FINAL PUBLISHABLE JRP REPORT

JRP-Contract number	NEW04
JRP short name	Uncertainty
JRP full title	Novel mathematical and statistical approaches to uncertainty evaluation
Version numbers of latest contracted Annex Ia and Annex Ib against which the assessment will be made	Annex Ia: V1.1 Annex Ib: V1.1
Period covered (dates)	From 1st August 2012 To 31st July 2015
JRP-Coordinator	
Name, title, organisation	Prof. Dr. Markus Bär (PTB)
Tel:	+49 (0) 30 34817687
Email:	markus.baer@ptb.de
JRP website address	https://www.ptb.de/emrp/new04-home.html
Other JRP-Partners	
Short name, country	CMI, Czech Republic FORCE, Denmark INRIM, Italy JV, Norway LGC, United Kingdom LNE, France NPL, United Kingdom SP, Sweden VSL, The Netherlands
REG1-Researcher (associated Home Organisation)	
Dr. Carlo PALMISANO (Universita degli Studi di Torino (UNITO), Italy)	Start date: 1st October 2013 Duration: 12 months

Report Status: PU Public

EMRP
European Metrology Research Programme
Programme of EURAMET
The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union

TABLE OF CONTENTS

1 Executive Summary

Introduction

Measurement uncertainty evaluation is fundamental to metrology, because without it no assessment of the reliability of a measurement can be given and no comparisons of measurement results are possible, either among themselves or with reference values given in a specification or a standard.

The Problem

The evaluation of measurement uncertainty is a critical part of quality management systems in all industries that employ measurement technology. Inaccurate evaluation of measurement uncertainty has important economic consequences for calibration and measurement activities. In calibration reports, the magnitude of the uncertainty is often taken as an indication of the quality of the laboratory, and smaller uncertainty values generally are of higher value and of higher cost.

In many cases best practice is covered by the existing GUM guidelines. There are, however, commonly encountered situations for which the GUM guidelines are not suitable. Development of guidance for these situations presents a significant mathematical challenge but brings benefit to a wide range of end users. The situations considered in this project were inverse and regression problems, uncertainty evaluation in computationally expensive problems and the role of measurement uncertainty in conformity assessment and decision making.

The current lack of guidelines means that in some cases uncertainty is not being considered, which can potentially lead to risky decisions. In other cases they are over-estimating uncertainty in order to be safe, and so better uncertainties would reduce costs. The lack of uncertainty evaluation can be a barrier to use of models as part of certification processes (e.g. aerospace) which would reduce R&D costs.

The specific needs require a co-ordinated effort to obtain reliable uncertainties, to ensure harmonisation and to develop a consistent application framework throughout Europe. A long-term infrastructure, such as a European Centre, to enable cooperation on mathematical and statistical topics relevant to metrology is needed to ensure dissemination of mathematical and statistical expertise for uncertainty evaluation for European NMIs and European industry.

The Solution

The project has established a Bayesian approach to treat uncertainties arising in regression problems, explored sampling methods and surrogate models in systems with computationally expensive model functions and addressed conformity assessment and reliable decision making through case studies including analytical technologies for biochemistry and biotechnology, transport processes, industry and regulation. The Good Practice Guide “A Guide to Bayesian inference for regression problems” [1] provides a general introduction about how to deal with uncertainty evaluation in regression. Methods and suggestions on how to select the prior distribution in relevant practical situations are given. Methods for uncertainty evaluation in computationally expensive problems cover smart sampling procedures and surrogate models and are summarized in the Good Practice Guide “Uncertainty evaluation for computationally expensive models” [2]. Finally, new perspectives on the role of measurement uncertainty in conformity assessment and decision making have been gained by extending existing approaches to multivariate problems, qualitative data, computationally expensive systems and the inclusion of measures of impact – the results were summarized in the Good Practice Guide “Decision making and conformity assessment” [3].

Impact

General dissemination has been achieved through 22 scientific journal papers, 47 conference contributions, and 8 trade journal articles. In summer 2015, a training course provided the first opportunity for experts from European NMIs, as well as interested stakeholders from industry, universities, regulatory bodies and NMIs outside Europe. This interaction will help encourage the application of mathematics and statistics to challenging uncertainty evaluation problems beyond the partners of the project. MATHMET2014, an International Workshop on Mathematics and Statistics for Metrology was held in PTB which gathered about 75 scientists from NMIs in Europe and overseas as well as colleagues from universities and research institutes in a lively exchange on metrology-related mathematical research themes.

Dissemination is also achieved by free distribution of 3 Best Practice Guides and 4 software packages through the project website and the MATHMET website. Best practice guides on the three topic areas give comprehensive summaries of the main findings in combination with tutorial introduction for practitioners, and are available for free download from the project website. The results of this project will strengthen European capabilities for innovation by enabling traceability for modern metrology and measurement techniques and by strengthening European NMIs role in international organisations (eg in the Joint Committee for Guides in Metrology (JCGM) – Working Group 1 on Uncertainty in Measurement). Product testing, safety regulations, medical diagnosis and drug testing will benefit from the procedures for reliable uncertainty evaluation, decision-making and conformity assessment that were developed in this project.

In the long term it is likely that this new guidance will be incorporated into GUM suite of documents, probably as a supplement; although the revision of the GUM is a long complicated process and will take many years.The BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, which are representative bodies for the standards in different sector of industry and research, had identified the need for this work for their sectors. Their nominated representatives constitute the JCGM Working Group 1 (JCGM/WG1) on the Expression of Uncertainty in Measurement.

The European NMI scientists involved in the project are in dialogue with international and national technical committees and the wider metrological community. The results of the project were forwarded regularly to stakeholders, which included large energy generators, industrial companies, SMEs, instrument makes and research groups, and feedback has been received. The case studies have been developed in intensive exchange and resulted in seven joint publications.

Collaboration between European NMIs with mathematical and statistical expertise is essential to ensure wide take-up of the project outputs and to maintain Europe's current leading role in mathematics for metrology. Four key members of the consortium founded *MATHMET – A European Centre for Mathematics and Statistics in Metrology* as a platform to support and encourage such collaboration. MATHMET will continue to disseminate results internationally through its webpage (www.mathmet.org), and via conferences and workshops.

2 Project context, rationale and objectives

Need for the project:

The evaluation of measurement uncertainty is a critical part of quality management systems in all industries that employ measurement technology. Inaccurate evaluation of measurement uncertainty has important economic consequences for calibration and measurement activities. In calibration reports, the magnitude of the uncertainty is often taken as an indication of the quality of the laboratory, and smaller uncertainty values generally are of higher value and of higher cost.

In many cases best practice is covered by the existing GUM guidelines. There are, however, commonly encountered situations for which the GUM guidelines are not suitable. Development of guidance for these situations presents a significant mathematical challenge but brings benefit to a wide range of end users in fields such as biochemistry, biotechnology, transport processes, industry and regulation. The situations considered in this project were inverse and regression problems, uncertainty evaluation in computationally expensive problems and the role of measurement uncertainty in conformity assessment (i.e. testing compliance with standards) and decision making.

The current lack of guidelines means that in some cases uncertainty is not being considered, which can potentially lead to risky decisions. In other cases they are over-estimating uncertainty in order to be safe, and so better uncertainties would reduce costs. The lack of uncertainty evaluation can be a barrier to use of mathematical models as part of certification processes (e.g. aerospace) which would reduce R&D costs.

The specific needs require a co-ordinated effort to obtain reliable uncertainties, to ensure harmonisation and to develop a consistent application framework throughout Europe. A long-term infrastructure, such as a European Centre, to enable cooperation on mathematical and statistical topics relevant to metrology is needed to ensure dissemination of mathematical and statistical expertise for uncertainty evaluation for European NMIs and European industry.

Regression and inverse problems

Regression and inverse problems arise when the quantity of interest cannot be measured directly, but has to be inferred from measurement data using a mathematical model that relates the quantity of interest to the data. Such problems occur in many applications from everyday calibration tasks to advanced techniques such as the enzyme-linked immunosorbent assay (ELISA) in biochemistry and scatterometry, a key technology in nanometrology, to name just two important examples from rapidly growing areas. There is a need for new tools as well as for guidelines and standards for uncertainty evaluation in such problems.

Computationally expensive systems

Many important applications in metrology are described by model equations whose numerical solutions are computationally expensive. Examples comprise the Navier-Stokes equations for fluid flows and other transport equations such as the heat equation. Many of these computationally expensive systems are strongly nonlinear. Linear approximations for uncertainty evaluation, as suggested by the GUM, can be applied, but since the results may be invalid, simulations using a Monte Carlo method are required, as is recommended by Supplement 1 to the GUM (GUM-S1). However, current Monte Carlo methods are based on a large number of model evaluations and consequently cannot be immediately applied to computationally expensive systems. As a result, uncertainty statements of measurements in such systems are usually either missing or based on crude approximations.

Conformity assessment and decision-making

Many measurements are made to provide an objective basis for decisions about a product or a process, for example compliance with a regulation related to environmental emissions, decisions as to whether or not a person is infected with a disease, assessment of compliance of a product with manufacturing tolerances, or the water content in oil for legal metrology and fiscal purposes. The inevitable presence of measurement uncertainty leads to the risk of incorrect conformity decisions for consumers, suppliers and the public at large. Hence, there is the need to make reliable decisions of conformity given relevant measurement results with associated uncertainty statements and to ensure the consistent application of decision-making techniques. This holds particularly for multivariate cases, which are currently not addressed in guidelines and standards. One example of such a multivariate case of conformity assessment is the decision as to whether a specimen meets its design, when this design is not described in terms of a single parameter, but in terms of several parameters (e.g. the several critical dimensions of periodic nanostructures).

Stakeholder requirements

At the international level, JCGM Working Group 1 (JCGM/WG1), responsible for the maintenance of the Guide to the Expression of Uncertainty in Measurement, has identified the need for research in the areas of Monte Carlo methods, regression and inverse problems, conformity assessment and the application of expert and prior knowledge [7]. JCGM/WG1 represents the organizations BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML. National, professional societies of engineers and accreditation bodies such as VDI/VDE (DE), IET (UK), IMDR (F), and DAkkS (DE) have recognised the need for further development of uncertainty evaluation methods. These bodies maintain specialised committees dealing with the topic of uncertainty evaluation that regularly seek advice from NMI experts. Pre-normative research about the treatment of measurement uncertainty has also been increasingly recognised as essential in several sector-specific international standardisation groups, ranging from fire safety (ISO TC 92) to healthcare (ISO TC 173).

European and international associations dealing with best-practice guides on the use of computationally expensive models, e.g., finite element models and models for turbulent flow have started to address questions of uncertainty. For example, NAFEMS has set up a special interest group dealing with stochastic aspects and uncertainty.

European metrology infrastructure

Without this JRP the required extension of the mathematical infrastructure for metrology in Europe would not have been possible, with no single European NMI being able to address the need on its own. Collaboration between European NMIs with mathematical and statistical expertise is essential to ensure wide take-up of the project outputs and to maintain Europe's current leading role in mathematics for metrology.

Objectives

Based on the requirements stated above the project addressed the following main scientific and technical objectives:

- To develop methods for the evaluation of uncertainty in the context of inverse problems and regression problems that arise throughout metrology.
- To develop methods for uncertainty evaluation in metrological applications with computationally expensive model functions
- To develop methods for conformity assessment and reliable decision-making that incorporate knowledge of the measurement uncertainties;

The project also focused on applying these methods to challenging applications by carrying out case studies where there was a pressing need for new uncertainty evaluation methods.

Finally, the project aimed to build on an existing network of mathematical experts active in the *EURAMET Focus Group Mathematical and Software Tools for Metrology* project and lay the foundation for a *European Centre for Mathematics and Statistics in Metrology.* The Centre would disseminate state of the art methods to European NMIs, industry and other organisations and ensure that the momentum developed by the project is carried forward and that impact can be realised way beyond the end of the project.

The project consortium comprised a mixture of NMIs with substantial track records in innovation in mathematics and statistics for metrology, and smaller NMIs, who have both benefitted from this expertise and provided example applications. Beyond the technical objectives listed above, the project has laid the foundations for MATHMET - The European Centre for Mathematics and Statistics in Metrology based on a subset of four members of the consortium that will continue to provide support in uncertainty evaluation to metrologists and other key tasks for mathematics in metrology throughout Europe in the future. The MATHMET organisation is open for further Euramet member states resp. national metrology institutes and designated institutes to join as members.

3 Research results

Objective 1: Regression and inverse problems (coordinated by PTB)

Regression and inverse problems arise when the quantity of interest cannot be measured directly but has to be inferred from measurement data using a mathematical model that relates the quantity of interest to the data. Novel statistical methods for regression and inverse problems were developed within this project in connection with real-life case studies. The methods are based on Bayesian statistics and address all relevant steps of a Bayesian analysis, namely the specification of a statistical model, the selection of a prior distribution, the numerical calculation of the results and the determination of their sensitivity. Software for Bayesian linear regression (lead author PTB with input by NPL) and the computation of calibration curves (author INRIM with initial inputs and software testing by CMI) was also created in this work package. The results were summarized in the Good-Practice Guide “A Guide to Bayesian inference for regression problems” [1], which contains the complete treatment of the considered case studies. In the following, these cases studies are introduced and the novel developments are described.

Case study 1: Normal linear regression (collaboration NPL and PTB)

In this case study the task of fitting a straight line to data has been considered. It is assumed that the data are normally distributed with unknown, constant variance. Figure 1 on the right shows a typical example of data for such a problem. Two properties of the underlying statistical model are important for its subsequent treatment: the linearity of the regression function with respect to the parameters (i.e. offset and slope) of the straight line and the fact that the observations are independent and normally distributed with constant variance. The use of conjugate priors is illustrated, including their assignment based on relevant prior knowledge. The advantage of using a conjugate prior is that the posterior can be given in analytical form which makes the calculation of the results simple. Furthermore, software based on MATLAB® [The MathWorks Inc., 2012] has been developed and made available.

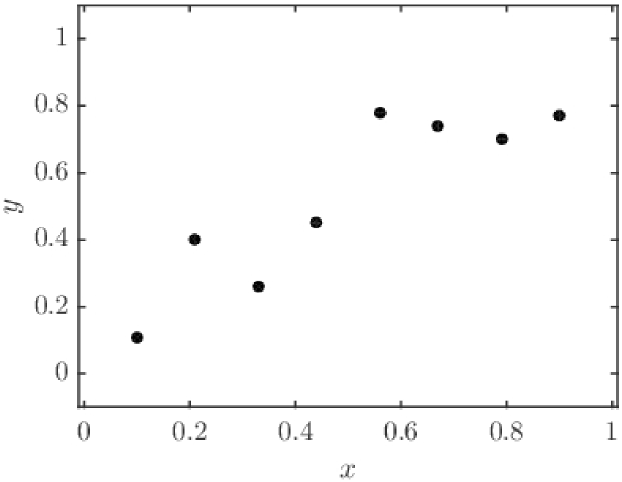


Figure 1: Example data for the normal linear regression task.

The treatment of the normal straight line regression model was done in a general way such that any normal linear regression model (e.g. higher-order polynomial models) is included. Explicit formulas were given that allow the calculations to be carried analytically. A procedure has been suggested, how the parameters of the conjugate prior are determined in view of relevant prior knowledge. Furthermore, the relationship to an inference based on so-called non-informative priors has been explored. Figure 2 shows results for the data of the Bayesian inference. Refer to [1] and [10] for more details.

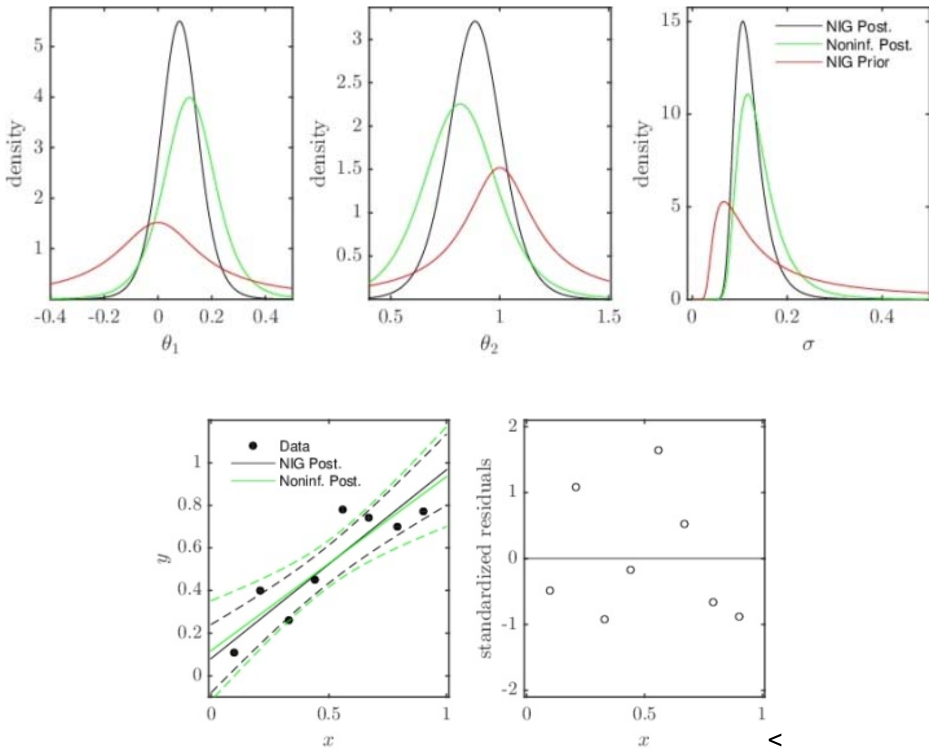


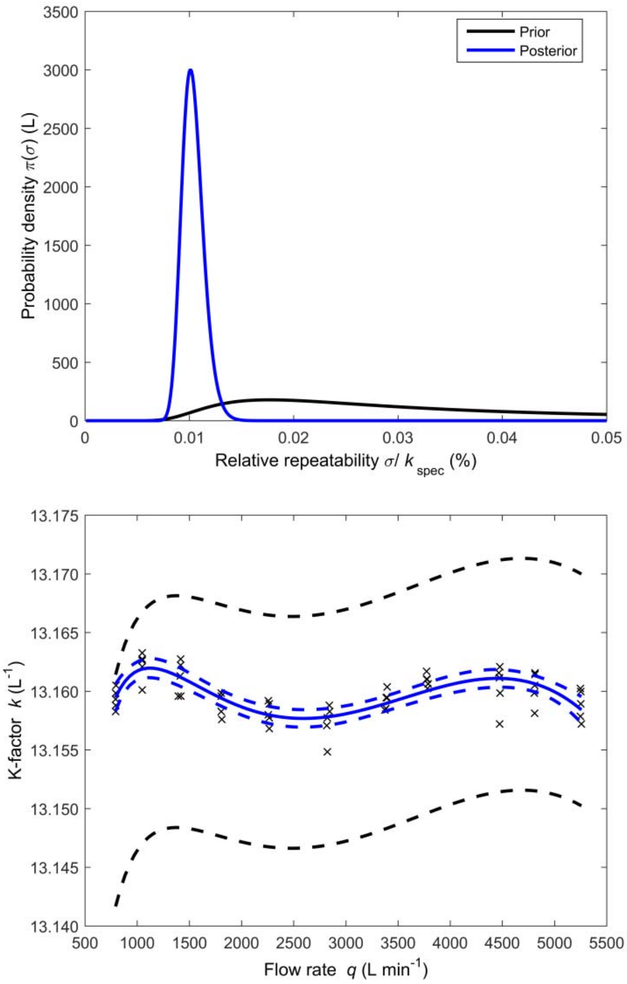
Figure 2: Top: posterior distributions for slope, offset (first two graphs) and standard deviation of the noise (third graph). The red lines show the employed priors, and the black lines the resulting posteriors. The green lines indicate the results when using non-informative priors. Bottom: fit of the straight line model including point-wise 95 % credible intervals for the data from Figure 1 (left) as well as standardized residuals (right).

Case study 2: Flow meter calibration - Normal linear regression with constraints (collaboration CMI, NPL, PTB, and VSL)

This case study concerns the determination of a calibration curve for a flow meter. The example is one of a normal linear regression model (see above), but the available prior knowledge includes a constraint on the values of the calibration curve. A result of this case study is the development of a numerical method for obtaining the results of a Bayesian inference when accounting for the given constraint. The method takes the form of a simple Monte Carlo procedure combined with an accept/reject algorithm. The procedure was illustrated in terms of its application to measured flow data.

The development of the numerical method makes use of the results derived in case study 1 for normal linear regression problems. Basically, the procedure works as follows. A prior distribution is assigned from the family of conjugate priors for the normal linear regression problem (cf. case study 1 above) that represents all available prior knowledge with the exception of the constraint on the calibration curve. The posterior distribution resulting for this prior distribution is then analytically available, and corresponding samples can be drawn in a simple way. By drawing many such samples, and by rejecting those for which the corresponding calibration curve does not satisfy the given constraint, samples from the sought posterior distribution are realized. Software based on MATLAB® [The MathWorks Inc., 2012] has been developed and made available.

Figure 3 shows results of the developed Bayesian analysis for master turbine flow meter from VSL's calibration facility for water flow meters. Refer to [1] and [17] for more details.



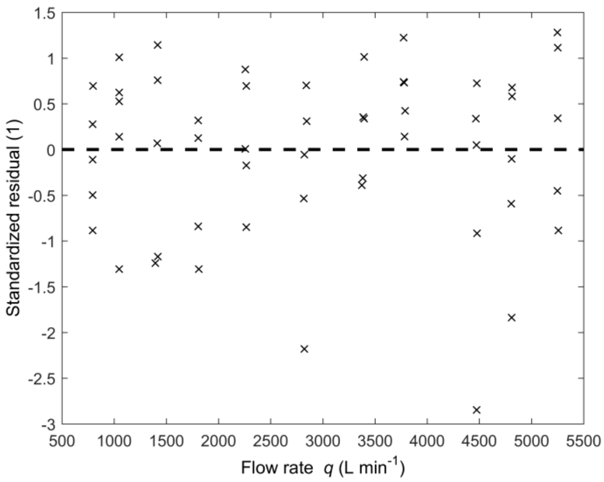


Figure 3: Results for the calibration of a flow meter. Top: prior and posterior distribution for the repeatability standard deviation. Middle: calibration data (black crosses), lower and upper bounds on the values of a calibration curve (black broken lines), the estimated calibration curve (blue solid line), and point-wise 95% credible intervals. Bottom: standardized residuals.

Case study 3: Inference of thermophysical properties - Use of hierarchical priors (collaboration LNE and NPL)

In this case study a method has been developed for the determination of the thermophysical properties of a material from transient temperature measurements. The case study constitutes a nonlinear regression model, for which hierarchical prior distributions were employed. A further feature of this example is that the regression function is not known explicitly but needs to be determined through the numerical solution of a partial differential equation. It is considered as an inverse problem in the sense that the quantity of interest, the thermal diffusivity, is not measured directly but is a characteristic of the material that is responsible for the observed temperature variations. Another result of the case study is a Metropolis-Hastings algorithm suitable for the numerical calculation of results.

The goal in this case study was to infer the thermal diffusivity of a specimen under test. Data for a thermogram obtained by the so-called FLASH technique is represented in Figure 4. The thermogram shows the (normalized) temperature rise on the back face of a thin disk induced by a short energy pulse appllied to the front face. In order to sufficiently model the data, their standard deviation σ needs to be modeled in addition, and Figure 4 also shows the prior distribution for this parameter used in the Bayesian inference. The observed temperature variations can be summarized with two parameters: the characteristic time τ which controls the increasing part of the curve and a so-called Biot number *Bi* which controls the decreasing part of the curve. The inference is performed for both parameters. Then, the thermal diffusivity *a* has been obtained from the thickness *e* of the specimen under test and the characteristic time according to

$$a = \frac{e^2}{\tau} \qquad (1)$$

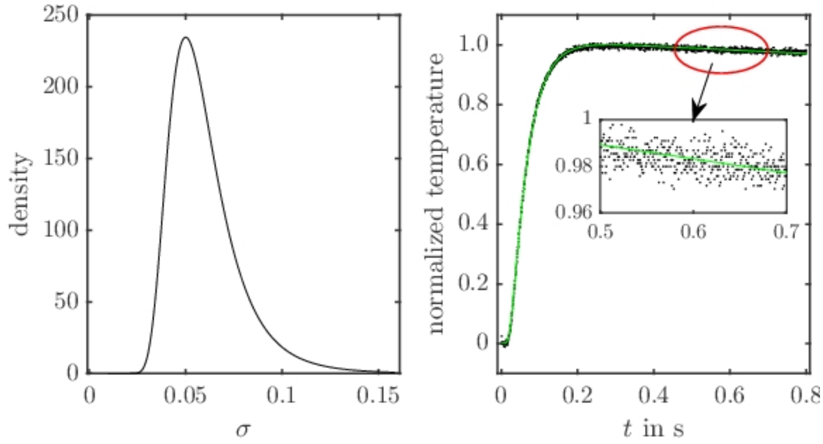


Figure 4: Right: Data (dotted points), together with the obtained mean regression curve (solid line). Left: Employed prior distribution for the standard deviation of the observations.

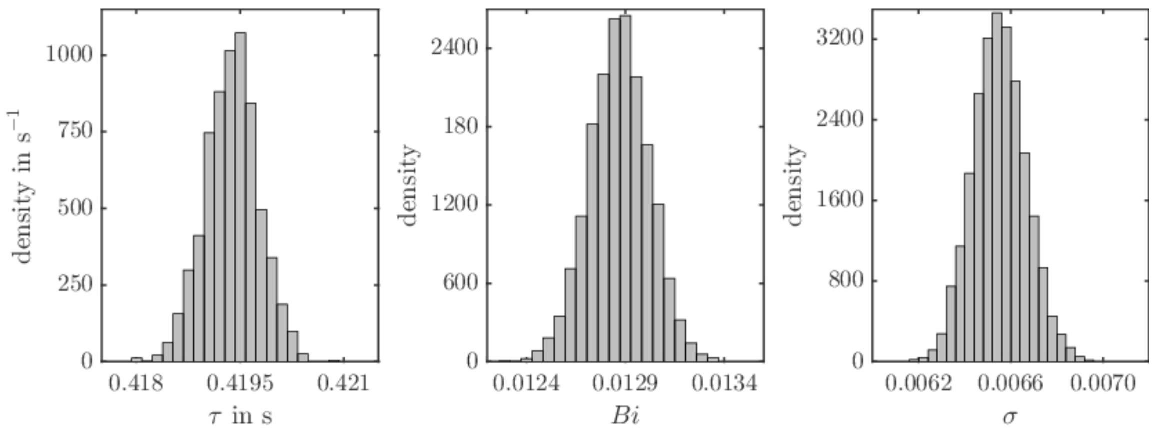


Figure 5: Posterior distributions of τ (left), *Bi* (middle) and σ (right) for the data for Figure 4.

Hierarchical priors for the characteristic time τ and the Biot number *Bi* were developed for the Bayesian analysis, and a Metropolis-Hastings algorithm was designed. Figure 5 shows results obtained for the data from Figure 4. Refer to [1] and [14] for further information.

Case study 4: Analysis of immunoassay data - A statistical calibration problem (PTB)

In this case study a novel method has been developed for the analysis of a bioanalytical test. The example illustrates a statistical calibration problem, in which the posterior distribution determined in the calibration stage is used as (part of) the prior distribution for the subsequent prediction. The main novelty achieved for this case study is the derivation of generally applicable prior distributions from historical measurements. Furthermore, software has been developed and implemented in BUGS (Bayesian inference Using Gibbs Sampling) [D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. Statistics in Medicine, 28:3049 -3067, 2009.].

An immunoassay is a type of bioanalytical test that exploits the highly specific binding between antibodies and antigens. Immunoassays are able to measure very small concentrations of a substance in solutions and have an immense range of applications, for example, to detect the presence of an infection, of hormones or drugs.

Figure 6 shows calibration data for an ELISA (Enzyme-Linked Immunosorbent Assay) where fluorescence intensity has been measured for known values of concentration. The novel prior distributions developed in the project for the (parameters of the) model are illustrated in Figure 7. Figure 8, finally, shows resulting posterior distributions for the unknown concentrations.

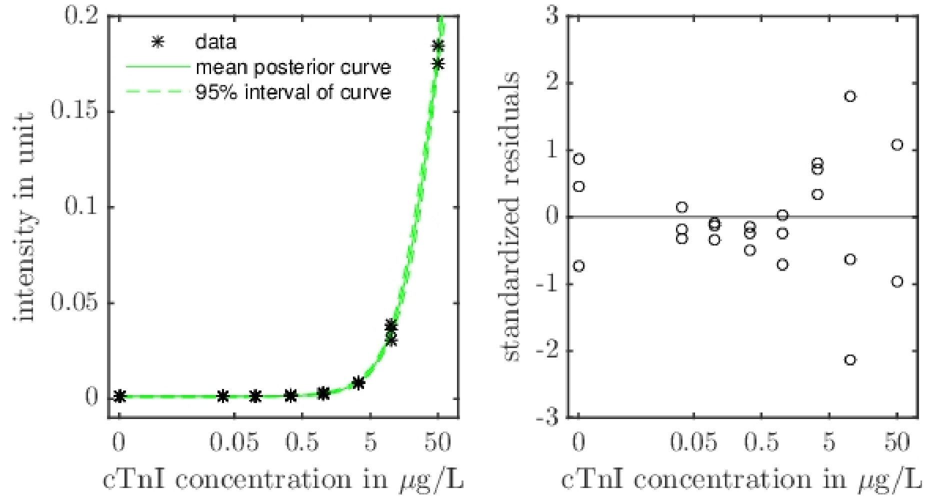


Figure 6: Results for the analysis of ELISA data. Left: data (dotted points) together with estimated regression function and point-wise 95% credible intervals. Right: Residual plot.

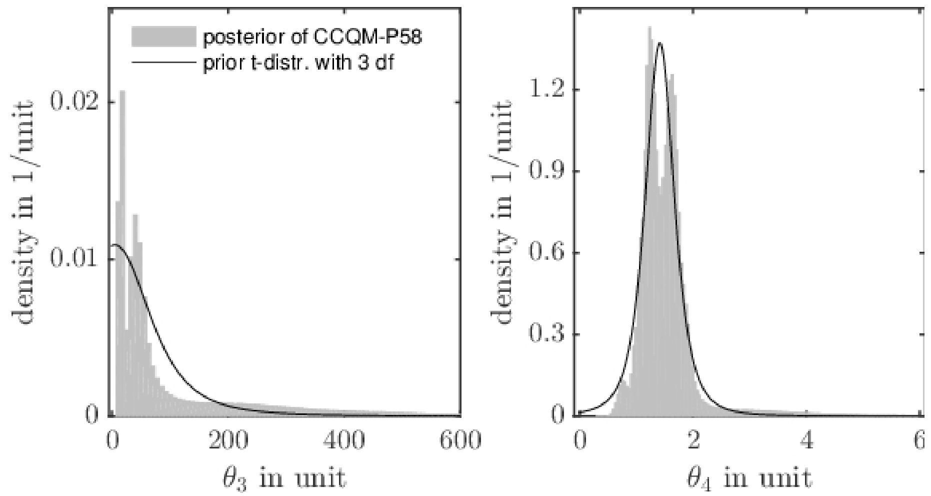


Figure 7: Prior distributions (solid lines) for two parameters of the parametric model used for the analysis of the ELISA data. The histograms show results from historic ELISA data that were used to determine these prior distributions.

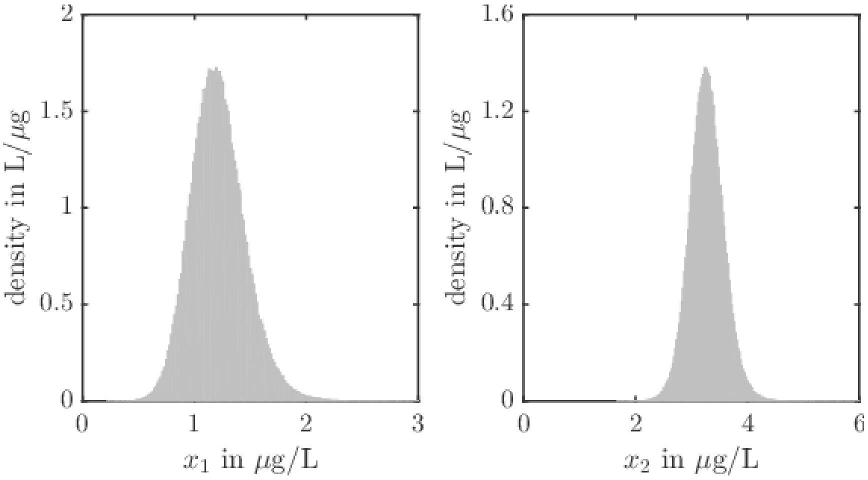


Figure 8: Posterior distributions of two unknown concentrations x1 (left) and x2 (right) that were estimated.

It has been demonstrated in the project that the novel prior distributions provide a more reliable analysis of the bioanalytical test that is also insensitive against model violations. Furthermore, the novel priors are generally applicable for the analysis of ELISAs, thereby ensuring that the advantages of the achievements made in this project can be easily shared by users of ELISA tests. Refer to [1] and [11] for further information.

Case study 5: Analysis of digital PCR data - A parametric inverse problem (LGC)

In this case study a novel method for the analysis of digital PCR data has been developed. The task constitutes a parametric inverse problem rather than a regression task. It was demonstrated that such problems can be treated by the very same Bayesian methodology as the preceding regression tasks. The example also illustrates the selection of different families of prior distributions to reflect different parameter domains. Furthermore, a Metropolis-Hastings algorithm was developed for numerical calculations together with a variable transformation enhancing convergence.

Digital PCR (dPCR) is a relatively recent measurement method for the absolute determination of the copy number concentration in a DNA or RNA sample. The method works by partitioning the sample into small volumes with each containing up to a few molecules on a chip and amplifying the DNA over a number of cycles. The DNA is measured using a fluorescent marker, and the cycles at which the fluorescence reaches a threshold, called quantification cycles or Cq, can be used to estimate the copy number concentration.

The developed Bayesian inference has been applied to digital PCR data (see Figure 9), and the results are shown in Figure 10. These results were obtained by a Markov chain Monte Carlo algorithm designed in the project, and Figure 11 shows the chains of this Monte Carlo method. Refer to [1] and [5] for further information.

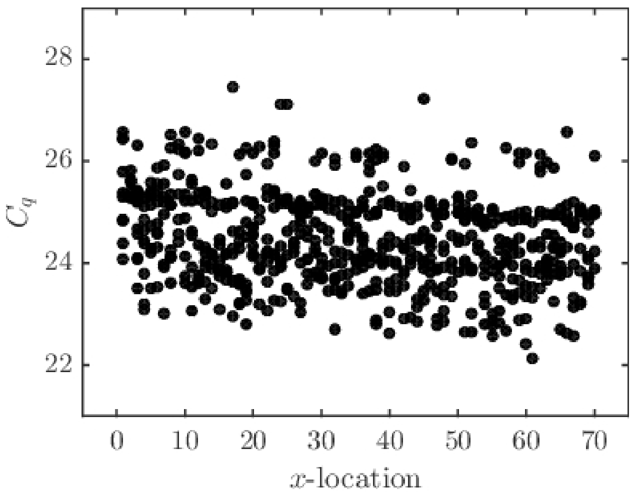


Figure 9: Example data for the digital PCR problem. The plot shows the quantification cycle values C_q.

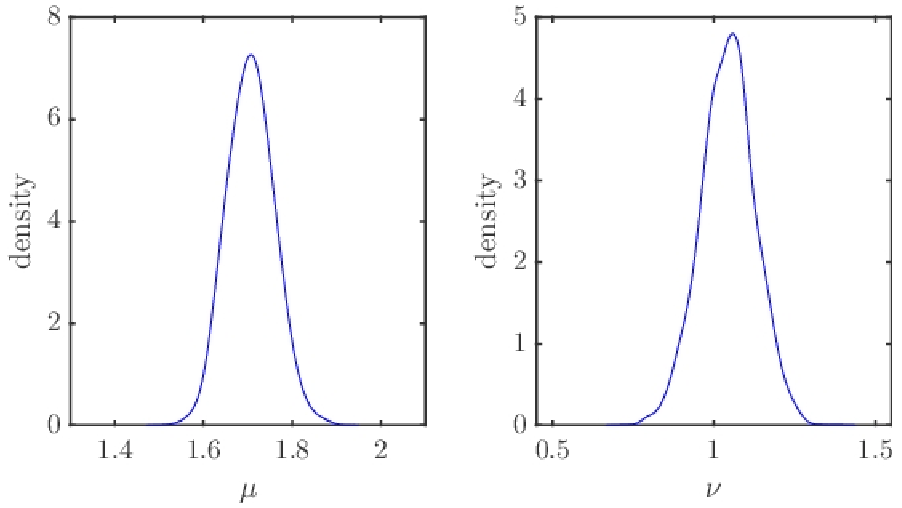


Figure 10: Results obtained by the analysis of the data from Figure 9. Displayed are the posterior distributions for the mean copy number μ (left) and the dispersion parameter ν (right).

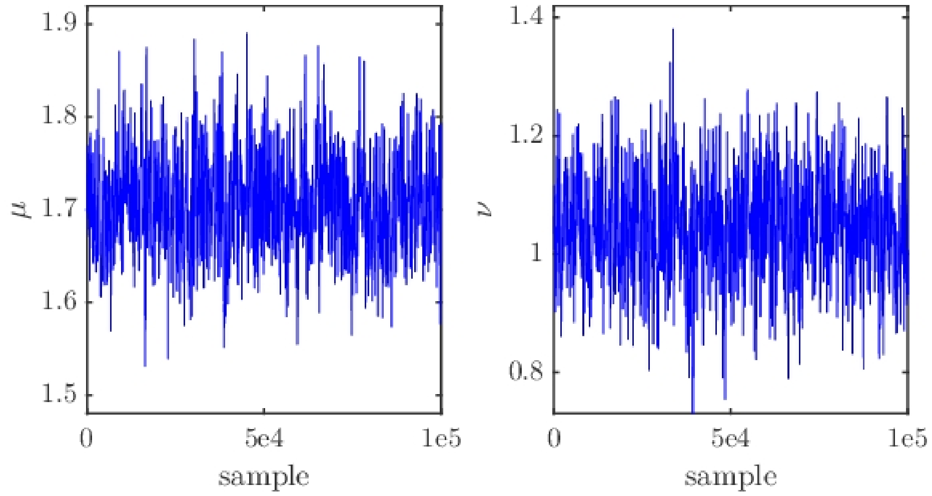


Figure 11: Markov chain Monte Carlo runs for numerically determining mean copy number er μ (left) and the dispersion parameter ν (right).

Objective 2: Computationally expensive models

Computationally expensive models, such as finite element (FE) models, are used in a wide range of industries beyond metrology. FE is commonly used in design and assessment of products in sectors ranging from aerospace and automotive to consumer goods and medical devices. There is a growing awareness amongst FE users that the decisions made in the design and assessment processes should be supported by an understanding of the uncertainties associated with the results of such models. The audience for the research in this topic therefore goes beyond metrologists and includes engineers and designers. The good practice guide was written with this audience in mind, and provides examples and case studies to support easy uptake of the outputs.

In all cases the main extension of the state of the art has been achieved through the case studies. Three metrology problems have been studied:

- determination of the geometric properties of nano-structured surfaces through scatterometry
- flow measurement in pipes and nozzles, and

- determination of the thermal conductivity of the components of a layered sample through the laser flash thermal diffusivity experiment.

In the first two case studies, the existing uncertainty evaluation methodology was replaced by a more efficient state-of-the art approach. In the third case study no uncertainty evaluation methodology was in place because the experiment is still being developed for layered materials, so the state of the art was established by this research. The case studies also provided the main vehicle for collaboration in this topic, with PTB, FORCE and VSL collaborating on the second listed and LNE and NPL collaborating on the third.

It was noted above that the approaches recommended by the GUM and its supplements are either potentially inaccurate due to the nonlinearity of the models (and often complicated if the software used to solve the model is a black box), or not feasible due to the number of model evaluations required for an accurate estimate of the uncertainty. The challenge for uncertainty evaluation for computationally expensive models is to minimise the number of model evaluations required to obtain a given level of accuracy of the estimated uncertainty.

The research undertaken to address this challenge was split into three parts: model preparation, smart sampling methods, and surrogate models. In all cases, the methods were assessed for repeatability as well as accuracy, and the dependence of the results on the number of model evaluations was investigated. This research enables the user to have confidence in their chosen method and to select an appropriate sample size for their needs.

The model preparation aspect focused on sensitivity analysis for input quantity screening. Sensitivity analysis of the input quantities of the model makes it possible to identify input quantities that have an insignificant effect on the output quantities of interest. Treating these inputs as constants and omitting them from the full uncertainty evaluation reduces the dimensionality of the uncertainty evaluation problem, improving the computational efficiency. The approach was also found to identify cases where combinations of a subset of input quantities have a strong effect on a given output quantity, potentially providing physical insight and allowing the problem size to be reduced further. Software implementing the most useful sensitivity analysis methods has been created by LNE using the statistical language R and made freely available from https://CRAN.R-project.org/package=ATmet.

The smart sampling methods work compared different methods to select the values of the input quantities used when models are evaluated as part of the uncertainty calculation. Random (Monte Carlo) sampling requires a large number of samples to produce an accurate uncertainty estimate, but selecting samples in a more controlled way that is still consistent with the probability distributions of the input quantities can lead to better uncertainty estimates, particularly for small numbers of model evaluations. Five methods were investigated and results are discussed in more detail below.

Surrogate models construct a model that is computationally cheap to evaluate by interpolating or otherwise manipulating a set of results of the full model. The surrogate model can then be used to estimate the required uncertainties by using a sampling method such as random sampling, because the computational expense is so much smaller. This approach requires selection of a set of input values to generate the results used in construction of the surrogate, and the sensitivity of the results to the selection of these input values has been investigated. Four classes of surrogate model have been investigated, and the results are discussed in more detail below.

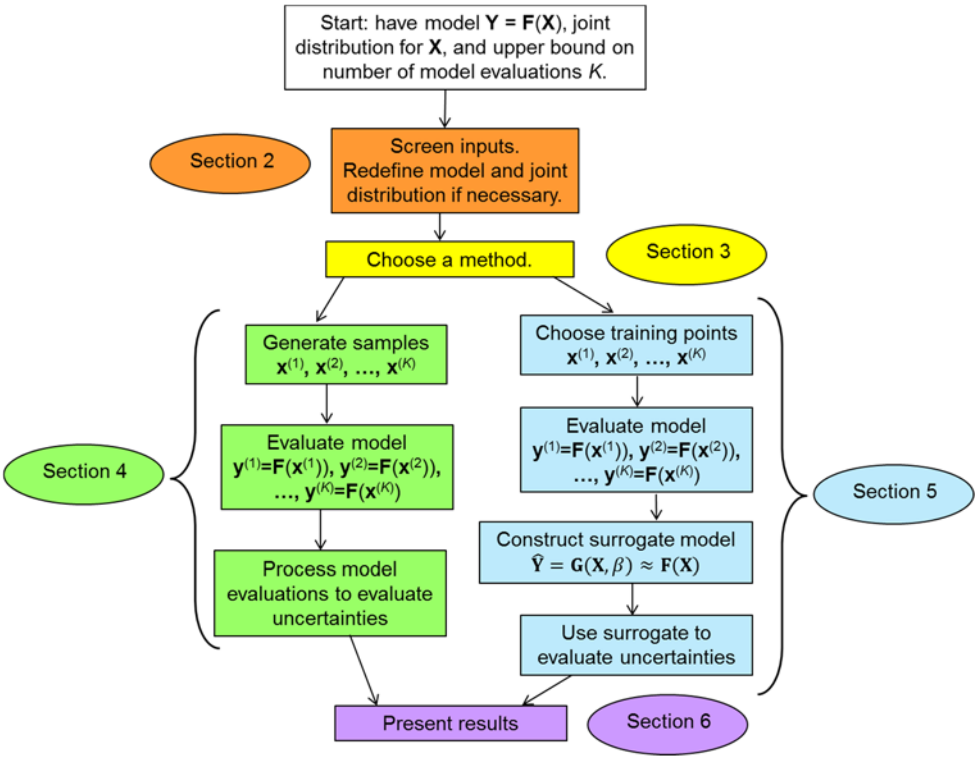


Figure 12: Flow chart illustrating the steps when evaluating the uncertainties associated with the results of a computationally expensive model, addressed in the good practice guide. Flow chart illustrating the steps when evaluating the uncertainties associated with the results of a computationally expensive model, addressed in the respective good practice guide [2].

It is not possible to state that one method is the best uncertainty evaluation method under all circumstances. All methods have their strengths and weaknesses. Some methods are not applicable to all problems, and some are more technically challenging to implement. Section 3 of the good practice guide [2] contains general discussion regarding points to consider when selecting an evaluation method, and discusses how to assess and compare methods in a fair way.

Communication of results to non-experts is key to integration of uncertainty evaluation within decision-making processes, and visual presentation of results can greatly enhance the understanding of a general audience. Therefore, section 6 of the good practice guide [2] provides some guidance on presentation and analysis of results including assessment of repeatability of methods.

The case studies have advanced metrology beyond the state of the art and so are discussed in some detail.

Case study 1: Scatterometry (PTB)

Scatterometry is a surface measurement technique that evaluates the geometric parameters that define periodic nanosurfaces. The surface is illuminated by monochromatic light and measurements are made of the intensity of diffracted light at different scattering angles. An FE model can be used to predict the scattered intensities, and an inverse modelling approach was applied to evaluate the three unknown geometric parameters. The uncertain input quantities represent the noise terms associated with the measured light intensities; material properties, geometric parameters not determined by the model, and optical parameters are assumed to have a negligible effect on the uncertainty. Figure 13 shows the geometric parameters of interest. The model is further complicated by the wave lengths dependence of the noise term, meaning that there are two noise parameters to be

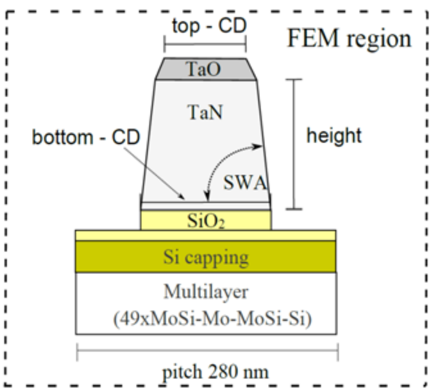


Figure 13: Geometric parameters of the scatterometry case study.

estimated.

The work carried out in this case study compared three different surrogate modelling approaches: nearest neighbour interpolation, response surface methodology, and polynomial chaos. The surrogate models were used to replace the computationally expensive FE model throughout. Each of the surrogate models were constructed using between 600 and 750 evaluations of the FE model.

As an initial test, the approximation errors of each type of surrogate model were estimated by comparing the outputs of the surrogate model evaluated for 100 randomly generated geometries to the outputs of the FE model evaluated at the same geometries. These estimates showed that the polynomial chaos surrogate was two orders of magnitude more accurate than the response surface, and that the response surface was more accurate than nearest neighbour interpolation. Figure 14 shows a comparison of the results generated for a typical geometry for each surrogate modelling method, clearly showing that polynomial chaos (PC, right hand plot) is more accurate than other methods.

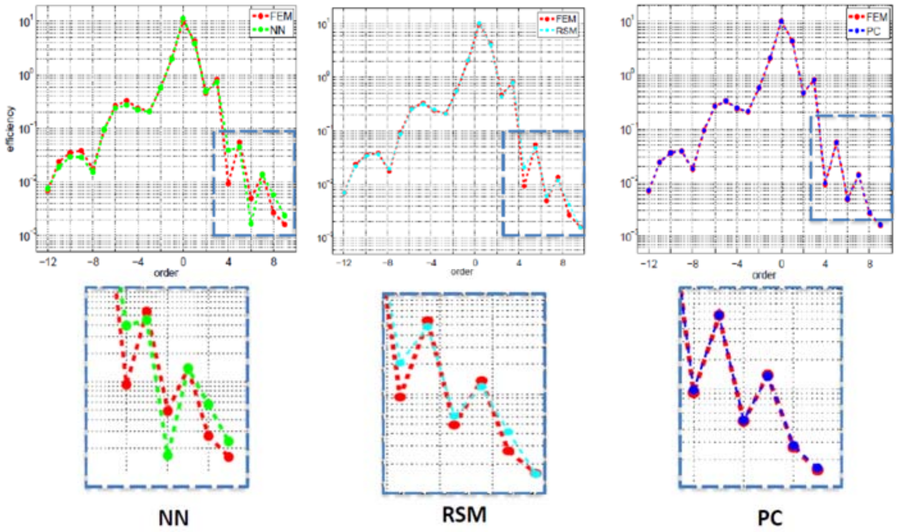


Figure 14: Comparison of surrogate models. From left to right: Nearest-neighbor interpolation, response surface methodology, polynomial chaos approach.

A set of test results were generated for a single geometry and a specified distribution of the parameters of the noise term. The surrogate models were then used to estimate the distributions of the geometric parameters using a Markov Chain Monte Carlo (MCMC) method. The MCMC approach evaluated the surrogate model more than 100,000 times. The surrogate models typically took fractions of a second to evaluate, whereas the FE model took two minutes which would render MCMC unusable.

The results showed that polynomial chaos gave the most accurate estimates of the geometric parameters with lowest uncertainties. Nearest neighbour interpolation gave biassed estimates, and the response surface methodology gave slightly inaccurate results with higher uncertainties than polynomial chaos. Finally, the surrogate models were used to estimate the parameters of the noise distribution, with the results showing the polynomial chaos gives the better estimates.

The research in this case study has developed an uncertainty evaluation methodology for scatterometry that uses polynomial chaos and has demonstrated that the approach is accurate and computationally efficient despite the complexity of the noise term. The evaluation approach advances the state of the art of scatterometry by producing accurate uncertainties for the technique. We refer to [2], [13] and [23] for further information.

Case study 2: Fluid flow (collaboration between FORCE, PTB and VSL)

The work in this case study was split into two parts, addressing uncertainty evaluation for flow meter calibration and uncertainty evaluation for Venturi nozzle flow, used in calibration of gas metering systems. Calibration of flow meters and characterisation of flow measurement devices usually require the flow conditions to be in a steady "plug flow" state to give good results. The state of flow within a pipe is difficult to obtain directly, so it is important to have an understanding of how upstream bends and constrictions can affect the flow conditions to ensure that measurements are taken sufficiently far from such obstacles. Computational fluid dynamics (CFD) models can be used to obtain this understanding, even for turbulent flow in rough pipes, but such models are very computationally expensive. The work reported in this case study has used a family of skew inflow profiles, where the position of the maximum is assumed to be uncertain. The output quantities describe the flow profile at a series of points downstream from the point at which the inflow profile is defined. An understanding of the uncertainties associated with the flow profile downstream allows the metrologist to identify a suitable separation between the location of any bends and the point at which measurements are made.

The work in this section focused on polynomial chaos, and looked at how the order of the polynomial chaos expansion, which affects the number of model evaluations required for uncertainty evaluation, affected accuracy. The problem was too computationally expensive to use reference results to test the accuracy of the method, so the convergence of the method with increasing order of polynomial was assumed to be equivalent to its accuracy. It was found that the method rapidly converged to a stable solution, with results suggesting that for the case studied, a polynomial requiring 5 model evaluations gave converged results. This number of model evaluations represents a huge reduction when compared to random sampling, making uncertainty evaluation using CFD models computationally feasible for pipe flow and thus improving the accuracy of the uncertainties obtained and advancing the state of the art. Refer to [2], [9] and [12] for more details. The methodology has been successfully applied to estimating systematic uncertainty contributions and biases stemming from installation effects on ultrasonic and electromagnetic flow meters [6].

In the second part of the case study on fluid flow, a CFD model of a sonic nozzle was studied. This CFD model may be used to improve the understanding of the physical conditions (i.e. temperature, pressure and flow speed distribution) on the flow within sonic nozzles so that appropriate environmental control can be achieved during calibrations. The model had seven input quantities with a variety of distributions. This part of the case study compared multiple uncertainty evaluation methods, and provided an additional comparison to the Law of Propagation of Uncertainty (LPU, as described in the GUM). The sampling methods compared included Monte Carlo sampling, stratified sampling, Latin hypercube sampling, and polynomial chaos. Two different sample sizes (8 and 128) were studied. Random sampling with approximately 2,000 samples was used to generate a set of reference results. The results showed that the LPU was very accurate for this problem, and that the smart sampling methods led to no improvement in accuracy over LPU. The LPU method is only valid for models that are linear in the uncertain quantities, which cannot be known a priori for most computationally expensive.

Case study 3: Thermophysical properties (collaboration between LNE and NPL)

The laser flash thermal diffusivity technique can be used to calculate the thermal conductivity of a homogeneous material. Optimisation methods can be used to extend the technique to measurement of the components of layered systems. Layered and coated systems are used in a variety of high-temperature applications, an important example being thermal barrier coatings. Thermal barrier coatings are multi-layer coatings that are used in turbines and engines to protect blades from high temperature environments. An understanding of the uncertainties associated with the properties of each layer enables more effective and cost-efficient coating systems to be developed.

The work reported in this case study uses a finite volume model within an optimisation algorithm to obtain estimates of the thermal conductivity of the components of a layered material by matching model predictions to measured values, and evaluates the uncertainties associated with the estimates. The initial model uses nine uncertain input quantities associated with various aspects of the experiment and has two output quantities whose uncertainties are of interest.

An initial sensitivity analysis enabled the model to be simplified, thus reducing the computational expense of the uncertainty calculation. The analysis showed that the geometric parameters in the model had no effect on the results and so could be regarded as fixed values. This simplification reduced the number of uncertain quantities to six. The sensitivity analysis also indicated that each of the output quantities was strongly associated with two of the input quantities (a different two in each case).

The work carried out compared three sampling methods (random sampling, stratified sampling and Latin hypercube sampling) and three surrogate modelling methods (nearest neighbour interpolation, response surface methodology, and Gaussian process modelling also known as kriging). Different methods of selecting the input quantity values for the surrogate model construction (known as the training points) were tested to ensure that the method used is not a source of inaccuracy. A set of reference results was generated using random sampling with over 2,000 samples so that accuracy of the methods could be assessed. Each of the methods assessed was run for several different sample sizes, and for each sample size ten separate samples were run to assess the repeatability of the methods.

The results suggested that Latin hypercube sampling gave the lowest sample-to-sample variability, particularly for small sample sizes, and gave good accuracy, and so would be the smart sampling method of choice for this problem. The assessment of different methods for training point selection suggested that for the problem studied, the method made little difference to the accuracy. The results of the surrogate models suggested that Gaussian process modelling performs well, particularly for output quantities that are not well approximated with a Gaussian distribution.

Objective 3: Decision-making and conformity assessment

Decisions based on measurement, such as tests of the significance of differences and of conformity, are currently made in many important application areas without a clear and harmonised basis for assessing the impact and sharing the risks that arise from measurement uncertainty. The need for improved insight into setting a 'fit-for-purpose' level of measurement effort *prior* to performing a given task is increasingly appreciated on all sides – from industrial, academic & societal end users as well as national metrology institutes. Better mutual understanding between the metrologist and those ordering measurement tasks about the import and limitations of the measurements when making decisions of conformance *after* a measurement is also needed. Rules for generic guidance were worked out in this project and address the role of measurement uncertainty in decision-making and conformity assessment for multivariate cases, regression and computationally expensive models illustrated for a number of case studies such as fire engineering, healthcare and electricity energy metering. The work in this area was led by SP and carried out together with JV, LNE and INRIM. Other partners added input and advice during project meetings, workshops and discussions by e-mail.

Measurement uncertainty in a test result on inspection of an entity (measurement object) can

- lead to incorrect estimates of consequences of entity error,
- aid in assessment of significance of apparent differences between repeated test results,
- and increase risk of making incorrect decisions, such as failing conforming entity or passing non-conforming entity when test result close to tolerance limit.

Decisions of conformity are currently made in many important application areas, such as environmental monitoring, the health sector and product safety testing, but without a clear and harmonised basis for sharing the risks that arise from measurement uncertainty between the consumer and the supplier.

Conformity assessment involves a set of processes that show your product, service or system meets the requirements of a standard, i.e. it

- provides confidence for consumer that requirements on products and services are met,
- provides the producer and supplier with useful tools to ensure product quality,
- and is often essential for reasons of public interest, public health, safety & order, protection of environment and consumer and of fair trading.

Since neither the production nor measurement processes are perfect, there will always be some dispersion in the observed product value either for repeated measurements of one item or for measurements of a series of items. Practically in cases where measurement dispersion is of comparable size to actual product variation, it can be difficult to separate these. This separation is however essential: If one overestimates measurement dispersion, then actual dispersion in product values will not be detected and thus lead to poorer product quality. Additional costs will also be incurred if it is decided, on the basis of estimated poor measurement quality, to spend more on (unnecessarily) improving measurement resources. An underestimation of measurement dispersion will lead to unnecessary adjustment of the production process and thus to increased production costs together with poorer product quality, where spurious measurement dispersion is transferred to product dispersion.

There are several ways of ensuring that limited measurement quality does not adversely affect Conformity Assessment. Risks and the consequences of incorrect decision-making in Conformity Assessment can be minimized with the following three steps: (i) set limits on maximum permissible measurement uncertainties (equivalently, minimum measurement capability) and on maximum permissible consequence costs at the specification stage of the measurement task, (ii) agree on acceptable locations of the uncertainty interval with respect to a specification limit, and (iii) optimize measurement uncertainty proactively, ahead of a series of measurements, by designing experiments so that the sum of costs of measurement and of incorrect decisions of conformity is at a minimum.

The JCGM 106:2012 Guide on Conformity Assessment document for instance addresses the technical problem of calculating the conformance probability and the probabilities of the two types of incorrect decisions – that is, supplier and consumer risks expressed in percentages, – given a probability density function (PDF) for the measurand, the tolerance limits and the limits of the acceptance interval. The treatment, however, was limited to a single measurement quantity for which the PDF is known and easily accessible. The work in the project treated three important extensions beyond the state-of-the-art: (i) Conformity Assessment in computationally expensive system within a case study on fire engineering, (ii) Conformity Assessment based on multivariate measurements exemplified in a bivariate case describing health-care products, and (iii) Conformity Assessment including an optimization of associated costs, where e. g. the cost for testing were balanced against the costs associated with the consumer risks of paying too much as a result of erroneous measurement. In addition a software for conformity in multivariate cases was provided by LNE.

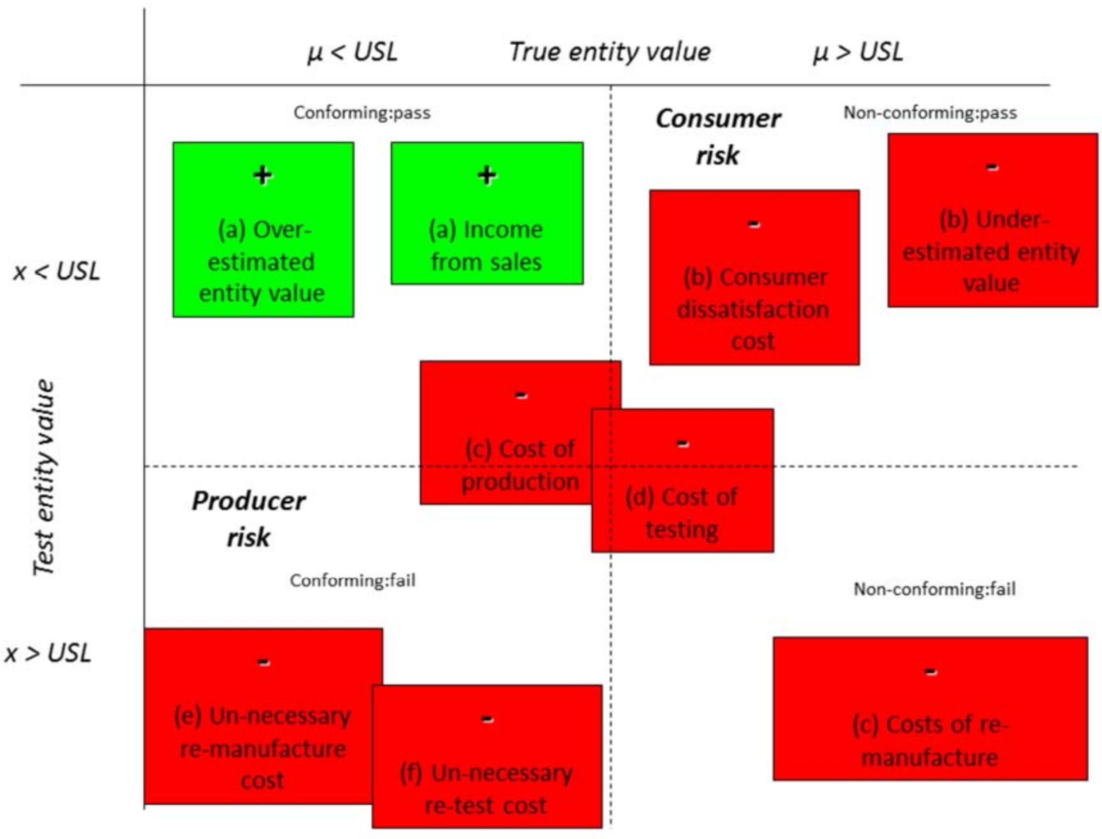


Figure 15: Different costs and income of conformity-assessed entity from the point of view of the supplier. Green represents a profit, while red represents a loss.

Conformity assessment for computationally expensive systems:

LNE, with input from other project partners, e. g. NPL, PTB and SP, have developed new methods for the conformity assessment of computationally expensive systems.

In a study of statistical methods addressing conformity assessment for computationally expensive systems, recent work has focused on the probability of conformity that output quantities exceed a regulatory threshold using in particular Monte Carlo methods, as well as importance sampling and sequential sampling. A method to combine expensive runs with fast approximations has also been addressed. The strengths, limitations and applicability of each method are stressed and key papers cited. A case study on fire engineering employs these new methods.

When dealing with a computational code, one may be interested in propagating the uncertainties related to the input variables to estimate the uncertainty associated with the output variable.

In the decision theoretical framework pertaining to conformity assessment, one is interested also in the position of the output variable with respect to a given threshold (regulatory threshold...). The problem of knowing whether a computationally expensive model exceeds a given threshold is very common for reliability analysis and safety-critical applications such as aerospace, nuclear power stations and civil engineering (models of bridges and buildings etc). The threshold can be interpreted as a lower specification limit defining a one-sided tolerance interval.

The statistical methods that are usually used to deal with the propagation of uncertainties may need to be adapted to compute the probability of exceeding a threshold. This is particularly the case when the probability is small or when the code is computationally expensive. These methods were initially developed for reliability assessment based on computer experiments. In this part of the project the probability of conformity was assessed as the probability that output quantities exceed a regulatory threshold. The evaluation of uncertainties focused on Monte Carlo methods based on a large number of calls of either the code or a metamodel of the code. Other methods addressed are importance sampling which is a variance reduction alternative to Monte Carlo sampling of the code and sequential sampling which builds sequentially a learning database for metamodelling by sampling points in the (unknown) conformity domain. A method to combine expensive runs with fast approximations is also addressed. Table 1 summarizes recommendations for different practically relevant cases.

Table 1: Recommendations for the choice of statistical methods to compute the probability of conformity

	direct sampling	**surrogate sampling**
target region sampling	importance sampling	metamodel, optimized learning design
whole domain sampling	Monte Carlo sampling	metamodel, Latin hypercube sampling learning design

Case study 1 - Fire engineering (LNE):

At the center of this case study was conformity assessment of computationally expensive systems using cheaper approximations for the model as described above. Modelling the relationship between the two codes (the expensive and the cheap one) and the behaviour of the codes viewed as black boxes evaluated at input points using the so-called co-kriging method, provides a probabilistic framework to assess the probability of conformity of the expensive system with an associated uncertainty. In order to increase the accuracy of the probability of non-conformity with a limited number of new runs, a procedure is derived to sequentially sample smart new points targeting the non-conformity domain. The approach is illustrated on a fire engineering example, in which the probability of conformity of a smoke control device is assessed by combining expensive CFD simulations (code FDS) with cheap code approximations (code CFAST). The use of such computational codes in Conformity Assessment has become common practice when experiments are not feasible or when their number is too parsimonious. The statistical modelling of numerical experiments with kriging models yields a probabilistic decision framework to assess the probability of failure of the system. Combining fast low-fidelity simulations with costly high-fidelity simulations has proved an efficient method to decrease the burden of costly simulations when predicting the output of a system. This studied was carried out by LNE together with application partners.

The fire engineering case study aims at computing the probability that the upper layer temperature in a room exceeds the 200°C threshold in the case of fire. A sequential co-kriging procedure returns an estimate of the probability of failure (non-conformity) of FDS bases of a very limited number of FDS runs thanks to numerous cheap approximations provided a valid relationship between the two codes. A flowchart of the complete sequential co-kriging procedure is displayed in Figure 16.

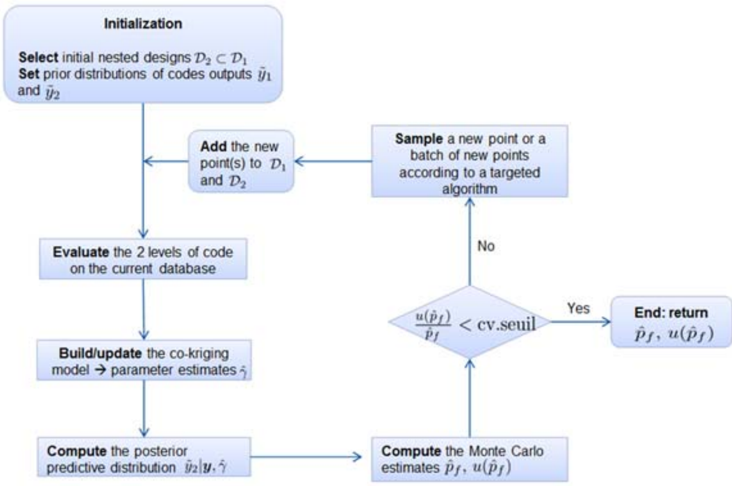


Figure 16: Flowchart of the sequential co-kriging procedure

A comparison of various kriging based methods has been carried out on the fire engineering case study to show the influence of the number of points in the database and the influence of their location on the estimates of the probability of conformity and its accuracy.

The methodology presented to assess the conformity of an expensive computational code has proved efficient to overcome too parsimonious evaluations from an expensive code when fast approximations are available. This method builds on the kriging models usually used to model code outputs, to improve the predictions of the kriging model of the expensive code. Another desirable feature is that the method allows smart sampling of new points targeting the non-conformity domain in order to further reduce the uncertainty of the probability of non-conformity. Although demonstrated in a simple but realistic case study, the method is flexible to take into account more complex relationships between the two codes in higher dimensional problems. For more details, see [3].

Case study 2: Health-care products (SP with stakeholder SCA)

Various healthcare studies were chosen to illustrate novel approaches to multivariate conformity assessment and decision-making. Such cases are characterised as dealing with properties important for consumers, such as smoothness of material surfaces (skin) as perceived in measurements by human panellists, which in turn depend on the surface topography, friction and hardness, of interest to the manufacturer wishing to fashion the product to the consumer's satisfaction.

Mathematical and statistical approaches to uncertainty evaluation are introduced in studies where typically there is no simple theory of how the perceived ('response') properties **Z** depend on the physical material ('explanatory') properties **X** of the material, that is, how the expression $\mathbf{Z} = f(\mathbf{X})$ is to be formulated.

Material and perceived quantities may be correlated in various ways and, in addition, limited measurement quality may enter on both sides of the expression. Principal component regression, consisting of an initial principal component analysis (PCA), can tackle correlation amongst the explanatory variables (**X**), thereby enabling both conformity assessment against explanatory variable **X** specification limits and multivariate regression including **Z.** A second challenge is that some measurements will be qualitative – such as the perceived 'slipperiness' in a current healthcare study – and require analysis on an ordinal scale, where many of the traditional tools of statistical analysis cannot be applied.

From a typical healthcare case study, a set of topical formulations (skin creams), perception of various consumer-related quantities, **Z**, such as slipperiness and smoothness, when each cream is applied on human skin, is related to a number of physiochemical properties, **X**, measured when the same creams are applied on artificial skin substrates.

The project introduced the study by making firstly a multivariate analysis of the measured physiochemical properties, **X**, such as friction adhesion, plotted in Figure 17 (a) together with corresponding upper and lower specification limits (*USL* and *LSL*), respectively, as set by the manufacturer with due consideration to both what is feasible to manufacture but ultimately what is required by the consumer.

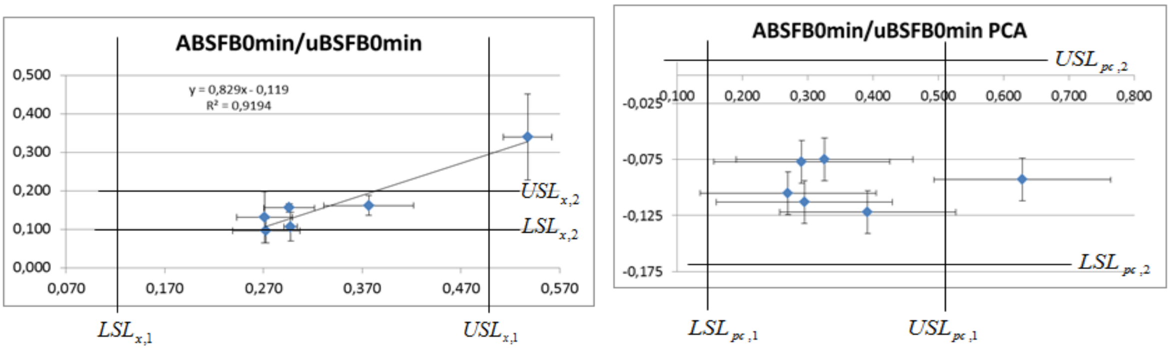


Figure 17: Correlation plot between a pair of explanatory variables X of six topical formulations [physical material properties: (x-axis) $\mu_{\mathrm{BS}}\mathrm{FB}_{0\min}$ (friction) and (y-axis) $A_{\mathrm{BS}}\mathrm{FB}_{0\min}$ (adhesion)]: (a) corresponding specification limits, (b) PC corresponding specification limits for the healthcare case studies

An initial study of specific conformity assessment - inspection of single items – was made in this bivariate healthcare case, where conformance probabilities associated with measurement uncertainty were calculated using the software *MathCad*.

To eliminate the covariance, evident in, which confounds the calculation of conformance probabilities, a PCA analysis in the current healthcare case was carried out. A resulting multivariate PC specification region $S_{PC1} = \left\{ y_i \in \mathbf{X}' \middle| \mathbf{LSL}_{PC} \le y_i \le \mathbf{USL}_{PC} \right\}$ is shown in Figure 17 (b). Most of the variability is concentrated in the first principal component, which of course reflects – as observed in Figure 17 (a) – that the observations amongst the six samples are clustered in essentially only two distinct points, thus providing a challenge to reliable correlation estimation.

Note finally that the above PC conformity assessment has so far neglected measurement uncertainty. In Figure 18 typical results of principal component regressions (PCR) for (a) traditional test theory and (b) Rasch analysis, providing explicit relations between sensory ('slipperiness') and explanatory variables, are shown.

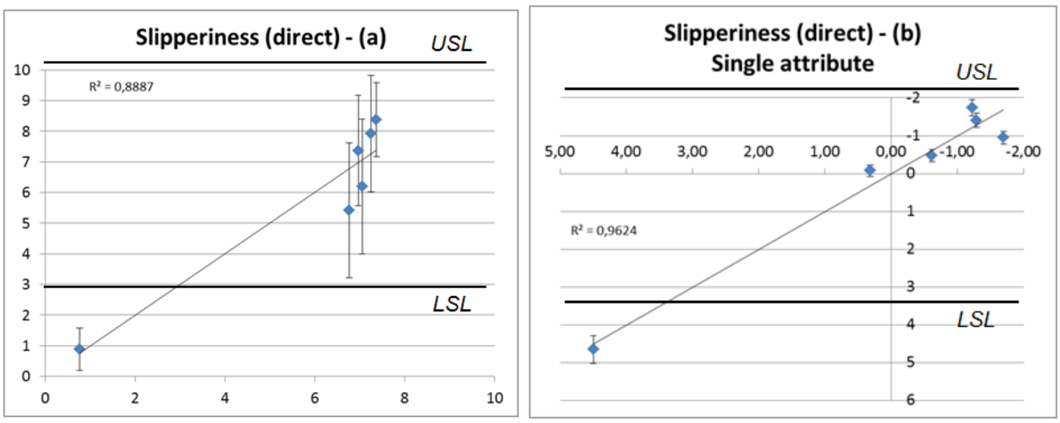


Figure 18: Relations in the healthcare study between sensory response (Slipperiness) (a) conventional test theory, (b) Rasch item parameters Θ) for the 6 samples, measured (y-axis) and predicted (x-axis) from explanatory variables *(ForceBoard, BIOSKIN)* from PCR analyses.

It is evident from Figure 18 that the Rasch analysis of the sensory data yields lower measurement uncertainties than conventional test theory, thanks to Rasch analysis' ability to treat ordinal data properly as well as to provide separate estimates of item (sample cream) and person (sensory panellist) attribute values. This work resulted partially from a cooperation of SP with application partner SCA. For more details, see also [3], [7], [15] and [18].

Case study 3: Utility meters - Conformity assessment including an optimization of associated costs (collaboration of JV and SP)

A case study of more effective decision rules for the conformance assessment of electrical energy meters in private households in Norway was carried out. It proposed how to use a specific risk analysis in order to set the time for the next meter test. The MID regulation today prescribes conformance assessment of electrical energy meters based on ISO standards for attribute sampling where decision rules are purely statistical decision rules and economic consequences are not explicitly taken into account. The risk analysis we introduce calculates the risks involved for erroneous decisions, either rejecting a conforming batch of meters (the producer risk) or accepting a non-conforming batch (the consumer risk). The consumer risk is sensitive to the period until the next test which becomes a quality characteristic of each batch. This time interval can be optimized by balancing the consumer risk against the producer risk. When the quality drops, the period until the next test will need to become shorter. But at a certain level of quality, the energy net supplier would rather replace the complete batch, than continue testing at such short intervals. This work was done by JV, who proposed and drove it throughout the project, and SP.

In 2011, the Norwegian Ministry of Industry issued a press release that pending the introduction across the country of smart meters for electric energy, there would be no replacement of electricity energy meters in private households in Norway during a transition period of up to 8 years. The argument was that all electricity meters anyway will be subsequently replaced by smart meters, so this decision would save money. The price of purchase and replacement of an electricity meter in a private household was around NOK 2000 (200€) at the time. However, the costs of not replacing non-conforming batches of meters were not calculated. The project completed a case study including a risk analysis of the conformance assessment of electrical energy meters in private households in Norway, taking into account also the costs of erroneous decisions associated with measurement uncertainty.

The cost of rejection is the cost of replacement of all units in the batch, NOK 2000, per unit. The cost of accepting the batch is the sum of all costs of a continuation of erroneous measurements. The accumulated cost error was calculated for a random meter in the batch due to measurement errors by a Monte Carlo approach.

The annual cost error for a random meter in a particular batch is calculated in a Monte Carlo simulation with a large number of randomly selected sets of data. These sets of data are:

1) Actual measurement errors for a randomly selected measuring device contained in the sample.
2) Actual frequency of use, one hour resolution, randomly drawn from a set of typical power profiles. Price of utility during one year, one hour resolution.

In this case, it is important to study the risks involved with two opposite and excluding decisions, accept or reject, which were about to be taken after the measurements have finished. At this moment, the costs connected to both decisions is known, as is the estimated error rate. The probability density function for the true error rate, p, is a normalized binomial distribution. In Figure 19, consumer and producer risks are plotted as functions of the number of non-conforming units d (instead of $\hat{p} = d/n$) and for a number of sampling plans with sample size n. These risk curves display the risks as functions of the observed error rate, $\hat{p}$, whereas risk operational characteristic curves display the risk as a function of the true error rate, p. Dashed lines indicate risk curves for the first stage of the sampling plan, and unbroken lines indicate risk curves for the second stage. These risk curves are discontinuous, because d can only have integer values. Indexes 1 and 2 refer to first and second part of the sampling plan in table 4.1.

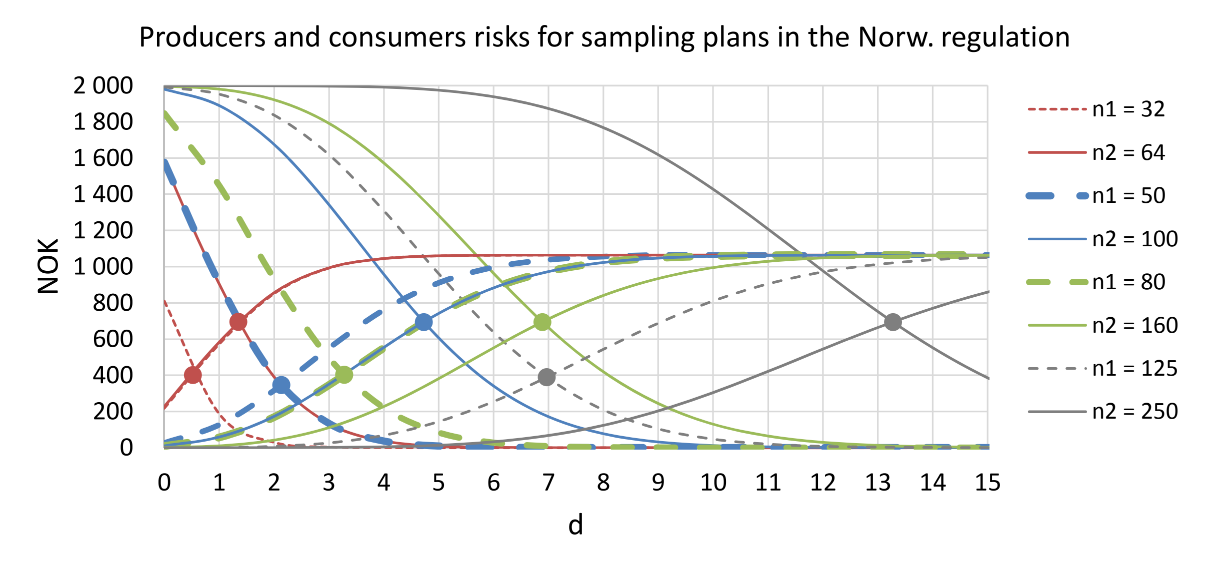


Figure 19: Average risk per instrument for producer (decreasing curves) and consumer (increasing curves) for the first (dashed lines) and second (unbroken lines) part of the sampling plan using watershed specifications. Explanations n_1 = 32, n_2 = 64, etc, refer to the sampling plans in the Norwegian regulation, see table 4.1.

In the example of n_1 = 50 (blue dashed lines) and n_2 = 100 (blue unbroken lines), the risk is higher for the producer than the consumer at d_1 = 2 ($\hat{p}_1$ = 4%) and d_2 = 4 ($\hat{p}_2$ = 4%), in which case the batch should be accepted at such observations of *d*. If the cost of replacing the meters were higher, the producer's risk curve would increase, and there would be even more reason to accept the batch. However, if the cost because of measurement errors were higher (due to larger measuring errors, higher price on the utility or longer time until the next test) the consumer risk might rise above the producer risk at this value of *d*. The conclusions of this case study have been published recently [27].

After a preliminary study of Bayesian methods for incorporating the model uncertainty in the uncertainty budget, the REG from the University of Torino studied numerical recipes for model selection and hypothesis testing. Subsequently, the REG wrote prototype MATHEMATICA (Wolfram) scripts to calculate multidimensional integrals. The numerical methods, procedures, and software developed were applied to a case study consisting of a selection between two different hypothesis about the peaks detected by a counter of photons. A paper about Bayesian statistics has been published [21] and three conference posters have been presented by the REG: SIGMAPHI 2014, CPEM 2014 (Rio de Janeiro – Brazil, 24-29 August 2014), DICE 2014 (Castiglioncello (LI) – Italy, 15-19 September 2014). Two proceedings and an article are currently under preparation. The REG-Researcher, the other Participants and INRIM remained in contact the entire time of grant sharing experience and knowledge, according to the assigned research topic. The REG has also contributed software on ellipsoidal nested sampling to the project results.

Summary

The three objectives have been completed and three Good Practice Guides have been produced and the MATHMET centre has been established.

Novel method for uncertainty evaluation in regression and inverse problems

Inverse problems are cases where the quantity of interest to the metrologist is not measured directly, but is instead a parameter in some relationship between the control and response variables of the measurement process. For example, immunoassay tests in biochemistry where testing for a molecule in solution is done through the use of an antibody. Regression problems are a type of inverse problem where the quantities of interest are parameters of a function linking the control and response variables that will then be used to estimate values of one variable at points where measurements have not been made. Calibration problems are the most commonly used instances of regression problems in metrology, and this project allows more accurate uncertainty calculations to be made.

A Good Practice Guide "A Guide to Bayesian inference for regression problems" has been developed and is freely available for download. It provides a general introduction about how to deal with uncertainty evaluation in regression and inverse problems and suggests a Bayesian treatment based on conditional probability.

The guidelines were used to develop a novel statistical method for the analysis of immunoassay tests. As a result, novel informative prior distributions have been developed which allow for a significantly improved analyses of such tests. These are often used for diagnosis of potential diseases in routine check-up tests of apparently healthy people as well as in diagnosis of specific diseases eg. myocardial infarction for which troponin is an indicator that is commonly tested with the enzyme-linked immunoassay tests studied in this project. In addition, the method serves as an illustration for the application of Bayesian inferences for non-linear regression in metrology. This is an example of a regression problem, not covered by GUM.

Novel methods for computationally expensive systems

Uncertainty evaluation for applications which are described by equations, such as fluid or heat flow, are often non-linear and can require considerable computational resource. Using the GUM gives a linear approximation which is often crude or inaccurate. A model was used to assess and compare different sampling methods and different surrogate modelling methods.

Work on uncertainty evaluation for computationally expensive models focused on smart sampling methods and surrogate models. Smart sampling methods aim to capture the behaviour of the quantity of interest in a small number of model evaluations by careful choice of input parameter values. Surrogate models build an approximation to the full model based on a small number of model evaluations.

A Best Practice Guide "Uncertainty evaluation for computationally expensive models" was produced and published in 2015 and is available for free download. The guide provides a walk-through of the steps in the uncertainty evaluation process and compares multiple smart sampling methods and surrogate models using a simple test problem. Software has been created and tested for implementation of two of the sampling methods investigated. The guidance was illustrated using a set of real-world case study problems; fluid flow through an installation of pipes and through a nozzle, thermal diffusivity and scatterometry methods important in the manufacture of a wide range of products such as photomasks that are widely used in microelectronic element production where a better analysis of measurement uncertainty supports the optimisation or improvement of manufacturing processes by better quality control. The fluid flow case study led to a much improved estimation of systematic errors in flow meters in everyday installations and is therefore beneficial to consumers as well as water suppliers who could get a more accurate estimate of the water consumption, if the findings are applied to the positioning and correction of flow meters in practical situations.

Novel methods for conformity assessment and reliable decision-making that incorporate knowledge of the measurement uncertainties

New perspectives on measurement uncertainty in conformity assessment and decision making have been gained by extending existing approaches to multivariate, qualitative data, computationally expensive systems and the inclusion of measures of impact.

A Best Practice Guide "A guide to decision-making and conformity assessment" and new software for assessing the conformance probability in different circumstances was published is freely available for download. Quality-assured measurements based on traceability and accurate evaluation of the measurement uncertainty are a valuable contribution to the technological infrastructure throughout the innovation process for many products and services.

Best-practice in making multivariate conformity assessment and decision-making was illustrated in case studies of healthcare products. Mathematical and statistical approaches to uncertainty evaluation were introduced in studies where typically there was no simple theory of how the perceived ('response') properties depend on the physical material ('explanatory') properties of the material. This is needed when dealing with properties important for consumers, such as smoothness of material surfaces (skin) as perceived in measurements by human panellists, which in turn depend on the surface topography, friction and hardness, of interest to the manufacturer wishing to fashion the product to the consumer's satisfaction.

The research on a case study on the conformance assessment of electrical utility meters contained a description of the impact of today´s regulation and possible proposed changes. The case study on fire engineering provided a scientifically challenging example for conformity assessment based on a computationally expensive model and employs a hybrid approach. Decisions on conformity are important for environmental monitoring and product safety testing, but there is often no clear or harmonised basis for sharing the risks that arise from measurement uncertainty between the consumer and the supplier. Measurements requiring multivariate approaches (e.g. in healthcare products) are commonly required in conformity assessment. In these cases, two or more quantities and the associated probability density functions are used in conformity assessment and decision making, and current guidelines do not address these cases.

European Centre for Mathematics and Statistics in Metrology

The European Centre for Mathematics and Statistics in Metrology was set up with the support from the NMIs for a long term commitment to scientific collaboration between the members of the consortium. It is a European platform for metrologists, academia and industry for mathematical statistical research in metrological areas to meet future industrial needs.

4 Actual and potential impact

4.1 Dissemination activities

General dissemination has been achieved through 22 scientific journal papers, 47 conference contributions, and 8 trade journal articles. In addition, conference papers that typically addressed wider audience were aimed at wider parts of the metrology community. Journal papers with a tutorial character have been written on Bayesian linear regression and the Markov chain Monte Carlo method that is often required to compute the posterior probability distribution in the Bayesian approach. Software for the general least-squares adjustment of calibration curves, for ellipsoidal nested sampling (REG from University of Torino) and for Bayesian regression has been developed and successfully validated. Software has been created and tested for implementation of two of the sampling methods investigated. The software also carries out sensitivity analysis. A survey paper on sampling methods and their performance in a simple model problem and a computationally expensive application appeared in a metrology journal.

A training course provided by the project provided the first opportunity for application experts from European NMIs and DIs as well as interested stakeholders to learn about the results of the project in early summer 2015. This will help develop the capacity in the application of mathematics and statistics to challenging uncertainty evaluation problems beyond the partners of the project.

The results of this project will strengthen European capabilities for innovation by enabling traceability for modern metrology and measurement techniques and by strengthening European NMIs role in international organisation (e. g. in the Joint Committee for Guides in Metrology (JCGM) – Working group 1 on Uncertainty at the BIPM). Product testing, safety regulations, medical diagnosis and drug testing shall benefit in the future from the procedures for reliable uncertainty evaluation, decision-making and conformity assessment that were developed in this project. Good practice guides on the three topical areas described above give comprehensive summaries of the main findings in combination with tutorial introduction for practitioners and are available for free download on the project webpage.

At the international level the JCGM Working Group 1 (JCGM/WG1) on the Expression of Uncertainty in Measurement, which represents BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, had identified the need for research in the areas of Monte Carlo methods, regression and inverse problems, conformity assessment and the application of expert and prior knowledge. European and international associations dealing with best-practice guides to the use of computationally expensive models had started to address questions of uncertainty. The European NMI scientists involved in the project have started an intensive dialogue with international and national technical committees and provided them with results. The project has contributed invited lectures to the 25 Year Celebration Event of the Guide for the Expression of Uncertainty in Measurement (GUM) at NPL in 2013 as well as dedicated session on selected results to the AMCTM conference of IMEKO TC21 in St. Petersburg in 2014 and to the BIPM workshop on Uncertainty in 2015, where the results were well received by a wider metrological community.

Dissemination events played an important role throughout the project. MATHMET2014, an International Workshop on Mathematics and Statistics for Metrology was held in PTB in March, 2014. This meeting gathered about 75 scientists for NMIs in Europe and overseas as well as colleagues from universities and research institutes in a lively exchange on metrology-related mathematical research themes. Altogether, 34 presentations including more than ten invited talks related to research in the project were made. The combined stakeholder workshop and training course mentioned above took place in Berlin in April 2015 and attracted 50 participants, see website (see http://www.french-metrology.com/workshop/workshop-

NEW04.asp). Dissemination is also achieved by free distribution of 3 Best Practice Guides and 5 Software packages through the project web site.

4.2 Early Impact

23 stakeholders had expressed their strong interest in this project and are regularly informed about its progress and meetings. The list includes industry, universities, professional societies, regulatory bodies, international organisation and NMIs outside Europe. Several stakeholders as well as NMI experts for many of the applications covered in NEW04 have formed an *Application Partner and Stakeholder Committee* (APSC) in the Midterm Meeting in 2014. The role of the APSC was to ensure that the outputs of the product have clear relevance to NMI experimentalists and to industrial stakeholders and end users. Committee members were involved in the stakeholder and training workshops that the NEW04 consortium will organise in 2015. The results of the project were forwarded regularly to several stakeholders and feedback has been received. The case studies have been developed in intensive exchange and resulted in seven joint publication of mathematicians active in NEW04 with application partners, i. e. experimentalists from NMIs. An early uptake of NEW04 outcomes was the decision of the organisers of the BIPM workshop on "Uncertainty" in June 2015 to have a session on uncertainty evaluation in regression problems that contained all the case studies in these area. Another early uptake is represented the use of the Bayesian regression method developed for ELISA in the CCQM-P58.1 Pilot study on Immunoassay Quantitation of Human Cardiac Troponin (D. Bunk et al., Metrologia 52, Technical Supplement, 08006, 2015). The newly founded European Centre has, moreover, adopted the three good practice guides coming out of the project as recommended guides of the Centre on its website.

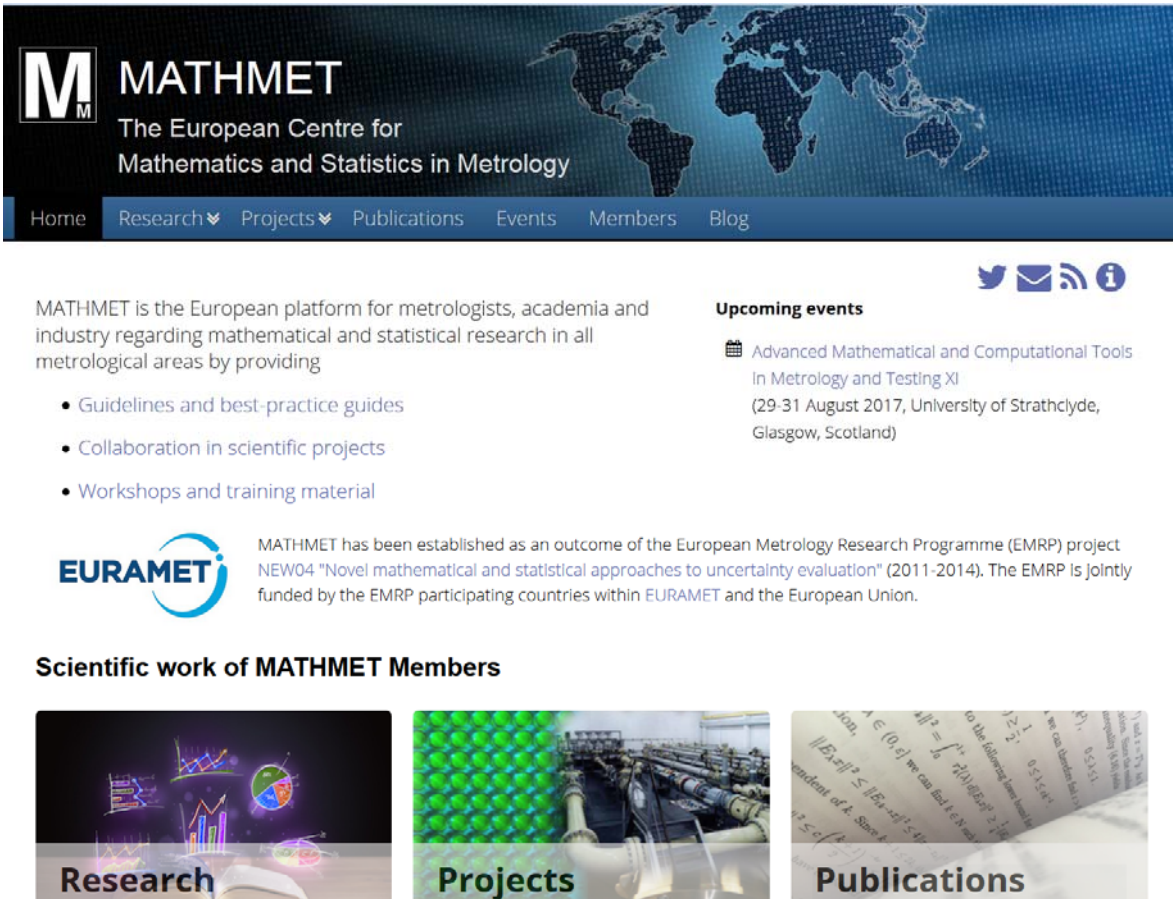


Figure 20: Website of European Centre MATHMET (see www.mathmet.org).

4.3 Effective collaboration between NMIs

This project was the first project that involved all European NMIs with substantial mathematical research activities (INRIM, LGC, LNE, NPL, PTB, SP and VSL) as well as a number of smaller NMIs that contributed applications which required mathematical expertise that provided applications as input for the case studies (CMI, FORCE and JV). Collaborations are documented by 12 publications and 9 conference contributions with joint authors from different partners. The collaboration was facilitated by regular project meetings (the three planned meetings – kick-off in 08/12 at PTB, midterm in 01/14 at VSL and the final meeting in conjunction with the stakeholder workshop and training course at PTB in 04/15 as well as additional working meetings in 04/13 at LNE and in 11/14 were usually very well attended with typically 15 – 20 consortium members and 2 – 10 application partners and stakeholders.

4.4 Contribution to the European metrology infrastructure: MATHMET – The European Centre for Mathematics and Statistics in Metrology

This project aimed at a substantial extension of the mathematical infrastructure for metrology in Europe. Collaboration between European NMIs with mathematical and statistical expertise is essential to ensure wide take-up of the project outputs and to maintain Europe's current leading role in mathematics for metrology. The established *MATHMET – A European Centre for Mathematics and Statistics in Metrology* was carried by four NMIs that were also key members of the consortium. MATHMET will continue to disseminate results internationally through its webpage (www.mathmet.org), via a conference series (next event in November 2016), sessions to other metrology conference (Metrologie Congress, IMEKO events) and dedicated topical workshops and is open for application of further European NMIs for membership. 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. It is organised as non-registered association based on a memorandum of understanding. MATHMET is coordinated by a chair and a co-chair along with a secretary elected by the general assembly for a three-year term. The following sketch shows the organisation structure.

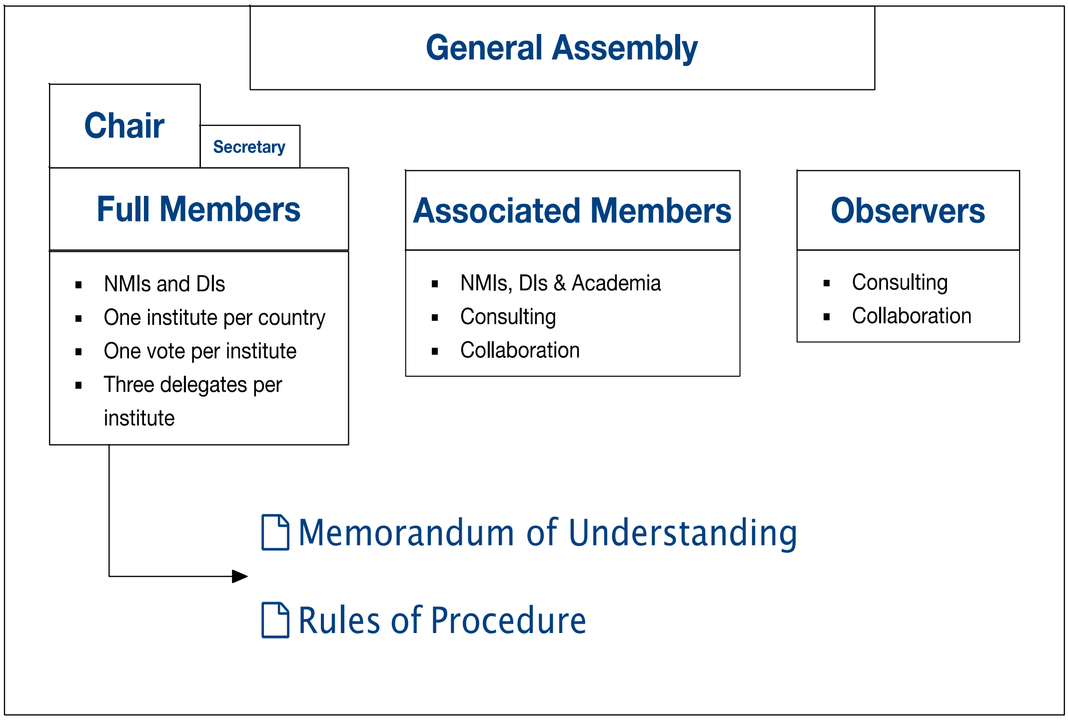


Figure 21: Organisation of European Centre MATHMET.

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 unique reference point of information. MATHMET is probably the single most important impact item delivered from this project and will therefore be described in some more detail in the following.

This project can be a considered a precursor to MATHMET by addressing issues with uncertainty measurements for newly evolving fields of metrology. The project recognised that modern metrology requires expert knowledge of advances in applied mathematics, statistics, numerical computation and related state-of-the-art computational tools to perform such tasks as uncertainty evaluation and provided a starting point for a close collaboration of mathematical experts from European NMIs culminating in the foundation of MATHMET. An appropriate treatment of mathematics and statistics, e. g. in uncertainty evaluation, for metrological applications requires consensus and harmonisation of methods. The most effective way of achieving this was to establish MATHMET as a more permanent body to continue to cooperate on and exchange new mathematical and statistical methods.

MATHMET activities are currently 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. The MATHMET website provides an overview on the activities of MATHMET and is a unique reference point for general information about research activities in mathematical and statistical methods for metrology.

Specifically, it aims to:

- undertake collaborative research in the field of mathematics and statistics for metrology, building new capabilities that are targeted towards support for the SI and traceability, industrial innovation and societal challenges.
- provide services for the metrology community to meet current requirements for mathematical and statistical capabilities.
- promote scientific collaboration between NMIs and provide a joint portal for cooperation with third parties.

MATHMET provides guidelines and best-practice guides, scientific collaboration, and workshops and training material. Joint work will be carried out in dedicated MATHMET projects funded by the member institutes and through externally funded research projects. The MATHMET website (www.mathmet.org) provides an overview on the activities of MATHMET and serves as a unique reference point for general information about research activities in mathematical and statistical methods for metrology.

MATHMET aims to engage with all relevant bodies to provide a focal point for good practice in mathematics and statistics for metrology, directly supporting the work of Government, commerce and industry, the enforcement of law, the maintenance of public health, the protection of the environment and the manufacture of quality products.

5 Website address and contact details

a. Project website: **https://www.ptb.de/emrp/new04-home.html**

b. Website of MATHMET – European Centre: **http://www.mathmet.org/**

c. Contact details: Prof. Dr. Markus Bär

Physikalisch-Technische Bundesanstalt (PTB)

Abbestraße 2-12, 10587 Berlin, Germany

E-mail: **markus.baer@ptb.de**

Tel.: +49-(0)30-3481-7687

6 List of publications

A. Good practice guides:

1. C. Elster, K. Klauenberg, M. Walzel, G. Wuebbeler, 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. R. Pendrill. A Guide to Bayesian Inference for Regression Problems, Good practice guide (WP1), available on https://www.ptb.de/emrp/new04-home.html (2015).

2. K. Rasmussen, J. B. Kondrup, A. Allard, S. Demeyer, N. Fischer, E. Barton, D. Partridge, L. Wright, M. Bär, A. Fiebach, H. Gross, S. Heidenreich, M.-A. Henn, R. Model, S. Schmelter, G. Kok and N. Pelevic. Novel mathematical and statistical approaches to uncertainty evaluation: Best practice guide to uncertainty evaluation for computationally expensive models, Good practice guide (WP2), available on https://www.ptb.de/emrp/new04-home.html (2015).

3. L. R. Pendrill, H. Karlsson, N. Fischer, S. Demeyer and A. Allard. A guide to decision-making and conformity assessment, Good practice guide (WP3), available on https://www.ptb.de/emrp/new04-home.html (2015).

B. Refereed Journal Papers:

1. A. Weissenbrunner, A. Fiebach, S. Schmelter, M. Bär, P. Thamsen, and T. Lederer. Simulation-based determination of systematic errors of flow meters due to uncertain inflow conditions. **Flow Measurement and Instrumentation**, approved for publication (2016).

   DOI: 10.1016/j.flowmeasinst.2016.07.011.

2. S. Heidenreich, H. Gross, M. Bär, and L. Wright. Uncertainty propagation in computationally expensive models: A survey of sampling methods and application to scatterometry.

   Measurement, approved for publication (2016). DOI: 10.1016/j.measurement.2016.06.009.

3. K. Klauenberg and C. Elster. Markov chain Monte Carlo methods: an introductory example

   Metrologia 53 (1), 32--39 (2016).

4. L. R. Pendrill and N. Petersson. Metrology of human-based measurements. Meas. Sci. Tech. 27, 094003 (2106). DOI: 10.1051/metrology/20150017001

5. S. Schmelter, A. Fiebach and A. Weissenbrunner. Polynomchaos zur Unsicherheitsquantifizierung in Strömungssimulationen für metrologische Anwendungen. **tm-Technisches Messen** 83 (2), 71--76 (2016).

6. K. Klauenberg, G. Wübbeler, B. Mickan, P. Harris, and C. Elster. A tutorial on Bayesian Normal linear regression. **Metrologia** 52 (6), 878-892 (2015). DOI: 10.1088/0026-1394/52/6/878.

7. K. Klauenberg, M. Walzel, B. Ebert and C. Elster. Informative prior distributions for ELISA analyses **Biostatistics** 16 (3), 454--464 (2015). DOI: 10.1093/biostatistics/kxu057.

8. S. Schmelter, A. Fiebach, R. Model and M. Bär. Numerical prediction of the influence of uncertain inflow conditions in pipes by polynomial chaos. **Int. J. Comp. Fluid. Dyn.** 29 (6-8), 411-422 (2015). DOI: 10.1080/10618562.2015.1112899.

9. S. Heidenreich, H. Gross and M. Bär. Bayesian approach to the statistical inverse problem of scatterometry: Comparison of three surrogate models. **Int. J. Uncertainty Quantification** 6, 511-526 (2015).

10. A. Allard, N. Fischer, G. Ebrard, P. Harris, L. Wright, D. Rochais and J. Mattout. A multi-thermogram-based Bayesian model for the determination of the thermal diffusivity of a material. **Metrologia** 53, 1--9 (2015). DOI: 10.1088/0026-1394/53/1/S32.

11. L. R. Pendrill and W.P. Fisher. Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. **Measurement** 71, 46--55 (2015).

12. H. Gross, S. Heidenreich, A. Rathsfeld, M. A. Henn and M. Bär. Modeling aspects to improve the solution of the inverse problem in scatterometry. **Cont. Dyn. S.** (8), 497--519 (2015).

13. G. Kok, A. van der Veen, P. Harris, I. Smith and C. Elster. Bayesian analysis of a flow meter calibration problem. **Metrologia** 52 (2), 392--399 (2015).

14. L. R. Pendrill. Man as a measurement instrument. **NCSLI Measure J. Meas. Sci.** 9, 24--35 (2014)

15. O. Bodnar and C. Elster. On the adjustment of inconsistent data using the Birge ratio **Metrologia** 51 (5), 516--521 (2014).

16. L. R. Pendrill. Using measurement uncertainty in decision-making and conformity assessment. **Metrologia** 51 (4), 191--196 (2014)

17. G. Mana and C. Palmisano. Interval estimations in metrology. **Metrologia** 51 (3), 191--196 (2014).

18. H. Gross, S. Heidenreich, M. A. Henn, F. Scholze and M. Bär. Modelling line edge roughness in periodic-line space structures by Fourier optics to improve scatterometry. **JEOS** 9, 14003 (10pp) (2014). DOI: 10.2971/jeos.2014.14003.

19. S. Heidenreich, H. Gross, M. A. Henn, C. Elster and M. Bär. A surrogate model enables a Bayesian approach to the inverse problem of scatterometry. **J. Phys.: Conf. Ser.** 490, 012007(4pp) (2014). DOI: 10.1088/1742-6596/490/1/012007.

20. G. Wübbeler and C. Elster. Simplified evaluation of magnetic field fluctuation thermometry. **Meas. Sci. Tech.** 24, 115004 (8pp) (2013). DOI: 10.1088/0957-0233/24/11/115004.

21. A. Malengo and F. Pennecchi. A weighted total least-square algorithm for any fitting model with correlated variables. **Metrologia** 50, 654-662 (2013). DOI: 10.1088/0026-1394/50/6/654

22. M. A. Henn, S. Heidenreich, H. Gross, C. Elster and M. Bär. Improved grating reconstruction by determination of line roughness in extreme ultraviolet scatterometry. Opt. Lett. 37 (24), 5229-5231 (2012). DOI: 10.1364/OL.37.005229.

C. Trade Journal and Conference Papers

23. H. Karlsson, A.A. Falnes Olsen and L. R. Pendrill. Conformance assessment of electrical energy meters investigated by risk analysis – a case study. OIML - Bulletin LVII (2) (2016).

24. P.G. Spazzini, F. Pennecchi, E. Pessana and A. Piccato. Analysis of Flow Meters Calibration Proceedings of the 9th ISFFM Symposium Arlington (2015). http://library.ceesi.com/...&orgid=16&eid=424

25. S. Demeyer, N. Fischer, F. Didieux and M. Binacchi Statistical methods for conformity assessment when dealing with computationally expensive systems: application to a fire engineering case study. Advanced Mathematical and Computational Tools in Metrology and Testing X. F. Pavese et al. (Eds.), World Scientific, Singapore (2015).

26. A. Weissenbrunner, A. Fiebach, S. Schmelter, M. Straka, M. Bär and T. Lederer. Numerical prediction of the flow rate through a flow meter with uncertain inflow profile Proceedings of Imeko 2015 XXI World Congress Measurement in Research and Industry (2015). https://www.imeko.org/pub...MEKO-WC-2015-TC21-389.pdf.

27. L. R. Pendrill and N. Petersson. Metrology of human-based measurement (2015). DOI: http://dx.doi.org/10.1051/metrology/20150017001;http://cfmetrologie.edpsciences.org/articles/metrology/abs/2015/01/metrology_metr2015_17001/metrology_metr2015_17001.html.

28. S. Heidenreich, H. Gross and M. Bär. Alternative methods for uncertainty evaluations in EUV scatterometry. Proc. SPIE: Modeling aspects in Optical metrology IV 8789, 87890T-1 (8pp) (2013) DOI: 10.1117/12.2020677.

29. M. Bär, C. Elster, S. Heidenreich, C. Matthews, L. R. Pendrill and L. Wright. Novel mathematical and statistical approaches to uncertainty evaluation: introducing a new EMRP research (2013). DOI: 10.1051/metrology/201304010; http://www.metrologie2013.com/index_en.php.

30. C. Elster, K. Klauenberg, M. Bär, A. Allard, N. Fischer, G. Kok, A. van der Veen, P. Harris, I. Smith, L. Wright, S. Cowen, P. Wilson and S. Ellison. Novel mathematical and statistical approaches to uncertainty evaluation in the context of regression and inverse problems (2013). DOI: 10.1051/metrology/201304003; http://www.metrologie2013.com/index_en.php.

C. Papers submitted to Refereed Journals

1. L. Wright, A. Allard, L. Chapman, S. Demeyer, N. Fischer, and D. Partridge. Laser flash experiment on layered materials: parameter estimation and uncertainty evaluation. Submitted, 2015.

2. S. Demeyer, D. Marquis, and N. Fischer. Surrogate-model based sequential sampling estimation of conformance probability for computationally expensive systems: application to fire safety science. Submitted, 2015.

3. P. Wilson and S. L. R. Ellison. Extending digital PCR analysis by modelling quantification cycle data. Submitted, 2015.