

Novel mathematical and statistical approaches to uncertainty evaluation in the context of regression and inverse problems

Clemens Elster¹, Katy Klauenberg¹, Markus Bär¹, Alexandre Allard², Nicolas Fischer^{2,a}, Gertjan Kok³, Adriaan van der Veen³, Peter Harris⁴, Maurice Cox⁴, Ian Smith⁴, Louise Wright⁴, Simon Cowen⁵, Philip Wilson⁵ and Stephen Ellison⁵

¹Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2-12, 10587 Berlin, Germany

²Laboratoire national de métrologie et d'essais (LNE), 1 rue Gaston Boissier, 75724 Paris cedex 15, France

³VSL, Thijsseweg 11, 2629 JA Delft, the Netherlands

⁴National Physical Laboratory (NPL), Hampton Road, Teddington, Middlesex TW11 0LW, UK

⁵LGC Ltd., Queens Road, Teddington, Middlesex TW11 0LY, UK

Abstract. The European Metrology Research Programme (EMRP) is currently funding project EMRP-NEW04 on novel mathematical and statistical approaches to uncertainty evaluation. One focus of the project is uncertainty evaluation in the context of regression and parametric inverse problems. The development of methods for such problems will be carried out in close connection with four application examples. Here we outline these application examples, and present some first results.

1 Introduction

Harmonized and reliable uncertainty evaluation is fundamental to metrology. Without it, measurement results cannot be compared among themselves, or made traceable to primary standards provided by the National Metrology Institutes (NMIs). In the 1990s the *Guide to the expression of uncertainty in measurement* (GUM) [1] was released that has meanwhile established a harmonized evaluation of uncertainty in metrology.

To evaluate the uncertainty associated with an estimate y of a measurand Y , a model of the form

$$Y = f(X_1, X_2, \dots) \quad (1)$$

is central to the GUM. The quantities X_1, X_2, \dots are the so-called input quantities. Model (1) is assumed to be known a priori, and it is important that the model contains all quantities that influence the uncertainty associated with the resulting estimate of the measurand.

The rationale for using a model of the form (1) is the assumption that prior to the analysis knowledge about the input quantities is available, but not (directly) about the measurand. The knowledge about the input quantities is assumed to be expressed in terms of estimates and an associated covariance matrix (GUM), or by a joint probability density function (PDF) (GUM Supplements 1

and 2 [2, 3]). Model (1) is then used to propagate this knowledge to the measurand, either through a linearization using estimates and the covariance matrix as in the GUM, or by (numerical) application of probability calculus in the case of PDFs (GUM Supplements 1 and 2).

The assumptions underlying the GUM procedure are, however, not always fulfilled. For example, prior knowledge about the measurand may be given, or the measurand is known to satisfy certain constraints (such as non-negativity). Furthermore, a model of the form (1) in which the measurand is a (unique) function of the input quantities may not always be adequate. Regression and parametric inverse problems are one particular exception. Here one is faced with observation equations [4] rather than with a model such as (1). There is thus a need for the development of methods for uncertainty evaluation that extend the present GUM and its supplements.

The European Metrology Research Programme (EMRP) is currently funding project EMRP-NEW04 on novel mathematical and statistical approaches to uncertainty evaluation, and one task is the development of such methods in the context of regression and parametric inverse problems. This development will largely be based on Bayesian methods (see, e.g., [5]), is in accordance with the intended revision of the GUM [6], and will be demonstrated by four application examples. These examples comprise the analysis of ELISA

^a Corresponding author: nicolas.fischer@lne.fr

(Enzyme-Linked Immunosorbent Assay) data, the inference of thermophysical properties of materials, the determination of calibration curves for flow meters, and the analysis of digital polymerase chain reaction (PCR) data. The first three applications include regression problems, and the fourth application deals with a parametric inverse problem. Here we introduce these example applications, discuss the challenges for their proper treatment, and present some first results for the ELISA application.

2 ELISA

ELISAs are biochemical tests applied to measure very small concentrations using the highly specific bindings between an antibody and its antigen [7]. Immunoassays have a wide range of applications, for example, to detect the presence of an infection, hormone or drug.

In fluorescent sandwich ELISAs the concentration of a solution is estimated from a set of fluorescence measurements that are generated by repeatedly diluting and chemically preparing the original solution. To determine the relation between concentration and fluorescence intensity, a calibration is performed for each ELISA, i.e. the same protocol steps are performed with a solution of known concentrations.

The relation between concentration and fluorescence intensity is often modeled by logistic functions. Figure 1 shows an example of ELISA calibration data, along with a Bayesian inference of the relation between concentration and fluorescence intensity. The concentrations $x_i, i=1, \dots, n$, were treated as exactly known, and the observed fluorescence intensity measurements were modelled by a Gaussian distribution with mean function

$$g_i(\theta) = g(x_i, \theta) = \theta_1 + \frac{\theta_2 - \theta_1}{1 + (x_i / \theta_3)^{\theta_4}}, \quad (2)$$

and covariance matrix

$$V(\mu) = \text{diag}(\mu_1 + \mu_2 x_1, \dots, \mu_1 + \mu_2 x_n). \quad (3)$$

For all parameters but θ_1 flat nonnegative priors were employed. The prior distribution for the upper asymptote θ_1 was constrained to be below an upper bound that was informed by additional control samples. A Markov chain Monte Carlo method was used to sample from the posterior distribution of the parameters, and Figure 1 shows the logistic function obtained, cf. also [8].

Research within the project will formalize additional and alternative prior knowledge into prior distributions, and investigate the plausibility of these priors as well as models such as that described above. The sensitivity of resulting estimates and associated uncertainties to models

and priors will be examined. Goodness-of-fit tests, residual diagnostics and posterior model probabilities will be used to judge the appropriateness of the likelihoods and priors employed. Although generic, these methods will be applied specifically to the case study of ELISA tests and thereby improve their reliability and comparability. Ultimately, our research will support traceability in biochemistry and simplify decision-making in medical diagnostics.

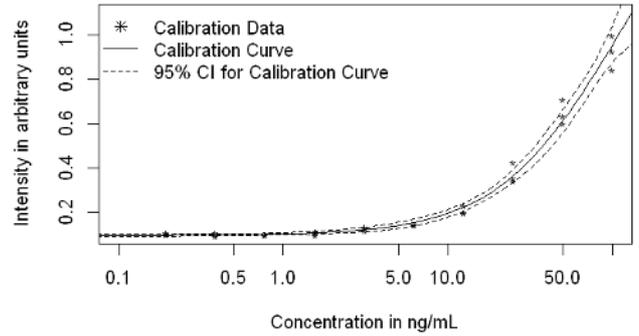


Figure 1. ELISA calibration data together with a Bayesian inference. For the estimated relationship between fluorescence intensity and concentration the solid line indicates the estimate and the dashed lines the (pointwise) 95% credible intervals.

3 Flow meter calibration

A flow meter is a device for measuring the quantity of fluid (e.g., water) transported through a pipeline, both measured as transported volume in a given time and as transported volume per unit of time (flow rate) [9]. A flow meter has typically (at least) two outputs: one analogue output for flow rate and one pulse or frequency output for (total) volume. For the latter output a conversion factor or *pulse factor* (often referred to as a “K-factor”) is specified, meaning that one output pulse is equivalent to a specified volume that has passed the meter (e.g., 10 pulses/l or alternatively 100 ml/pulse). In a flow meter calibration this pulse output is usually calibrated, i.e., the pulse factor is the explained variable in the regression model, and flow rate the regressor variable. For these calibrated pulse factors a regression curve is determined as a function of the (reference) flow rate. This curve can then be used to correct the flow meter reading at a measured flow rate. A model of the form

$$k_i = \beta_0 + \beta_1 q_i^{r_1} + \dots + \beta_p q_i^{r_p} + \varepsilon_i, \quad (4a)$$

$$\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \quad (4b)$$

is often used to describe flow meter calibration measurements. The $k_i, i=1, \dots, n$, denote measurements of the K-factor for the corresponding flow rates q_i , and r_1, \dots, r_p denote powers (e.g., $-1, 1, 2, 3$). The powers are known from previous validation work. The reference in a flow meter calibration can be a volume prover (often a

steel cylinder of known volume), a vessel on a weighing scale or another calibrated flow meter. Figure 2 shows one example.

Initial research within the project will study model (4) from a Bayesian point of view and compare the results with those obtained using a classical approach. In a further stage we may relax the conditions of model (4), i.e. we will consider heteroscedastic models and allow for temporal drifts of the regression curves. Prior knowledge will be explored and exploited in a Bayesian inference for the calibration model. Goodness-of-fit tests, residual diagnostics and posterior model probabilities will be used to judge the appropriateness of the likelihoods and priors employed. A further challenge is statistical calibration [10], i.e., the inverse use of the estimated calibration model to predict the flow rate given observation(s) of the K-factor.

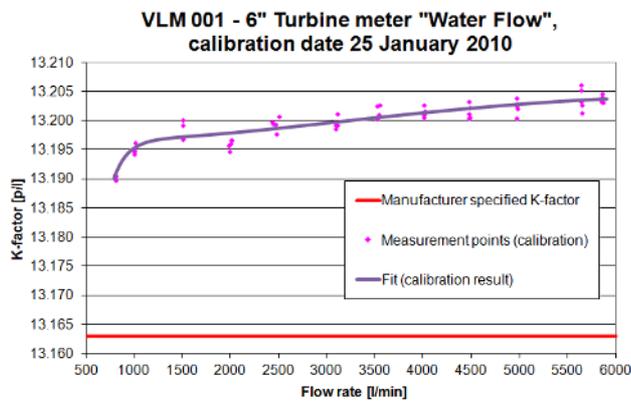


Figure 2. Example of flow meter calibration data and fitted model curve.

4 Thermophysical properties

The flash (or laser flash) technique proposed by Parker et al. [11] is now widely used by NMIs for the measurement of thermal conductivity and thermal diffusivity. In this application we focus on the measurement of thermal diffusivity. The state of the art for this measurement consists of the comparison using the least squares method of an experimental thermogram and a theoretical one obtained as a numerical solution of the heat equation,

$$\lambda \frac{\partial^2 T(z,t)}{\partial z^2} + Q = \rho c_p \frac{\partial T(z,t)}{\partial t} \tag{5}$$

A single value (estimate) for the thermal diffusivity

$$a = \frac{\lambda}{\rho c_p} \tag{6}$$

is obtained, so far without any information about its associated uncertainty. An evaluation of measurement uncertainty according to the GUM [1] may be performed but it is not straightforward [12].

Figure 3 shows the experimental layout in which a measurement is made of the temperature rise on the back face of a thin disk sample resulting from a short energy pulse applied to the front surface. The sample is placed in a furnace and isothermally heated to a uniform temperature. Then, a short laser pulse irradiates one side of the sample. The transient temperature rise on the opposite face is measured by an infra-red detector and is recorded as a function of time. The laser pulse should have the same diameter as that of the sample, and its duration should be extremely short in comparison to the characteristic time of the subsequent heat diffusion in the sample. To avoid the evaluation of the energy absorbed into the sample, a thermogram is normalized to take values in the interval [0,1] as illustrated by Figure 4.

The main focus of research will be the Bayesian inference of a model for the thermogram. Like the ELISA and the flow meter calibration problems this constitutes a regression problem. However, in contrast to those problems no analytical expression for the model is available which makes difficult the numerical calculation of posterior distributions. For this reason we will therefore focus on a homoscedastic error model and employ conjugate priors that facilitate the calculation of the posterior. Both vague or almost noninformative priors and informative prior distributions will be considered.

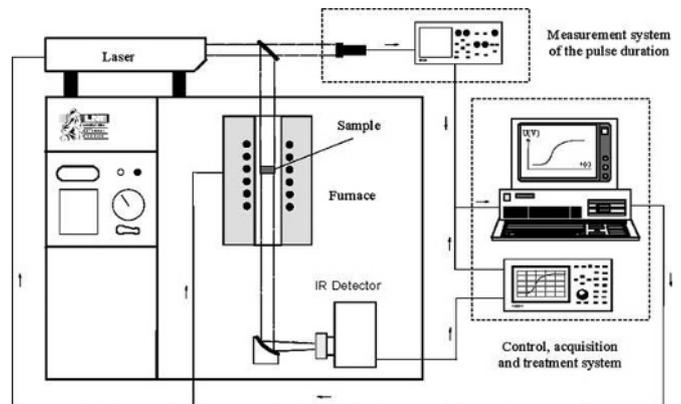


Figure 3. Measurement layout of the laser flash method.

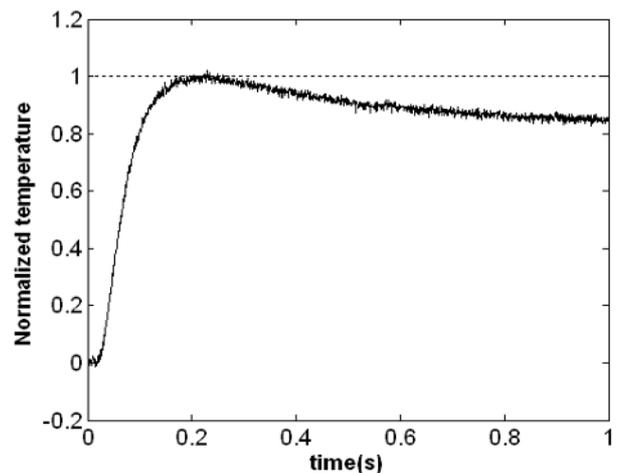


Figure 4. A normalized thermogram.

5 Digital PCR

Digital PCR (dPCR) [13] is a relatively recent measurement method for the absolute determination of the copy number concentration in a DNA or RNA sample. It is quite simple in principle, and works by partitioning the sample into small volumes and taking advantage of the resulting small-sample stochastic variation.

Two major advantages are that no calibration data are needed for the measurement, which is not the case with methods such as real-time quantitative PCR (qPCR), and, moreover, dPCR appears to offer better measurement precision [14]. Typical applications are those involving minority targets, for example single nucleotide polymorphisms (SNPs), mutations and the presence of particular expressed genes in clinical samples [15].

The basic principle of the dPCR method is dilution of the sample and partition of the reaction volume into small enough wells so that some but not all wells will contain one or more molecules. PCR amplification then produces sufficient product to be detected via a fluorescent label, and the sample can be quantified by counting the number of partitions producing a positive response and applying a correction for stochastic effects. Modelling the number of molecules in a well as a Poisson variable relates the probability p of observing a positive well response to the sample concentration as

$$p = 1 - e^{-\lambda} \quad (7)$$

where λ is the mean number of copies per well. An estimate of p follows from the number of wells k showing a positive signal response out of a total n , and by rearranging Equation (7) we obtain an estimate $\hat{\lambda}$ for λ as follows:

$$\hat{\lambda} = -\ln(1 - k/n) \quad (8)$$

In this model, k follows a binomial distribution, and it is implicitly assumed that if a molecule is present, a positive signal will result. The sample concentration then follows from the result of applying equation (8). Since the mean and variance of the Poisson and binomial distributions are well defined, estimates of the uncertainty in $\hat{\lambda}$ based on a GUM-type approach are fairly straightforward to derive [16]. However, as more sources of uncertainty are taken into account and as models become more complex, different approaches to measurement uncertainty become necessary.

We take a Bayesian approach to the analysis of the dPCR method. In the simple case of direct copy number measurement the likelihood function reads

$$l(\lambda; k, n) \propto (1 - e^{-\lambda})^k (e^{-\lambda})^{n-k} \quad (9)$$

which is just the probability mass function of a binomial distribution with k out of n trials resulting in success, $1 - e^{-\lambda}$ is the probability of success, and the normalizing constant has been dropped. The likelihood (9) is of fairly limited interest, but is useful to test the consistency of the Bayesian approach with calculations based on existing theory, and serves as a base upon which more complex models can be constructed.

Within project EMRP-NEW04 we will compare the above simple model with more complex likelihoods that account for additional uncertainty sources. For instance, we (i) will relax the assumption that all wells have the same volume, (ii) account for small changes in the nucleic acid sequences resulting in copy number variation, and (iii) allow for simple failure to detect signals (molecular dropout) whose treatment also requires a duplex assay format. These models will be combined with prior distributions we develop that account for available knowledge about dPCR.

6 Summary and outlook

We have outlined the activities within EMRP project NEW04 to develop methods for uncertainty evaluation in regression and parametric inverse problems, and we have introduced four example applications. For each of the example applications particular challenges and directions of future research were identified, and we have presented first results for the ELISA application.

The four problems address both generic issues as well as specific requirements of the applications. We believe that project EMRP-NEW04 will advance reliable uncertainty evaluation in metrology for regression and parametric inverse problems, and that the case studies produced by our applications will prove useful as templates for other similar problems. The results will also be offered to the JCGM Working Group on the Expression of Uncertainty in Measurement to be considered in the development of future supplements to the GUM.

6 Acknowledgement

This work is part of EMRP-project NEW 04 “Novel mathematical and statistical approaches to uncertainty evaluation“. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

References

1. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*, Joint Committee for Guides in Metrology, JCGM 100:2008 (2008).

2. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, *Evaluation of measurement data – Supplement 1 to the Guide to the expression of uncertainty in measurement – Propagation of distributions using a Monte Carlo method*, Joint Committee for Guides in Metrology, JCGM 101:2008 (2008).
3. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, *Evaluation of measurement data – Supplement 2 to the Guide to the expression of uncertainty in measurement – Extension to any number of output quantities*, Joint Committee for Guides in Metrology, JCGM 102:2011 (2011).
4. A. Possolo, B. Toman, *Assessment of measurement uncertainty via observation equations*, *Metrologia*, **44**, 464-475 (2007)
5. A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, *Bayesian Data Analysis – Second edition*, (Chapman and Hall, London, 2003).
6. W. Bich, M. G. Cox, R. Dybkaer, C. Elster, W. T. Estler, B. Hibbert, H. Imai, W. Kool, C. Michotte, L. Nielsen, L. Pendrill, S. Sidney, A. M. H. van der Veen W. Wöger, *Revision to the “Guide to the Expression of Uncertainty in Measurement”*, *Metrologia*, **49**, 702-705 (2012)
7. D. Wild, editor, *The Immunoassay Handbook*, Elsevier (Third edition, Amsterdam, 2005).
8. K. Klauenberg, B. Ebert, J. Voigt, M. Walzel, J. E. Noble, A. E. Knight, C. Elster, *Bayesian analysis of an international ELISA comparability study*, *Clin. Chem, Lab. Med.*, **49**, 1459-1468 (2011).
9. R. C. Baker, *Flow Measurement Handbook*, Cambridge University Press (2000).
10. P. J. Brown, *Measurement, Regression, and Calibration*, Oxf. stat. sc. ser. (Clarendon Press Oxford, 1993).
11. W. J. Parker, W. J. Jenkins, C. P. Butler, G. L. Abbott, *Flash method of determining thermal diffusivity, heat capacity and thermal conductivity*, *J. Appl. Phys.*, **32**, 1679-1684 (1961).
12. B. Hay, J. R. Filtz, J. Hameury, L. Rongione, *Uncertainty of thermal diffusivity measurements by laser flash method*, *Int. J. Therm.*, **26**, 1883-1898 (2005).
13. B. Vogelstein, K. W. Kinzler, *Digital PCR*, *Proc. Natl. Acad. Sci.*, **96**, 9236-9241 (USA, 1999).
14. R. Sanders, J. F. Huggett, C. A. Bushell, S. Cowen, D. J. Scott, C. Foy, *Evaluation of Digital PCR for Absolute DNA Quantification*, *Anal. Chem.*, **83**, 6474-6484 (2011).
15. G. Pohl, Ie-Ming Shih, *Principle and applications of digital PCR*, *Expert Rev. Mol. Diagn.*, **4**, 41-47 (2004).
16. A. S. Whale, J. Huggett, S. Cowen, V. Speirs, J. Shaw, S. Ellison, C. A. Foy, D. J. Scott, *Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation*, *Nucl. Acids Res.*, **40**, 1-9, (2012).