rheometry refers to the determination of properties of materials such as stresses and strains in various types of flow conditions with many applications such as in industrial process control and process modelling. Even for the simplest case of uniform flow, there are many unknowns and influence quantities that are not reflected fully in existing mathematical models. The example presented took account of all quantities judged to be influential and showed adequate statistical tools must be applied to evaluate the measurement uncertainty. The use of conventional recipes might bring erroneous results, which can have implications in the oil and mining industries or in health applications. The example developed by IPQ, NEL and VSL, with support from LNEC and NPL, applies to rheometer calibration with a highly non-linear viscosity model derived for rotational rheometer equipment. It related to the measurement according to DIN standard 53019 using GUM and GUM-S1. Regarding GUM--S1 as the 'gold standard', the GUM overstated the uncertainty in the calculated viscosity by 'only' 4 %. It was emphasized that this result applies only to the data in hand. The viscosity uncertainty could change drastically according to the measurement conditions such as torque, temperature stability and outer gap radius. An uncertainty budget such as that derived can assist greatly in identifying the most significant contributions to the combined uncertainty.

Quantifying extremely small flow rates. Measurement quality is vital in medical applications, in dealing with life issues and the well-being of society from oncology to newborns, and more recently to patients of the Covid-19 pandemic. In such situations the accuracy of the quantity of fluid delivered according to a prescribed dose can be critical. Microflow applications are growing in importance for a wide variety of scientific fields, namely, drug development and administration, Organ-on-a-Chip, or bioanalysis, with accurate and reliable measurements, a tough challenge for very small flow rates, from 1000 µL h^{-1} down to 1 µL h^{-1}. One challenge is dealing with the measurement of a quantity such as flow rate that is inherently positive, especially when using a measuring system that due to inevitable errors can provide negative values. For a measurement model of a syringe pump, with 12 influence quantities, IPQ, NEL and VSL, with input from LNE, LNEC and NPL, analyzed two sets of actual data using the GUM, GUM-S1 and Bayesian methods. The methods performed comparably when there were few negative readings. When there were a significant number of negative readings, GUM and GUM--S1 performed poorly, under-estimating the measured flow by 40 %, whereas Bayes, using a prior distribution that respected the positivity of the measurand, provided a meaningful result (Figure 18).

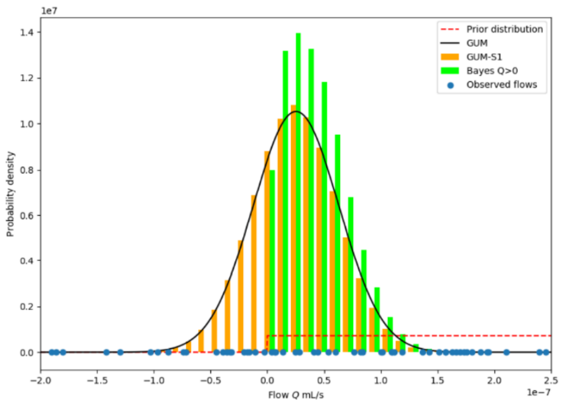

Figure 18 Probability density functions for all three approaches illustrating the importance of using a Bayesian approach to treat a positive quantity

Specification requirements of temperature in medical applications. Measurement guides and standards are an essential source of information in most areas of testing. However, the terminology used is often inadequate in situations where conformity decisions are to be made that are consistent with the GUM or standards used to support accreditation such as international standards ISO/IEC 17025 and ISO 15189. Using data provided by Health Facilities Scotland, two examples were developed by UKAS and NPL uncertainty experts to illustrate how such guidance and standards requirements might be interpreted. These were not selected as 'bad' examples but chosen since they show that even in otherwise authoritative and carefully drafted guidance there can still be issues of interpretation. The aim was to support drafting of future revisions and to indicate options when working with existing versions. The first example concerned the decontamination of medical devices by steam sterilization. In this case, allowable limits on measurement uncertainty were defined. The only issue was to establish or formalize a clear and unambiguous rule to explain the role that measurement uncertainty must take in the decision process. The second example arose during the mapping and monitoring of cold storage systems for blood products, where temperature is measured to decide whether an alarm condition is met. In this example, what at first may seem to be a clear 'accuracy' requirement was shown to be incomplete, requiring further information or assumptions before conformity can be decided.

Industry and society

Exhaust thrust in a turbofan jet engine. Acoustic sensor systems used for measuring exhaust thrust (and four other measurands) in a turbofan jet engine are more resistant than other measuring systems to the harsh conditions (high temperatures, pressures, etc.) within the engine. The measurands, shown as rectangular boxes in Figure 19 (right), were expressed by NPL experts, aeronautical engineers at RR and academics at collaborator Virginia Tech in terms of the geometrical configuration of the sensors, times of flight between sensor sources and receivers, and properties of the medium – Figure 19 (left). The measurement model comprised five sub-models, as implied by the computational workflow in Figure 19 (right). For the first time, a full uncertainty analysis of such a set-up was produced. The GUM was applied but, since derivatives were very difficult to obtain analytically, requiring repeated application of the chain rule of differential calculus to ensure all dependencies were fully considered, an automatic method of differentiation (the complex step- method) was used. Estimates of the measurands were obtained for representative test data from RR, and the associated uncertainties evaluated. It was found that by modifying the sensor configuration geometry, uncertainties could be reduced by 60 %. All calculations were confirmed using GUM-S1.

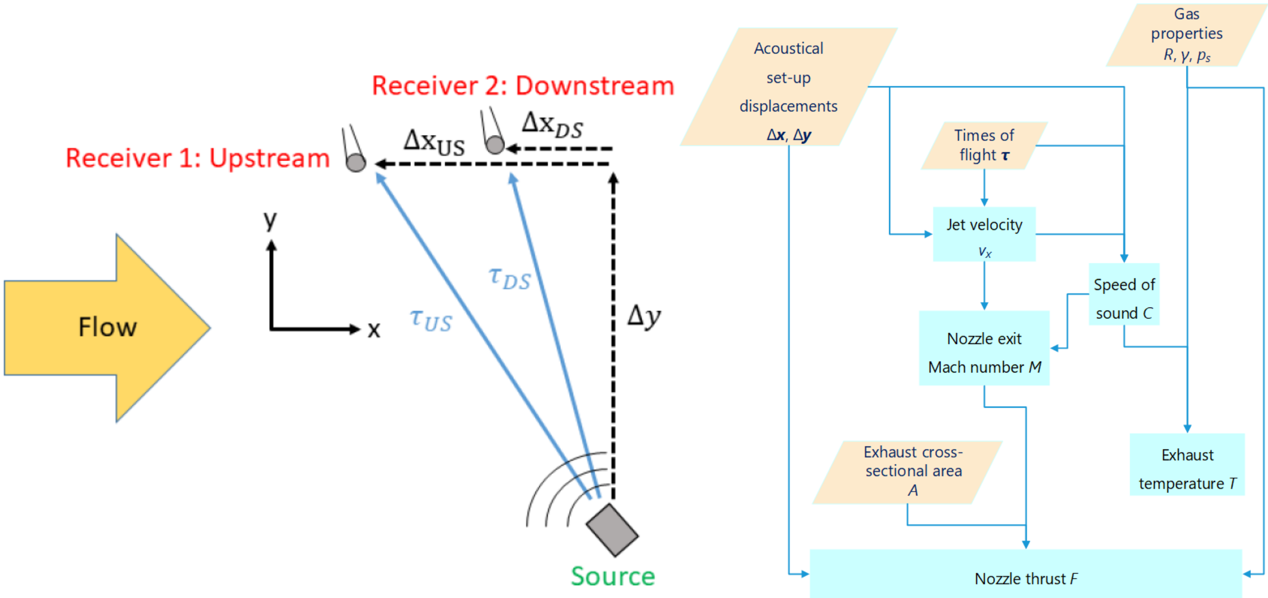

Figure 19 Upstream and downstream acoustic component displacements and times of flight and (right) computational workflow with input quantities shown in trapezoidal boxes and output quantities (measurands) as rectangular boxes

Calibration of a torque measuring system. Torque measurement is required in many fields, particularly in the automotive industry where rotary torque sensors are used for strain measurement, and testing of clutches, gearboxes and suspension systems. This example addressed the straight-line calibration of a torque measuring sensor against a reference system using measurements taken at various torque values. For each torque value, a single measurement result of the reference system was available, together with results of repeated measurements of a sensor to be calibrated. The goal was to determine a linear relationship that relates results of the torque measuring sensor with those of the reference system. The data were analyzed by PTB with support from ACCREDIA and LNE by applying (i) ordinary and weighted least-squares estimation in combination with an uncertainty evaluation following the GUM, and (ii) Bayesian inference. Analytic expressions were given for the Bayesian uncertainty analysis to simplify its application. The results obtained by the different approaches were discussed and recommendations given. The data used were partly taken from the guideline VDI/VDE 2600 part 2. The example constituted one of the *Straight-line calibration* examples in Objective 1.

Harmonizing documentary standards for hardness testing. Hardness testing is a valuable tool for the assessment of materials, quality control of manufacturing processes and in research and development work. It gives an indication of a material's properties, such as strength, ductility and wear resistance, applying to all areas where metals are used. The evaluation of uncertainty in hardness verification was treated considering that repetition of tests at the same location is in general not possible. In order to improve the methods provided by standards for the Vickers, Knoop, Rockwell and the Brinell Hardness Test, and the Instrumented Indentation Test, BAM, IMBiH and ACCREDIA with input from NPL and LNE developed a harmonized GUM compliant approach for uncertainty evaluation for application to all hardness tests. The GUM was used for the uncertainty evaluation and the results confirmed using GUMS1. Further, proposals were made on how to use measurement uncertainty in hardness testing in conformity assessment ensuring consistent alignment with the provisions of the GUM.

Mobile optical measurement systems (MOMS) are used in the automotive, aerospace, and structural engineering industries. The measuring system studied by LNEC and NPL with input from IPQ contained linear charge-coupled device (CCD) cameras, in various spatial positions, permitting the simultaneous observation of infrared light emitting diodes (LEDs) located in a region of interest. By applying triangulation, the system determined the spatial location of the LEDs. MOMS permit (in situ non-contact manual or automatic) positional measurement of complex geometrical objects. Using actual measurement data, the example determined the LED locations and used the GUM approach for uncertainty evaluation. It was shown that the reference standards used made the largest contribution to the calibration uncertainty.

Quantification of ephedrine in anti-doping testing. Many stimulants are prohibited in sport. An anti-doping laboratory's evaluation of the standard uncertainty associated with the measured concentration in urine samples of the stimulant ephedrine at levels close to the threshold was described by WADA regulators, VSL chemists and NPL uncertainty experts with input from collaborator INMETRO. For this purpose, laboratory validation data for intermediate precision and bias were used. The evaluation was verified through the laboratory's participation in three rounds of WADA's proficiency testing programme, EQAS. The study concluded with a statement concerning the acceptability of the evaluated standard uncertainty in accordance with WADA requirements.

Multiplexed data-acquisition systems. The magnitude of the impedance of an electrical component is often measured to characterize behaviour at various operating frequencies. A device under test can be used as a current shunt to convert an unknown AC current into an AC voltage signal and then its impedance magnitude Z_X must be known at the frequencies of the current. The measurement of an electrical impedance by comparison to a standard resistor was considered (Figure 20) by ACCREDIA with support from NPL and collaborator BIPM. The voltage signals across the impedance and the resistor were acquired using a multiplexed data-acquisition (DAQ) board. Measurement uncertainty was evaluated according to the GUM. The proposed measurement technique was effective since, using a low-cost commercial multiplexed DAQ board, it was possible to obtain a relative uncertainty of the measurand Z_X that was only an order of magnitude larger than that of the standard resistor R_S used, which is the most accurate and expensive component of the proposed circuit. To obtain a relative standard uncertainty comparable to that of R_S, it would be necessary to use a board with independent (non-multiplexed) channels, thus minimizing cross-talk effects, and with a programmable gain amplifier characterized by a higher common mode- rejection ratio. The cost of such a system would become an order of magnitude higher.

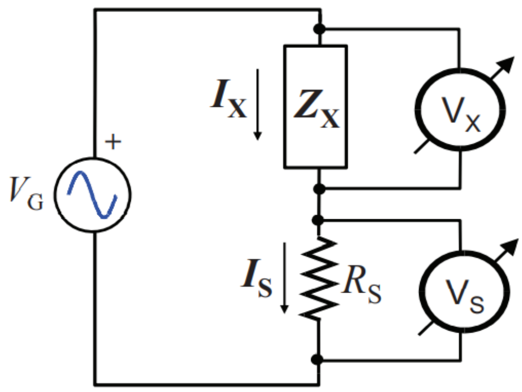

Figure 20 The basic circuit for measurement of the impedance magnitude

Temperature measurement with a micro-controller-based board. The application of thermal cycles inside a climatic chamber constitute common tests to verify the correct behaviour of a device in the operating temperature range stated by the manufacturer. During these tests, the air temperature inside the climatic chamber and the temperature of various parts of the device under test are measured using sensors such as thermistors, resistive temperature detectors, thermocouples or integrated electronic devices. The measurement of the temperature inside a climatic chamber was undertaken by ACCREDIA with input from NPL by means of a measurement chain that included a Resistive Temperature Detector (RTD), a simple conditioning circuitry and a commercial micro-controller (µ-C) based board. The advantages of such a system relate to the characterization of the Analogue-to-Digital Converter internal to the µ-C board and the implementation of a dithering technique. Measurement uncertainty was evaluated according to the GUM. For actual data relating to the use of Pt100 as a sensor, temperature standard uncertainties were quantified over the full operating range 10 °C to 70 °C.

Objective 2: Summary of key outputs and conclusions

Key outputs and conclusions for Objective 2, which was fully achieved, are given below. The outputs added considerable value beyond the state of the art through extensive collaboration between the national metrology institutes, designated institutes, regulators and industry in the project.

Choice of method of uncertainty evaluation. The avoidance of the dangers of blind adherence to the GUM approach to uncertainty evaluation, as opposed to the use of the Monte Carlo or Bayesian approach was emphasized throughout. Clear guidance on the issue was given in many examples such as those concerned with air monitoring, energy efficiency and thermal comfort in buildings, and jet engine thrust.

Sources of correlation. As in Objective 1, correlation was treated in many of the examples in Objective 2. Instances include correlation caused by (i) use of a single calibration gas mixture from which dilutions are made in the calibration of an analyzer for nitrogen oxides, (ii) use of the same sensor in temperature observations in pressure drop measurement, (iii) application of the same gas equation in pressure drop measurement, and (iv) the likeness (to four or five significant digits) of activity data arising in greenhouse gas emissions' inventories. Instance (iv) constituted *perceived* correlation in data of questionable provenance.

Different measurands or measuring systems. Whilst some of the examples in Objective 1 (and in Objective 2) compared the application of different methods for uncertainty propagation (GUM, GUM-S1 and Bayes), some examples in Objective 2 compare the use of different measurands or measuring systems. For instance, in estimating average areal rainfall in a region, the arithmetic mean method, Thiessen polygons' method and the isohyetal method, which use distinct interpretations of the physical quantity in relation to the geometric context, were compared. Also, the use of weighing gauges and tipping-bucket gauges were contrasted. The differences observed across measurands and measuring systems in terms of the provided estimates and uncertainties are relevant to making choices of measurand or measuring system.

Skewed probability distributions. The differences between results obtained by applying GUM and MC are most marked when the probability distribution for the measurand is highly skewed. Relevant instances are the examples on low masses of benzo[a]pyrene and thermal comfort in buildings. In such cases the estimate and uncertainty provided by MC are much to be preferred; MC is regarded as a gold standard for uncertainty evaluation when prior information on the measurand is not available or used.

Including prior information. Ignoring physical information available prior to measurement can produce unreliable results. An important instance arises where instrument imperfections yield negative readings, but the quantity is known to be positive. For instance, it is shown in the healthcare example on the quantification of extremely small flow rates that a Bayesian treatment gives meaningful results in cases where GUM and MC fail. The relevance of a weakly informative Bayesian prior as used in the example of inverse modelling in greenhouse gas emissions' inventories is emphasized. Here, the old edict of 'let the data speak' applies but the weak prior stabilizes the solution, just as regularization does in frequentist statistical calculations.

Simple formulae. Simplicity is often desirable in metrology calculations and in putting messages across to practitioners. In the example of totalization of volume measurements in drinking water supply networks, the simple formulae derived are readily capable of spreadsheet implementation. Especially important is that the assumptions on which the formulae are based are clearly stated.

Clarity in standards and regulations is a further important consideration, an aspect covered by several examples. In the example on specification requirements of temperature in medical applications, current guidance is open to interpretation. Advice was given on clarification in the context of supporting drafting of future revisions and in indicating clear options when working with existing versions.

Supporting documentary standards. Many of the examples are in a form that can be considered by standardization committees (i) to extend current provisions in standards and guides to include measurement uncertainty evaluation, (ii) to incorporate uncertainty considerations where none existed before, or (iii) to support standards in their current statements of measurement uncertainty. A key contribution in category (iii) is the example of flowmeter calibration using an orifice plate. Detailed analysis of a large database confirmed the statement in the current standard ISO 5167-2 that the standard uncertainty in the discharge coefficient should not exceed 0.1 %, and hence increased confidence in this important standard.

5. Impact

Over 20 presentations were made on project activities at international conferences. The compendium containing 42 uncertainty examples is freely available on the website. To support these documented examples, a 2-day workshop was held in January 2020 and a 2-day online measurement uncertainty training course in March 2021. The training course was open to all but particularly targeted at Western Balkan countries. It assisted in empowering training courses run by LNE, NPL, PTB and UKAS. An unforeseen benefit of the Coronavirus pandemic was that the course was held online, which permitted an attendance of over 160 rather than the envisaged 25. These activities were extremely helpful in extending the understanding of measurement uncertainty evaluation to a wider circle of end-users. Training was also delivered at eight smaller events. There will be specific longer-term impact in many areas including greenhouse gas emission inventories, energy efficiency and thermal comfort in buildings, neonatology and cancer treatments, and doping tests. The project website includes information on the scope and objectives of the project, and the compendium, as well as uploaded presentations and links to open-access papers published as part of the project.

Four popular articles were published to bring the outputs of the project to a wider audience:

A. Ribeiro and M.G. Cox, EUROLAB supports the EMPIR Special Project “Examples of Measurement Uncertainty Evaluation” (EMUE), EUROLAB Newsbriefing, April 2021.

M.G. Cox, EMPIR project EMUE Examples of Measurement Uncertainty Evaluation, ILAC Newsletter, April 2021.

Francesca R. Pennecchi, Michela Sega and Adriaan van der Veen, EMPIR PROJECT EMUE: EXAMPLES OF MEASUREMENT UNCERTAINTY EVALUATION, CITAC News, April 2021, pages 55-57.

A. Carullo, S. Corbellini and A. Vallan, Misura di temperatura con scheda a micro-controllore (Temperature measurement with micro-controller based board), Tutto Misure https://www.tuttomisure.it/, Issue 2/21, pages 19-25.

Impact on industrial and other user communities

Work on the uncertainty analysis related to the thrust generated by a turbofan jet engine (and other measurands) is regarded as ground-breaking by Rolls-Royce and their academic subcontractor Virginia Tech with substantial improvements possible in the uncertainty associated with exhaust thrust, exhaust temperature and other quantities.

A study of an anti-doping laboratory’s evaluation of the uncertainty associated with the measured concentration of the stimulant ephedrine at levels close to the threshold in urine samples was completed. That stimulant is prohibited for use in sport. The evaluation was verified through the laboratory’s participation in three rounds of the World Anti-Doping Agency’s (WADA) proficiency testing programme. The study concluded with the acceptance of the evaluated standard uncertainty in accordance with WADA requirements.

An example on electric property tomography, a quantitative imaging method based on magnetic resonance imaging (MRI), has been developed in cooperation with Philips, a leading manufacturer of MRI scanners. Philips has been given access to the outcomes including the full quantification of the repeatability of the results.

The work on mobile optical measurement systems (MOMS) and on measurement uncertainty in a multiplexed data-acquisition system has been communicated to practitioners in the automotive and aeronautical industries and in the electronic industries.

Impact on the metrology and scientific communities

The compendium of 42 examples produced in the project has been made available in its entirety to the JCGM, the material within it relating to the further development of the JCGM guides on conformity and interlaboratory comparisons and the JCGM examples document. In addition, the compendium contains supporting tutorial material on Monte Carlo, Bayesian statistics, reporting measurement results, and handling correlation. JCGM/WG1 has formed a sub-group to decide the parts of the compendium to include in three of its guidance documents.

It is expected that use will be made of the compendium, which is fully indexed for reference purposes, by the many metrology and scientific sectors addressed by the project. Its availability through the JCGM will ensure considerable use noting that the number of downloads of the JCGM guides is some 5 000 per month. It is anticipated that the number of downloads of the compendium will be comparable.

Impact on relevant standards

The project made input to over ten national and international standards committees, many of which have made statements of need for improved examples, and to Eurachem, European Accreditation, EUROLAB and UKAS. Several partners are members of ISO/TC 69, the international standards committee concerned with the application of statistical methods, to which presentations on project progress have been made at its annual plenary meetings and its uncertainty working group. ISO/TC 69 is engaged on many standards involving statistical methods for uncertainty evaluation and already technology developed within the project is having an impact, The committee intends to make use of the project's results in upgrading one of its technical specifications, on straight-line calibration, to full international standard.

The coordinator of the project is the liaison officer between ISO/TC 69 and the JCGM. He has regularly reported on JCGM activities, especially those activities to which this project makes input, at TC 69 plenary meetings.

Several presentations have been made to national and international standards committees, which in many cases will have impact on the standards they are producing

Specific input made to standards bodies included an example relating to gas analysis, an example contained in a guide on analytical chemistry for evaluation of risks of conformity assessment of a multicomponent material or object due to measurement uncertainty.

Longer-term economic, social and environmental impacts

In implementing methods of uncertainty evaluation, many practitioners benefit from learning from example. By providing over 40 worked examples in diverse areas of application and in the energy, environment and health care sectors, and in industry and society, in a categorized 600-page compendium, the project will assist end-users in this way. Since correlation between quantities is a topic that many find difficult to understand and handle, the considerable emphasis placed on this topic will be of particular value to the community. The training-course material provided and delivered in the project will be of much interest to many organizations. Particular benefit will accrue to some ten international standards' committees to which project outputs have been provided; many of these committees will use that material in improving their current standards and in developing new standards that take measurement uncertainty into consideration. In particular, the Joint Committee for Guides in Metrology, responsible for the GUM suite of documents, will use a sizeable fraction of the material in the compendium in its guidance documents.

6. List of publications

J.A Sousa, E. Batista, O Pellegrino, A S Ribeiro, L L Martins. Method selection to evaluate measurement uncertainty in microflow applications, Journal of Physics: Conf. Series, 1379 (2019), 012033 (doi: 10.1088/1742-6596/1379/9/012033).

A.M.H. van der Veen, M.G. Cox. Getting started with uncertainty evaluation using the Monte Carlo method in R. Accred Qual Assur 26, 129–141 (2021). https://doi.org/10.1007/s00769-021-01469-5

A.S. Ribeiro, M.C. Almeida, M.G. Cox, J.A. Sousa, L. Martins, D. Loureiro, R. Brito, M. Silva, A.C. Soares. Role of measurement uncertainty in the comparison of average areal rainfall methods, Metrologia 58 (2021), 044001, https://doi.org/10.1088/1681-7575/ac0d49

J. Pétry, B. De Boeck, N. Sebaïhi, M. Coenegrachts, T. Caebergs, M. Dobre. Uncertainty evaluation in atomic force microscopy measurement of nanoparticles based on statistical mixed model in a Bayesian framework, Meas. Sci. Technol. **32** (2021) 085008, https://doi.org/10.1088/1361-6501/abe47f

J.A. Sousa, E. Batista, S. Demeyer, N. Fischer, O. Pellegrino, A.S. Ribeiro, L.L. Martins, Uncertainty calculation methodologies in microflow measurements: Comparison of GUM, GUM-S1 and Bayesian approach, Measurement 181 (2021), 109589, https://doi.org/10.1016/j.measurement.2021.109589.

This list is also available here: https://www.euramet.org/repository/research-publications-repository-link/