On use of Bayesian statistics to make the *Guide to the Expression of Uncertainty in Measurement* consistent

2003 Metrologia 40 235

(http://iopscience.iop.org/0026-1394/40/5/305)

INSTITUTE OF PHYSICS PUBLISHING METROLOGIA
Metrologia **40** (2003) 235–248 PII: S0026-1394(03)66815-5

On use of Bayesian statistics to make the *Guide to the Expression of Uncertainty in Measurement* consistent

R Kacker and A Jones

National Institute of Standards and Technology, Gaithersburg, MD 20899-8910, USA

E-mail: raghu.kacker@nist.gov

Received 9 May 2003
Published 2 September 2003
Online at stacks.iop.org/Met/40/235

Abstract

The International Organization for Standardization (ISO) *Guide to the Expression of Uncertainty in Measurement* is being increasingly recognized as a *de facto* international standard. The ISO *Guide* recommends a standardized way of expressing uncertainty in all kinds of measurements and provides a comprehensive approach for combining information to evaluate that uncertainty. The ISO *Guide* supports uncertainties evaluated from statistical methods, Type A, and uncertainties determined by other means, Type B. The ISO *Guide* recommends classical (frequentist) statistics for evaluating the Type A components of uncertainty; but it interprets the combined uncertainty from a Bayesian viewpoint. This is inconsistent. In order to overcome this inconsistency, we suggest that all Type A uncertainties should be evaluated through a Bayesian approach. It turns out that the estimates from a classical statistical analysis are either equal or approximately equal to the corresponding estimates from a Bayesian analysis with non-informative prior probability distributions. So the classical (frequentist) estimates may be used provided they are interpreted from the Bayesian viewpoint. The procedure of the ISO *Guide* for evaluating the combined uncertainty is to propagate the uncertainties associated with the input quantities. This procedure does not yield a complete specification of the distribution represented by the result of measurement and its associated combined standard uncertainty. So the correct coverage factor for a desired coverage probability of an expanded uncertainty interval cannot always be determined. Nonetheless, the ISO *Guide* suggests that the coverage factor may be computed by assuming that the distribution represented by the result of measurement and its associated standard uncertainty is a normal distribution or a scaled-and-shifted t-distribution with degrees of freedom determined from the Welch–Satterthwaite formula. This assumption may be unjustified and the coverage factor so determined may be incorrect. A popular convention is to set the coverage factor as 2. When the distribution represented by the result of measurement and its associated standard uncertainty is not completely determined, the 2-standard-uncertainty interval may be interpreted in terms of its minimum coverage probability for an applicable class of probability distributions.

1. Introduction

The *Guide to the Expression of Uncertainty in Measurement* is published by the International Organization for Standardization (ISO) in the names of seven international scientific organizations [1][1]. It is based on specific recommendations of the International Committee for Weights and Measures (CIPM). The ISO *Guide* has been adopted by most of the national metrology institutes (NMIs) and regional metrology

[1] The following seven organizations supported the development of this Guide, which is published in their names: International Bureau of Weights and Measures (BIPM), International Electro-technical Commission (IEC), International Federation of Clinical Chemistry (IFCC), International Organization for Standardization (ISO), International Union of Pure and Applied Chemistry (IUPAC), International Union of Pure and Applied Physics (IUPAP), and International Organization of Legal Metrology (OIML).

organizations (RMOs) in the world. The ISO *Guide* is the basis of the Eurachem/CITAC *Guide on Quantifying Uncertainty in Analytical Measurement* [2]. The American National Standards Institute (ANSI) and the National Conference of Standards Laboratories International (NCSLI) have adopted the ISO *Guide* as the US *Guide* [3][2]. The ISO *Guide* is being increasingly recognized in industrial laboratories.

1.1. Background of the ISO Guide

Before publication of the ISO *Guide*, many different expressions of uncertainty were in vogue. Sometimes this led to difficulties in interpretation of published expressions of uncertainty, in comparing similar results whose uncertainties were expressed differently, and in using published expressions of uncertainty as inputs for other uncertainty evaluations. The ISO *Guide* recommends that all uncertainties must be expressed as *standard uncertainties* or as *expanded uncertainties*. These terms are based on the following concepts introduced by the ISO *Guide* (section 2 and annex B). A *measurand* is a particular quantity subject to measurement. A *result of measurement* is a value attributed to the measurand. *Uncertainty* is a parameter associated with the result of measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand. *Standard uncertainty* is uncertainty expressed as a standard deviation. *Expanded uncertainty* is a stated multiple of the standard uncertainty that defines an interval about the result of measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand (the ISO *Guide*, sections 2.3.5 and 3.3.7).

The ISO *Guide* is based on the concept of a measurement equation. A *measurement equation* is a mathematical description of the process that is used for determining the result of measurement and its associated standard uncertainty from various input quantities[3]. The result of measurement is determined by substituting the estimates of input quantities in the measurement equation. The uncertainties associated with the input estimates are components of uncertainty in determining the result of measurement. The *combined standard uncertainty* associated with the result of measurement is determined by propagating the components of uncertainty through a linear approximation of the measurement equation about the input estimates.

Some components represent uncertainty arising from random effects and others represent uncertainty arising from systematic (non-random) effects. With advancements in measurement science and in instruments of measurement, the uncertainty arising from random effects in physical measurements continues to decrease. Consequently, the importance of uncertainty arising from systematic effects continues to increase. Before publication of the ISO *Guide*, there was no generally accepted approach to account for the uncertainty arising from systematic effects. The ISO *Guide* (section 3.2) recommends the following approach. Correct each result of measurement for all recognized systematic effects and make every effort to identify such effects. Quantify the uncertainties associated with the corrections applied for systematic effects. Include the uncertainties associated with the corrections in the combined standard uncertainty associated with the corrected result of measurement (NIST Technical Note 1297 [4], sections D.2 and D.3). Some or all of the corrections may be set to zero as expression of ignorance. However, the uncertainties associated with the corrections may be significant. The ISO *Guide* treats the uncertainty components from random effects and from the corrections applied for systematic effects in exactly the same way.

The ISO *Guide* recognizes that some components of uncertainty may be evaluated from statistical methods and some must be determined by other means, meaning scientific judgment. The components of uncertainty evaluated from statistical analysis of current measurements are referred to as Type A, and the components of uncertainty determined by scientific judgement based on other measurements or published data are referred to as Type B. Generally, some components are Type A and some Type B. So Type A uncertainties need to be combined with Type B uncertainties to determine the combined uncertainty associated with the result of measurement. Before publication of the ISO *Guide*, this had been a sticky issue for many metrologists. The ISO *Guide* resolved the issue by declaring that no distinction should be made between Type A and Type B components of uncertainty.

The uncertainties associated with the corrections for systematic effects are often determined by scientific judgement, so they are Type B evaluations. However, there is not always a simple correspondence between the classification of uncertainties as Type A and Type B and the classification of uncertainties as arising from random and systematic effects (the ISO *Guide*, section 3.3.3, and NIST Technical Note 1297 [4], section 2.3). This point is widely misunderstood.

1.2. Review of the ISO Guide

The ISO *Guide* summarizes the steps for evaluating the combined uncertainty in its section 8. Here is a brief review of these steps.

Step 1. Express mathematically the relationship between the measurand Y and the input quantities X_i on which Y depends: $Y = f(X_1, \ldots, X_n)$. This function is the *measurement equation*.

Step 2. Determine x_i, the estimated value of input quantity X_i, either on the basis of the statistical analysis of a series of measurements or by other means.

Step 3. Evaluate the standard uncertainty $u(x_i)$ of each input estimate x_i from the statistical analysis of a series of measurements or by other means.

Step 4. Evaluate the correlation coefficients associated with any pairs of input estimates that are correlated.

Step 5. Calculate the result of measurement y for Y from the measurement equation using for the input quantities X_i the estimates x_i. Thus, $y = f(x_1, \ldots, x_n)$.

Step 6. Determine the combined standard uncertainty $u(y)$ of the measurement result y from the standard uncertainties and correlation coefficients associated with the input estimates using the law of propagation of uncertainties:

[2] The US Guide is available from NCSLI, 1800, 30th Street, Suite 305, Boulder, CO 80301, USA (http://www.ncslinternational.org).

[3] Reference on Constants, Units, and Uncertainty, Physics Laboratory, National Institute of Standards and Technology, 2000 (http://physics.nist.gov/cuu/Uncertainty/basic.html).

$u^2(y) = \sum_i c_i^2 u^2(x_i) + 2\sum_{(i<j)} c_i c_j u(x_i) u(x_j) r(x_i, x_j)$, where $c_1, \ldots, c_n$ are partial derivatives of Y with respect to $X_1, \ldots, X_n$ evaluated at $x_1, \ldots, x_n$, respectively, and $r(x_i, x_j)$ is the *correlation coefficient* between X_i and X_j for $i, j = 1, \ldots, n$ and $i \neq j$. The partial derivatives $c_1, \ldots, c_n$ are *sensitivity coefficients*.

Step 7. If it is necessary to give an expanded uncertainty U, whose purpose is to provide an interval $[y \pm U]$ that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand Y, multiply the combined standard uncertainty $u(y)$ by a coverage factor k to obtain $U = ku(y)$. Select k on the basis of level of confidence (coverage probability) required of the interval. According to Taylor [5], the interval $[y \pm U] = [y \pm ku(y)]$ should be referred to as an *expanded uncertainty interval*. However, some metrologists prefer the term *coverage interval*.

Step 8. Report the result of measurement y together with its combined standard uncertainty $u(y)$ or expanded uncertainty U. Describe how y, $u(y)$, and U were obtained.

The ISO *Guide* prescribes a procedure for evaluating the result of measurement y for the value Y of the measurand and its associated combined standard uncertainty $u(y)$ from the estimated values $x_1, \ldots, x_n$ of the input quantities $X_1, \ldots, X_n$ and their associated standard uncertainties $u(x_1), \ldots, u(x_n)$ when the measurement equation $Y = f(X_1, \ldots, X_n)$ is specified. This procedure does not use any information that might be available on the probability distributions for the input quantities $X_1, \ldots, X_n$ and it does not yield a complete specification of the probability distribution represented by y and $u(y)$. In this sense, the ISO *Guide* propagates uncertainties rather than probability distributions.

The ISO *Guide* (section 6.2.2) defines the coverage probability (level of confidence) of an expanded uncertainty interval $[y \pm U] = [y \pm ku(y)]$ as follows:

> More specifically, U is interpreted as defining an interval about the measurement result that encompasses a large fraction p of the probability distribution characterized by that result and its combined standard uncertainty, and p is the *coverage probability* or *level of confidence* of the interval.

The word 'characterize' has a very specific and different meaning in statistics than the way it is used in the ISO *Guide*. So in this paper (and elsewhere), we use the word 'represent' instead. According to the ISO *Guide* (section 2.3), the result of measurement y and its associated combined standard uncertainty $u(y)$ represent a probability distribution for the values that could reasonably be attributed to the value Y of the measurand. The ISO *Guide* is vague about the relationship between the pair y and $u(y)$ and the statistical parameters of the probability distribution they represent. We interpret that the distribution represented by y and $u(y)$ is a distribution that has expected value y and standard deviation $u(y)$.

The ISO *Guide*'s definition of coverage probability requires that the value Y of the measurand be treated as a random variable with a specified probability distribution. The probability distributions, such as normal (Gaussian) distribution, Student's t-distribution, and rectangular distribution, are described by a probability density function (pdf). The coverage probability is defined with respect to a pdf for Y that has the expected value $E(Y) = y$ and standard deviation $S(Y) = u(y)$. The coverage probability p of the interval $[y \pm ku(y)]$ is the probability $\Pr[y - ku(y) \leqslant Y \leqslant y + ku(y)]$ with respect to the specified pdf for Y, where y, $u(y)$, and k are constants. Alternatively, $\Pr[-k \leqslant (Y - y)/u(y) \leqslant k] = p$. In the vernacular of the ISO *Guide*, the coverage probability p is the fraction of the pdf represented by y and $u(y)$ that is encompassed by the interval $[y \pm ku(y)]$. When the pdf for Y is a normal distribution, the coverage probability p of the interval $[y \pm 2u(y)]$ is about 95% (the ISO *Guide*, table G.2). The coverage probability of the interval $[y \pm 2u(y)]$ with respect to another pdf for Y may be less or more than 95%.

The ISO *Guide* (sections 2.3.5 and 6.3.2) recognizes that the coverage factor k such that the interval $[y \pm ku(y)]$ has a desired coverage probability p (or the coverage probability p of the interval $[y \pm ku(y)]$ for a given coverage factor k) cannot be determined unless the distribution represented by y and $u(y)$ and its relevant probabilities and percentiles are known. Nonetheless, the ISO *Guide* (section 6.3.3 and annex G) suggests that the coverage factor k such that the interval $[y \pm ku(y)]$ has a desired coverage probability p may be determined by assuming that the pdf represented by y and $u(y)$ is approximately normal or a scaled-and-shifted t-distribution with degrees of freedom determined from the Welch–Satterthwaite formula (the ISO *Guide*, equation (G.2*b*)).

1.3. Propagation of probability distributions by numerical simulation

An alternative to propagating uncertainties is propagating probability distributions by numerical simulation of the measurement equation $Y = f(X_1, \ldots, X_n)$. When the probability distributions for all input variables $X_1, \ldots, X_n$ are fully specified and the measurement equation $Y = f(X_1, \ldots, X_n)$ is computable, a distribution for Y can be simulated. The simulated distribution for Y can then be used to determine the result y, uncertainty $u(y)$, and an interval $[y \pm ku(y)]$ with a desired coverage probability p. There is growing interest among metrologists in propagating distributions by numerical simulation as an alternative to propagating uncertainties.

1.4. Relationship between the ISO Guide and the classical and Bayesian statistics

In classical statistics, the value of the measurand is assumed to be an unknown constant, often called the true value, and the measurement data are random variables each with a sampling probability distribution. Assume that the conditions of measurement are somehow fixed at the intended levels. Assume that the process of making measurements is repeated infinitely many times in the fixed conditions producing independent measurements. Assume that the dispersion of data from repeat measurements arises from random effects occurring in the fixed measurement conditions. Now think of the relative frequency of realizing data in the neighbourhood of a given measurement. A *sampling distribution* is a probability distribution that describes the relative frequencies of occurrence for all possible measurements in the fixed measurement conditions. A sampling distribution has

an expected value and a standard deviation among other parameters. A classical statistical analysis starts with an assumed sampling distribution that applies to each measurement in a set. The measurements are generally assumed to be independent. The metrologist relates the expected value of the sampling distribution of measurement data to the value of the measurand. The standard deviation of the sampling distribution quantifies the dispersion of all possible measurements in the fixed conditions. An estimate for the value of the measurand is determined from a statistical theory. The assumed sampling distribution for the measurement data induces a sampling distribution for the estimated value of the measurand. The primary outputs of a classical statistical analysis are (1) an estimate for the value of the measurand and (2) an estimate of the standard deviation of the sampling distribution for the estimated value of the measurand. The difference between the expected value of the sampling distribution for the estimated value of the measurand and the true value of the measurand is systematic error (bias). The classic (frequentist) estimates are based on one occurrence in measurement conditions assumed to be fixed at the intended levels. The classical estimates do not apply to other conditions of measurement. In classical statistics, one cannot use scientific judgement based on other measurements and published data.

In Bayesian statistics, the measurement data are constants and the value of the measurand is a random variable. The probability distribution for the value of the measurand is a *state of knowledge distribution* that describes the degrees of belief about all possible values that could be attributed to the value of the measurand. The degrees of belief are based on all available information including scientific judgement, current measurements, and ancillary knowledge. Similar state of knowledge probability distributions apply to the other unknown quantities involved in measurement. When we say that a variable represents state of knowledge we mean that its probability distribution is a state of knowledge distribution that describes belief about all possible values of the variable based on available information. A Bayesian analysis starts with prior distributions, which represent the states of knowledge before measurements are made, for the values of unknown quantities. Negligible prior knowledge is expressed by using non-informative probability distributions. The measurements are then used to update the prior distributions using Bayes' theorem [6, 7] to obtain posterior distributions. The posterior distribution for the value of the measurand is a probability distribution that could reasonably be attributed to the value of the measurand after measurements are made. A measure of centrality (such as the expected value) and a measure of dispersion (such as the standard deviation) of the Bayesian posterior distribution quantify, respectively, the result of measurement and its associated standard uncertainty. Thus, Bayesian analysis is a statistical method for updating the state of knowledge. In Bayesian statistics, one uses both measurement data and scientific judgement. As additional information becomes available, Bayesian estimates are updated via Bayes' theorem. A more thorough discussion of the classical and Bayesian statistics from the viewpoint of quantifying measurement uncertainty is given in references [8–10].

It turns out that the estimates from a classical statistical analysis are either equal or approximately equal to the corresponding estimates from a Bayesian analysis with the same sampling distributions and non-informative prior (proper or improper) probability distributions [6, 7]. Therefore, classical (frequentist) estimates may be used in a Bayesian framework provided they are interpreted from the Bayesian viewpoint.

The ISO *Guide* does not follow completely either the classical or the Bayesian statistics. For example, the ISO *Guide* (section 4.2 and annex G) recommends the use of classical (frequentist) statistics for evaluating the result x_i and its associated standard uncertainty $u(x_i)$. However, the ISO *Guide*'s (section 6.2.2) definition of coverage probability corresponds to the Bayesian view of uncertainty. The ISO *Guide* is very clear that the interval $[y \pm ku(y)]$ is not to be interpreted as a confidence interval and the coverage probability (level of confidence) is not the confidence level of classical statistics. Some metrologists, however, believe that when all components of y and $u(y)$ are Type A and they are evaluated from classical statistics, the ISO *Guide* permits the interval $[y \pm ku(y)]$ to be interpreted as a confidence interval of classical statistics.

The ISO *Guide* (section 4.1.6) gives its interpretation of the Type A evaluations from classical statistics and Type B evaluations from scientific judgement as follows:

> Each input estimate x_i and its associated standard uncertainty $u(x_i)$ are obtained from a distribution of possible values of the input quantity X_i. This probability distribution may be frequency based, that is, based on a series of observations $X_{i,k}$ of X_i, or it may be an *a priori* distribution. Type A evaluations of standard uncertainty components are founded on frequency distributions while Type B evaluations are founded on *a priori* distributions. It must be recognized that in both cases the distributions are models that are used to represent the state of our knowledge.

Thus, the ISO *Guide* interprets the Type A evaluations from classical (frequentist) statistics as expressions of the state of knowledge and Type B distributions as prior probability distributions. Hence, the ISO *Guide* treats all input and output quantities of the measurement equation $Y = f(X_1, \ldots, X_n)$ as variables representing states of knowledge. The ISO *Guide* does not justify its interpretation.

The main point of this paper is that the ISO *Guide*'s interpretation can be justified if the Type A estimates from classical statistical analyses are interpreted as approximations to the corresponding estimates from Bayesian analyses with non-informative prior distributions[4] and the Type B distributions are interpreted as Bayesian prior distributions. This interpretation would make the ISO *Guide* internally consistent and consistent with Bayesian statistics.

1.5. Problems with the ISO Guide and proposed resolutions

Many researchers are uneasy with the ISO *Guide*. Statistician Gleser [11] writes, 'The ISO recommendation has been of

[4] In some cases, the classical (frequentist) estimates may be equal to the corresponding Bayesian estimates for certain prior distributions.

concern to many statisticians because it appears to combine frequentist performance measures and indices of subjective distributions in a way that neither frequentists nor Bayesians can fully endorse'. Physicists Weise and Wöger [8] from the Physikalisch-Technische Bundesanstalt (PTB), the national metrology institute of Germany, do not support the use of classical (frequentist) statistics for quantifying uncertainty in measurement.

In this paper, we address the following problems with the ISO *Guide*.

1. The probabilistic interpretation of the Type A component of the combined standard uncertainty evaluated from classical statistics is not consistent with the ISO *Guide*'s definition of coverage probability.
2. The measurement equation and the procedure of the ISO *Guide* for evaluating the combined uncertainty are not consistent when the Type A component of the combined standard uncertainty is evaluated from classical statistics.
3. The ISO *Guide*'s suggestion that the coverage factor for a desired coverage probability may be determined from a normal distribution or a scaled-and-shifted t-distribution with degrees of freedom determined from the Welch–Satterthwaite formula may be unjustified and the coverage factor so determined may be incorrect.

We suggest that the first two problems would disappear if all Type A components are determined through Bayesian statistics or interpreted as approximations of Bayesian estimates. The third problem stems from the fact that the ISO *Guide*'s method of propagating uncertainties does not yield a complete specification of the probability distribution represented by the result of measurement and its associated standard uncertainty. Therefore, the correct coverage factor k for a given coverage probability p cannot always be determined (the ISO *Guide*, section 6.3.2). A popular convention is to set $k = 2$. When the probability distribution represented by the result y and uncertainty $u(y)$ is not completely determined, the 2-standard-uncertainty interval $[y \pm 2u(y)]$ may be interpreted in terms of its *minimum coverage probability*. The minimum coverage probability of the interval $[y \pm ku(y)]$ for a given k is the lower bound of coverage probabilities for a class of probability distributions each of which has the expected value y and standard deviation $u(y)$.

1.6. Outline of this paper

In section 2, we illustrate that when the Type A component of the combined standard uncertainty $u(y)$ is determined from classical statistics, the probabilistic interpretation of the interval $[y \pm ku(y)]$ is not consistent with the ISO *Guide*'s definition of coverage probability. Then, we show that when Bayesian statistics is used, the interval $[y \pm ku(y)]$ has well-defined coverage probability. In section 3, we illustrate that the measurement equation and the procedure of the ISO *Guide* for evaluating the combined standard uncertainty $u(y)$ are consistent when Bayesian statistics is used for Type A evaluations but not when classical statistics is used. Then we propose a revision of the ISO *Guide*'s steps for evaluating the combined standard uncertainty that would make it consistent. In section 4, we illustrate that the coverage factor k such that the interval $[y \pm ku(y)]$ has a desired coverage probability p may be incorrect when annex G of the ISO *Guide* is used. When the distribution represented by the result y and uncertainty $u(y)$ is not completely determined, a popular convention is to select the coverage factor as 2 ([4], appendix C). We will describe the minimum coverage probability of the interval $[y \pm 2u(y)]$ for the following two classes of probability distributions: those that have expected value y and standard deviation $u(y)$ and those that are symmetric, unimodal, and have expected value and mode y and standard deviation $u(y)$. In section 5, we address practical issues in using the ISO *Guide*. In section 6, we address practical issues in justifying the statistical assumptions about measurement data. A summary appears in section 7.

2. Proposal to make the Type A uncertainties consistent with the ISO *Guide*'s definition of coverage probability

In this section, we show that the probabilistic interpretation of the interval $[y \pm ku(y)]$ determined from classical statistics is not consistent with the ISO *Guide*'s definition of coverage probability. However, the coverage probability of the interval $[y \pm ku(y)]$ is well defined when Bayesian statistics is used.

2.1. Uncertainty evaluated from classical statistics

Let $z_1, \ldots, z_m$ be a series of measurements that are subject to both random and systematic effects for evaluating the unknown value $\mathcal{Y}$ of the measurand. In classical statistics, $\mathcal{Y}$ is an unknown constant and the uncertainty is about the measurements $z_1, \ldots, z_m$. We use the script symbols $\mathcal{Y}$ and $\mathcal{Z}$ for the unknown constant parameters of classical statistics. We assume the following about random effects in $z_1, \ldots, z_m$:

Assumption 1. The measurements $z_1, \ldots, z_m$ are independent random variables having the same normal distribution with expected value $\mathcal{Z}$ and standard deviation σ. Here, both $\mathcal{Z}$ and σ are unknown constants and σ represents the dispersion of all possible measurements in the given conditions. The dispersion of measurements is presumed to arise from random effects. The common probability distribution for $z_1, \ldots, z_m$ is a sampling distribution.

Let z_A be the arithmetic mean of $z_1, \ldots, z_m$, i.e. $z_\mathrm{A} = (1/m)\sum_i z_i$. Let $s(z)$ be the experimental standard deviation of $z_1, \ldots, z_m$, i.e.

$$s(z) = \sqrt{\frac{\sum_i (z_i - z_\mathrm{A})^2}{m-1}}.$$

Let $s(z_\mathrm{A}) = s(z)/\sqrt{m}$ be the experimental standard deviation of the mean z_A. Then, z_A is the best statistical (Type A) estimate of $\mathcal{Z}$ in the sense that the expected value $E(z_\mathrm{A})$ is equal to $\mathcal{Z}$ and the standard deviation $S(z_\mathrm{A})$, which is equal to $\sigma/\sqrt{m}$, is smallest among all estimates whose expected value is $\mathcal{Z}$. Assumption 1 implies that $s(z)$ is representative of the standard deviation σ. According to the ISO *Guide* (section 4.2), the Type A standard uncertainty associated with z_A from classical statistics is $u(z_\mathrm{A}) = s(z_\mathrm{A}) = s(z)/\sqrt{m}$. The uncertainty $u(z_\mathrm{A}) = s(z_\mathrm{A}) = s(z)/\sqrt{m}$ is an estimate of the standard deviation $S(z_\mathrm{A}) = \sigma/\sqrt{m}$ of the sampling distribution for z_A. When the number m of measurements $z_1, \ldots, z_m$ is small, the

classical uncertainty $s(z_A)$ is uncertain. The uncertainty in $s(z_A)$ arising from small m is a statistical uncertainty. The statistical uncertainty in $s(z_A)$ is accounted for by the degrees of freedom (the ISO *Guide*, section E.4.3). The degrees of freedom associated with $s(z_A)$ are $(m-1)$.

The error in z_A with respect to $\mathcal{Y}$ is $(z_A - \mathcal{Y}) = (z_A - \mathcal{Z}) + (\mathcal{Z} - \mathcal{Y})$, where $(z_A - \mathcal{Z})$ is random error and $(\mathcal{Z} - \mathcal{Y})$ is systematic error (bias). The uncertainty $u(z_A) = s(z_A) = s(z)/\sqrt{m}$ accounts for the uncertainty arising from random effects. In classical statistics, the bias $(\mathcal{Z} - \mathcal{Y})$ is a constant; therefore, one cannot deal with the uncertainty arising from bias. In the framework of the ISO *Guide* (section 3.2), a correction is required to counter the bias when it may not be negligible. A typical assumption about the required correction is as follows:

Assumption 2. The correction c and its associated standard uncertainty $u(c)$ are determined from a rectangular distribution on the interval $(-a, a)$ specified by scientific judgement (Type B). The correction c is identified with the expected value and the uncertainty $u(c)$ is identified with the standard deviation of the rectangular distribution on the interval $(-a, a)$. Thus, $c = 0$ and $u(c) = a/\sqrt{3}$ (the ISO *Guide*, section 4.3.7).

The ISO *Guide* (section G.4.2) introduced the concept of degrees of freedom associated with Type B uncertainties and proposed an approach to specify them. The degrees of freedom associated with $u(c) = a/\sqrt{3}$ are often assumed to be infinity, which means that there is no doubt that the rectangular distribution on the interval $(-a, a)$ is sufficiently wide to account for possible systematic error in z_A (the ISO *Guide*, section G.4.3). The claim of infinite degrees of freedom for the standard uncertainty evaluated from such rectangular distribution is a sticky issue for some metrologists.

According to the ISO *Guide* (section 5), the result of measurement y for the unknown value $\mathcal{Y}$ of the measurand is $y = z_A + c = z_A + 0 = z_A$ and its associated standard uncertainty is

$$u(y) = \sqrt{u^2(z_A) + u^2(c)} = \sqrt{s^2(z_A) + \frac{a^2}{3}}.$$

Following the ISO *Guide* (annex G), the effective degrees of freedom ν_{eff} associated with $u(y) = \sqrt{s^2(z_A) + a^2/3}$ are determined from the Welch–Satterthwaite formula (the ISO *Guide*, equation (G.2*b*)). Thus,

$$\nu_{\text{eff}} = \frac{u^4(y)}{s^4(z_A)/(m-1)}.$$

According to the ISO *Guide*, an expanded uncertainty interval for $\mathcal{Y}$ with coverage probability p is

$$[y \pm ku(y)] \equiv [y \pm t_{(1/2)(1+p)}(\nu_{\text{eff}})u(y)] \equiv \left[z_A \pm t_{(1/2)(1+p)}(\nu_{\text{eff}})\sqrt{s^2(z_A) + \frac{a^2}{3}}\right],$$

where the coverage factor $k = t_{(1/2)(1+p)}(\nu_{\text{eff}})$ is the $\frac{1}{2}(1+p) \times 100$th percentile of the t-distribution with degrees of freedom ν_{eff}.

It follows from assumption 1 that the sampling distribution of $(z_A - \mathcal{Z})/s(z_A)$ is the t-distribution with degrees of freedom $(m-1)$ [12]. Thus, $[z_A \pm t_{(1/2)(1+p)}(m-1)s(z_A)]$ is a confidence interval for $\mathcal{Z}$ with confidence level p, where $t_{(1/2)(1+p)}(m-1)$ is the $\frac{1}{2}(1+p) \times 100$th percentile of the t-distribution with degrees of freedom $(m-1)$. The confidence level p is not a statement about the computed interval. It is a statement about the statistical procedure. Suppose the process of making the measurements and computing confidence intervals could be repeated infinitely many times under exactly the same experimental conditions and the same sampling distributions continued to apply, then p specifies the fraction of such intervals that would include $\mathcal{Z}$ [12].

In classical (frequentist) statistics, z_A and $s(z_A)$ are random variables with sampling distributions. Since $c = 0$ and $u(c) = a/\sqrt{3}$ are constants, the result $y = z_A + c = z_A + 0 = z_A$ and uncertainty

$$u(y) = \sqrt{u^2(z_A) + u^2(c)} = \sqrt{s^2(z_A) + \frac{a^2}{3}}$$

are random variables. However, in the ISO *Guide*'s definition of coverage probability, the result y and uncertainty $u(y)$ are constants that represent a probability distribution for the values that could reasonably be attributed to the value of the measurand. Therefore, the probabilistic interpretation of the interval

$$[y \pm ku(y)] \equiv [y \pm t_{(1/2)(1+p)}(\nu_{\text{eff}})u(y)] \equiv \left[z_A \pm t_{(1/2)(1+p)}(\nu_{\text{eff}})\sqrt{s^2(z_A) + \frac{a^2}{3}}\right]$$

determined from classical statistics is not consistent with the ISO *Guide*'s definition of coverage probability.

Note 1. The choice of rectangular distribution for the correction may be justified by the *principle of maximum entropy*. Metrologists can sometimes specify the minimum and the maximum possible correction for systematic error. When only the minimum and maximum of a random variable are known and nothing else, the maximum entropy distribution is rectangular distribution. When the expected value and standard deviation of a distribution are known and nothing else, the maximum entropy distribution is normal distribution. When the expected value, standard deviation, minimum, and maximum of a distribution are known and nothing else, the maximum entropy distribution is a truncated normal distribution [13].

2.2. Uncertainty evaluated from Bayesian statistics

In Bayesian statistics, the measurements $z_1, \ldots, z_m$ and hence the statistics z_A and $s(z)$ are constants. The uncertainty is in the state of knowledge about the value of the measurand. We will use the symbols Y and Z for the random variables that represent the states of knowledge about the value $\mathcal{Y}$ of the measurand and the expected value $\mathcal{Z}$ of the common sampling distribution for $z_1, \ldots, z_m$, respectively. We use the symbol σ for both the unknown constant representing the standard deviation of the common sampling distribution for $z_1, \ldots, z_m$ and the random variable that represents the state of knowledge about the standard deviation. The states of knowledge about $\mathcal{Z}$ and σ before measurements are made are expressed by *prior distributions* of Z and σ. Negligible prior knowledge

about $\mathcal{Z}$ and σ is expressed by using *non-informative prior distributions*. The measurements $z_1, \ldots, z_m$ are then used to update the prior distributions to obtain *posterior distributions* of Z and σ. Assumption 1 concerning the common sampling distribution for $z_1, \ldots, z_m$ provides the *likelihood function* of Z and σ conditional on the given measurements $z_1, \ldots, z_m$. The mechanism for updating is Bayes' theorem. *Bayes' theorem states that the posterior distribution is proportional to the product of the likelihood function and the prior distribution* [6, 7]. When the prior distributions of Z and σ are non-informative then under assumption 1 the posterior distribution of the ratio $(Z - z_A)/s(z_A)$ is the t-distribution with degrees of freedom $(m-1)$ ([7], theorem 2.4.1). Here, Z is a variable and z_A and $s(z_A)$ are constants. In determining the distribution of $(Z-z_A)/s(z_A)$, the posterior distribution of σ is integrated out. Thus, the Bayesian probability distribution of Z is a scaled-and-shifted t-distribution with degrees of freedom $(m-1)$ that has been scaled by $s(z_A)$ and shifted by z_A.

The expected value and standard deviation of the t-distribution with degrees of freedom ν are zero and $\sqrt{\nu/(\nu-2)}$, respectively [14]. So the expected value and standard deviation of Z are

$$E(Z) = z_A \qquad \text{and} \qquad S(Z) = \sqrt{\frac{m-1}{m-3}} \times s(z_A),$$

respectively. Hence, the Bayesian evaluation of Type A uncertainty arising from random effects in z_A is

$$u_{\text{Bayes}}(z_A) = S(Z) = \sqrt{\frac{m-1}{m-3}} \times s(z_A).$$

As the number m of mutually independent and normally distributed measurements increases, the t-distribution tends to normal distribution. The uncertainty $u(z_A) = s(z_A)$ from classical statistics may be interpreted as an approximation to the Bayesian uncertainty

$$u_{\text{Bayes}}(z_A) = \sqrt{\frac{m-1}{m-3}} \times s(z_A).$$

The approximation is poor when m is small but improves as m increases.

In order to account for the uncertainty arising from possible bias in z_A, a measurement equation is needed. A simple measurement equation with additive correction C is $Y = Z + C$ [10]. The bias in z_A is the constant $(\mathcal{Z} - \mathcal{Y})$. The correction C for the bias in z_A is a variable. In the 'negative of bias' $(Y - Z)$, both Y and Z are treated as variables representing states of knowledge. In order to specify a probability distribution for C, the variable Z is replaced with its expected value z_A. Thus, the probability distribution for C represents belief about the possible values of $(Y - z_A)$, where z_A is a constant and Y is the variable. The belief about possible values of Y is based on all available information including measurements and scientific judgement. The correction C is assumed to be distributed independently of Z. According to assumption 2, the probability distribution for C is rectangular on the interval $(-a, a)$ with expected value $E(C) = c = 0$ and standard deviation $S(C) = u(c) = a/\sqrt{3}$. Thus, the Type B component of uncertainty associated with the correction c for possible bias in z_A is $u(c) = a/\sqrt{3}$.

According to the ISO *Guide*, the result of measurement based on the measurement equation $Y = Z + C$ is $y = z_A + c = z_A + 0 = z_A$, where $z_A = E(Z)$ and $c = E(C) = 0$. The combined standard uncertainty associated with y is $u_{\text{Bayes}}(y) = \sqrt{S^2(Z) + S^2(C)}$, which is equal to $\sqrt{(m-1)/(m-3) \times s^2(z_A) + a^2/3}$. Since the measurement equation is linear, the distribution represented by y and $u_{\text{Bayes}}(y)$ is the probability distribution of Y. Since the ratio $(Z - z_A)/s(z_A)$ has the t-distribution with degrees of freedom $(m-1)$ and C has rectangular distribution on the interval $(-a, a)$, the distribution of $Y = Z + C$ is identical to that of the random variable $z_A + s(z_A)T + aR$, where T has the t-distribution with degrees of freedom $(m-1)$ and R has rectangular distribution on the interval $(-1, 1)$. Thus, the coverage probability of the interval $[y \pm ku_{\text{Bayes}}(y)]$ is the fraction p of the distribution for the random variable $z_A + s(z_A)T + aR$ that is encompassed by $[y \pm ku_{\text{Bayes}}(y)]$. This example illustrates that when the measurement equation for Y is a linear function of the input quantities, the distribution represented by the result y and uncertainty $u_{\text{Bayes}}(y)$ is the probability distribution of Y. So the coverage probability of the interval $[y \pm ku_{\text{Bayes}}(y)]$ is well defined with respect to the distribution of Y.

Now suppose that the measurement equation $Y = f(X_1, \ldots, X_n)$ is non-linear. The result y and uncertainty $u_{\text{Bayes}}(y)$ are determined from a linear approximation of the measurement equation (see section 4.2 of this paper). According to the ISO *Guide*, the result y and uncertainty $u_{\text{Bayes}}(y)$ represent a probability distribution for the values that could reasonably be attributed to Y. The coverage probability is well defined with respect to the distribution represented by y and $u_{\text{Bayes}}(y)$.

Note 1. A Bayesian uncertainty itself has no statistical uncertainty. In classical statistics, the standard uncertainty associated with the arithmetic mean z_A of independent measurements $z_1, \ldots, z_m$ having the same normal sampling distribution is $u(z_A) = s(z_A)$. The classical (frequentist) uncertainty $s(z_A)$ is uncertain when m is small. The statistical uncertainty in $s(z_A)$ is accounted for by degrees of freedom. The corresponding Bayesian uncertainty is

$$u_{\text{Bayes}}(z_A) = \sqrt{\frac{m-1}{m-3}} \times s(z_A).$$

The factor $\sqrt{(m-1)/(m-3)}$ built into the Bayesian uncertainty accounts for the statistical uncertainty that arises from a small number of measurements. So the Bayesian uncertainty $u_{\text{Bayes}}(z_A)$ has no statistical uncertainty. However, a classical uncertainty as well as a Bayesian uncertainty may have non-statistical uncertainty. The sources of non-statistical uncertainty include inadequate measurement equation, unjustified statistical assumptions (Type A), and unreasonable state of knowledge distributions (Type B). The non-statistical uncertainty may be larger than the statistical uncertainty in uncertainty. Therefore, evaluation of measurement uncertainty involves more than statistics.

Note 2. An expression for the standard uncertainty associated with the arithmetic mean z_A is required when the corresponding variable Z may be used as an input

for evaluating another quantity. The Bayesian standard uncertainty

$$u_{\text{Bayes}}(z_{\text{A}}) = \sqrt{\frac{m-1}{m-3}} \times s(z_{\text{A}})$$

requires at least $m = 4$ independent measurements. Sometimes it is not practical to realize more than two or three independent measurements. What options do we have for quantifying the standard uncertainty associated with z_{A} when $m = 2$ or $m = 3$? The first option is to investigate reasonable prior distributions that might yield a useful expression for $u_{\text{Bayes}}(z_{\text{A}})$. The second option is to determine and use a Type B evaluation of the standard uncertainty associated with z_{A}. These options are mathematically sound; however, they may not be practical in certain applications. So, here, we propose *ad hoc* expressions for the standard uncertainties associated with z_{A} when $m = 2$ and 3. The factor $\sqrt{(m-1)/(m-3)}$ built into the Bayesian uncertainty $u_{\text{Bayes}}(z_{\text{A}})$ accounts for the statistical uncertainty that arises when the number of independent measurements is small. The factor $\sqrt{(m-1)/(m-3)}$ is approximately equal to the ratio of the percentile for coverage probability p of the t-distribution with degrees of freedom $(m-1)$ to the corresponding percentile of the normal distribution (the ISO *Guide*, section G.3.4). The conventional value of the coverage factor is 2, which corresponds to p of about 95% for normal distribution. For $p = 95\%$, the percentiles of the t-distribution for $m = 2$ and $m = 3$ are 12.71 and 4.30, respectively (the ISO *Guide*, table G.2). The corresponding percentile of the normal distribution is 1.96. The ratios $12.71/1.96 = 6.48$ and $4.30/1.96 = 2.19$ may be taken as factors that account for the statistical uncertainty in the uncertainty that arises from a small number of measurements when $m = 2$ and $m = 3$, respectively. Therefore, one may use $u^*(z_{\text{A}}) = 6.48 \times s(z_{\text{A}})$ and $u^*(z_{\text{A}}) = 2.19 \times s(z_{\text{A}})$ as *ad hoc* standard uncertainties associated with z_{A} when $m = 2$ and $m = 3$, respectively. Some metrologists do not support these *ad hoc* expressions because an expression for standard uncertainty should not depend on probabilistic considerations associated with uncertainty intervals.

Note 3. According to the ISO *Guide* (sections 3.3.7 and 8, step 7), an expanded uncertainty interval $[y \pm ku(y)]$ is determined *after* the expected value $E(Y) = y$ and standard deviation $S(Y) = u(y)$ have been evaluated. An interval of the type $[y \pm ku(y)]$ can sometimes be constructed with respect to a probability distribution for Y where y is the expected value of Y but $u(y)$ is *not* the standard deviation of Y. For example, the ratio $(Z - z_{\text{A}})/s(z_{\text{A}})$ has the t-distribution with degrees of freedom $(m-1)$. So the interval $[z_{\text{A}} \pm t_{(1/2)(1+p)}(\nu) \times s(z_{\text{A}})]$, where $\nu = (m-1)$ has coverage probability p but $s(z_{\text{A}})$ is not the standard deviation of Z. It is not clear that this interval agrees with the ISO *Guide*. However, the interval $[z_{\text{A}} \pm k\sqrt{(m-1)/(m-3)} \times s(z_{\text{A}})]$ fully agrees with the ISO *Guide*.

Note 4. The Bayesian result that the ratio $(Z - z_{\text{A}})/s(z_{\text{A}})$ has the t-distribution with degrees of freedom $(m-1)$ is of importance when m is small. This result requires that z_{A} and $s(z_{\text{A}}) = s(z)/\sqrt{m}$ have independent sampling distributions ([7], theorem 2.4.1). It turns out that the normal distribution is the only distribution for which z_{A} and $s(z)$ are independently distributed [15]. The ISO *Guide* (section G.5.4) acknowledges the need of normal distribution. The claim of t-distribution for $(Z - z_{\text{A}})/s(z_{\text{A}})$ may, therefore, be sensitive to some kinds of departures from the assumed normal distribution for the measurements $z_1, \ldots, z_m$.

Note 5. The sampling distribution of z_{A} is sometimes claimed to be approximately normal with standard deviation $\sigma/\sqrt{m}$ by the central limit theorem (CLT). Here, σ is the standard deviation of the common sampling distribution for $z_1, \ldots, z_m$. The original CLT is as follows. Let $z_1, z_2, \ldots$ be a sequence of identically distributed random variables with expected value μ and standard deviation σ (both finite), any finite number of which are independent. Let $z_{\text{A}} = (1/m)\sum_i z_i$, then the distribution of the ratio $(z_{\text{A}} - \mu)/(\sigma/\sqrt{m})$ tends to normal distribution with expected value zero and standard deviation one as m tends to infinity [16]. The CLT is also true if the z_i have different distributions, μ then being replaced by the mean of individual μs and $\sigma/\sqrt{m}$ being replaced by the standard deviation of z_{A} [16]. However, the number m of summands required for convergence may be very large. The critical assumptions in the CLT are that the measurements $z_1, \ldots, z_m$ must be independently distributed, their standard deviations must be known and of similar magnitude, and the number m of summands must be sufficiently large for the given standard deviations. When the CLT is invoked, these assumptions must be justified.

3. Proposed revision of the ISO *Guide* to make it consistent

In section 3.1, we illustrate that the measurement equation and the procedure of the ISO *Guide* for evaluating the combined uncertainty are consistent only when Bayesian statistics is used for the Type A evaluations. In section 3.2, we propose a revision of the eight steps (the ISO *Guide*, section 8) for evaluating the combined uncertainty that would make the ISO *Guide* consistent.

3.1. Measurement equation

The concept of a measurement equation is, perhaps, the most fundamental technical contribution of the ISO *Guide*. In its simplest form, a measurement equation is expressed as a mathematical function $Y = f(X_1, \ldots, X_n)$ that represents the process that is used to determine the result of measurement and its associated standard uncertainty. All input and output quantities are regarded as random variables representing states of knowledge (the ISO *Guide*, section 4.1.6). Some of the input variables $X_1, X_2, \ldots, X_n$ may themselves be viewed as measurands and functions of additional input variables [4]. For example, we may have $X_1 = f_1(W_{11}, W_{12}, \ldots, W_{1K})$ and $X_2 = f_2(W_{21}, W_{22}, \ldots, W_{2L})$, etc. In such cases, the measurement equation is actually a system of equations. Such a system could be used to link a hierarchy of measurements, partition a complex measurement problem into smaller more manageable components, or to combine a number of input variables. Sometimes the function f may be determined experimentally or exist only as an algorithm that must be evaluated numerically (the ISO *Guide*, section 4.1.2).

A measurement equation includes influence quantities that take into account sources of variability, such as different observers, instruments, samples, laboratories, and the times at which measurements are made as well as corrections for systematic effects. The measurement equation should include all influence quantities that could contribute a significant component to the uncertainty in assigning a value to the measurand [4]. The measurement equation $Y = f(X_1, \ldots, X_n)$ should be exhaustively complete in the sense that one of the input variables should represent a correction for possible bias in the rest of the function.

3.1.1. Type A evaluations from Bayesian statistics. The measurement equation $Y = f(X_1, \ldots, X_n)$ is well defined from the following Bayesian viewpoint. The Type A evaluations are parameters of Bayesian posterior distributions and Type B evaluations are parameters of Bayesian prior distributions. They have the same probabilistic interpretations. In both cases the expected values $x_1, \ldots, x_n$ and standard deviations $u(x_1), \ldots, u(x_n)$ of $X_1, \ldots, X_n$ are constants. The pair x_i and $u(x_i)$ represents a probability distribution for the values that could reasonably be attributed to the input quantity X_i for $i = 1, 2, \ldots, n$. When the measurement equation $Y = f(X_1, \ldots, X_n)$ is non-linear, it is approximated by a first-order (linear) Taylor series for evaluating the combined standard uncertainty $u(y)$. The first-order Taylor series approximation is

$$Y = f(X_1, \ldots, X_n) \approx f(x_1, \ldots, x_n) + \sum_i c_i(X_i - x_i)$$
$$= y + \sum_i c_i(X_i - x_i),$$

where $c_1, \ldots, c_n$ are partial derivatives of Y with respect to $X_1, \ldots, X_n$ evaluated at $x_1, \ldots, x_n$. The partial derivatives $c_1, \ldots, c_n$ are sensitivity coefficients[5]. The linear approximation provides the law of propagation of uncertainties (section 1.2, step 6). Thus, the standard uncertainty $u(y)$ can be determined from the standard deviations and pairwise correlation coefficients of $X_1, \ldots, X_n$. The result y and uncertainty $u(y)$ determined in this way are constants that represent a probability distribution for the values that could reasonably be attributed to Y. Thus, the measurement equation $Y = f(X_1, \ldots, X_n)$ and the procedure of the ISO *Guide* for evaluating the result y and uncertainty $u(y)$ are well defined and consistent from the Bayesian viewpoint.

In the example of section 2, the measurement equation is $Y = Z + C$, where Z is a random variable with Bayesian posterior distribution, which is a scaled-and-shifted t-distribution with degrees of freedom $(m - 1)$ that has been scaled by $s(z_A)$ and shifted by z_A. The correction C is a variable representing the state of knowledge interpreted as a Bayesian prior distribution (the ISO *Guide*, section 4.1.6), which is assumed to be rectangular on the interval $(-a, a)$. Thus, the measurement equation $Y = Z + C$ is well defined and the distribution of Y is identical to that of the random variable $z_A + s(z_A)T + aR$, where T has the t-distribution with degrees of freedom $(m - 1)$ and R has rectangular distribution on the interval $(-1, 1)$ (section 2.2). Following the procedure of the ISO *Guide*, the result of the measurement is $y = z_A + c = z_A + 0 = z_A$ and its associated standard uncertainty is

$$u(y) = \sqrt{S^2(Z) + S^2(C)} = \sqrt{\frac{m-1}{m-3} \times s^2(z_A) + \frac{a^2}{3}}.$$

Since the measurement equation $Y = Z + C$ is linear, the result y is equal to $E(Y)$ and uncertainty $u(y)$ is equal to $S(Y)$, and the distribution represented by y and $u(y)$ is the probability distribution of Y.

3.1.2. Type A evaluations from classical statistics. Let us return to the example discussed above and previously in section 2. Now suppose that the Type A component of the combined standard uncertainty is evaluated from classical statistics. Consider the equation $Y = z_A + C$, where z_A is a random variable having sampling distribution (Type A) with expected value $E(z_A) = \mathcal{Z}$ and standard deviation $S(z_A) = \sigma/\sqrt{m}$. The distribution of z_A is not a state of knowledge distribution because $\mathcal{Z}$ and $\sigma/\sqrt{m}$ are unknown constants. The sum $Y = z_A + C$ is a random variable with unknown expected value $E(Y) = \mathcal{Z} + c = \mathcal{Z} + 0 = \mathcal{Z}$ and unknown standard deviation $S(Y) = \sqrt{\sigma^2/m + a^2/3}$. Therefore, the probability distribution of Y is not a state of knowledge distribution. In order to estimate $E(Y)$ and $S(Y)$, one would substitute the estimate z_A for $\mathcal{Z}$ and the estimate $s(z_A) = s(z)/\sqrt{m}$ for $\sigma/\sqrt{m}$. Then, z_A would be an estimate of $E(Y)$ and $\sqrt{s^2(z_A) + a^2/3}$ would be an estimate of $S(Y)$. The equation $Y = z_A + C$ and this procedure are not consistent with the ISO *Guide*. Also, the uncertainty $\sqrt{s^2(z_A) + a^2/3}$ determined from classical statistics does not agree with the uncertainty $\sqrt{(m-1)/(m-3) \times s^2(z_A) + a^2/3}$ determined from Bayesian statistics.

3.2. Proposed revision of the steps for evaluating the combined uncertainty

We suggest that the following revision of the eight steps (the ISO *Guide*, section 8) for evaluating the combined uncertainty would make the ISO *Guide* consistent:

Step 1. Describe the measurement process as a measurement equation $Y = f(X_1, \ldots, X_n)$, where the inputs $X_1, \ldots, X_n$ may have their own measurement equations. All input and output quantities are variables with probability distributions representing states of knowledge. The Type A state of knowledge distributions are interpreted as Bayesian posterior distributions and the Type B state of knowledge distributions are interpreted as Bayesian prior distributions.

Step 2. Determine the results $x_1, \ldots, x_n$ for the input variables $X_1, \ldots, X_n$, respectively. All Type A results are interpreted as expected values of Bayesian posterior distributions or their approximations. All Type B results are interpreted as expected values of Bayesian prior distributions.

Step 3. Determine the standard uncertainties $u(x_1), \ldots, u(x_n)$ for the input variables $X_1, \ldots, X_n$, respectively. All Type A standard uncertainties are interpreted as standard deviations of Bayesian posterior distributions or their

[5] When the function f is determined experimentally or exists only as an algorithm, the sensitivity coefficients may be determined experimentally: one measures the change in Y produced by a change in a particular X_i while holding the remaining input quantities constant.

approximations. All Type B standard uncertainties are interpreted as standard deviations of Bayesian prior distributions.

Step 4. Determine the correlation coefficients $r(x_i, x_j)$ for all pairs X_i and X_j of the input variables that may be significantly correlated, for $i, j = 1, \ldots, n$, $i < j$. These evaluations are Type A or Type B depending on whether they are determined from statistical analysis of current measurements or by scientific judgement based on other measurements or published data.

Step 5. Determine the combined result of measurement y for the value Y of the measurand by substituting the results $x_1, \ldots, x_n$ for the input variables $X_1, \ldots, X_n$ in the measurement equation $Y = f(X_1, \ldots, X_n)$. Thus, $y = f(x_1, \ldots, x_n)$.

Step 6. Determine the combined standard uncertainty $u(y)$ associated with the measurement result y by propagating the standard uncertainties $u(x_1), \ldots, u(x_n)$ and correlation coefficients $r(x_1, x_2), \ldots, r(x_{(n-1)}, x_n)$. A first-order (linear) Taylor series approximation of the measurement equation $Y = f(X_1, \ldots, X_n)$ provides the following law of propagation of uncertainties:

$$u^2(y) = \sum_i c_i^2 u^2(x_i) + 2 \sum_{(i<j)} c_i c_j u(x_i) u(x_j) r(x_i, x_j),$$

where $c_1, \ldots, c_n$ are partial derivatives of Y with respect to $X_1, \ldots, X_n$ evaluated at $x_1, \ldots, x_n$, respectively. The product $u(x_i)u(x_j)r(x_i, x_j)$ is equal to the covariance $u(x_i, x_j)$ for $i, j = 1, \ldots, n$ and $i \neq j$.

Step 7. If it is necessary to express the uncertainty as an interval, multiply the combined standard uncertainty $u(y)$ by a coverage factor k to obtain the interval $[y \pm ku(y)]$. Generally, set the coverage factor k as 2. Use values of k other than 2 in special cases only.

Step 8. Report the result of measurement y together with its associated combined standard uncertainty $u(y)$ or expanded uncertainty interval $[y \pm ku(y)]$. Describe how y and $u(y)$ were obtained. Describe how k was chosen when k is not 2.

4. The coverage factor determined from annex G of the ISO *Guide* may be incorrect

Since the procedure of the *Guide* does not yield a complete specification of the distribution represented by y and $u(y)$, the coverage factor k such that the interval $[y \pm ku(y)]$ has a desired coverage probability p, such as 95%, cannot always be determined. We will illustrate that the coverage factor k determined from a normal distribution or a scaled-and-shifted t-distribution with degrees of freedom determined by the Welch–Satterthwaite formula, as suggested by the ISO *Guide* (annex G), may be incorrect.

A convention that has gained popularity is to set the coverage factor k as 2. The NIST Policy ([4], appendix C) on the coverage factor is as follows:

> Use expanded uncertainty U to report the results of all NIST measurements other than those for which u_c has traditionally been employed. To be consistent with current international practice, the value of k to be used at NIST for calculating U is, by convention, $k = 2$. Values of k other than 2 are only to be used for specific applications dictated by established and documented requirements.

When the distribution represented by y and $u(y)$ is not completely determined, the 2-standard-uncertainty interval $[y \pm 2u(y)]$ may be interpreted in terms of its minimum coverage probability. We will describe the minimum coverage probability of the interval $[y \pm 2u(y)]$ for two classes of probability distributions represented by y and $u(y)$.

4.1. Linear measurement equation

Suppose the measurement equation $Y = f(X_1, \ldots, X_n)$ is a linear function $Y = \sum_i a_i X_i$, where $a_1, \ldots, a_n$ are constants and not all pairs of the input variables $X_1, \ldots, X_n$ may be independently distributed. If each of the variables $X_1, \ldots, X_n$ has a normal distribution, then Y has a normal distribution. It turns out that the variable Y has a normal distribution *if and only if* each of the input variables $X_1, \ldots, X_n$ has a normal distribution ([15], theorem 1.1). The input variables that are corrections for systematic effects are often assumed to have rectangular or other non-normal distributions. Thus, not all of the variables $X_1, \ldots, X_n$ may be normally distributed and hence the distribution of Y need not be normal. Consequently, the coverage factor k determined from normal distribution may be incorrect.

The distribution of $Y = \sum_i a_i X_i$ is sometimes claimed to be approximately normal by the CLT. The CLT applies only when $a_1 X_1, \ldots, a_n X_n$ (or $X_1, \ldots, X_n$) are all mutually independent, the standard deviations of $a_1 X_1, \ldots, a_n X_n$ are of similar magnitude, and n is sufficiently large (the ISO *Guide*, section G.2). The CLT does not apply in the absence of mutual independence. Also, the CLT does not apply when the number n of summands is not sufficiently large for the given standard deviations of $a_1 X_1, \ldots, a_n X_n$.

The Welch–Satterthwaite formula was developed in a classical (frequentist) setting. However, the ISO *Guide* (annex G) uses it from the Bayesian viewpoint. The Welch–Satterthwaite formula does not apply when $Y = \sum_i a_i X_i$, where $i = 1, \ldots, n$, and some pairs of the variables $X_1, \ldots, X_n$ are correlated.

Let us discuss the case where $Y = \sum_i a_i X_i$ and $X_1, \ldots, X_n$ are all mutually independent. Suppose $X_1, \ldots, X_r$, for $r < n$, are evaluated from statistical methods (Type A) and $X_{r+1}, \ldots, X_n$ are evaluated by scientific judgement (Type B). Suppose X_i is evaluated from a series of m_i measurements that are mutually independent and have the same normal distribution, where $i = 1, \ldots, r$ for $r < n$. Suppose the mean and standard deviation of the m_i measurements are x_i and $s(x_i)$ with degrees of freedom $(m_i - 1)$, respectively. Then, with non-informative prior distributions, the variable X_i has a scaled-and-shifted t-distribution with degrees of freedom $(m_i - 1)$ that has been scaled by $s(x_i)$ and shifted by x_i (section 2.2). Thus,

$$E(X_i) = x_i \qquad \text{and} \qquad S(X_i) = \sqrt{\frac{m_i - 1}{m_i - 3}} \times s(x_i).$$

Suppose X_i, for $i = r + 1, \ldots, n$, is evaluated by scientific judgement (Type B) with expected value x_i and standard deviation $u(x_i)$ with infinite degrees of freedom each. Then,

$$E(Y) = \sum_i a_i x_i$$

and

$$S(Y) = \sqrt{\sum_i a_i^2 \frac{m_i - 1}{m_i - 3} s^2(x_i) + \sum_i a_i^2 u^2(x_i)},$$

where the first sum is for $i = 1, \ldots, r$ and the second sum is for $i = r + 1, \ldots, n$. According to the ISO *Guide* (section 6.3.3 and annex G), the distribution of $Y = \sum_i a_i X_i$ can be approximated by a scaled-and-shifted t-distribution with degrees of freedom ν_{eff} that has been scaled by $\sqrt{\sum_i a_i^2 s^2(x_i) + \sum_i a_i^2 u^2(x_i)}$ and shifted by $\sum_i a_i x_i$, where ν_{eff} is determined from the Welch–Satterthwaite formula (the ISO *Guide*, equation (G.2*b*)). Thus,

$$\nu_{\text{eff}} = \frac{\left\{\sqrt{\sum_i a_i^2 s^2(x_i) + \sum_i a_i^2 u^2(x_i)}\right\}^4}{\sum_i [|a_i| s(x_i)]^4/(m_i - 1)}.$$

The standard deviation of this distribution is $\sqrt{\nu_{\text{eff}}/(\nu_{\text{eff}} - 2)} \times \sqrt{\sum_i a_i^2 s^2(x_i) + \sum_i a_i^2 u^2(x_i)}$. This may differ from the standard deviation

$$S(Y) = \sqrt{\sum_i a_i^2 \frac{m_i - 1}{m_i - 3} s^2(x_i) + \sum_i a_i^2 u^2(x_i)}.$$

Thus, the tail-percentiles of the scaled-and-shifted t-distribution with degrees of freedom determined from the Welch–Satterthwaite formula may differ from those of the distribution of $Y = \sum_i a_i X_i$. Hence, the coverage factor k determined from the Welch–Satterthwaite formula may be incorrect.

4.2. Non-linear measurement equation

Suppose the measurement equation $Y = f(X_1, \ldots, X_n)$ is non-linear. The ISO *Guide*'s rule for determining the result y is to substitute the results $x_1, \ldots, x_n$ for the variables $X_1, \ldots, X_n$ in the measurement equation $Y = f(X_1, \ldots, X_n)$. That is, $y = f(x_1, \ldots, x_n)$. The ISO *Guide*'s rule for determining the uncertainty $u(y)$ is to propagate the uncertainties $u(x_1), \ldots, u(x_n)$. That is,

$$u^2(y) = \sum_i c_i^2 u^2(x_i) + 2 \sum_{(i<j)} c_i c_j u(x_i) u(x_j) r(x_i, x_j),$$

where $r(x_i, x_j)$ is the correlation coefficient between X_i and X_j for $i, j = 1, \ldots, n$ and $i \neq j$. The result y and uncertainty $u(y)$ are, respectively, the expected value and standard deviation of a linear approximation Y_{linear} of Y, where

$$Y_{\text{linear}} = f(x_1, \ldots, x_n) + \sum_i c_i (X_i - x_i) = y + \sum_i c_i (X_i - x_i)$$

and $c_1, \ldots, c_n$ are, respectively, the partial derivatives of Y with respect to $X_1, \ldots, X_n$ evaluated at $x_1, \ldots, x_n$. The coverage probability of the interval $[y \pm ku(y)]$ is defined with respect to the distribution of Y_{linear}. The coverage factor k determined from annex G of the ISO *Guide* may be incorrect for the reasons discussed in section 4.1.

4.3. Minimum coverage probability of the 2-standard-uncertainty interval

According to the *Bienayme–Chebyshev inequality* [16], the coverage probability of the interval $[y \pm ku(y)]$ is at least $1 - 1/k^2$ for any distribution that has expected value y and standard deviation $u(y)$. Thus, the minimum coverage probability of the 2-standard-uncertainty interval $[y \pm 2u(y)]$ is 75% for any distribution that has expected value y and standard deviation $u(y)$.

According to the *Gauss inequality* [16], when a distribution with expected value y and standard deviation $u(y)$ is symmetric-unimodal with mode at y, the coverage probability of the interval $[y \pm ku(y)]$ is at least $1 - 4/(9k^2)$. Thus, the minimum coverage probability of the interval $[y \pm 2u(y)]$ is 89% for any symmetric-unimodal distribution that has expected value and mode at y and standard deviation $u(y)$.

Thus, the 2-standard-uncertainty interval $[y \pm 2u(y)]$ encompasses a large fraction of the probability distribution represented by y and $u(y)$. Since the procedure of the ISO *Guide* does not yield a complete specification of the distribution represented by y and $u(y)$, the concept of minimum coverage probability of the interval $[y \pm 2u(y)]$ fits with the ISO *Guide*.

5. Practical issues in using the ISO *Guide*

In this section, we discuss some of the practical issues in using the ISO *Guide*.

5.1. Propagation of uncertainties when the measurement equation is non-linear

In some applications, the measurement equation is of the type $Y = X^a Z^b$. Suppose $E(X) = x$, $E(Z) = z$, and $y = x^a z^b$. A linear Taylor series approximation of the measurement equation is

$$\frac{Y - y}{y} \approx \frac{a}{x}(X - x) + \frac{b}{z}(Z - z).$$

Suppose $S(Y)$, $S(X)$, and $S(Z)$ are, respectively, the standard deviations of Y, X, and Z and $C(X, Z)$ is the covariance between X and Z. The ratios $S(Y)/y$, $S(X)/x$, and $S(Z)/z$ are called the relative standard uncertainties denoted by $u_r(y)$, $u_r(x)$, and $u_r(z)$, respectively, and the ratio $C(X, Z)/(x \times z)$ is called the relative covariance denoted by $u_r(x, z)$. From the linear approximation, we have

$$u_r^2(y) \approx a^2 u_r^2(x) + b^2 u_r^2(z) + 2ab u_r(x, z).$$

A measurement equation of the type

$$Y = \frac{X_1 + X_2}{X_3 + X_4}$$

can be simplified as $Z_1 = X_1 + X_2$, $Z_2 = X_3 + X_4$, and $Y = Z_1/Z_2$. A measurement equation of the type $Y = X_1/X_2 + X_3/X_4$ can be simplified as $Z_1 = X_1/X_2$, $Z_2 = X_3/X_4$, and $Y = Z_1 + Z_2$. Then, the simplified equations can be used to propagate the uncertainties.

5.2. Evaluation of the correlation coefficient between two measurands

When some of the input quantities are common in evaluating the results and uncertainties for two measurands of values Y_1

and Y_2, their correlation coefficient $r(y_1, y_2)$ must also be evaluated. Suppose y_1 and $u(y_1)$ are the result and uncertainty for Y_1 and suppose y_2 and $u(y_2)$ are the result and uncertainty for Y_2. Suppose $Y_1 = f(X_1, \ldots, X_n)$, $Y_2 = g(X_1, \ldots, X_n)$, where not all of $X_1, \ldots, X_n$ may contribute to Y_1 and not all of $X_1, \ldots, X_n$ may contribute to Y_2 but some contribute to both Y_1 and Y_2. Let $c_1, \ldots, c_n$ be the partial derivatives of Y_1 and $d_1, \ldots, d_n$ be the partial derivatives of Y_2 with respect to $X_1, \ldots, X_n$ evaluated at their expected values $x_1, \ldots, x_n$. Then,

$$Y_1 \approx f(x_1, \ldots, x_n) + \sum_i c_i(X_i - x_i) = y_1 + \sum_i c_i(X_i - x_i)$$

and

$$Y_2 \approx g(x_1, \ldots, x_n) + \sum_i d_i(X_i - x_i) = y_2 + \sum_i d_i(X_i - x_i).$$

Denoting the covariance between Y_1 and Y_2 by $u(y_1, y_2)$, we have $u(y_1, y_2) \approx \sum_i \sum_j c_i d_j u(x_i, x_j)$, where $u(x_i, x_j)$ is the covariance between X_i and X_j for $i, j = 1, \ldots, n$ and $i \neq j$. Then, the correlation coefficient $r(y_1, y_2)$ is approximately equal to $u(y_1, y_2)/[u(y_1)u(y_2)]$. When more than two measurands are involved it is easier to do the calculations in matrix algebra.

5.3. Requirements of the ISO Guide's formula for a second-order approximation

The ISO *Guide* (section 5.1.2) gives a formula for a second-order approximation of the measurement equation $Y = f(X_1, \ldots, X_n)$. However, it does not state the required conditions. The formula requires three conditions: (1) the input variables $X_1, \ldots, X_n$ are independently distributed; (2) the distributions of $X_1, \ldots, X_n$ are symmetric about the corresponding expected values $x_1, \ldots, x_n$; (3) the *coefficients of kurtosis* of $X_1, \ldots, X_n$ are all equal to 3. The coefficient of kurtosis is the ratio of the fourth central moment to the fourth power of standard deviation $E[(X_i - E(X_i))^4]/[\sqrt{E[(X_i - E(X_i))^2]}]^4$ [14]. These conditions are satisfied when $X_1, \ldots, X_n$ are independent and normally distributed. When these conditions are not satisfied, the formula given in the ISO *Guide* may be a crude approximation.

5.4. Coverage probability, level of confidence, and confidence level

The ISO *Guide* uses the term level of confidence as a synonym for coverage probability. We strongly believe that the term level of confidence should not be used for the coverage probability. This nomenclature has no precedence in statistical literature and it is easy to gloss over the difference between level of confidence and confidence level[6]. The ISO *Guide*'s definition of coverage probability p is a statement about the computed interval $[y \pm ku(y)]$. The confidence level of classical (frequentist) statistics is not a statement about the computed interval $[y \pm ku(y)]$ (section 2.1).

[6] Lira and Wöger [11] note that in Spanish and Portuguese the terms 'level of confidence' and 'confidence level' are synonyms. Therefore, it was difficult to translate the ISO *Guide*, which is an international document.

5.5. Requirement of 95% coverage probability

Many metrologists believe that the ISO *Guide* requires that the coverage factor k should be determined such that the coverage probability of the interval $[y \pm ku(y)]$ is 95% (the ISO *Guide*, section 8, step 7). The ISO *Guide*'s method of propagating uncertainties does not yield a complete specification of the distribution represented by the result y and uncertainty $u(y)$. Therefore, the coverage factor k cannot often be determined such that the coverage probability of the interval $[y \pm ku(y)]$ is 95%. We suggest that one should not claim 95% coverage probability when such a claim cannot be justified.

Note 1. According to some metrologists, having computed the result y and standard uncertainty $u(y)$, one is free to view this information as the only existing information about Y. In that case, the maximum entropy distribution for Y is the normal distribution with expected value y and standard deviation $u(y)$. Then, the coverage factor $k = 2$ corresponds to about 95% coverage probability with respect to the maximum entropy distribution for Y. Mathematically, the maximum entropy distribution does not apply as long as the data used to compute y and $u(y)$ are available.

6. Practical issues in justifying the statistical assumptions about measurement data

An expression of uncertainty is always conditional on the underlying statistical assumptions. Therefore, the assumptions must be justified. In this section, we address practical issues in justifying assumption 1 of section 2, which consists of the following parts: (1) the measurements $z_1, \ldots, z_m$ are independently distributed; (2) the experimental standard deviation $s(z)$ is representative of the standard deviation σ of the common sampling distribution of $z_1, \ldots, z_m$; (3) the common sampling distribution of $z_1, \ldots, z_m$ is approximately normal. These assumptions are special properties that do not just happen; they must be designed into the measurement data.

6.1. Independence

Repeat measurements made close together in either time or in close proximity on a material often tend to be correlated. Therefore, the measurements should be made sufficiently far apart in time or on sufficiently distanced samples of the material to prevent correlation. A useful strategy in some applications is to turn the instruments of measurements off and on between successive measurements to induce independence [17]. Another strategy is to identify important random influence quantities, treat them as factors of an experiment, and use a multi-factor experiment plan to make a list of conditions for measurement. Such a list of conditions is called noise-array [18, 19]. Then exactly one measurement is made for each condition in noise-array. The measurements so obtained may be treated as independent.

6.2. Representative standard deviation

The dispersion of measurements $z_1, \ldots, z_n$ as quantified by the experimental standard deviation

$$s(z) = \sqrt{\frac{\sum_i (z_i - z_A)^2}{m - 1}}$$

must be representative of the standard deviation σ of all possible measurements in the given conditions. If the influence quantities such as time at which measurements are made, operator, location, measuring instrument, and conditions of use are to be treated as random, then they must be changed to make the measurements representative. We quote Youden [20], 'Repeat measurements cannot reveal the vicissitudes of measurement making unless the operator gives the vicissitudes a chance to occur'. When it is not practical to change the random influence quantities, they become systematic influence quantities that must be identified and then accounted for in the combined standard uncertainty as recommended by the ISO *Guide*.

6.3. Normality

Justification of the same normal sampling distribution for $z_1, \ldots, z_n$ requires a demonstration that the measurement process is in a state of statistical control and that the measurements appear as random drawings from a normal distribution. In many physical experiments, it is impractical or impossible to demonstrate statistical control. The CLT is often invoked to claim normality of measurements. Such invocation must be justified (section 2.2, note 5).

7. Summary

The ISO *Guide* recommends a standardized way of expressing uncertainty in all kinds of measurements and provides a comprehensive approach for combining information to evaluate that uncertainty. The ISO *Guide* is, however, not fully consistent. It recommends the use of classical (frequentist) statistics for evaluating the Type A component of the combined standard uncertainty $u(y)$ associated with the result of measurement y. The probabilistic interpretation of the corresponding uncertainty interval $[y \pm ku(y)]$ is not consistent with the ISO *Guide*'s definition of coverage probability (level of confidence). In addition, the measurement equation and the procedure of the ISO *Guide* for evaluating the combined standard uncertainty are not consistent when the Type A component of the combined standard uncertainty is evaluated from classical statistics. We suggest that both problems would disappear if all Type A components are determined through Bayesian statistics. It turns out that the estimates from a classical statistical analysis are either equal or approximately equal to the corresponding estimates from a Bayesian analysis with non-informative prior probability distributions. Therefore, the classical (frequentist) estimates may be used provided they are interpreted from the Bayesian viewpoint. This proposal is entirely consistent with the ISO *Guide* because Bayesian statistical analyses are fully consistent with the definition of Type A evaluations.

Since by propagating uncertainties one does not obtain a complete specification of the probability distribution represented by the result of measurement y and its associated standard uncertainty $u(y)$, the coverage factor k such that the interval $[y \pm ku(y)]$ has a desired coverage probability p, such as 95%, cannot often be determined. The coverage factor determined from a normal distribution or a scaled-and-shifted t-distribution with degrees of freedom determined by the Welch–Satterthwaite formula as recommended by the ISO *Guide* may be incorrect. When the distribution represented by y and $u(y)$ is not completely determined, the interval $[y \pm ku(y)]$ may be interpreted in terms of its minimum coverage probability for an applicable class of probability distributions.

An alternative to propagating uncertainties is propagating distributions by numerical simulation of the measurement equation. Simulation is attractive when it can be done and is justified. However, it is important to note that when the measurement equation $Y = f(X_1, \ldots, X_n)$ is non-linear, the distribution for Y determined by numerical simulation is different from the distribution determined by propagating uncertainties through a linear approximation of Y.

Acknowledgments

We thank the referee for a careful and thorough review and for making suggestions for improvement. The following provided useful comments on an earlier draft of this paper: Bob Frenkel, Ron Boisvert, Don Malec, Steve Phillips, Fern Hunt, and Geoff McFadden.

References

[1] *Guide to the Expression of Uncertainty in Measurement* 2nd edn (Geneva: International Organization for Standardization) 1995 ISBN 92-67-10188-9

[2] Ellison S L R, Rosslein M and Williams A (ed) 2000 *Quantifying Uncertainty in Analytical Measurement* 2nd edn (Eurachem and Cooperation on International Traceability in Analytical Chemistry CITAC) http://www.measurementuncertainty.org/

[3] *US Guide to the Expression of Uncertainty in Measurement* (American National Standards Institute) 2000 ANSI/NCSL Z540-2-1997

[4] Taylor B N and Kuyatt C E 1994 *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results* NIST Technical Note 1297, US Department of Commerce, National Institute of Standards and Technology, http://physics.nist.gov/Document/tn1297.pdf

[5] Taylor B N 1998 Personal communication, National Institute of Standards and Technology, Gaithersburg, MD

[6] Lee P M 1997 *Bayesian Statistics* 2nd edn (New York: Oxford University Press)

[7] Box G E P and Tiao G C 1973 *Bayesian Inference in Statistical Analysis* (Reading, MA: Addison-Wesley)

[8] Weise K and Wöger W 1992 A Bayesian theory of measurement uncertainty *Meas. Sci. Technol.* **3** 1–11

[9] Estler W T 1999 Measurement as inference: fundamental ideas *Ann. CIRP* **48** 611–31

[10] Lira I and Wöger W 2001 Bayesian evaluation of the standard uncertainty and coverage probability in a simple measurement model *Meas. Sci. Technol.* **12** 1172–9

[11] Gleser L J 1998 Assessing uncertainty in measurement *Stat. Sci.* **13** 277–90

[12] Hoel P G 1971 *Introduction to Mathematical Statistics* 4th edn (New York: Wiley)

[13] Tribus M 1969 *Rational Descriptions, Decisions and Designs* (New York: Pergamon)
[14] Evans M, Hastings N and Peacock B 2000 *Statistical Distributions* 3rd edn (New York: Wiley)
[15] Mathai A M and Pederzoli G 1977 *Characterizations of the Normal Probability Law* (New York: Wiley)
[16] Stuart A and Ord J K 1987 *Kendall's Advanced Theory of Statistics, Distribution Theory* vol 1 (New York: Oxford University Press)
[17] Easterling R G 2002 Personal communication, Sandia National Laboratories, Albuquerque, NM
[18] Taguchi G 1987 *System of Experimental Design* (White Plains, NY: UNIPUB Kraus International Publications)
[19] Kackar R N 1985 Off-line quality control, parameter design, and the Taguchi method *J. Quality Technol.* **17** 176–209
[20] Youden W J 1962 Realistic estimates of error *ISA J.* Publication of Instrument Society of America, 57–8