XVIII IMEKO WORLD CONGRESS
Metrology for a Sustainable Development
September, 17 – 22, 2006, Rio de Janeiro, Brazil

MEASUREMENT SYSTEMS ANALYSIS: CONCEPTS AND COMPUTATIONAL APPROACHES

Alistair Forbes [1], *Christophe Perruchet* [2]

[1] National Physical Laboratory, Teddington, UK, alistair.forbes@npl.co.uk
[2] UTAC, Montlhery, France, christophe.perruchet@utac.com

Abstract: In this paper, we review measurement systems analysis from the point of view of the Guide to the Expression of Uncertainty in Measurement and ISO 5725: Accuracy (Trueness and Precision) of Measurement Methods and Results. We also consider other computational approaches such as Markov chain Monte Carlo simulations that could be useful for uncertainty evaluation.

Keywords: measurement uncertainty, statistical modelling, computational methods.

1. INTRODUCTION

Both the Guide to the Expression of Uncertainty in Measurement (GUM) [1] and ISO 5725 [2] are concerned with the evaluation of uncertainty associated with measurements. The GUM gives guidelines on how to evaluate and express the uncertainty associated with the output of a system, given estimates of the inputs, their associated uncertainties, and a model specifying how the output depends on the inputs. The uncertainty quantifies the likely variation in the output values if all inputs were allowed to vary in accordance with their uncertainties.

The uncertainty associated with an input can be assigned according to a Type A or Type B evaluation. A Type A evaluation is usually provided through analyzing repeated measurements and inferring the uncertainty from the standard deviation of the measurement values. In a Type B evaluation, the uncertainty is assigned by other means, such as expert judgement, past experience, calibration history, or manufacturer's specifications.

The approach to measurement systems analysis in ISO 5725 addresses measurement uncertainty in terms of *trueness* and *precision*. Trueness relates to the difference between the expected measurement result using the system (and method) and an accepted or reference value. Precision relates to the closeness of agreement between repeated measurements under stipulated conditions: the more the conditions are allowed to vary, the larger the expected variation of results. The use of trueness and precision estimates for evaluating uncertainties is presented in ISO/TS 21748 [3], recently published by ISO/TC 69.

One of the differences between the GUM and ISO 5725 is in the treatment of systematic effects. A main thrust, rightly, of the GUM is that the uncertainty due to an unknown systematic effect should be aggregated in exactly the same way as other uncertainty contributions. One could argue that this philosophy has been interpreted by some to mean that it is incorrect to use terms such as trueness or bias. By their nature, systematic effects do not contribute to the variation in repeated measurements so their contribution has to be evaluated by other means, hence Type B evaluations. In ISO 5725, there is an underlying assumption that the measurement systems/methods are being tested against a more accurate source of information, e.g., through using a more accurate measuring system/method or using accurately characterized reference artefacts or materials. In this context, it seems perfectly reasonable to use terms such as bias and trueness.

In section 2 we consider the analysis on interlaboratory comparison data, in particular, looking at what information can be derived using standard least squares methods about laboratory systematic effects. In section 3, we consider a more general approach involving prior information that encodes degrees of belief associated with the uncertainty parameters and describe maximum likelihood and posterior sampling methods to determine parameter estimates. Our concluding remarks are given in section 4.

2. INTERLABORATORY COMPARISONS

A major concern of ISO 5725 is the analysis of interlaboratory comparison (ILC) measurements using analysis of variance (ANOVA) concepts. Since the signing of the Mutual Recognition Arrangement (MRA), interlaboratory (key) comparisons are also a strong focus of attention in the metrology community. The GUM has no direct guidance on the design and analysis of ILCs and metrologists involved in such activity should be interested in ISO 5725.

2.1 Role of ILC exercises

ILCs have different uses. In a validation role, the ILC can confirm that the measurements (values and uncertainties) provided by different laboratories form a self-consistent set of information through showing that the variation in the measurement values can be explained by chance, taking into account the uncertainty statements. Given such a validation, users are provided with quantitative evidence that the laboratories involved in the exercise are implementing procedures that give reliable uncertainty

statements, and hence can judge the suitability of a laboratory.

In a validation role, an ILC is largely passive in that no change in the activities of the laboratories is anticipated assuming consistency of the measurements. (If there is a lack of consistency, then the problem of finding out why can often be difficult.) ILCs can also be used in a more active, information-gathering role.

2.2 Analysis of ILC data

Suppose in an ILC involving the measurement of a single artefact, the measurement model for the *i*th laboratory is

$$y_i = \alpha + \delta_i + \varepsilon_i \qquad (1)$$

where α represents the measurand associated with the artefact, δ_i a systematic (laboratory) effect present in all measurements from that laboratory, and ε_i a random effect for that particular measurement. As part of the uncertainty budgeting activity, the laboratory estimates that $\varepsilon_i \sim N(0,\sigma_i^2)$ and assigns a distribution $\delta_i \sim N(0,\rho_i^2)$ for the systematic effect parameter. Here, we regard ρ_i and σ_i as known *a priori*. In the "standard" model of ISO 5725 it is assumed that ρ_i and σ_i are constant over *i* but are to be estimated as part of the analysis involving multiple measurements; see section 3. For a single measurement of α, the laboratory quotes a measurement result y_i with standard uncertainty $u_i = \left(\rho_i^2 + \sigma_i^2\right)^{1/2}$. Let $\mathbf{A}$ be the $m\times 1$ matrix with $\mathbf{A}_{i1} = 1/u_i$ and $\mathbf{b}$ the $m\times 1$ vector with $b_i = y_i/u_i$. The reference value for α is estimated by the solution a of the linear least squares problem $\mathbf{A}a = \mathbf{b}$ so that

$$a = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}\,.$$

The uncertainty $u(a)$ associated with a is given by $u^2(a) = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}$.

The data also contains information about $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_m)^{\mathrm{T}}$. Let $D_{\mathbf{v}}$ be the $m\times m$ diagonal matrix with *i*th diagonal element $v_i = 1/\sigma_i$, $D_{\mathbf{w}}$ similarly defined with *i*th diagonal element $1/\rho_i$ and set

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{v} & D_{\mathbf{v}} \\ 0 & D_{\mathbf{w}} \end{bmatrix}, \qquad \tilde{b}_i = v_i y_i\,.$$

Then α and $\boldsymbol{\delta}$ are estimated by a and $\mathbf{d}$, respectively, that solve the linear least squares problem

$$\tilde{\mathbf{A}}\begin{bmatrix} a \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{b}} \\ \mathbf{0} \end{bmatrix}.$$

The solution a and its uncertainty $u(a)$ are as for the previous calculation. More simply, if $r_i = y_i - a$ and $r_i = d_i + e_i$, then $d_i/e_i = \rho_i^2/\sigma_i^2$ so that δ_i can be estimated by d_i from knowledge of y_i, σ_i, ρ_i and a. All this information will be available to the *i*th laboratory if the reference value a is published. This means that once a is made known, each laboratory can then adjust subsequent measurements, $y_i := y_i - d_i$, on the basis of this knowledge.

The uncertainties $u(\mathbf{d})$ associated with the estimates $\mathbf{d}$ are found from the uncertainty matrix $\tilde{\mathbf{U}} = \left(\tilde{\mathbf{A}}^{\mathrm{T}}\tilde{\mathbf{A}}\right)^{-1}$. If $u(d_i)$ is known to the *i*th laboratory (from knowledge of $\tilde{\mathbf{A}}$ for example), it can also adjust its evaluated uncertainty for subsequent measurements using the model

$$y_i = \alpha + \delta_i + \varepsilon_i\,,\ \ \varepsilon_i \sim N(0,\sigma_i^2)\,,\ \ \delta_i \sim N(0,u^2(d_i))\,,$$

with adjusted uncertainty $\tilde{u}_i^2 = u^2(d_i) + \sigma_i^2$. The improved knowledge of δ_i gained from the comparison will lead to smaller uncertainties.

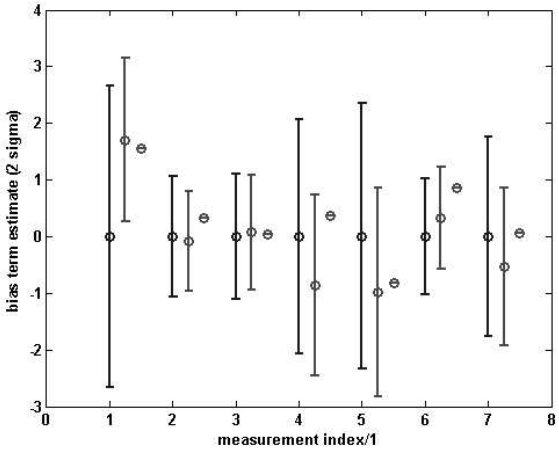

Fig. 1. Prior estimates $0 \pm 2\rho_i$ (left hand error bar) associated with the systematic effects δ_i (rightmost circle), and posterior estimates $d_i \pm 2u(d_i)$ (right hand error bar) for data generated with σ_i, $\rho_i \approx 1.0$.

Fig. 1 shows prior estimates $0 \pm 2\rho_i$, posterior estimates $d_i \pm 2u(d_i)$ and actual values δ_i for data generated according to the model in Eq. 1, with σ_i, $\rho_i \approx 1.0$. The improved knowledge of δ_i gained from the comparison will lead to smaller uncertainties. If the prior model for laboratories is known to be correct, then a legitimate aim of the ILC is to provide improved estimates of the systematic effect parameters. Note that if two laboratories participating in the same ILC adopt this procedure, their subsequent measurements will be correlated since the measurements will depend on common information derived from the ILC.

Fig. 2 shows a set of subsequent measurements $y_i \pm 2u_i$ and adjusted measurements $y_i - d_i \pm 2\tilde{u}_i$. In this case, there is a modest reduction in the uncertainty associated with δ_i which leads to a even more modest reduction in the uncertainty associated with adjustments of a subsequent set of measurements. Figures 3 and 4 show the same information for a simulation with $\rho_i \approx 1.0$ and $\sigma_i \approx 0.2$. In this case, the smaller uncertainty associated with the random component means that there is more definite information

about δ_i leading to significant reductions in posterior uncertainties.

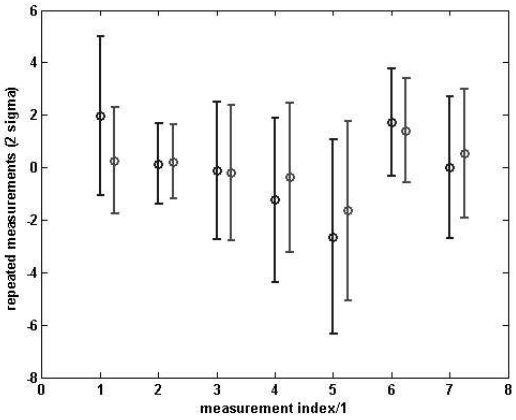

Fig. 2. Simulated subsequent measurements $y_i \pm 2u_i$ (left hand error bar) and adjusted subsequent measurements $y_i - d_i \pm 2u(d_i)$ (right hand error bar) for simulated data generated with σ_i and $\rho_i \approx 1.0$.

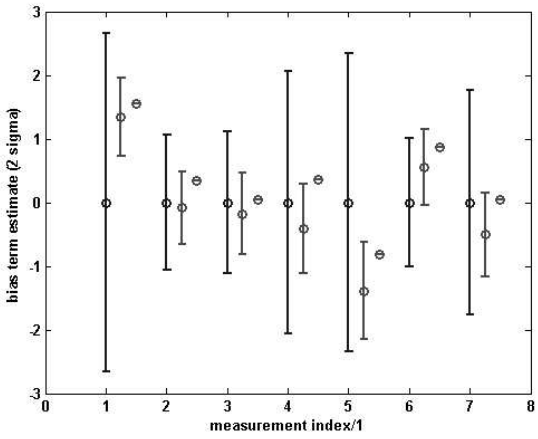

Fig. 3. Same as Fig. 1 but for $\rho_i \approx 1.0$ and $\sigma_i \approx 0.2$.

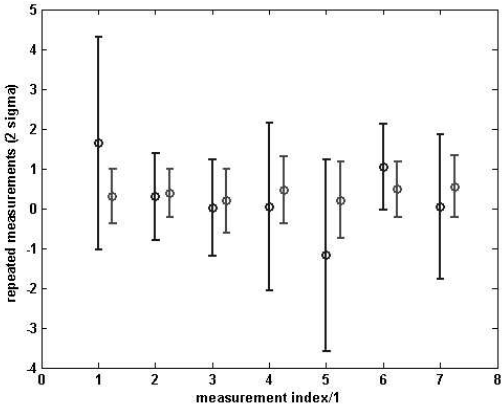

Fig. 4. Same as Fig. 2 but for $\rho_i \approx 1.0$ and $\sigma_i \approx 0.2$.

3. PARTIAL INFORMATION ABOUT VARIANCES

The analysis above assumes that the standard deviations ρ_i and σ_i are known *a priori*. More realistically, these statistical parameters will be estimated with varying degrees of belief. The ISO 5725 approach applies in the case where they are constant across laboratories but for which there is no prior information. For example, consider the model ,

$$y_{i,q} = \alpha + \delta_i + \varepsilon_{i,q}, \quad \varepsilon_{i,q} \in N(0,\sigma^2), \quad \delta_i \in N(0,\rho^2), \quad (2)$$

where the *i*th laboratory, $i = 1,...,L$, performs n repeat measurements, $q = 1,...,n$. Letting

$$\bar{y}_{i.} = \frac{1}{n}\sum_{q=1}^{n} y_{i,q} \quad \text{and} \quad \bar{y}_{..} = \frac{1}{nL}\sum_{i=1}^{L}\sum_{q=1}^{n} y_{i,q},$$

then the expected value of $\mathrm{MSB} = \frac{n}{L-1}\sum_{i=1}^{L}(\bar{y}_{i.} - \bar{y}_{..})^2$ is $\sigma^2 + n\rho^2$ and that for $\mathrm{MSW} = \frac{1}{L(n-1)}\sum_{i=1}^{L}\sum_{q=1}^{n}(y_{i,q} - \bar{y}_{i.})^2$ is σ^2. This means that σ and ρ can be estimated by *s* and *r*, respectively, where $s^2 = \mathrm{MSW}$ and

$$r^2 = (\mathrm{MSB} - \mathrm{MSW})/n.$$

Unfortunately, for small samples there is no guarantee that r^2 so calculated will be positive. Figure 5 shows the histogram of the calculated r^2 for 100,000 Monte Carlo simulations for data generated with n = L = 4 and $\sigma = \rho$ =1.0. Approximately 10 % of the calculated values are negative.

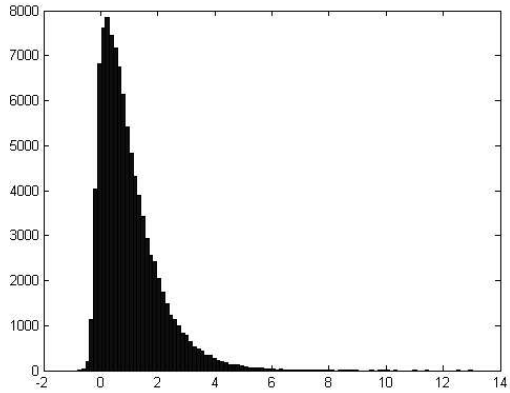

Fig. 5. Histogram of calculated values for r^2 for 100,000 Monte Carlo simulations.

Our analysis so far has considered the cases where either σ and ρ are known exactly or there is no information about them. We now consider the case where prior information about both is given in terms of prior distributions for variables $\phi = 1/\sigma^2$ and $\psi = 1/\rho^2$ in the form

$$m_0\sigma_0^2\phi \sim \chi^2_{m_0}, \quad L_0\rho_0^2\psi \sim \chi^2_{L_0}. \quad (3)$$

These distributions encode a degree of belief in the prior estimate σ_0 for σ, for example, equivalent to that gained by observing the sum of squares $\sigma_0^2 = \sum_{q=1}^{m_0} z_q^2$ where each

z_q is sampled from $N(0,\sigma^2)$. The pdf $p(\phi)$ for ϕ is such that $p(\phi) \propto \phi^{m_0/2-1} \exp\{-\phi m_0 \sigma_0^2/2\}$, $\phi \geq 0$. As m_0 increases, the greater the belief in the estimate σ_0. In the limiting case, $m_0 \to 0$, $p(\phi) \propto 1/\phi$, a non-informative prior for the parameter ϕ (but one which encodes the constraint $\phi \geq 0$).

3.1 Maximum likelihood estimation and Gaussian approximations

We let $\boldsymbol{\beta}$ represent the four sets of parameters α, $\boldsymbol{\delta}$, $\log\phi$ and $\log\psi$. The probability $p(\mathbf{y} \mid \boldsymbol{\beta})$ of observing data **y**, given the parameters is

$$p(\mathbf{y} \mid \boldsymbol{\beta}) = \phi^{nL/2} \exp\{-\phi F(\alpha, \boldsymbol{\delta})/2\},$$

where

$$F(\alpha, \boldsymbol{\delta}) = \sum_i \sum_q (y_{i,q} - \alpha - \delta_i)^2.$$

Similarly, the prior distribution $p(\boldsymbol{\delta})$ for the fixed effects is

$$p(\boldsymbol{\delta}) = \psi^{L/2} \exp\{-\psi S(\boldsymbol{\delta})/2\},$$

where $S(\boldsymbol{\delta}) = \sum_i \delta_i^2$. Applying Bayes theorem, the posterior distribution $p(\boldsymbol{\beta} \mid \mathbf{y})$ for the parameters, given the measurements **y**, is such that

$$p(\boldsymbol{\beta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}) p(\boldsymbol{\delta}) p(\phi) p(\psi),$$

so that $p(\boldsymbol{\beta} \mid \mathbf{y})$ is proportional to

$$\phi^{(m_0+nL)/2-1} \exp\{-\phi \bar{F}(\alpha, \boldsymbol{\delta})/2\} \psi^{(L_0+L)/2-1} \exp\{-\psi \bar{S}(\boldsymbol{\delta})/2\},$$

where $\bar{F}(\alpha, \boldsymbol{\delta}) = m_0\sigma_0^2 + F(\alpha, \boldsymbol{\delta})$ and $\bar{S}(\boldsymbol{\delta}) = L_0\rho_0^2 + S(\boldsymbol{\delta})$.

Point estimates **b** of the parameters $\boldsymbol{\beta}$ can be found by maximizing the posterior distribution. The point estimate *r* of ρ will necessarily be feasible (i.e., positive) by virtue of the constraint imposed implicitly by the prior distribution for ψ. In fact, if *a*, **d**, *s* and *r* present the estimates of α, $\boldsymbol{\delta}$, σ and ρ determined from **b**, then the conditions for the optimality of **b** imply that

$$s^2 = \frac{m_0\sigma_0^2 + F(a, \mathbf{d})}{m_0 + nL - 2} \text{ and } r^2 = \frac{L_0\rho_0^2 + S(\mathbf{d})}{L_0 + L - 2}.$$

It is seen that both *s* and *r* are determined from weighted averages of a prior sum of squares (represented by σ_0^2 and ρ_0^2) and sums of squares relating to the fit of the model to the data (represented by *F* and *S*).

Given the point estimates **b,** the posterior distribution $p(\boldsymbol{\beta} \mid \mathbf{y})$ can be approximated by $\hat{p}(\boldsymbol{\beta} \mid \mathbf{b}, V)$, the pdf for the multivariate Gaussian distribution $N(\mathbf{b}, V)$ with mean **b** and variance matrix $V = H^{-1}$ where H is the Hessian matrix of second partial derivatives of

$$Q(\boldsymbol{\beta}) = -\log p(\boldsymbol{\beta} \mid \mathbf{y}) \approx p(\mathbf{b}) + \frac{1}{2}(\boldsymbol{\beta} - \mathbf{b})^{\mathrm{T}} H (\boldsymbol{\beta} - \mathbf{b}).$$

In other words, the Gaussian approximation $\hat{p}(\boldsymbol{\beta} \mid \mathbf{b}, V)$ is derived from a quadratic approximation to $\log p(\boldsymbol{\beta} \mid \mathbf{y})$ about its maximum **b**. The main reason for using the parameters $\log\phi$ and $\log\psi$ rather than ϕ and ψ in $\boldsymbol{\beta}$ is that their distributions are more likely to be approximately Gaussian.

3.2 Markov chain Monte Carlo simulation

While maximum likelihood estimation can be used to determine parameter estimates and associated uncertainties, it is based on a quadratic estimate of the posterior distribution and may not be adequate, especially for nonlinear models. An alternative approach is to use Markov chain Monte Carlo (MCMC) simulation methods to create a set of points $\{\mathbf{b}_t\}$ sampled from the posterior distribution $p(\boldsymbol{\beta} \mid \mathbf{y})$ and then base estimates, uncertainties and coverage intervals on information derived straightforwardly from $\{\mathbf{b}_t\}$. The Metropolis-Hastings MCMC algorithm [4] is in principle easy to implement. Suppose we wish to sample $\{\mathbf{b}_t\}$ from a distribution $p(\boldsymbol{\beta})$, $\boldsymbol{\beta} = (\beta_1 ..., \beta_k)^{\mathrm{T}}$. Given a draw $\mathbf{b}_{t-1}$, a proposed new draw $\mathbf{b}^*$ for the next member of the sequence is drawn at random from a *proposal distribution* $q(\boldsymbol{\beta} \mid \mathbf{b}_{t-1})$. Then $\mathbf{b}_t$ is set to $\mathbf{b}^*$ with acceptance probability

$$p_t = \min\left\{1, \frac{p(\mathbf{b}^*) q(\mathbf{b}_{t-1} \mid \mathbf{b}^*)}{p(\mathbf{b}_{t-1}) q(\mathbf{b}^* \mid \mathbf{b}_{t-1})}\right\}.$$

After a number of iterations that allow the Markov chain to converge, the sampled $\mathbf{b}_t$ are drawn from the target distribution. The number of iterations necessary to ensure convergence is difficult to predict and most implementations perform a number of repeat simulations to gauge if the chains have converged to the target distribution [4, section 11.6].

To implement the algorithm it is necessary to generate the random draw from $q(\boldsymbol{\beta} \mid \mathbf{b})$ and evaluate the acceptance probability. If $\hat{p}(\boldsymbol{\beta})$ is a distribution that approximates $p(\boldsymbol{\beta})$ then we can set $q(\boldsymbol{\beta} \mid \mathbf{b}_{t-1}) = \hat{p}(\boldsymbol{\beta})$ (so the draws $\mathbf{b}^*$ are independent of the current step), in which case

$$p_t = \min\left\{1, \frac{p(\mathbf{b}^*) \hat{p}(\mathbf{b}_{t-1})}{p(\mathbf{b}_{t-1}) \hat{p}(\mathbf{b}^*)}\right\}.$$

If $p(\boldsymbol{\beta}) = \hat{p}(\boldsymbol{\beta})$, then $p_t = 1$ and the proposed $\mathbf{b}^*$ is always accepted. For $p(\boldsymbol{\beta}) \neq \hat{p}(\boldsymbol{\beta})$, the role of p_t is to modify the draws from the proposal distribution so that they become draws from the target distribution. Note that $p(\boldsymbol{\beta})$, $q(\boldsymbol{\beta} \mid \mathbf{b})$ (or $\hat{p}(\boldsymbol{\beta})$) need only be known up to constant in order to able to evaluate the acceptance probability. For our application we can take for $\hat{p}(\boldsymbol{\beta})$ the Gaussian approximant $\hat{p}(\boldsymbol{\beta} \mid \mathbf{b}, V)$ determined as described above. The complete simulation scheme can be implemented as follows:

1) Minimise $Q(\boldsymbol{\beta}) = -\log p(\boldsymbol{\beta} \mid \mathbf{y})$ to determine estimate $\mathbf{b}$ and Hessian matrix H of second order partial derivatives evaluated at $\mathbf{b}$.
2) Calculate the Cholesky factorization $H = R^{\mathrm{T}}R$ (see, e.g., [5]) and set $B = R^{-1}$. Then the variance matrix for the Gaussian approximant is $V = H^{-1} = BB^{\mathrm{T}}$.
3) Draw $\mathbf{e}_0 \in N(0, I)$, so that $\mathbf{e}_0$ is a k-vector of independent, normally distributed random numbers and set $\mathbf{b}_0 = \mathbf{b} + B\mathbf{e}_0$, $Q_0 = Q(\mathbf{b}_0)$, and $\hat{Q}_0 = \mathbf{e}_0^{\mathrm{T}}\mathbf{e}_0 / 2$.
4) For t=1,2,…
   a. Draw $\mathbf{e}^* \in N(0, I)$ and set $\mathbf{b}^* = \mathbf{b} + B\mathbf{e}^*$, $Q^* = Q(\mathbf{b}^*)$ and $\hat{Q}^* = (\mathbf{e}^*)^{\mathrm{T}}\mathbf{e}^* / 2$.
   b. Evaluate the ratio $r_t = \dfrac{p(\mathbf{b}^* \mid \mathbf{y})\,\hat{p}(\mathbf{b}_{t-1} \mid \mathbf{b}, V)}{p(\mathbf{b}_{t-1} \mid \mathbf{y})\,\hat{p}(\mathbf{b}^* \mid \mathbf{b}, V)} = \exp\{Q_{t-1} - Q^* + \hat{Q}^* - \hat{Q}_{t-1}\}$, and draw $u \in R[0,1]$.
   c. If $u < r_t$, set $\mathbf{b}_t = \mathbf{b}^*$, $Q_t = Q^*$ and $\hat{Q}_t = \hat{Q}^*$. Otherwise set $\mathbf{b}_t = \mathbf{b}_{t-1}$, $Q_t = Q_{t-1}$ and $\hat{Q}_t = \hat{Q}_{t-1}$.

At steps 3) and 4b), $\mathbf{b}_0$ and $\mathbf{b}^*$ are draws from $N(\mathbf{b}, V)$. At steps 3) and 4a), $\hat{Q}_0 = -\log \hat{p}(\mathbf{b}_0 \mid \mathbf{b}, V)$ and $\hat{Q}^* = -\log \hat{p}(\mathbf{b}^* \mid \mathbf{b}, V)$, up to the same additive constant. At step 4c), the test on $u \in R[0,1]$ drawn from the rectangular distribution defined on the interval [0,1] ensures that $\mathbf{b}^*$ is accepted with probability $p_t = \min\{1, r_t\}$.

3.3 Example calculations

We have generated data according to the model in Eq. 2 with $n = L = 4$ and $\sigma = \rho = 1$. The data is shown in Fig. 6.

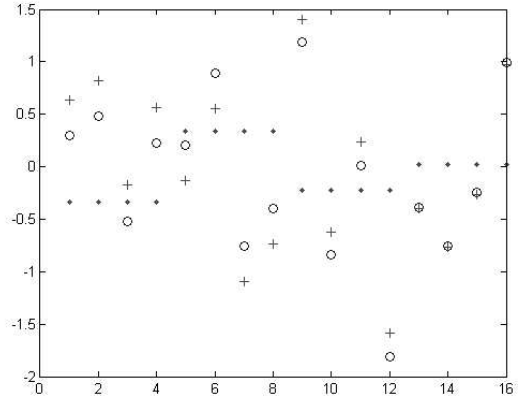

Fig. 6. Data y generated according to the model in Eq. 1: 'o' = $y_{i,q}$, '+' = $\varepsilon_{i,q}$ and '.' = δ_i with n=L=4.

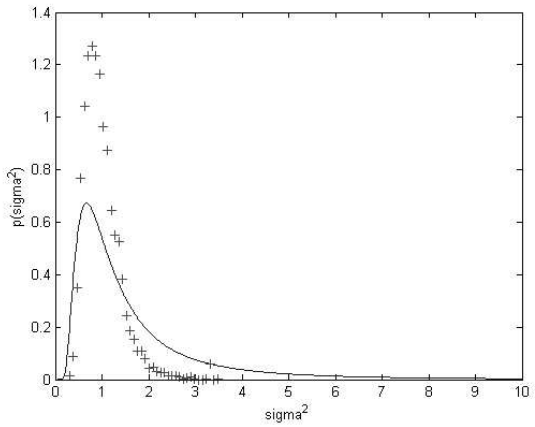

Fig. 7. Prior distribution (solid curve) and estimated posterior distribution '+' for σ^2 determined from MCMC simulations with n=L=4.

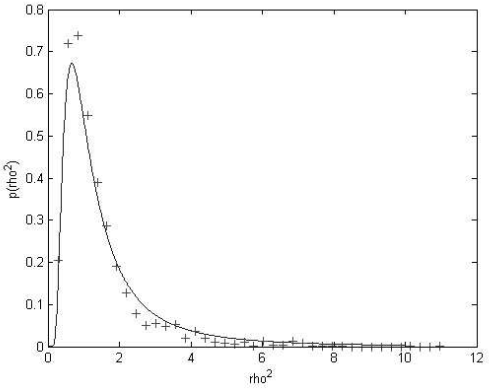

Fig. 8. Prior distribution (solid curve) and estimated posterior distribution for ρ^2 determined from MCMC simulations with n=L=4.

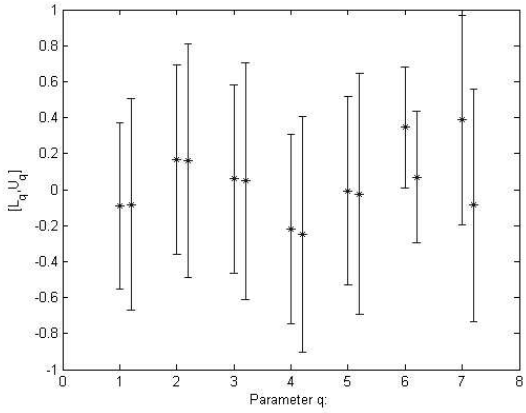

Fig. 9. Maximum likelihood estinates (lefthand intervals) and MCMC estimates (righthand intervals) for $\boldsymbol{\beta}$ for the case n=L=4.

By chance, the variation in the data $\mathbf{y}$ is less than the variation in the random effects so that the ANOVA estimate r^2 of ρ^2 is –0.14. Maximum likelihood estimates $\mathbf{b}$ of the seven parameters α, $\delta_i, i = 1,2,3,4$, $\log\phi$ and $\log\psi$ have been determined using a prior distibution for $\phi = 1/\sigma^2$ and $\psi = 1/\rho^2$ as in Eq. 3 with $m_0 = L_0 = 4$ and $\sigma_0^2 = \rho_0^2 = 1$. The corresponding prior distributions for σ^2 and ρ^2 are

shown in Figs. 7 and 8 (they are the same). In Fig. 9, the lefthand intervals are the 95 % coverage intervals for the seven parameters determined from a Gaussian approximation to the posterior distribution about its mode. The righthand intervals are those determined using the Metropolis-Hastings MCMC algorithm determined from 4 sets each of 8000 simulations from which the latter 4000 simulations were used to the provide estimates and 95 % coverage intervals. For the parameters α and $\boldsymbol{\delta}$, both MLE and MCMC estimates are similar, but the MCMC intervals are larger as the true posterior distribution for these parameters is similar to a student-*t* distribution and has heavier tails than a Gaussian. The two sets of intervals for $\log\phi$ and $\log\psi$ (parameters 6 and 7) are less similar. Fig. 7 gives the prior and posterior distribution for σ^2, the latter estimated from the MCMC samples, and shows that the data has provided new information about σ^2. By contrast, the prior and posterior distributions for ρ^2 are similar.

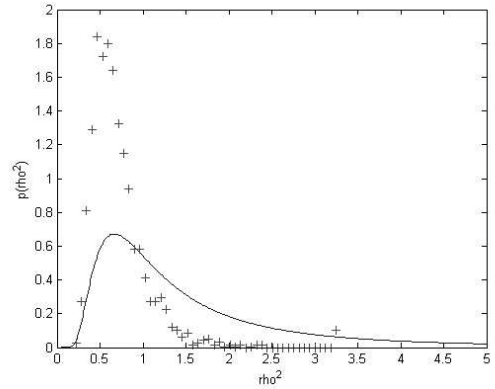

Fig. 10. Prior distribution (solid curve) and estimated posterior distribution for ρ^2 determined from MCMC simulations with *n*=*L*=16**.**

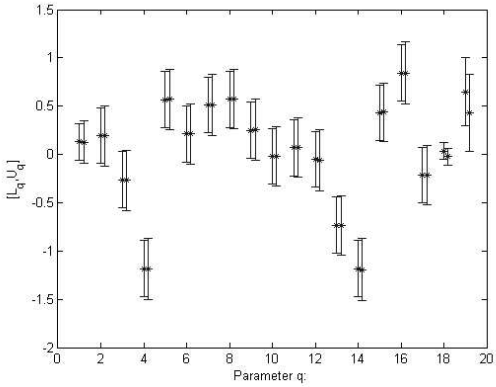

Fig. 11. Maximum likelihood estinates (lefthand intervals) and MCMC estimates (righthand intervals) for $\boldsymbol{\beta}$ for the case *n*=*L*=16.

We have repeated the calculations for a second data set, this time generated with *n*=*L*=16. In this case, the posterior distribution for ρ^2 reflects the significant contribution of the measurement data; Fig. 10. The MLE and MCMC estimates shown in Fig. 11 show good agreement as should be the case: as the number of data points increases, the posterior distribution approaches a multivariate Gaussian.

4. CONCLUDING REMARKS

Much if not all of the GUM and ISO 5725 rely on computational methods that can be implemented by hand (or at least on a calculator) supported by statistical tables. Relatively recent developments such as Markov chain Monte Carlo simulation along with the vast computing power available on desk top PCs, permit much more computationally intensive approaches to be applied to uncertainty evaluation and measurement systems analysis. In particular, it is possible to remove assumptions about normality and employ hierarchical models to model more comprehensively what is known about the measurement system. In ANOVA, as classically implemented in paper or electronic spreadsheets, for small sample experiments negative estimates of variance components can arise. Using statistical software it is possible to solve this problem as well as introducing multiple effects in the model, as long as enough data is available. Using Bayesian approaches it is possible to impose natural constraints in the model so that these events cannot arise.

To improve good practice in measurement uncertainty evaluation and the analysis of interlaboratory comparisons (ILCs) it is necessary that concepts from both the GUM and ISO 5725 are used appropriately. The role of ILCs as validation tools and in providing new information needs to be developed and managed. Newer computational methods allow for more comprehensive models to be analysed. There is a risk that current models are unnecessarily simple due to perceived limitations in what is computationally feasible.

ACKNOWLEDGMENTS

This work was partially supported by the UK's Department of Trade and Industry's Software Support for Metrology Programme. We thank Dr P M Harris, NPL, for a careful reading of an earlier version of this paper.

REFERENCES

[1] ISO The Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization, Geneva, 1995.

[2] ISO 5725 Accuracy (Trueness and Precision) of Measurement Methods and Results, Parts 1 to 6, International Organization for Standardization, Geneva, 1995.

[3] ISO/TS 21748 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, International Organization for Standardization, Geneva, 2004.

[4] Bayesian Data Analysis, A Gelman, J B Carlin, H S Stern and D B Rubin, Chapman Hall/CRC, Boca Raton 2004.

[5] Matrix Computations, G H Golub and C Van Loan, Johns Hopkins University Press, 3rd edn, 1996.