Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate



2011 Metrologia 48 233

(http://iopscience.iop.org/0026-1394/48/5/001)

IOP PUBLISHING METROLOGIA
Metrologia **48** (2011) 233–240 doi:10.1088/0026-1394/48/5/001

# Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate

**Clemens Elster**[1] **and Blaza Toman**[2]

[1] Physikalisch-Technische Bundesanstalt, Abbestrasse 2-12, 10587 Berlin, Germany
[2] Statistical Engineering Division, Information Technology Laboratory, National Institute of Standards and Technology, US Department of Commerce, Gaithersburg, MD, USA

Received 7 March 2011, in final form 5 May 2011
Published 16 June 2011
Online at stacks.iop.org/Met/48/233

**Abstract**
Application of least-squares as, for instance, in curve fitting is an important tool of data analysis in metrology. It is tempting to employ the supplement 1 to the GUM (GUM-S1) to evaluate the uncertainty associated with the resulting parameter estimates, although doing so is beyond the specified scope of GUM-S1. We compare the result of such a procedure with a Bayesian uncertainty analysis of the corresponding regression model. It is shown that under certain assumptions both analyses yield the same results but this is not true in general. Some simple examples are given which illustrate the similarities and differences between the two approaches.

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

Since its release in 2008 supplement 1 to the GUM (GUM-S1) [1] has found widespread use in the evaluation of measurement uncertainty. The supplement proposes a Monte Carlo method to approximate a probability density function (PDF) which is viewed to encode the state-of-knowledge about the measurand. Standard and expanded uncertainties are then calculated in terms of this PDF. Like the GUM [2], the starting point is a measurement model

$$\theta = f(\mu_1, \ldots, \mu_M) \tag{1}$$

between the input quantities $\mu_1, \ldots, \mu_M$ and the measurand $\theta$. (In contrast to GUM-S1 we use $\theta$ instead of $Y$ to denote the measurand.) GUM-S1 assumes that a joint PDF $p(\mu_1, \ldots, \mu_M)$ has been assigned which encodes the knowledge about the input quantities. The proposed Monte Carlo procedure then approximates the sought PDF $p(\theta)$ for $\theta$ which is uniquely determined by the rules of probability theory. Since it treats the derived PDF as one which expresses the state-of-knowledge about the value of the measurand, GUM-S1 employs the same perspective as Bayesian evaluation of measurement uncertainty. This is in contrast to the GUM itself which contains procedures based on both frequentist and Bayesian statistics [3].

Even though the GUM-S1 method is compatible with the Bayesian approach, it does not employ Bayes' theorem. Furthermore, many features which are a natural part of a Bayesian analysis such as the incorporation of prior knowledge about the measurand, or the treatment of simultaneous observation equations [4], are not addressed in GUM-S1. As GUM-S1 is now widely used, it should be clarified whether it can be adapted to these cases, and if so, to what extent the results are equivalent to a traditional Bayesian analysis. This is important from a conceptual point of view and, perhaps, also for the design of future guidelines.

It has been shown in [5–8] that the PDF obtained by GUM-S1 for the particular case of independent Gaussian measurements on one or several input quantities is the same as the PDF obtained employing Bayes' theorem with specific non-informative prior distributions.

A simple modification [9] of the GUM-S1 procedure allows for the incorporation of prior knowledge about the measurand in the case when no data are actually measured

  

and all of the distributions of the parameters of model (1) are in a sense 'prior'. The procedure corresponds to a consensus density derived via the logarithmic opinion pool described, for example, in [10, 11]. In the presence of data, and using the assignment of PDFs on input quantities as proposed in GUM-S1, application of the simple modification yields a PDF which is the same as that obtained via standard Bayesian analysis only when model (1) is linear [7].

This paper addresses the treatment of regression models, another important application in metrology [12]. We propose the standard Bayesian analysis for uncertainty evaluation of the estimates of the regression coefficients and we show under what conditions the resulting PDFs are the same as those that would be obtained by a proceeding in line with GUM-S1. We focus on the case when no prior knowledge about the measurand is available, for which the Bayesian analysis is carried out using (standard) non-informative priors. We consider regression models of the kind

$$y(x) = g_{\boldsymbol{\theta}}(x) + \varepsilon(x) \tag{2}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ are *unknown* parameters of a *given* model, $g_{\boldsymbol{\theta}}(x)$, which describes the relationship between an independent variable $x$ and a dependent (univariate) variable $y$. The $\varepsilon(x)$ is the error term, usually assumed to be the realization of a Gaussian random variable with mean 0. It is important to keep in mind that $\varepsilon(x)$ is also unknown; only the measurement values $y(x)$ at some $x$ are known in (2). For ease of notation, $x$ is also taken to be univariate. We assume that measurement results for $y$ are available at different fixed, exactly known values of $x$. For example, $x$ may be the temperature and $y$ the indication of an employed thermometer, in which case $g_{\boldsymbol{\theta}}(x)$ is the thermometer's calibration curve. If this curve can be assumed to be a straight line then

$$g_{\boldsymbol{\theta}}(x) = \theta_1 + \theta_2 x. \tag{3}$$

According to its specifications, GUM-S1 is not designed for application to models of this kind. This is because model (2) cannot in general be uniquely transformed into a single measurement model of the kind (1), cf the discussion in section 5.

Nevertheless, GUM-S1 may be applied to a particularly selected measurement model such as that defined by the usual least-squares estimation. This way of estimating the parameters is often applied and we consider this as the 'GUM-S1 approach' to the evaluation of the uncertainties associated with the resulting parameter estimates. We compare the resulting PDF with that obtained via standard Bayesian uncertainty evaluation.

Standard Bayesian uncertainty analysis of regression models is widely available. In all but the simplest cases it requires the use of Markov chain Monte Carlo [13], a different procedure from that required for the implementation of GUM-S1. As GUM-S1 is a current guideline intending the Bayesian point of view, and since many metrologists have established their software implementation for it, it is relevant to consider whether it can be extended to regression models. Related studies have already been undertaken in [14–16]. The conclusions are that application of GUM-S1 and Bayesian uncertainty evaluation in this context are different methods of inference. Yet both yield the same PDFs when the model is linear in all of its parameters, the error variance is known, and the Bayesian analysis employs constant priors for the parameters. This paper confirms these results but shows that when the error variance is unknown, the two methods produce different densities even for linear models and that the GUM-S1 method may in fact fail to produce a result in certain cases. In the non-linear case the densities also usually differ.

The paper is organized as follows. Section 2 specifies all notation and gives general expressions for the Bayesian posterior and the PDF obtained by the GUM-S1 approach. The linear case is further studied in section 3. It shows the equivalence of the GUM-S1 approach applied to the least-squares estimate and a Bayesian analysis using Bayes' theorem in the absence of prior knowledge about the measurand and when the error variance is known; for unknown variance, generally different results are reached. Section 4 discusses two simple non-linear examples which demonstrate that there is a structural difference between the two approaches in general. Section 5 presents a discussion followed by conclusions in section 6.

## 2. Assumptions and notations

Assume that pairs of measurements $(x_i, y_{ij})$ from model (2) are available for $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, where

$$y_{ij} = g_{\boldsymbol{\theta}}(x_i) + \varepsilon_{ij} \tag{4}$$

and the $\varepsilon_{ij}$ are independent realizations from the density $N(0, \sigma^2)$. Thus, the $y_{ij}$ are independently drawn from Gaussian distributions with unknown means $g_{\boldsymbol{\theta}}(x_i)$ and variance $\sigma^2$. We assume that the $x_i$ are fixed and exact. We will consider both the case when the variance $\sigma^2$ is known and the case when it is unknown.

We use upper case letters for random variables and lower case letters for realizations or possible values of them. For instance, $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ denotes the random variables for the parameters and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ possible values of them. Likewise, $y_{ij}$ are realizations from the random variables $Y_{ij}|\boldsymbol{\Theta} \sim N(g_{\boldsymbol{\Theta}}(x_i), \sigma^2)$. In some cases the notation is simpler if we reduce the problem using sufficient statistics [17], that is, use $\bar{Y}_i = (1/n_i)\sum_{j=1}^{n_i} Y_{ij}$ instead of the individual $Y_{ij}$. Then $\bar{Y}_i|\boldsymbol{\Theta} \sim N(g_{\boldsymbol{\Theta}}(x_i), \sigma^2/n_i)$. We specify the reference of a PDF by the use of lower case letters of the associated random variable, e.g. $p(\boldsymbol{\theta})$ denotes a PDF for $\boldsymbol{\Theta}$.

In the Bayesian analysis we determine the posterior PDF $p(\boldsymbol{\theta}|\text{data})$ via Bayes' theorem as

$$p(\boldsymbol{\theta}|\text{data}) \propto p(\boldsymbol{\theta}) l(\boldsymbol{\theta}; \text{data}) \tag{5}$$

when $\sigma$ is known, where $p(\boldsymbol{\theta})$ denotes the prior and $l(\boldsymbol{\theta}; \text{data})$ the likelihood. When $\sigma$ is unknown, the posterior PDF $p(\boldsymbol{\theta}|\text{data})$ is determined as an integral over $\sigma$ of

$$p(\boldsymbol{\theta}, \sigma|\text{data}) \propto p(\boldsymbol{\theta}, \sigma) l(\boldsymbol{\theta}, \sigma; \text{data}), \tag{6}$$

where again $p(\boldsymbol{\theta}, \sigma)$ denotes the corresponding prior. The likelihood $l(\boldsymbol{\theta}, \sigma; \text{data})$ is in our case given by

$$l(\boldsymbol{\theta}, \sigma; \text{data}) \propto \frac{1}{\sigma^N} \mathrm{e}^{-\chi^2(\boldsymbol{\theta},\sigma)/2} \tag{7}$$

where $N = \sum_{i=1}^{n} n_i$ and

$$\chi^2(\boldsymbol{\theta}, \sigma) = \sum_{i=1}^{n} \frac{n_i(\bar{y}_i - g_{\boldsymbol{\theta}}(x_i))^2 + (n_i - 1)s_i^2}{\sigma^2}. \tag{8}$$

In (8), $\bar{y}_i = (1/n_i)\sum_{j=1}^{n_i} y_{ij}$ and $s_i^2 = 1/(n_i - 1) \times \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$. For known $\sigma$ the likelihood $l(\boldsymbol{\theta}; \text{data})$ used in (5) is also given by (7) with $\sigma$ held fixed.

The posterior PDF $p(\boldsymbol{\theta}|\text{data})$ then encodes the knowledge about all parameters $\boldsymbol{\theta}$. For a particular parameter, $\theta_1$ say, the marginal posterior $p(\theta_1|\text{data})$ is determined according to

$$p(\theta_1|\text{data}) = \int p(\boldsymbol{\theta}|\text{data})\, \mathrm{d}\theta_2 \cdots \mathrm{d}\theta_p. \tag{9}$$

An estimate of $\theta_1$ is then calculated as the expectation, and the standard uncertainty as the standard deviation of $p(\theta_1|\text{data})$. This PDF is also used to calculate a (highest posterior density (HPD)) credible interval which is termed (shortest) coverage interval in [1].

The GUM-S1 approach is characterized here as follows: we define $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ as the random variable obtained by minimizing

$$\chi^2(\tilde{\tilde{\boldsymbol{Y}}}) = \sum_{i=1}^{n} \left( \frac{\tilde{\tilde{Y}}_i - g_{\boldsymbol{\theta}}(x_i)}{w_i} \right)^2 \tag{10}$$

with respect to $\boldsymbol{\theta}$. This is formally expressed by the measurement model

$$\boldsymbol{\Theta} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left( \frac{\tilde{\tilde{Y}}_i - g_{\boldsymbol{\theta}}(x_i)}{w_i} \right)^2. \tag{11}$$

In (10) and (11) the $\tilde{\tilde{Y}}_i$ denote random variables whose distribution expresses the knowledge about the true value $g_{\boldsymbol{\theta}}(x_i)$ at the exactly known $x_i$, given the observations $y_{ij}$. Its exact form depends on our knowledge of $\sigma$. The $w_i^2 = 1/n_i$ specify weights according to the number of repeated measurements at $x_i$.

First, consider $\sigma$ to be a known value. Even though GUM-S1 does not explicitly consider this case, when the variance of the sampling distribution is known exactly, $N(\bar{y}_i, \sigma^2/n_i)$ as a distribution of $\tilde{\tilde{Y}}_i$ appears to be in line with GUM-S1. Next, consider the case when $\sigma$ is unknown, here GUM-S1 is clear. The distribution of $\tilde{\tilde{Y}}_i$ is to be taken as Student $t$ with $\gamma_i = n_i - 1$ degrees of freedom and parameters $\bar{y}_i$ and $s_i^2/n_i$, which will be denoted by $t_1(\gamma_i, \bar{y}_i, s_i^2/n_i)$. The joint PDF for $\tilde{\tilde{\boldsymbol{Y}}} = (\tilde{\tilde{Y}}_1, \ldots, \tilde{\tilde{Y}}_n)^{\mathrm{T}}$ is the product of the individual PDFs as the $\tilde{\tilde{Y}}_i$ are independent. Note that for this density to exist the requirement that *all* $n_i \geqslant 2$ must be satisfied. It is possible to relax this and require that *at least one* of the $n_i \geqslant 2$. In that case, the joint distribution of $\tilde{\tilde{\boldsymbol{Y}}} = (\tilde{\tilde{Y}}_1, \ldots, \tilde{\tilde{Y}}_n)^{\mathrm{T}}$ can be taken to be multivariate Student $t$ [18] with degrees of freedom $\gamma_n = \sum_{i=1}^{n} (n_i - 1) = N - n$, and parameters $\bar{\boldsymbol{y}} = (\bar{y}_1, \ldots, \bar{y}_n)^{\mathrm{T}}$ and $s_y^2 \boldsymbol{\Sigma}_n$, where $s_y^2 = (1/\gamma_n) \sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$, see [17]; $\boldsymbol{\Sigma}_n = \mathrm{diag}(1/n_1, \ldots, 1/n_n)$ denotes a diagonal matrix with corresponding elements. This density will be denoted by $t_n(\gamma_n, \bar{\boldsymbol{y}}, s_y^2 \boldsymbol{\Sigma}_n)$. In this case, the $\tilde{\tilde{Y}}_i$ are not independent. For the marginal PDF of a single $\tilde{\tilde{Y}}_i$ a $t$-distribution $t_1(\gamma_n, \bar{y}_i, s_y^2/n_i)$ is obtained. The two marginal PDFs are quite different. We will discuss both variants.

After specifying the joint PDF for $\tilde{\tilde{\boldsymbol{Y}}}$, a realization of $\boldsymbol{\Theta}$ is obtained by drawing a set of 'new data' $\tilde{\tilde{y}}_i, i = 1, \ldots, n$ and by determining that $\boldsymbol{\theta}$ for which $\chi^2(\tilde{\tilde{\boldsymbol{y}}})$ in (10) is minimum. In repeatedly doing this, the PDF $p_{\mathrm{LS}}(\boldsymbol{\theta})$ for $\boldsymbol{\Theta}$ is approximated as described in GUM-S1.

## 3. Linear case

The case of a linear regression model can be treated analytically since in this case the PDF $p_{\mathrm{LS}}(\boldsymbol{\theta})$ obtained by application of GUM-S1 to the least-squares estimate can be given in closed form. The reason is that in this case the corresponding measurement model (11) has a closed form solution. In the linear case the regression model (4) can be written as

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{12}$$

with $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^{\mathrm{T}}$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})$, $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \ldots, \varepsilon_{nn_n})^{\mathrm{T}}$, and the $N \times p$ design matrix with $N = \sum_{i=1}^{n} n_i$ is

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 \otimes \mathbf{1}_{n_1} \\ \vdots \\ \boldsymbol{C}_n \otimes \mathbf{1}_{n_n} \end{pmatrix},$$

where $\boldsymbol{C}_i = (C_{i1}, \ldots, C_{ip})$ and $C_{i\alpha} = \partial g_{\boldsymbol{\theta}}(x)/\partial\theta_\alpha|_{x=x_i}$. The notation $\otimes$ denotes the Kronecker product, and $1_{n_i}$ is a $n_i$ dimensional vector whose elements are all equal to one. The vector $\boldsymbol{\varepsilon}$ is drawn from a $N(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$ distribution where $\boldsymbol{I}_N$ is the identity matrix of size $N$ and $\mathbf{0}$ a vector of length $N$ containing zeros.

### 3.1. Known $\sigma$

*3.1.1. Bayesian analysis.* When $\sigma$ is known, it is simpler to write the model in terms of the sufficient statistics, the $\bar{y}_i$, that is,

$$\bar{\boldsymbol{y}} = \boldsymbol{C}_S\boldsymbol{\theta} + \boldsymbol{\varepsilon}_S, \tag{13}$$

where $\boldsymbol{C}_S = [\boldsymbol{C}_1^{\mathrm{T}} \cdots \boldsymbol{C}_n^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{\varepsilon}_S$ is distributed $N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_n)$. The posterior density $p(\boldsymbol{\theta}|\text{data})$ is

$$p(\boldsymbol{\theta}|\text{data}) \propto p(\boldsymbol{\theta}) \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathrm{T}} \boldsymbol{V}^{-1} (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})} \tag{14}$$

where

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{C}_S^{\mathrm{T}} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{C}_S)^{-1} \boldsymbol{C}_S^{\mathrm{T}} \boldsymbol{\Sigma}_n^{-1} \bar{\boldsymbol{y}} \tag{15}$$

and

$$\boldsymbol{V} = (\boldsymbol{C}_S^{\mathrm{T}} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{C}_S)^{-1}. \tag{16}$$

For the linear model and in the absence of prior knowledge $p(\boldsymbol{\theta}) \propto 1$ may be employed to express prior ignorance about the parameters [19]. The posterior is then given by

$$p(\boldsymbol{\theta}|\text{data}) = \frac{1}{\sigma^p\sqrt{(2\pi)^p|\boldsymbol{V}|}}\mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})} \tag{17}$$

where $|\boldsymbol{V}|$ stands for the determinant of $\boldsymbol{V}$.

*3.1.2. GUM-S1 applied to the least-squares estimate.* The random variables $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ defined in (11) are readily obtained as

$$\boldsymbol{\Theta}_{\mathrm{LS}} = (\boldsymbol{C}_S^{\mathrm{T}}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{C}_S)^{-1}\boldsymbol{C}_S^{\mathrm{T}}\boldsymbol{\Sigma}_n^{-1}\tilde{\tilde{\boldsymbol{Y}}}. \tag{18}$$

Since $\tilde{\tilde{\boldsymbol{Y}}}$ is distributed as $N(\bar{\boldsymbol{y}}, \sigma^2\boldsymbol{\Sigma}_n)$ for known $\sigma$, cf section 2, the PDF $p_{\mathrm{LS}}(\boldsymbol{\theta})$ immediately follows to be

$$p_{\mathrm{LS}}(\boldsymbol{\theta}) = \frac{1}{\sigma^p\sqrt{(2\pi)^p|\boldsymbol{V}|}}\mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})}, \tag{19}$$

where $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{V}$ are given by (15) and (16), respectively. Hence, the GUM-S1 PDF (19) is the Bayesian posterior (14) when a constant prior $p(\boldsymbol{\theta}) \propto 1$ is employed for the latter.

### 3.2. Unknown $\sigma$

*3.2.1. Bayesian analysis.* When $\sigma$ is unknown, the non-informative prior density

$$p(\boldsymbol{\theta}, \sigma) \propto 1/\sigma$$

can be employed [19]. The posterior density is $t_p(\gamma_{\mathrm{B}}, \hat{\boldsymbol{\theta}}, s_{\mathrm{B}}^2\boldsymbol{V})$ with $\gamma_{\mathrm{B}} = N - p$,

$$s_{\mathrm{B}}^2 = \frac{1}{\gamma_{\mathrm{B}}}(\boldsymbol{y} - \boldsymbol{C}\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{C}\hat{\boldsymbol{\theta}}),$$

and $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{V}$ given by (15) and (16), respectively. This density is

$$p(\boldsymbol{\theta}|\text{data}) = \frac{\Gamma\left[\frac{1}{2}(\gamma_{\mathrm{B}} + p)\right]|\boldsymbol{V}|^{-1/2}\,s_{\mathrm{B}}^{-p}}{\Gamma\left(\dfrac{\gamma_{\mathrm{B}}}{2}\right)(\pi\gamma_{\mathrm{B}})^{p/2}} \times \left[1 + \frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{\gamma_{\mathrm{B}}s_{\mathrm{B}}^2}\right]^{-\frac{1}{2}(\gamma_{\mathrm{B}}+p)}. \tag{20}$$

Note that this is a proper density for any set of $n_i \geqslant 1$, even for the case where all of the $n_i$ equal 1, as long as $N > p$.

*3.2.2. GUM-S1 analysis.* The random variables $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)$ defined in (11) are obtained as in (18) with the joint PDF for $\tilde{\tilde{\boldsymbol{Y}}}$ taken as either a product of the $n$ Student $t$-distributions $t_1(\gamma_i, \bar{y}_i, s_i^2/n_i)$ (clearly the GUM-S1 prescription), or a multivariate $t$-distribution $t_n(\gamma_n, \bar{\boldsymbol{y}}, s_y^2\boldsymbol{\Sigma}_n)$.

Consider the case when all of the sample sizes $n_i$ are equal to 1. In that situation the GUM-S1 procedure fails because there are not enough degrees of freedom to estimate $\sigma$ and so, clearly, the GUM-S1 procedure does not produce the Bayes PDF given in (20). This illustrates that one difference between the two approaches is in the estimation of the unknown $\sigma$. Further evidence is provided by the case of $n_i = 2$ for all of the $i$. When assigning independent Student $t$ densities with 1 degree of freedom to the $\tilde{\tilde{Y}}_i$, the solution of (18) is a linear function of the $\tilde{\tilde{Y}}_i$ and it was shown in [20] that the marginal distribution for a single parameter, $\Theta_i$, has a Cauchy distribution. Suppose that instead we use the $t_n(\gamma_n, \bar{\boldsymbol{y}}, s_y^2\boldsymbol{\Sigma}_n)$ for $\tilde{\tilde{\boldsymbol{Y}}}$, then (18) implies that $\boldsymbol{\Theta}_{\mathrm{LS}}$ follows the multivariate $t$-distribution $t_p(\gamma_n, \boldsymbol{A}\bar{\boldsymbol{y}}, s_y^2\boldsymbol{A}\boldsymbol{\Sigma}_n\boldsymbol{A}^{\mathrm{T}})$ with $\gamma_n = n$ degrees of freedom and $\boldsymbol{A} = (\boldsymbol{C}_S^{\mathrm{T}}\boldsymbol{\Sigma}_n^{-1}\boldsymbol{C}_S)^{-1}\boldsymbol{C}_S^{\mathrm{T}}\boldsymbol{\Sigma}_n^{-1}$. Consequently, the marginal distribution for a single parameter, $\Theta_i$, also follows a $t$-distribution with $n$ degrees of freedom. In contrast, the Bayesian PDF for $\Theta_i$ is the Student $t$ density (20) with $\gamma_{\mathrm{B}} = 2n - p$ degrees of freedom.

For GUM-S1, pooling of the variances leads to (significantly) larger degrees of freedom for the resulting $t$-distributions of the $\Theta_i$ than no pooling. Therefore, assigning a multivariate $t$-distribution to $\tilde{\tilde{\boldsymbol{Y}}}$ may be deemed to be more advantageous. But neither way will GUM-S1 generally produce the Bayesian posterior, and the degrees of freedom $\gamma_{\mathrm{B}}$ always exceed $\gamma_n$ unless $p = n$.

The more replicates that are obtained at each value of $x$, the closer the results of the Bayesian posterior and the GUM-S1 approach with pooled variances are expected to be. The reason is that the (different) degrees of freedom in both cases are large and the $\bar{y}_i$ will eventually converge to the $g_{\boldsymbol{\theta}}(x_i)$ and these in turn are points of a *known* function. This behaviour would not occur if the $g_{\boldsymbol{\theta}}(x_i)$ was simply an *approximation* of an unknown underlying function, another common application of least-squares estimation.

### 3.3. Example

In order to illustrate the estimation of regression parameters for the linear case we consider the following regression model

$$g_{\boldsymbol{\theta}}(x) = \theta_1 + \theta_2 x^2. \tag{21}$$

Let the parameter of interest be $\theta_2$. We assume first that $\sigma$ is known and that no prior knowledge is available which is expressed by the prior $p(\theta_1, \theta_2) \propto 1$. Figure 1 shows a set of simulated data for the case when all of the $n_i$ are equal to 1. The parameters underlying the simulation are $(\theta_1, \theta_2) = (1, 1)$ and $\sigma = 0.2$. In addition to the data, figure 1 also shows the curve (21) with $(\theta_1, \theta_2) = (1, 1)$, and the curve $y(x) = \hat{\theta}_1 + \hat{\theta}_2 x^2$ where $\hat{\boldsymbol{\theta}}$ denotes the expectation of the Bayesian posterior $p(\boldsymbol{\theta}|\text{data})$ (17) which coincides with $p_{\mathrm{LS}}(\boldsymbol{\theta})$ (19) for the case of known $\sigma$. The distribution for $\theta_2$ is obtained by marginalization

$$\begin{aligned} p_{\mathrm{LS}}(\theta_2) = p(\theta_2|\text{data}) &= \int p(\theta_1, \theta_2|\text{data})\,\mathrm{d}\theta_1 \\ &= \int \frac{1}{\sigma^2\sqrt{(2\pi)^2|\boldsymbol{V}|}}\mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})}\,\mathrm{d}\theta_1 \\ &= \frac{1}{\sigma\sqrt{2\pi\,V_{22}}}\mathrm{e}^{-(\theta_2-\hat{\theta}_2)^2/(2\sigma^2V_{22})} \end{aligned} \tag{22}$$

where $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{V}$ are given by (15) and (16).

The PDF for $\theta_2$ is Gaussian, and for these data, the mean and standard deviation are 1.21 and 0.18, respectively.

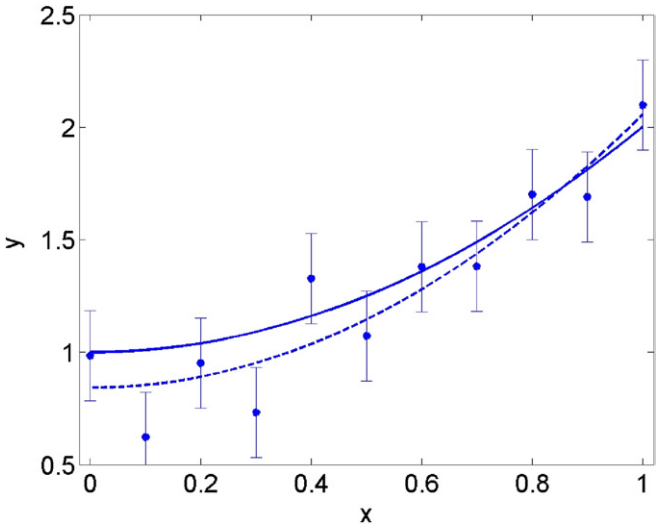


**Figure 1.** Simulated data for regression model (21) including their standard deviation together with the underlying model (solid line) and the estimated model obtained by least-squares adjustment (dashed line).

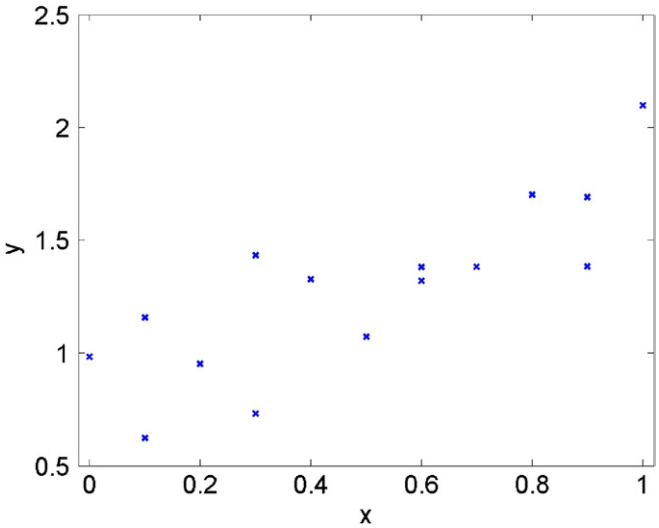


**Figure 2.** Data from figure 1, amended by one replicated measurement at $x = 0.1, 0.3, 0.6, 0.9$.

Next, consider the case when $\sigma$ is unknown. The posterior density for $\theta_2$ is obtained by marginalization from (20) as

$$p(\theta_2|\text{data}) = \frac{\Gamma\left(\frac{N-1}{2}\right)(s_B^2 V_{22})^{-1/2}}{\Gamma\left(\frac{N-2}{2}\right)(\pi(N-2))^{1/2}} \times \left[1 + \frac{(\theta_2 - \hat{\theta}_2)^2}{(N-2)s_B^2 V_{22}}\right]^{-\frac{1}{2}(N-1)}. \quad (23)$$

This is the Student $t$ density with $N-2$ degrees of freedom where $N = n$. For the data shown in figure 1 we obtain a mean of 1.21 and a standard deviation of 0.18.

As was discussed in 3.2.2, the GUM-S1 PDF does not exist for the case of unknown $\sigma$ and one observation at each value of $x$. Suppose that we obtain another replicate at $x = 0.1$, 0.3, 0.6 and 0.9, see figure 2. Now it is possible to employ

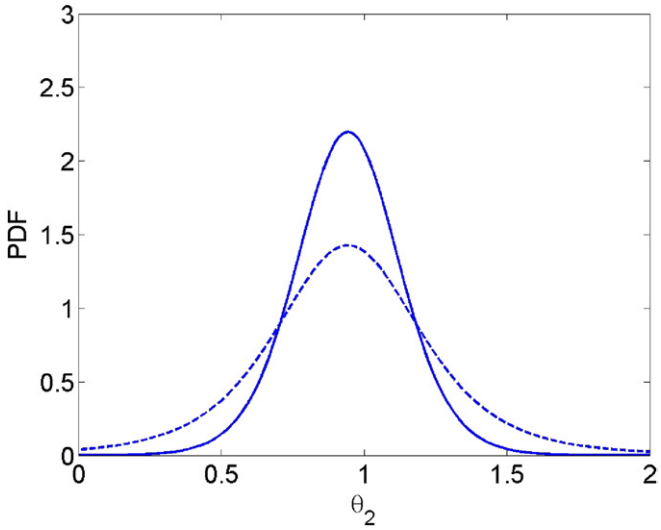


**Figure 3.** GUM-S1 PDF (dashed), and Bayesian posterior (solid) for the data from figure 2.

the GUM-S1 procedure. We did so using the pooled variance variant, leading to a $t$-distribution for $\Theta_2$ with $\gamma_n = 4$ degrees of freedom. For this data set we obtain a mean of 0.94 and a standard deviation of 0.37. For comparison, the Bayes posterior given in (23) is a Student $t$ density with 13 degrees of freedom and produces a mean of 0.94 and a standard deviation of 0.19. Figure 3 shows the two PDFs.

The difference between the GUM-S1 analysis and the Bayesian analysis is mainly in the treatment of the estimation of the unknown $\sigma$, with the GUM-S1 analysis using the deviations from the $\bar{y}_i$, while the Bayesian estimation uses deviations from the estimated function. Both approaches result in $t$-distributions. The difference becomes particularly clear when comparing the corresponding degrees of freedom: while the GUM-S1 variant (considering pooling the variances) results in 4 degrees of freedom, the Bayesian posterior has 13 degrees of freedom and decays much more rapidly.

## 4. Non-linear models

Two examples are used to illustrate differences between the estimation approaches in the non-linear case. The first example shows the importance of the choice of a prior for the Bayesian approach, and the second example demonstrates the effect of constraints on the two solutions.

### *4.1. Example 1*

Let

$$g_\theta(x) = \theta^3, \quad (24)$$

that is, the data $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ are realizations from a $N(\theta^3, \sigma^2)$ distribution which do not depend on any covariates $x$. We also assume that the value of $\sigma$ is known. This example is somewhat artificial since the model becomes linear if we let $g_{\tilde{\theta}}(x) = \tilde{\theta} = \theta^3$; nevertheless, some interesting features can be demonstrated. In this case, for a prior $p(\theta)$, the Bayesian

posterior is given by

$$p(\theta|\text{data}) \propto p(\theta)\frac{\mathrm{e}^{-\frac{1}{2\sigma^2/n}(\theta^3-\bar{y})^2}}{\sqrt{2\pi\sigma^2/n}} \tag{25}$$

where $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$. The least-squares solution is given by $\theta_{\mathrm{LS}} = (\bar{y})^{1/3}$ and hence (11) by

$$\Theta_{\mathrm{LS}} = \left(\tilde{\bar{Y}}\right)^{1/3}, \tag{26}$$

where $\tilde{\bar{Y}} \sim N(\bar{y}, \sigma^2/n)$ with $\tilde{\bar{Y}} = (1/n)\sum_{i=1}^{n} \tilde{Y}_i$. Application of the *change-of-variables formula* [4] then yields for the GUM-S1 PDF

$$p_{\mathrm{LS}}(\theta) = \frac{3\theta^2\mathrm{e}^{-\frac{1}{2\sigma^2/n}(\theta^3-\bar{y})^2}}{\sqrt{2\pi\sigma^2/n}}. \tag{27}$$

In the absence of prior knowledge, if $p(\theta) \propto 1$ is employed, the two densities are clearly different. However, for this example $p(\theta) \propto 1$ might not be a good choice. Indeed, $\theta^3$ is a location parameter and formal procedures [21] would yield a constant prior for $\theta^3$. Application of the change-of-variable formula then yields $p(\theta) \propto \theta^2$ as a non-informative prior for $\theta$. This prior density yields a Bayesian posterior which matches the GUM-S1 PDF.

This simple example illustrates the fact that the choice of non-informative prior is consequential for Bayesian analysis of non-linear models. In the linear case, a constant prior can be expected to be a good selection. For non-linear models, different criteria for choosing a prior can result in different (and non-constant) priors. But if the procedures described in [21] are used, the Bayesian posterior and the GUM-S1 PDF may in fact be identical.

### 4.2. Example 2

The second example is of measurement (such as in the inference of fundamental constants) under the constraint that the quantity is non-negative. The model is

$$g_\theta(x) = \theta, \tag{28a}$$

for

$$\theta \geqslant 0. \tag{28b}$$

The data $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ are realizations from a $N(\theta, \sigma^2)$ distribution with a known $\sigma$; again, the mean does not depend on any covariates. The example is viewed as non-linear due to the constraint (28*b*). When applying an estimation procedure such as least squares, this constraint would need to be taken into account.

Constraint least-squares estimation results in

$$\theta_{\mathrm{LS}} = \begin{cases} \bar{y}, & \bar{y} \geqslant 0 \\ 0, & \text{otherwise}, \end{cases} \tag{29}$$

which depends non-linearly on the data. Working out the GUM-S1 analysis then yields the density

$$p_{\mathrm{LS}}(\theta) = \Phi\left(\frac{-\bar{y}}{\sigma/\sqrt{n}}\right)P_{\theta,0} + \lambda(\theta)\frac{\mathrm{e}^{-\frac{1}{2\sigma^2/n}(\theta-\bar{y})^2}}{\sqrt{2\pi\sigma^2/n}}. \tag{30}$$

which is a mixture of a discrete density with a positive weight at $\theta = 0$, and a continuous distribution. In (30), $\Phi(\cdot)$ denotes the distribution function of $N(0, 1)$, $P_{\theta,0}$ puts probability 1 at $\theta = 0$, and the function $\lambda(\cdot)$ is given by

$$\lambda(\theta) = \begin{cases} 1, & \theta \geqslant 0 \\ 0, & \text{otherwise}. \end{cases} \tag{31}$$

Since $\Phi(-\bar{y}/(\sigma/\sqrt{n})) > 0$ for any value of $\bar{y}$, the GUM-S1 density (30) assigns a non-zero probability that $\theta$ equals zero (which is in contrast to the Bayesian posterior (33) below). The reason for this is that for all $\bar{y} < 0$ the same constrained least-squares estimate $\theta_{\mathrm{LS}} = 0$ results, and GUM-S1 assigns a non-zero probability to $\tilde{\bar{Y}} < 0$.

For Bayesian analysis, unless there is a reason to believe that $\theta = 0$, the usual non-informative prior is

$$p(\theta) \propto \begin{cases} 1, & \theta \geqslant 0 \\ 0, & \text{otherwise}. \end{cases} \tag{32}$$

This results in the posterior density

$$p(\theta|\text{data}) = \frac{\lambda(\theta)}{1 - \Phi\left(\dfrac{-\bar{y}}{\sigma/\sqrt{n}}\right)}\frac{\mathrm{e}^{-\frac{1}{2\sigma^2/n}(\theta-\bar{y})^2}}{\sqrt{2\pi\sigma^2/n}}. \tag{33}$$

Clearly, the two PDFs (30) and (33) are different in structure, and they do not coincide for any sensible choice of prior that we considered. Of course, this difference is relevant only when $\bar{y}$ is negative, or not too large when compared with $\sigma/\sqrt{n}$, as for $\bar{y} \gg \sigma/\sqrt{n}$ the two PDFs coincide. But when there is a difference, the Bayesian posterior is more conservative, i.e. its HPD regions are always larger than those obtained for the GUM-S1 density. Note that for the GUM-S1 density a HPD region may split into two parts (and is expected to do so for $\bar{y} \gg \sigma/\sqrt{n}$) as $\theta = 0$ is always part of the HPD region. Figure 4 shows a set of $n = 5$ simulated data points for $\theta = 0.1$ and $\sigma = 1$. Figure 5 shows the PDFs together with 95% credible intervals. As expected, the credible interval of the Bayesian posterior is larger than that for the GUM-S1 PDF.

## 5. Discussion

Application of GUM-S1 requires a measurement model of the kind (1). This is not available for a regression problem. The reason is that a regression model of the kind (2) holds simultaneously for all (or many) $x$ and the corresponding values of $y$, and it can thus in general not be uniquely transformed into a single measurement model of the kind (1). For example, any mapping which defines $\boldsymbol{\theta}$ as the parameters which minimize some norm $||[y_1, \ldots, y_n]^{\mathrm{T}} - [g_{\boldsymbol{\theta}}(x_1), \ldots, g_{\boldsymbol{\theta}}(x_n)]^{\mathrm{T}}||$ provides a measurement model in terms of the 'input quantities' $y_1, \ldots, y_n$ (provided this is a unique mapping). However, different norms imply different measurement models. If one chooses a particular norm according to reported uncertainties of measurement results (as is done in (10)), then the actual measurement model depends on the statistical model for the data rather than being chosen *a priori* as assumed in the GUM or GUM-S1. Hence,

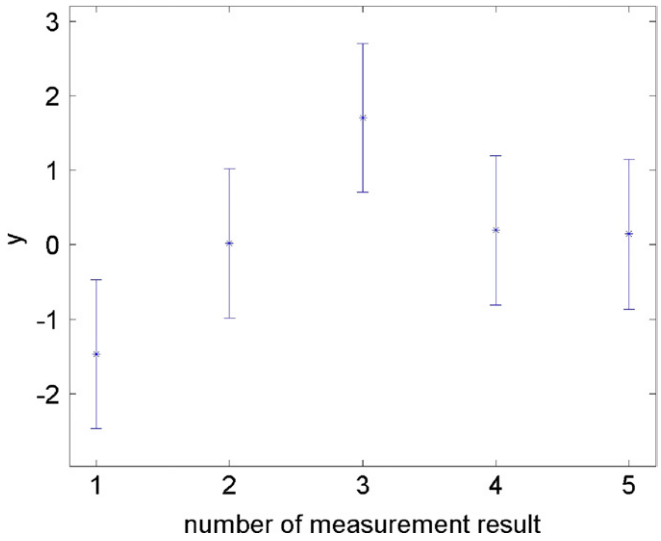


**Figure 4.** Simulated data for regression model (28*a*) and (28*b*) including their standard deviation.

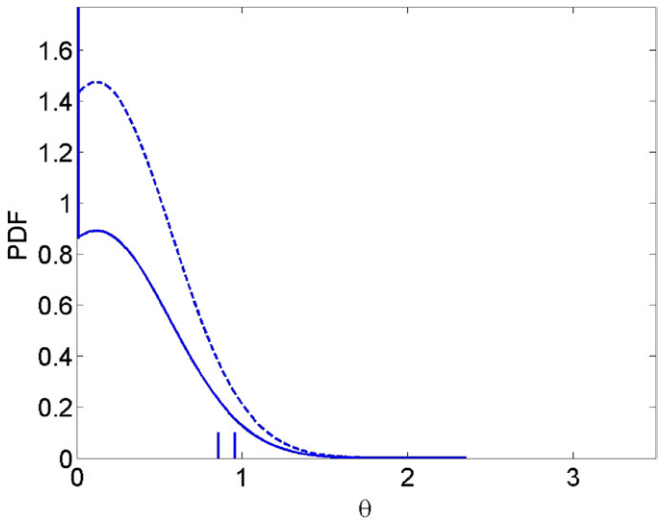


**Figure 5.** GUM-S1 PDF (solid line) and Bayesian posterior (dashed line). The bar at $\theta = 0$ indicates the discrete part of the GUM-S1 distribution with $P(\theta = 0) = 0.40$. The two bars on the abscissa show the upper bound $\bar{\theta}$ of 95% credible intervals of the form $[0, \bar{\theta}]$ for the GUM-S1 PDF (left bar at $\bar{\theta} = 0.85$) and the Bayesian posterior (right bar at $\bar{\theta} = 0.95$).

a regression model cannot be uniquely transformed into a measurement model. Therefore, this is different from the single observation equation considered in [7]. There, the observation equation model and the measurement model were in a one-to-one correspondence. It should be noted that this is the problem considered in GUM-S1.

When $\sigma$ is known, the least-squares parameter estimate minimizing (8) corresponds to the maximum likelihood estimate. Hence in the linear case with known $\sigma$ the norm in (10) is in accordance with the likelihood which is the reason for the observed equivalence between the GUM-S1 and the Bayesian approach in the absence of prior knowledge about the measurand. In the case of unknown $\sigma$ the least-squares solution no longer corresponds to maximum likelihood, and the GUM-S1 PDF is generally different from the Bayesian posterior.

As the interpretation of the GUM-S1 PDF is also that of a state-of-knowledge distribution, it is equivalent to a Bayesian posterior for a *linear* regression model, known variance and the absence of prior knowledge about the measurand. Hence GUM-S1 provides an easy numerical tool for calculation of the Bayesian posterior in these cases. In the presence of prior knowledge about the measurand (and when the variance is known), the Bayesian posterior (14) is proportional to the GUM-S1 PDF (19) (ignoring the prior knowledge) and the prior. In such a situation a simple modification of the GUM-S1 procedure [9] also allows one to produce the Bayesian PDF (14).

When the variance is unknown, and there are no replicated measurements, the GUM-S1 method fails to produce a PDF for the measurand while the Bayesian posterior density does exist. With some replication the GUM-S1 method can be applied to produce a PDF. But this will generally not match the usual Bayesian posterior.

The least-squares estimate does not depend linearly on the data when the regression is not linear. In such cases the GUM-S1 density is usually not the same as a Bayesian posterior, but it can be, as was seen in the example in section 4.1. The example in section 4.2 shows that the Bayesian posterior and the GUM-S1 PDF may differ in structure, a difference not simply due to the choice of underlying prior. The reason is that the measurement model defined by (11) considers a projection of the model on $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$, whereas the Bayesian posterior rests on the norm between the *observed* data $(y_1, \ldots, y_n)$ and the model. It is only in the linear case with a known variance that these different proceedings are equivalent.

We conclude that for the general regression model, it is better to perform a Bayesian analysis rather than to apply GUM-S1. As a consequence, other numerical methods such as Markov Chain Monte Carlo methods [13] need to be employed. As non-linear least-squares problems are encountered in metrology, employment of such methods is recommended. But note that when no prior knowledge about the measurand is available, and the regression model depends non-linearly on its parameters, the selection of a prior is not straightforward. Simply choosing a constant prior may not be a good choice [7, 21].

Model (3) was motivated as one which describes the calibration curve of a thermometer. It was assumed that the abscissa data are exact, often not the case as these data would also be the result of measurement. When the uncertainties of the abscissa data cannot be ignored, minimization of (10) is no longer adequate. Instead, also the abscissa data need to be adjusted, and the resulting least-squares estimate then depends non-linearly on the data even though the regression model is linear w.r.t. the model parameters, cf [22].

We focused here on the case when no prior knowledge about the measurand is given as that case is addressed by GUM-S1. However, often some prior knowledge about the measurand is available or physical constraints are known as in the example in section 4.2, and application of Bayes' theorem is strongly recommended then.

## 6. Conclusions

Application of GUM-S1 for evaluating the uncertainty associated with a least-squares estimate in a regression model has been compared with a Bayesian uncertainty analysis. When the least-squares estimate depends linearly on the data and the variance is known, the GUM-S1 procedure in general yields a density identical to the Bayesian posterior for the usual non-informative prior, otherwise not. In the case of unknown variance and for non-linear problems, the two methods generally produce different results and application of Bayes' theorem is strongly recommended then.

## References


[1] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 *Evaluation of Measurement Data—Supplement 1 to the Guide to the Expression of Uncertainty in Measurement—Propagation of distributions using a Monte Carlo method* Joint Committee for Guides in Metrology, JCGM 101:2008

[2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML 2008 *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement* Joint Committee for Guides in Metrology, JCGM 100:2008

[3] Kacker R and Jones A 2003 On use of Bayesian statistics to make the *Guide to the Expression of Uncertainty in Measurement* consistent *Metrologia* **40** 235–48

[4] Possolo A and Toman B 2007 Assessment of measurement uncertainty via observation equations *Metrologia* **44** 464–75

[5] Elster C, Wöger W and Cox M G 2007 Draft GUM Supplement 1 and Bayesian analysis *Metrologia* **44** L31–2

[6] Kyriazis G A 2008 Comparison of GUM Supplement 1 and Bayesian analysis using a simple calibration model *Metrologia* **45** L9–11

[7] Elster C and Toman B 2009 Bayesian uncertainty analysis under prior ignorance of the measurand versus analysis using the Supplement 1 to the *Guide*: a comparison *Metrologia* **46** 261–6

[8] Lira I and Grientschnig D 2010 Equivalence of alternative Bayesian procedures for evaluating measurement uncertainty *Metrologia* **47** 334–6

[9] Elster C 2007 Calculation of uncertainty in the presence of prior knowledge *Metrologia* **44** 111–6

[10] Genest C and Zidek J V 1986 Combining probability distributions: a critique and an annotated bibliography *Stat. Sci.* **1** 114–48

[11] Willink R 2006 Forming a comparison reference value from different distributions of belief *Metrologia* **43** 12–20

[12] Willink R 2008 Estimation and uncertainty in fitting straight lines to data: different techniques *Metrologia* **45** 290–8

[13] Robert C P and Casella G 1999 *Monte Carlo Statistical Methods* (Berlin: Springer)

[14] Cox M G, Forbes A B, Harris P M and Sousa J 2007 Distributions associated with measurands and their estimates *PTB-BIPM Workshop on the Impact of Information Technology in Metrology (Berlin, Germany)*

[15] Cox M G, Forbes A B, Harris P M and Smith I M 2009 Measurement uncertainty evaluations associated with calibration functions *XIX IMEKO World Congress (Lisbon, Portugal)*

[16] Forbes A B 2009 Nonlinear least squares and Bayesian inference *Advanced Mathematical & Computational Tools in Metrology* ed F Pavese *et al* (Singapore: World Scientific) pp 104–12

[17] Lindley D V 1970 *Introduction to Probability and Statistics* (Cambridge: Cambridge University Press)

[18] Gelman A, Carlin J B, Stern H S and Rubin D B 2004 *Bayesian Data Analysis* (London/Boca Raton, FL: Chapman and Hall/CRC Press)

[19] Box G E P and Tiao G C 1992 *Bayesian Inference in Statistical Analysis* (New York: Wiley)

[20] Walker G and Saw J 1978 The distribution of linear combinations of $t$-variables *J. Am. Stat. Assoc.* **73** 876–78

[21] Kass R E and Wasserman L 1996 The selection of prior distributions by formal rules *J. Am. Stat. Assoc.* **91** 1343–70

[22] Lira I, Elster C and Wöger W 2007 Probabilistic and least-squares inference of the parameters of a straight-line model *Metrologia* **44** 379–84