

An improved procedure for combining Type A and Type B components of measurement uncertainty

R. Willink*

Received: 26 October 2012 / Accepted: 22 November 2012

Abstract. The *Guide to the Expression of Uncertainty in Measurement* describes a method of evaluating measurement uncertainty that involves attributing an underlying variance to each component of the total measurement error. Assessment of such a variance by statistical methods is known as Type A evaluation of uncertainty, while assessment by any other means is known as Type B evaluation. The involvement of degrees of freedom with Type A evaluation is in accordance with statistical theory. However, the use of “degrees of freedom” with Type B evaluation is an idea primarily developed for practicality, and might be improved upon. That idea ignores the fact that the underlying distribution of the corresponding error will often be assumed to be blunter than normal, as with a uniform distribution, and it can lead to underestimation of the total error variance. Consequently, an alternative concept for Type B evaluation is presented. The resulting analysis fits easily with the use of degrees of freedom in Type A evaluation and leads to an overall method of evaluation that often permits the proper quotation of shorter uncertainty intervals.

Keywords: Coefficient of excess; coefficient of kurtosis; confidence interval; error analysis

1 Introduction

The fundamental idea underlying the *Guide to the Expression of Uncertainty in Measurement* [1] (the *Guide*) is the idea that a statement of measurement uncertainty is obtained through explicit combination of different components of “error” or “uncertainty”. The *Guide* gives a description of how a component of uncertainty quantified by statistical analysis of data, i.e. by “Type A” evaluation, can be combined with a component of uncertainty quantified by any other means, i.e. by “Type B” evaluation. The result of this procedure is a single estimate of variance, u^2 , that is associated with the measurement result and a single number of degrees of freedom linked to u^2 . The fact that the different types of component were previously treated separately is indicative of the fact that the method of the *Guide* is not theoretically obvious and that it represents a pragmatic solution to a difficult problem. As such, it cannot be expected to be perfect or complete, especially when reasonable performance measures are proposed. This paper describes an alternative technique for combining the two types of uncertainty components. This technique is designed to strengthen the theoretical basis of the *Guide*’s approach and so to improve its practical performance.

To understand the method of the *Guide* as a whole, we must examine the classical idea of Type A evaluation of uncertainty, which usually relates to the spread of experi-

mental data. The basic method for estimating the variance in the output of a procedure from the spread of data is well known. For a sample of fixed size, an unbiased estimate is the familiar “sample variance” s^2 . When the sample size is $\nu + 1$, this estimate is said to be based on “ ν degrees of freedom” partly because if the data were independently drawn from a single normal distribution with variance σ^2 then the unknown figure $\nu s^2/\sigma^2$ was drawn from the chi-square distribution with ν degrees of freedom. So if a quantity is measured by taking the mean of n results obtained in repetition of some procedure then the associated standard uncertainty of measurement is $u = s/\sqrt{n}$ and the accompanying number of degrees of freedom is $\nu = n - 1$. A figure of $\nu = \infty$ represents exact knowledge of σ^2 .

Suppose that an overall measurement error is the sum of several components of error each independently assessed in this way, and suppose that the i th component has unknown variance σ_i^2/n_i . It follows that the variance of the total measurement error is $\sigma_A^2 \equiv \sum \sigma_i^2/n_i$ and that an unbiased estimate of this variance is

$$u^2 = \sum u_i^2. \quad (1)$$

By an approximation of Welch [2,3] and Satterthwaite [4], an effective number of degrees of freedom that can be associated with u^2 is

$$\nu_{\text{eff}} = \frac{u^4}{\sum u_i^4/\nu_i}, \quad (2)$$

with $\nu_i = n_i - 1$. The idea of Welch and Satterthwaite is that $\nu_{\text{eff}} u^2/\sigma_A^2$ can be thought of as having been drawn

* Correspondence: robin.willink@gmail.com

from a distribution well approximated by the chi-square distribution with ν_{eff} degrees of freedom. By standard theory in classical statistics, this enables the width of an uncertainty interval to be obtained by multiplying u by the appropriate quantile of Student's t -distribution with ν_{eff} degrees of freedom.

This idea was incorporated into the procedure of the *Guide* in its presentation of Type A evaluation [1, Sect. 4.2]. The essential ideas are that (i) s_i^2/n_i is an unbiased estimate of the variance of the distribution from which the corresponding component of error is drawn, i.e. the parent distribution of the component of error; and (ii) ν_i is a measure of the accuracy of this estimate.

The full procedure given in the *Guide* also involves the idea that parent distributions, and hence parent variances, can be attributed to all other components of measurement error. In Type B evaluation, the *Guide* recommends that a figure u_i^2 be chosen to represent the parent variance of the error and that a finite number of degrees of freedom ν_i be attached to u_i^2 if there is doubt about it [1, Eq. G.4.2]. More specifically, if a figure $\Delta u_i/u_i$ is put forward as the relative uncertainty in u_i then

$$\nu_i \approx \frac{1}{2} \left(\frac{u_i}{\Delta u_i} \right)^2. \quad (3)$$

(For example, as is described in the *Guide*, if it is judged that the value of u_i is reliable to about 25% then the number of degrees of freedom to be associated with u_i^2 is $\nu_i = 8$.) In this way, the results of Type A and Type B evaluations in the *Guide* are both represented in a single form for combination and propagation using (1) and (2).

The method of the *Guide* is simple and practical, and has been used widely. However, there remains the possibility that Type B evaluation can be given a stronger theoretical backing and that the overall method can have better performance. In particular we can identify two issues. First, Type B evaluation in the *Guide* implicitly treats the parent distributions attributed to the corresponding errors as being normal, i.e. Gaussian, whereas the distributions typically employed for this purpose have "lighter tails", like the uniform distribution. The result is that often the multiplier calculated with the help of (2) is larger than need be, so that the uncertainty interval is larger than need be. So some measurement capability is, in effect, being lost in the data analysis. Second, as shall be explained in Section 2.3, the way in which a finite number of degrees of freedom ν_i is associated with a Type B evaluation in practice implies that, on average, u_i^2 underestimates the corresponding variance, albeit by only a small amount. Through (1), this has the effect of biasing an overall estimate of variance when the measurand acts as an input quantity in another measurement. In this paper, we propose a method to overcome these concerns.

2 The proposed method

For simplicity, let us use the term "Type A error" to refer to a component of error whose parent variance was estimated by Type A evaluation and the term "Type B error"

to refer to a component of error whose parent variance was estimated by Type B evaluation. We now propose a method of analysis that entails a different treatment of Type B errors. The method is as simple as that of the *Guide* and is equivalent to it when all components of error are of Type A.

So our attention is now solely on Type B evaluation of uncertainty. The method of the *Guide* involves the idea that a Type B error is drawn from a probability distribution with a known form but with a scale factor that is possibly unknown. This might be thought to imply that we do not fully know the effective parent distribution of the error, but that is not correct. As we shall see more clearly in Section 2.3, to assume that an error comes from a distribution that is known apart from a scale factor and subsequently to give the scale factor a known distribution is, in effect, to assume that the error comes from a fully known distribution. So the idea of representing the corresponding component of uncertainty by some estimate of variance and an associated finite number of degrees of freedom, as in (2), becomes theoretically inappropriate.

Instead we propose that the parent distribution of the Type B error be represented by its known variance u_i^2 and by its known coefficient of excess, γ_i , which is an index of its shape only. The coefficient of excess is zero for a normal distribution and is negative for a distribution with lighter tails, such as the uniform distribution, which has $\gamma_i = -1.2$. When there are several independent Type B errors with parent variances $\{u_i^2\}$ and coefficients of excess $\{\gamma_i\}$, the sum of these errors has variance $u^2 = \sum u_i^2$, as in (1), and has coefficient of excess given exactly by [5, Sect. 2] [6, Sect. 7.3]

$$\gamma = \frac{\sum \gamma_i u_i^4}{u^4}. \quad (4)$$

This equation allows the propagation of shape information when independent symmetric random variables are summed.

Equation (4) has the same form as the equation

$$\frac{1}{\nu_{\text{eff}}} = \frac{\sum \frac{1}{\nu_i} u_i^4}{u^4},$$

which is a simple re-expression of (2). This suggests that, for a constant a yet to be determined, a simple means of treating the two types of errors together is to set $\tau_i = 1/\nu_i$ for a Type A error and $\tau_i = \gamma_i/a$ for a Type B error and then to apply (1) but calculate

$$\tau = \frac{\sum \tau_i u_i^4}{u^4}$$

instead of (2). If the result τ is positive then the half-width of the uncertainty interval is obtained by multiplying u by the appropriate quantile of the t -distribution with $\nu = 1/\tau$ degrees of freedom, but if τ is negative then the width of the uncertainty interval is to be the difference between the appropriate upper and lower tail quantiles of some symmetric distribution with variance u^2 and coefficient of excess $\gamma = a\tau$. The half-width of the interval will be $k(\tau)u$, where the coverage factor $k(\tau)$ is some continuous function

of τ . We call τ the *coverage index* and call the method the *coverage-index method*. If a and the family of symmetric distributions are chosen well then this method would be expected to perform better than the method of the *Guide*.

2.1 Optimizing performance

In order to say that one method can “perform better” than another, we must have one or more measures of performance in mind. Suppose the context involves generating uncertainty intervals that will be quoted as being 95% reliable. Then approximately 95% of the intervals obtained in a long series of different measurements are to enclose the true values of the measurands. This level of performance is ensured if, for every possible set of parameters defining a measurement, approximately 95% of the intervals calculated in repetition of that measurement enclose the true value of that measurand. It is this stronger condition that relates to the classical statistical concept of “95% confidence”. The approach taken here is to choose a and the corresponding function $k(\tau)$ so that this stronger condition is met. A second measure of performance is the shortness of the intervals generated. A method of analysis with at least 95% reliability that generates narrow intervals is superior to a method with at least 95% reliability that generates longer intervals: the client is being given a more informative statement about the measurand.

Consider some specific requirements for $k(\tau)$. For $\tau > 0$, it should evaluate to the 0.975 quantile of the t -distribution with $1/\tau$ degrees of freedom, which we denote by $t_{95,1/\tau}$ for consistency with the *Guide*. The method will then reproduce the interval of the *Guide* when there are only Type A errors. Similarly, we require $k(0) = 1.96$, this being the 0.975 quantile of the standard normal distribution. Also, we stipulate that $k(-1.2/a) \approx 1.645$, this being the 0.975 quantile of the standardized rectangular distribution. This is because the most common non-normal distribution used in Type B evaluation is the uniform distribution, and one component of error drawn from such a distribution could be dominant. Last, but not least, the expression for $k(\tau)$ should be reasonably simple.

After some experimentation, the value of $a = 100$ was chosen and a suitable candidate for $k(\tau)$ was found to be the piecewise but continuous function given by (7) below. These choices were seen to perform well in simulations described in Section 3. Figure 1 shows the function $k(\tau)$ and the quantile $t_{95,1/\tau}$ for $\tau = 0.05, 0.1, \dots, 1$. The function passes through these points almost exactly. Figure 2 focuses on the region with $\tau \approx 0$.

2.2 The general result

The basic measurement problem addressed in the *Guide* involves a measurand Y that is related to “input quantities” X_1, \dots, X_m by a function that is approximately linear in the vicinity of the measured values x_1, \dots, x_m . The necessary generalizations of our method from the linear

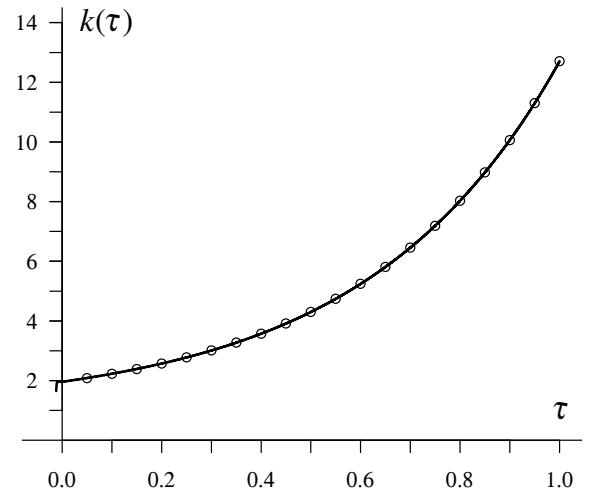


Fig. 1. The function $k(\tau)$ for $-0.012 \leq \tau \leq 1$ and the quantile $t_{95,1/\tau}$ for $\tau = 0.05, 0.1, \dots, 1$.

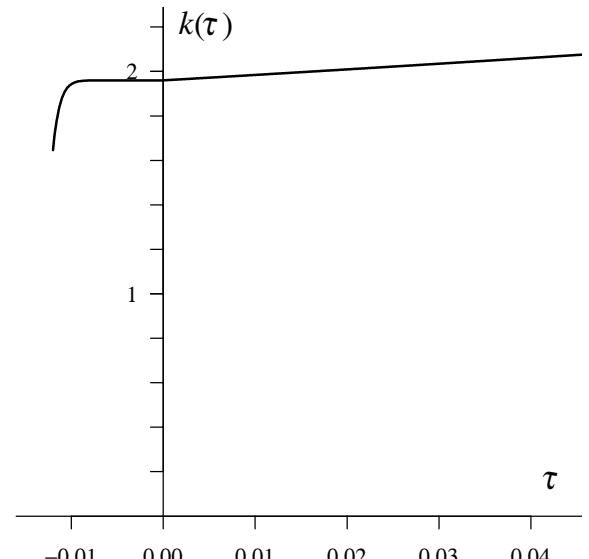


Fig. 2. The function $k(\tau)$ for $\tau \approx 0$.

situation addressed above are easily made. The proposed method can then be described in the following way.

Let the unique unknown value of the measurand be $Y = F(X_1, \dots, X_m)$ where F is a known function and X_1, \dots, X_m are the unknown values of input quantities. Also, suppose that x_1, \dots, x_m are estimates of X_1, \dots, X_m with corresponding standard uncertainties u_1, \dots, u_m . If u_i^2 is an estimate of variance obtained from a sample of size $\nu_i + 1$, so that it is based on ν_i degrees of freedom, then we set $\tau_i = 1/\nu_i$. Alternatively, if u_i^2 is the variance of a (parent) distribution attributed to the error $x_i - X_i$ by non-statistical means then we set $\tau_i = 0.01\gamma_i$, where γ_i is the coefficient of excess of that distribution.

Define $y = F(x_1, \dots, x_m)$ and $c_i = \partial y / \partial x_i$. So c_i describes the sensitivity of the overall estimate y to a small change in x_i . Then, if the errors $\{x_i - X_i\}$ were incurred independently, the standard uncertainty to be associated with y as an estimate of Y is

$$u = \sqrt{\sum_{i=1}^m c_i^2 u_i^2}, \quad (5)$$

the coverage index to accompany u is

$$\tau = \frac{\sum_{i=1}^m \tau_i c_i^4 u_i^4}{u^4}, \quad (6)$$

and a 95% interval of expanded uncertainty is $[y - k(\tau)u, y + k(\tau)u]$, where

$$k(\tau) = \begin{cases} 1.96 - 0.017(100\tau)^{16} & \text{for } -0.012 \leq \tau \leq 0 \\ \frac{a_0 + a_1\tau + a_2\tau^2 + a_3\tau^3}{1 - 0.473\tau} & \text{for } 0 \leq \tau \leq 1, \end{cases} \quad (7)$$

with $a_0 = 1.96$, $a_1 = 1.491$, $a_2 = 1.381$ and $a_3 = 1.864$. The figures y , u and τ can be used to represent Y in any subsequent measurement that involves Y as an input quantity.

This technique is primarily intended as a modification of the method of the *Guide* for the explicit admission of Type B errors drawn from non-normal distributions. Equation (5) is the law of propagation of uncertainty with uncorrelated errors, while (6) is a generalization of the “Welch-Satterthwaite formula” (2), which is recovered by setting $\nu_{\text{eff}} = 1/\tau$ and $\nu_i = 1/\tau_i$. Equation (7) removes the need for a table of the t -distribution when $\tau > 0$ and gives an appropriately small coverage factors when $\tau < 0$.

Section 2.4 will give an example of the use of the coverage-index method, while Section 3 will compare the performances of this method and the *Guide*'s method. Before that, however, we must discuss important points about the identity of u_i in Type B evaluation and the role of its square in (5). This discussion relates to the common situation where there is doubt about the scale factor of the distribution underlying the Type B error.

2.3 Doubt about a scale factor

Earlier we mentioned that a Type B error might be regarded as having been drawn from a distribution with a certain form but with an unknown standard deviation. One way to accommodate this lack of knowledge about the standard deviation is to see it as having arisen from some known distribution itself. If we take this step then the Type B error is, in effect, drawn from an overall parent distribution that is fully known [5, Sect. 4.3] [6, Theory 7.2]. That is, if $f(x; \sigma)$ is the probability density function of

the initial distribution when the standard deviation is σ and if the distribution from which σ is drawn has probability density function $g(\sigma)$ then the probability density function of the overall parent distribution is the known function

$$h(x) = \int_0^\infty f(x; \sigma) g(\sigma) d\sigma,$$

and the mean, variance and coefficient of excess of the overall distribution follow. This final distribution is called a *compound distribution*, because the distribution is formed when one of the original parameters is itself distributed. A useful notation exists to describe such a situation. Suppose, for example, that the initial (compounded) distribution is the normal distribution with mean μ and unknown standard deviation σ , denoted $N(\mu, \sigma^2)$, and that σ is treated as a random variable with the (compounding) uniform distribution with limits σ_{low} and σ_{high} , denoted $U(\sigma_{\text{low}}, \sigma_{\text{high}})$. Then the compound distribution is denoted

$$N(\mu, \sigma^2) \bigwedge_\sigma U(\sigma_{\text{low}}, \sigma_{\text{high}}).$$

The concept of a compound distribution is relevant when we come to interpret a statement such as “The estimated interval of -0.05 °C to $+0.05$ °C for the temperature difference... is believed to be reliable only to 50 percent”, which is a statement that is found in the first example in the *Guide* [1, Eq. H.1.6] and alluded to in its first supplement [7, Sect. 9.5.2.9]. If we are to form a parent distribution from this information then we suggest taking it to mean that the initial distribution is uniform with mean zero and with a width that is itself drawn from a uniform distribution with limits 0.25 °C to 0.75 °C. The compound distribution can then be represented as

$$U(-h, h) \bigwedge_h U(0.25, 0.75).$$

Equivalently, we take the statement to mean that the initial distribution is uniform with mean zero and standard deviation drawn from the uniform distribution with limits $\sigma_{\text{low}} = 0.25/\sqrt{3}$ °C and $\sigma_{\text{high}} = 0.75/\sqrt{3}$ °C. The compound distribution can then be represented as

$$U(-\sqrt{3}\sigma, \sqrt{3}\sigma) \bigwedge_\sigma U(0.25/\sqrt{3}, 0.75/\sqrt{3}),$$

with the units being degrees Celsius.

Many statements of experimental uncertainty will be interpretable in such a way. The initial distribution will have some form and will have a standard deviation best estimated by some figure $\hat{\sigma}$ but known only to a certain relative amount, $100\delta\%$. This will be taken to mean that the standard deviation of that initial distribution was drawn from a uniform distribution with limits

$$\sigma_{\text{low}} = \hat{\sigma}(1 - \delta)$$

and

$$\sigma_{\text{high}} = \hat{\sigma}(1 + \delta).$$

(So we can identify δ with $\Delta u_i/u_i$.) Also

$$\delta = \frac{\sigma_{\text{high}} - \sigma_{\text{low}}}{\sigma_{\text{high}} + \sigma_{\text{low}}} \quad (8)$$

and $\hat{\sigma}^2 = (\sigma_{\text{high}} + \sigma_{\text{low}})^2/4$.

Accordingly, the overall parent distribution for the Type B error is a compound distribution with a variance that can be shown to be given by

$$u_i^2 = \frac{(\sigma_{\text{high}} + \sigma_{\text{low}})^2}{4} + \frac{(\sigma_{\text{high}} - \sigma_{\text{low}})^2}{12} \quad (9)$$

or, equivalently, given by (10) below. The variance of this distribution therefore exceeds the figure $\hat{\sigma}^2$ unless there is no doubt about the scale factor, i.e. unless $\delta = 0$. (This can also be seen in the analogous analysis in the first supplement to the *Guide* [7, Sects. 6.4.3.3 and 9.5.2.9], where the variance exceeds the variance of the initial uniform distribution.) The figure $\hat{\sigma}$ will be the best single estimate of the scale factor, and so it corresponds to the figure called “ u_i ” in the method of the *Guide*. Therefore, (9) and (10) justify the claim made in Section 1 that the method of the *Guide* tends to underestimate the total measurement variance, albeit by an amount that will be small in practice.

Let us make this point again. With our method, the figure u_i^2 in Type B evaluation that is to appear in (1) and (5) is given by equation (9). However, with the *Guide*'s method, u_i^2 will be identified with $\hat{\sigma}^2$, which is simply the square of the best estimate of standard deviation, and so is not generally the best estimate of variance. Equations (9) and (10) show that, if we interpret statements like “the scale factor is deemed reliable to 10%” in the manner described above then the method of the *Guide* will lead to a negatively biased estimate of the overall measurement variance.

Now let the coefficient of excess for the form of the initial distribution be γ and, as above, let the standard distribution of that distribution be assumed to have been drawn from the uniform distribution with limits σ_{low} and σ_{high} , and so with δ given by (8). Then the parent distribution for the Type B error has variance u_i^2 given by (9) or (10) and has coefficient of excess γ_i given by (11) below [6, Result 7.3]. These are the figures to be used in (5) and (6) for this component of uncertainty.

Therefore, the following paragraph is to be appended to the indented text in Section 2.2.

The variance u_i^2 and coefficient of excess γ_i to be attributed to the parent distribution of an error $x_i - X_i$ by non-statistical means can often be found in the following way. Suppose that the error is thought to have been drawn from a distribution with zero mean, unknown standard deviation and known coefficient of excess γ . (For example, $\gamma = 0$ if the distribution is normal and $\gamma = -1.2$ if the distribution is uniform.) If the best estimate of the standard deviation is $\hat{\sigma}$ and if this figure is deemed to be accurate to $\pm 100 \delta\%$ then

$$u_i^2 = \hat{\sigma}^2 \left(1 + \frac{\delta^2}{3} \right) \quad (10)$$

and

$$\gamma_i = \frac{\left(1 + 2\delta^2 + \frac{\delta^4}{5} \right) \gamma + 4\delta^2 + \frac{4\delta^4}{15}}{\left(1 + \frac{\delta^2}{3} \right)^2}. \quad (11)$$

These are the figures to be used in (5) and (6) for this component of uncertainty.

2.4 Example

Let us apply the method to the data of Example H1 of the *Guide*, which describes the evaluation of the uncertainty in the length of a gauge. The six relevant sources of uncertainty are listed in Table 1, which is based on Table H.1 of the *Guide*. Each partial derivative c_i has been incorporated into the error $x_i - X_i$, so in effect $c_i = 1$.

The first three sources of uncertainty are associated with repeated measurements, and so they relate to Type A errors. Therefore, the method of the *Guide* and the proposed method treat these errors in the same way. The corresponding figures of u_i and ν_i in Table 1 are taken from the *Guide*, and because they are used in the same way by both methods there is little merit in representing these u_i figures to more significant figures.

The fourth variable relates to the systematic effects of the comparator. This has standard uncertainty 20 nm at the “three sigma level” [1, Eq. H.1.3.2], and this uncertainty is assumed to be reliable to 25% [1, Eq. H.1.6]. So the figure of standard uncertainty used in the *Guide* is $20/3 = 6.667$ and the number of degrees of freedom is, from (3), $1/2 \times 0.25^{-2} = 8$. In the absence of other information, we take the distribution of $x_4 - X_4$ to be that of a normal variable with mean zero and standard deviation uniformly distributed between $\sigma_{\text{low}} = 6.667 \times 0.75$ nm and $\sigma_{\text{high}} = 6.667$ nm $\times 1.25$ nm. So $\gamma = 0$ and $\delta = 0.25$. We obtain $u_i = 6.736$ nm and $\gamma_i = 0.24$.

The fifth variable relates to the difference in the thermal expansion coefficients. This is represented by a uniform variable with limits $\pm 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ multiplied by the factor $0.05 \text{ m} \times 0.1 \text{ }^{\circ}\text{C}$ that is seen in equation H.6a of the *Guide*. The limits for this variable are deemed to be reliable to 10% [1, Eq. H.1.6]. Therefore, the figure of standard uncertainty used in the *Guide* is $5/\sqrt{3} = 2.887$ nm and the number of degrees of freedom is $1/2 \times 0.1^{-2} = 50$. In the proposed method $\delta = 0.1$, $\sigma_{\text{low}} = 2.887 \times 0.9$ nm, $\sigma_{\text{high}} = 2.887 \times 1.1$ nm and $\gamma = -1.2$. We obtain $u_i = 2.892$ nm and $\gamma_i = -1.20$.

The sixth variable relates to the difference in temperatures. This is represented by a uniform variable with limits $\pm 0.05 \text{ }^{\circ}\text{C}$ multiplied by the factor $0.05 \text{ m} \times 11.5 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ that is also seen in equation H.6a of the *Guide*. The limits are only deemed to be reliable to 50% [1, Eq. H.1.6]. Therefore, the figure of standard uncertainty used in the *Guide* is $28.75/\sqrt{3} = 16.60$ nm and the number of degrees of freedom is $1/2 \times 0.5^{-2} = 2$. In the proposed method $\delta = 0.5$, $\sigma_{\text{low}} = 16.60 \times 0.5$ nm, $\sigma_{\text{high}} = 16.60 \times 1.5$ nm and $\gamma = -1.2$. We obtain $u_i = 17.28$ nm and $\gamma_i = -0.77$.

Applying the method of the *Guide* gives a standard uncertainty of $u = 31.7$ nm and a 95% uncertainty

Table 1. Summary of an example analysis for the method of the *Guide* and the coverage-index method.

Source of uncertainty	<i>i</i>	<i>Guide</i>		Coverage-index method		
		<i>u_i</i>	<i>v_i</i>	<i>u_i</i>	<i>v_i</i>	<i>γ_i</i>
Calibration of standard gauge	1	25 nm	18	25 nm	18	
Repeated observations	2	5.8 nm	24	5.8 nm	24	
Random effects of comparator	3	3.9 nm	5	3.9 nm	5	
Systematic effects of comparator	4	6.667 nm	8	6.736 nm		0.24
Difference in expansion coefficients	5	2.887 nm	50	2.892 nm		-1.20
Difference in temperatures	6	16.60 nm	2	17.28 nm		-0.77
Summary measure of scale		$u = 31.66 \text{ nm}$		$u = 32.03 \text{ nm}$		
Associated index		$v_{\text{eff}} = 16.73$		$\tau = 0.020$		
Coverage factor for 95% probability		$t_{95,16.73} = 2.113$		$k(\tau) = 2.010$		
Expanded uncertainty		$t_{95,16.73} u = 66.87 \text{ nm}$		$k(\tau)u = 64.37 \text{ nm}$		

interval with half-width of 66.9 nm. Applying the proposed method gives a larger standard uncertainty of $u = 32.0$ nm but a 95% uncertainty interval with the half-width of 64.4 nm, which is smaller by approximately 4%. The output effective number of degrees of freedom with the method of the *Guide* method is $v_{\text{eff}} = 16.7$ and the corresponding coverage factor is 2.11. In contrast, the output shape factor with the proposed method is $\tau = 0.002$, so the coverage factor is the smaller figure 2.01. The proposed method gives a shorter interval while having the same level of assurance, 95%, so the metrologist who uses the proposed method is justifiably giving a more precise statement about the length of the gauge block.

This example illustrates the two unfavourable aspects of the *Guide*'s method that are described in Section 1. First, it gives an interval that is longer because it cannot take into account the bluntness of the uniform distributions. Second, it has resulted in an estimate of variance $u^2 = 1002 \text{ nm}^2$ that is smaller than the figure $u^2 = 1026 \text{ nm}^2$ obtained using the proposed method. If what metrologists mean by “uncertainty in a scale factor” in Type B evaluation is adequately represented in Section 2.3 then the conclusion to be drawn is that the *Guide*'s estimate of variance is too low. Using that estimate in some future measurement where the length of the gauge was an input quantity will, through (5), lead to a negatively biased estimate of total measurement variance.

3 Performance assessment

To determine whether the coverage-index method is to be preferred to the *Guide*'s method in general use we must examine a range of situations, not just the situation described in the previous example. More importantly, we must be clear about what any method of uncertainty analysis is actually intended to achieve. Our basic performance characteristics for such a method have already been outlined in Section 2.1. We require that a method said to produce “95% uncertainty intervals” will generate intervals containing the value of the measurand at least 95 times out of 100 in a long sequence of measurements that is relevant to a client. This is ensured if, for every measurement, the method generates intervals containing the value of the measurand at least 95 times out of 100 when that same

measurement is repeated. Given two methods with this property, the one to be regarded as superior is the one that generates shorter intervals on average.

With these considerations in mind, the performances of the *Guide*'s method and the proposed method were examined by Monte-Carlo methods for the function $Y = X_1 + X_2$, where $x_1 - X_1$ is a Type A error and $x_2 - X_2$ is a Type B error. The parent standard deviation of the Type A error is given by σ_1^* , (which is $\sigma_1/\sqrt{n_1}$), and the accompanying number of degrees of freedom is v_1 . Without loss of generality for situations that involve a Type B error, the best estimate of the parent standard deviation of the Type B error, $\hat{\sigma}_2$, could be set to unity. The reliability of this estimate is represented by δ_2 . The parent distribution of the Type B error was either uniform or normal, so in the proposed method $\gamma_2 = -1.2$ or $\gamma_2 = 0$.

Without loss of generality, we set $X_1 = X_2 = 0$. So each measurement situation was then defined by $\gamma_2, \sigma_1^*, \delta_2$ and v_1 . For each setting of γ_2 in $\{-1.2, 0\}$, σ_1^* in $\{0.01, 0.1, 1, 10, 100\}$, δ_2 in $\{0, 0.1, 0.25, 0.5\}$ and v_1 in $\{2, 3, 8, 20\}$, the measurement was simulated 10^6 times, and the proportions of occasions on which the method of the *Guide* and the proposed method gave an interval containing the true value $Y = 0$ were recorded. Also, the ratio of the mean widths of the intervals generated by the two methods was calculated. The results are given in Table 2 for a Type B error based around the uniform distribution, i.e. for $\gamma = -1.2$, and in Table 3 for a Type B error based around the normal distribution, i.e. for $\gamma = 0$. To save space, the results for $\sigma_1^* = 100$ and for $\delta_2 = 0.1$ have been omitted. The columns headed P give the proportions for which the method of the *Guide* gave an interval containing the true value, while the columns headed P_τ give the proportions for which the proposed method gave an interval containing the true value. The columns headed \bar{w}_τ/\bar{w} give the ratio of the mean length of the intervals with the proposed method to the mean length of the interval with the *Guide*'s method.

For $\gamma = -1.2$ the proportions for the method of the *Guide* ranged from 0.938 to 1.000 while those for the proposed method ranged from 0.937 to 0.979. For $\gamma = 0$ the proportions for the method of the *Guide* ranged from 0.939 to 0.999 while those for the proposed method ranged from 0.938 to 0.951. The smallest proportions for

Table 2. Performance results when measuring $Y = X_1 + X_2$ where $x_1 - X_1$ is a Type A error and $x_2 - X_2$ is a Type B error arising from a uniform distribution.

σ_1^*	δ_2	ν_1	P	P_τ	\bar{w}_τ/\bar{w}
0.01	0	2	1.000	0.950	0.84
0.01	0	3	1.000	0.951	0.84
0.01	0	8	1.000	0.951	0.84
0.01	0	20	1.000	0.951	0.84
0.01	0.25	2	1.000	0.951	0.76
0.01	0.25	3	1.000	0.951	0.76
0.01	0.25	8	1.000	0.952	0.76
0.01	0.25	20	1.000	0.951	0.76
0.01	0.5	2	1.000	0.948	0.50
0.01	0.5	3	1.000	0.948	0.50
0.01	0.5	8	1.000	0.949	0.50
0.01	0.5	20	1.000	0.949	0.50
0.1	0	2	1.000	0.974	0.89
0.1	0	3	1.000	0.975	0.89
0.1	0	8	1.000	0.978	0.88
0.1	0	20	1.000	0.979	0.88
0.1	0.25	2	1.000	0.963	0.79
0.1	0.25	3	1.000	0.964	0.79
0.1	0.25	8	1.000	0.964	0.79
0.1	0.25	20	1.000	0.963	0.79
0.1	0.5	2	1.000	0.953	0.51
0.1	0.5	3	1.000	0.952	0.51
0.1	0.5	8	1.000	0.953	0.51
0.1	0.5	20	1.000	0.952	0.51
1	0	2	0.943	0.940	0.99
1	0	3	0.948	0.947	1.00
1	0	8	0.954	0.953	1.00
1	0	20	0.955	0.954	1.00
1	0.25	2	0.961	0.941	0.94
1	0.25	3	0.964	0.947	0.95
1	0.25	8	0.964	0.952	0.95
1	0.25	20	0.964	0.953	0.96
1	0.5	2	0.993	0.942	0.79
1	0.5	3	0.991	0.946	0.81
1	0.5	8	0.986	0.950	0.84
1	0.5	20	0.983	0.952	0.85
10	0	2	0.938	0.938	1.00
10	0	3	0.946	0.946	1.00
10	0	8	0.950	0.950	1.00
10	0	20	0.950	0.950	1.00
10	0.25	2	0.939	0.939	1.00
10	0.25	3	0.947	0.946	1.00
10	0.25	8	0.950	0.950	1.00
10	0.25	20	0.950	0.950	1.00
10	0.5	2	0.940	0.937	1.00
10	0.5	3	0.946	0.946	1.00
10	0.5	8	0.950	0.950	1.00
10	0.5	20	0.950	0.950	1.00

Table 3. Performance results when measuring $Y = X_1 + X_2$ where $x_1 - X_1$ is a Type A error and $x_2 - X_2$ is a Type B error arising from a normal distribution.

σ_1^*	δ_2	ν_1	P	P_τ	\bar{w}_τ/\bar{w}
0.01	0	2	0.950	0.950	1.00
0.01	0	3	0.950	0.950	1.00
0.01	0	8	0.950	0.950	1.00
0.01	0	20	0.950	0.950	1.00
0.01	0.25	2	0.975	0.948	0.86
0.01	0.25	3	0.975	0.948	0.86
0.01	0.25	8	0.975	0.948	0.86
0.01	0.25	20	0.975	0.948	0.86
0.01	0.5	2	0.999	0.942	0.52
0.01	0.5	3	0.999	0.942	0.52
0.01	0.5	8	0.999	0.941	0.52
0.01	0.5	20	0.999	0.943	0.52
0.1	0	2	0.950	0.950	1.00
0.1	0	3	0.950	0.950	1.00
0.1	0	8	0.950	0.950	1.00
0.1	0	20	0.950	0.950	1.00
0.1	0.25	2	0.975	0.948	0.86
0.1	0.25	3	0.975	0.948	0.86
0.1	0.25	8	0.974	0.948	0.86
0.1	0.25	20	0.975	0.948	0.86
0.1	0.5	2	0.998	0.942	0.53
0.1	0.5	3	0.998	0.942	0.53
0.1	0.5	8	0.998	0.942	0.53
0.1	0.5	20	0.998	0.942	0.53
1	0	2	0.941	0.941	1.00
1	0	3	0.945	0.945	1.00
1	0	8	0.949	0.949	1.00
1	0	20	0.950	0.950	1.00
1	0.25	2	0.959	0.941	0.95
1	0.25	3	0.960	0.945	0.95
1	0.25	8	0.959	0.949	0.96
1	0.25	20	0.958	0.949	0.96
1	0.5	2	0.989	0.942	0.80
1	0.5	3	0.986	0.945	0.81
1	0.5	8	0.980	0.948	0.84
1	0.5	20	0.977	0.948	0.85
10	0	2	0.939	0.938	1.00
10	0	3	0.946	0.946	1.00
10	0	8	0.950	0.950	1.00
10	0	20	0.950	0.950	1.00
10	0.25	2	0.939	0.938	1.00
10	0.25	3	0.946	0.946	1.00
10	0.25	8	0.949	0.950	1.00
10	0.25	20	0.950	0.950	1.00
10	0.5	2	0.940	0.938	1.00
10	0.5	3	0.946	0.946	1.00
10	0.5	8	0.950	0.950	1.00
10	0.5	20	0.950	0.950	1.00

the proposed method were increased to 0.946 and 0.941 respectively when attention was restricted to situations with $\nu \geq 3$. Broadly speaking, the proposed method performed much better than the method of the *Guide* when $\sigma_1^* \lesssim 1$, which is when a significant part of the overall uncertainty was due to the Type B component. The success rates were close to or greater than 0.95 and the intervals were up to 50% shorter. The performances of the two methods were comparable when the Type A error dominated. These results support the suggestion that the proposed method is to be preferred.

4 Comments

Type A evaluation of uncertainty and Type B evaluation of uncertainty are different in nature. Accordingly, there has been considerable historical and recent debate about the proper means of their treatment together. In this paper we have proposed a new way of combining uncertainty information obtained by Type A evaluation and Type B evaluation. This has involved replacing the *Guide's* idea of degrees of freedom for Type B evaluation with the idea of the coefficient of excess of the associated probability distribution.

Our method may be helpfully compared with others by identifying three broad approaches for the joint treatment of Type A and Type B components of uncertainty, as follows. The first is to treat Type B errors in a way that appeals to the theory of treating Type A errors. This is the approach taken in the *Guide* where, as we have seen, the idea of “degrees of freedom” is extended from its theoretical context of Type A evaluation to be used with Type B evaluation also.

Conversely, a second approach is to treat Type A errors in a way that fits Type B errors. This means attributing a known parent distribution to a Type A error. This approach was taken by this author in the relevant chapters of a recent text [6, Chaps. 6–8], where the analysis is carried out under the pretence that the parent distribution for a Type A error was a certain shifted and scaled *t*-distribution. One merit of that approach was its conservatism.

A third approach is to maintain the proper theoretical difference between Type A and Type B errors, which is a difference that relates to how the scale of the error is estimated. It is this approach that we have taken in this paper. This approach was the first listed in my book [6, Sect. 6.4], but there it was put to one side as being too difficult or potentially too complicated. The simplicity of the method obtained, as is reflected in (6), shows such a judgement to be too severe.

5 Summary

As in the *Guide*, our subject has been the measurement of an unknown quantity able to be written as $Y = F(X_1, \dots, X_m)$ where F is a known function

and X_1, \dots, X_m are the unknown values of input quantities. The procedure for the evaluation of measurement uncertainty advocated in the *Guide* involves the calculation of a measurement estimate y and the associated standard uncertainty u . It also involves the quotation of a number of degrees of freedom ν_{eff} of (2), which indexes the *t*-distribution whose tail quantile is used to multiply u when calculating an interval of uncertainty.

The figure ν_{eff} is necessarily positive, so the multiplier cannot be less than the corresponding quantile of the standard normal distribution, e.g. 1.96 for a 95% uncertainty interval. In effect, this paper has proposed that the number of degrees of freedom ν_{eff} be replaced by the coverage index τ of (6), which indexes a *t*-distribution if positive but which legitimately leads to the calculation of a narrower interval if negative. This modification will often improve performance when some of the components of error are drawn from non-normal distributions whose variances are assessed by non-statistical means, i.e. by Type B evaluation. In general, these distributions will have lighter tails than a normal distribution, so the method will lead to narrower uncertainty intervals. The method may therefore be seen as increasing measurement capability.

A second benefit of the method is a removal of bias from the overall estimate of measurement variance. The interpretation usually given to the idea that there is uncertainty about the scale of an error in Type B evaluation implies that the figure formed by squaring the best estimate of standard deviation, $\hat{\sigma}$, is negatively biased. The overall estimate of variance becomes negatively biased, and this bias propagates to future measurements. The proposed method avoids this effect through a simple calculation of variance.

References

1. Guide to the Expression of Uncertainty in Measurement (International Organization for Standardization, Geneva, 1995)
2. B.L. Welch, The significance of the difference between two means when the population variances are unequal, *Biometrika* **29**, 350–362 (1937)
3. B.L. Welch, The generalization of “Student’s” problem when several different population variances are involved, *Biometrika* **34**, 28–35 (1947)
4. F.E. Satterthwaite, An approximate distribution of estimates of variance components, *Biometrics* **2**, 110–114 (1946)
5. R. Willink, A procedure for the evaluation of measurement uncertainty based on moments, *Metrologia* **42**, 329–343 (2005)
6. R. Willink, *Measurement Uncertainty and Probability* (Cambridge University Press, 2013)
7. Joint Committee for Guides in Metrology, Evaluation of measurement data – Supplement 1 to the “Guide to the expression of uncertainty in measurement” – Propagation of distributions using a Monte Carlo method (2006)