Standard Practice for Characterizing Uncertainty in Air Quality Measurements[1]

This standard is issued under the fixed designation D7440; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

ε[1] NOTE—Editorial corrections were made throughout in July 2015.

1. Scope

1.1 This practice is for assisting developers and users of air quality methods for sampling concentrations of both airborne and settled materials in characterizing measurements as to uncertainty. Where possible, analysis into uncertainty components as recommended in the ISO Guide to the Expression of Uncertainty in Measurement (ISO GUM, (**1**)[2]) is suggested. Aspects of uncertainty estimation particular to air quality measurement are emphasized. For example, air quality assessment is often complicated by: the difficulty of taking replicate measurements owing to the large spatio-temporal variation in concentration values to be measured; systematic error or bias, both corrected and uncorrected; and the (rare) non-normal distribution of errors. This practice operates mainly through example. Background and mathematical development are relegated to appendices for optional reading.

1.2 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[3]

D1356 Terminology Relating to Sampling and Analysis of Atmospheres

D3670 Guide for Determination of Precision and Bias of Methods of Committee D22

D6246 Practice for Evaluating the Performance of Diffusive Samplers

D6552 Practice for Controlling and Characterizing Errors in Weighing Collected Aerosols

2.2 *Other International Standards:*

ISO GUM Guide to the Expression of Uncertainty in Measurement, ISO Guide 98, 1995 (See Ref (**1**), for an additional measurement uncertainty resource.)[4]

ISO 7708 Air Quality—Particle Size Fraction Definitions for Health-Related Sampling[4]

ISO 15767 Workplace Atmospheres—Controlling and Characterizing Errors in Weighing Collected Aerosol[4]

ISO 16107 Workplace Atmospheres—Protocol for Evaluating the Performance of Diffusive Samplers, 2007[4]

EN 482 Workplace Atmospheres—General Requirements for the Performance of Procedures for the Measurement of Chemical Agents[4]

3. Terminology

3.1 *Definitions*—For definitions of terms used in this practice, see Terminology D1356.

3.2 *Other terms defined as follows are taken from ISO GUM unless otherwise noted:*

3.2.1 *accuracy*—closeness of agreement between the result of a measurement and a true value of the measurand.

3.2.2 *combined standard uncertainty,* u_c—standard uncertainty of the result of a measurement when that result is obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities.

3.2.2.1 *Discussion*—As within ISO GUM, the "other quantities" are designated uncertainty components u_j from source *j*. The component u_j is taken as the standard deviation estimate from source *j* in the case of a source of random variation.

3.2.3 *coverage factor, k*—numerical factor used as a multiplier of the combined standard uncertainty (u_c) in order to obtain an expanded uncertainty (U).

[1] This practice is under the jurisdiction of ASTM Committee D22 on Air Quality and is the direct responsibility of Subcommittee D22.01 on Quality Control.

Current edition approved July 1, 2015. Published July 2015. Originally approved in 2008. Last previous edition approved in 2008 as D7440 – 08. DOI: 10.1520/D7440-08R15E01.

[2] The boldface numbers in parentheses refer to the list of references at the end of this standard.

[3] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

[4] BIPM version available for download from http://www.bipm.org/en/publications/guides/gum.html. ISO version available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

3.2.3.1 *Discussion*—The factor *k* depends on the specific meaning attributed to the expanded uncertainty *U*. However, for simplicity this practice adopts the now nearly traditional coverage factor as the value 2, determining the specific meaning of the expanded uncertainty *U* in different circumstances. Other coverage factors if needed are then easily implemented simply by multiplication of the traditional expanded uncertainty *U* (see 7.1 – 7.4).

3.2.3.2 *Discussion*—The use of a single coverage factor, often through approximation, avoids the overly conservative use of individual component confidence limits rather than root variance estimates as uncertainty components.

3.2.4 *error (of measurement)*—result of a measurement minus a true value of the measurand.

3.2.5 *expanded uncertainty, U*—quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.

3.2.5.1 *Discussion*—This definition has the breadth to encompass a wide variety of conceptions.

3.2.5.2 *Discussion*—The expanded uncertainty *U* in some cases is expressed in absolute terms, but sometimes as relative to the measurement result. What is meant is generally clear from the context.

3.2.6 *influence quantity*—quantity that is not the measurand but that affects the result of the measurement.

3.2.7 *measurand*—particular quantity subject to measurement.

3.2.8 *measurand value*—(adapted from ISO GUM), unknown quantity whose measurement is sought, often called the true value. Examples are the concentration (mg/m^3) of a substance in the air at a particular time and place, the time-weighted average of a concentration at a particular position, or the expected mean concentration estimate as obtained by a reference method at a specific time and position.

3.2.9 *(population) variance (of a random variable)*—the expectation of the square of the centered random variable.

3.2.10 *random error*—result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under the same (*repeatability*) conditions of measurement.

3.2.10.1 *Discussion*—Random error is equal to error minus systematic error.

3.2.11 *(sample) variance*—the sum of the squared deviations of observations from their average divided by one less than the number of observations.

3.2.11.1 *Discussion*—The sample variance is an unbiased estimator of the population variance.

3.2.12 *standard deviation*—positive square root of the variance.

3.2.13 *symmetric accuracy range A*—the range symmetric about (true) measurand values containing 95 % of measurement estimates. *A* is a specific quantification of *accuracy*. (**2**)
ISO 16107

3.2.14 *systematic error (bias)*—mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand.

3.2.15 *Type A evaluation (of uncertainty)*—method of evaluation of uncertainty by the statistical analysis of series of observations.

3.2.16 *Type B evaluation (of uncertainty)*—method of evaluation of uncertainty by means other than the statistical analysis of series of observations.

4. Background Information

4.1 Uncertainty in a measurement result can be taken as the range about an estimate, corrected for bias if known, containing the true, or mean reference value—in the language of ISO GUM, the *measurand* value at given confidence. Uncertainty accounts not only for variation in a method's results at application, but also for incomplete characterization of the method when evaluated. In accordance with ISO GUM, uncertainty may often usefully be analyzed into individual components.

4.2 There are several aspects of uncertainty characterization specific to air quality measurements. One of these aspects concerns known, that is, correctible, systematic error or mean bias of a measurement relative to a true measurand value. Several measurement methods exist with such bias left uncorrected because of policy, tradition, or other reason. *Uncertainty* deals only with what is unknown about a measurement, and as such does not include correctible (known) bias. The magnitude of the difference between estimate and measurand value is covered by *accuracy* as defined qualitatively in ISO GUM, rather than *uncertainty*, particularly when the bias is known, but uncorrected. Such methods require specification of both uncertainty and as much as is known of the uncorrected bias, or alternatively the adoption of an accuracy measure.

4.3 Often bias is known to exist, but with unknown value. In the case where only limits may be placed on the magnitude of the bias, ISO GUM generally recommends treating the bias as uniformly distributed within the known limits. Such a distribution refers to independent situations, for example, calibrations, where bias may arise (see 7.4 and Appendix X2), rather than variation at the point of method application. Even though such an equal-likelihood bias distribution may be unrealistic, nevertheless a standard deviation estimate may be made that reveals the limits on the bias. If the even-distribution approximation is clearly invalid for a relevant set of measurements, the procedure may be adjusted slightly by adopting an accuracy measure tailored to the assumed limits.

4.4 Another issue concerns the distribution of measurements. ISO GUM deals only with normally distributed first-order (that is, "small") variations relative to measurand values. An example to the contrary is afforded by normally distributed data confounded by a small number of apparent outliers (**3**), which may not detract from the method performance (see Appendix X4 for details). Another example is the determination of an aerosol concentration at one location (perhaps at a worker's lapel) as an estimate of the concentration at a separate point (such as a breathing zone). In this case the variations can be of the order of the estimate itself and may have the character of a log-normal distribution.

4.5 The spatial inhomogeneity alluded to in 4.4 relates to another point regarding the focus of this practice. The spatio-temporal variations in air quality characteristics are generally so large **(4)** as to preclude evaluation of a method during application through the use of replicate measurements. In this case, often an initial single method evaluation is undertaken with the purpose of determining uncertainty present in subsequent applications of the method. Confidence in such an evaluation can be specified and relates to the concept of *prediction-intervals* **(5)** (see 7.2).

4.6 A related subject is measurement system control. The measurement system must remain in a state of statistical control if an introductory evaluation is to characterize later practical applications of the method. Measurement system control is evaluated using an ongoing quality control program, testing critical performance aspects for detecting problems which may develop in the method.

5. Summary of Practice

5.1 The essential idea behind ISO GUM is the *analysis* to the fullest extent practical of the elemental sources of what is unknown in the estimate of a measurand value. This contrasts with a *global* or *top-down* determination of uncertainty, which could for example be done ideally by comparing replicate estimates to known measurand values over all conditions expected in application of the method. Although a global uncertainty evaluation may sometimes seem inexpensive, there is a difficulty in covering essential contingencies of the method application.

5.2 Uncertainty component analysis further has several specific advantages over global analysis. The results may be applicable to a variety of situations. For example, an aerosol sampler might be (globally) evaluated as to particle-size-dependent error by side-by-side comparison to a reference sampler in several coal mines. The knowledge obtained may not be as easily applied for sampler use in iron mines, for example, as more detailed information on how the sampler performs over given dust size distributions may be needed. Furthermore, specific problem areas of a given method may be pinpointed. The detailed itemization of uncertainty sources leads to a transparency in covering the essential problems of a measurement method. Examples of potentially significant uncertainty components are listed in Table 1.

5.3 *Type A and B Uncertainty Components:*

5.3.1 Components that have been statistically evaluated during method application may be classified as Type A. (See Section 7 for specific examples.)

5.3.2 Some components are often statistically evaluated during an initial method evaluation, rather than at application. Also acknowledged is a common situation that components may not have been characterized in a statistically valid manner and therefore may require professional judgment for itemizing. Such components are termed Type B uncertainties. Type B uncertainties are often associated with unknown systematic error or bias; however, random variation may also fall into this category. For example, a common assumption (see, for example, EN 482) regarding personal sampling in the workplace is that the relative standard deviation associated with personal sampling pump variations is <5 % at essentially 100 % confidence.

TABLE 1 Common Potential Uncertainty Components

Sampling
personal sampling pump flow rate: setting the pump and subsequent drift
sampling rate of diffusive sampler
sampler dimension (aerosol and diffusive sampling)
collection efficiency of a sampler or sampling medium (also, see **(6)**)
Analytical
aerosol weighing
recovery (for example, chromatographic or spectroscopic methods)
Poisson counting (for example, in XRD methods)
instrument or sensor variation
operator effects giving inter-lab differences (if data from several labs are to be used)
Sample
sample stability
sample preparation (for example, handling silica quasi-suspensions)
sample loss during transport or storage
Evaluation
calibration material uncertainty
evaluation chamber concentration uncertainty
other bias-correction uncertainty
Environmental Influence Parameters
temperature (inadequacy of correction, if correction is made as with diffusive samplers)
atmospheric pressure
humidity
aerosol size distribution (if not measured by a given aerosol sampling method)
ambient wind velocity
sampled concentration magnitude itself (for example, sorbent loading)

5.4 *Intrinsic versus Environmentally Associated Components: Influence Quantities:*

5.4.1 Some uncertainties may be intrinsic to a method. For example, estimates from aerosol samplers may depend critically on sampler dimensions, which if variable leads to intersampler estimate variation.

5.4.2 On the other hand, a sampler's performance may depend on the environment. For example, suppose a sampler is sensitive to temperature changes that are impractical to measure in the field; that is, sampler estimates are not temperature-corrected. Then measurement of this sensitivity during method evaluation together with knowledge of the temperature variation expected for a given field application can be used to determine the uncertainty associated with this effect.

5.4.3 A quantity such as the temperature is known as an *influence quantity*. A common example where influence variables are important involves diffusive monitors, where wind velocity, temperature, pressure, and fluctuating workplace concentrations can affect diffusive monitor uptake rates (Practice D6246, ISO 16107).

5.4.4 Situations exist for which the distribution of an influence quantity is unknown. For example, the deviation between aerosol concentration estimates and samples taken according to accepted convention (for example, ISO 7708) generally depend on the aerosol size distribution sampled. Only limits on the distribution of size distributions (the influence quantity) may be known. In this case, the ISO GUM approach is generally to assume a uniform distribution (see 7.4).

5.4.5 On the other hand, the size distribution may be known to be constant over a set of measurements. In this case, the constant-distribution assumption leads to an abstract performance characterization. Alternatively, a quantity known as the *symmetric accuracy range A* (Appendix X1 and Section X4.2) in the case of unknown, but large limited |*bias*|, may be used to establish intervals bracketing the (true) values of measurand and thus represents the *expanded uncertainty*.

5.5 *Combined and Expanded Uncertainty*—The essential ISO GUM approach then is to obtain estimates u_j of the standard deviation (often designated as *s* as computed on most handheld calculators) associated with the *j*th uncertainty source. The estimates u_j may be designated as *uncertainty components*. Then if the sources are independent, that is, if the variations are uncorrelated, a *combined standard uncertainty* u_c estimating the net standard deviation may be computed as:

$$u_c = \sqrt{\sum_j u_j^2} \quad (1)$$

5.5.1 Finally, an *expanded uncertaintyU* is calculated at *coverage factor k* as:

$$U = k \cdot u_c \quad (2)$$

5.5.2 The purpose of the expanded uncertainty *U* is to bracket the unknown measurand value (for example, unknown mass *M*) given an estimate *m*. For example, a coverage factor could be selected so that:

$$m - U < M < m + U \quad \text{for 95 \% of estimates } m \text{ of measurand value } M \quad (3)$$

5.5.3 However, this practice suggests use of the nearly traditional value $k = 2$, permitting the meaning in terms of confidence levels to float.

6. Significance and Use

6.1 A primary use intended for this practice is for qualifying ASTM International Standards as Standard Test Methods. In the past, a “Precision and Bias” report has been required. However, recently a statement of uncertainty has become an acceptable alternative to D3670 – 91: Guide for Determination of Precision and Bias of Methods of Committee D22. Inclusion of such a statement with a method description simplifies comparison of ASTM Test Methods to analogous ISO and CEN standards, now required to have uncertainty statements.

6.2 Standardizing the characterization of sampling/analytical method performance is expected to be useful in other applications as well. For example, performance details are a necessity for justifying compliance decisions based on experimental air quality assessments (**7**). Documented uncertainty can form a basis for specific criteria defining acceptable sampling/analytical method performance.

6.3 Furthermore, high quality atmospheric measurements are vital for making decisions as to how hazardous substances are to be controlled. Valid data are required for drawing reasonable epidemiological conclusions, for making sound decisions as to acceptable limits, as well as for determining the efficacy of a hazard control system.

6.4 Finally, because of developing world-wide acceptance of ISO GUM for detailing measurements when statistics are simple, the practice should be useful in comparing ASTM International Test Methods to others’ published methods. The codification of statistical procedures may in fact minimize the difficulty in interpreting a plethora of individual, albeit possibly valid, approaches.

7. Specific Examples

NOTE 1—Some of the above concepts can be illuminated through example. Application to more complicated situations is then possible.

7.1 *Standard Deviation σ Known Exactly:*

7.1.1 Suppose the method yields unbiased estimates *m* in measuring unknown *M* so that:

$$m = M + M \cdot \varepsilon \quad (4)$$

where ε is normally distributed about 0 with known standard deviation σ, sometimes designated the *true relative standard deviation TRSD*. For example, suppose the method has been evaluated with essentially an infinite number of measurements of a calibration standard, giving a tight estimate of σ. Then estimates *m* are distributed normally about *M* so that:

$$M - 1.960 \times M \cdot \sigma < m < M + 1.960 \times M \cdot \sigma \quad \text{at probability} = 95\ \% \quad (5)$$

7.1.2 Thus, to first order in σ, the true value *M* is bracketed by:

$$m - 1.960 \times m \cdot \sigma < M < m + 1.960 \times m \cdot \sigma \quad \text{at probability} = 95\ \% \quad (6)$$

7.1.3 Therefore, the (relative) *expanded uncertainty U* would be consistent with Eq 3, if the coverage factor *k* is chosen as:

$$k = 1.960 \quad (7)$$

as a factor of *combined standard uncertainty* u_c:

$$u_c = \sigma \quad (8)$$

in other words:

$$U = 1.960 \times \sigma \quad (9)$$

7.1.4 Eq 7 is consistent with the traditional selection $k = 2$.

NOTE 2—Although the measurement variation depicted in Eq 4 is very common in air quality measurements, at decreasing values of *M*, generally a constant variation (that is, independent of *M*) becomes significant, leading to non-zero limits of quantitation and detection. (See, for example, ISO 15767 and Practice D6552.)

7.2 *Standard Deviation σ Estimated Initially by n Replicates (Type B Uncertainty):*

7.2.1 Almost as simple as 7.1 is the situation in which a (relative) standard deviation estimate *s* is obtained through an initial *n* measurements of a calibration standard prior to the method’s multiple subsequent uses without re-evaluation. Variations of this situation are common in air quality measurement. For example, diffusive samplers may be evaluated initially by a vendor followed by many applications without re-evaluation (see ISO 16107 or Practice D6246). Suppose Eq 4-6 still hold, except that that now σ is unknown but is estimated by s with $\upsilon = n - 1$ degrees of freedom. What *is* known is that σ is limited by:

$$\sigma < (\upsilon/\chi^2_{\upsilon\ 0.05})^{1/2} s \tag{10}$$

at 95 % confidence in the evaluation/calibration experiment, where $\chi_\upsilon{}^2{}_{0.05}$ is the chi-square 5 % quantile at υ degrees of freedom (obtainable from statistics tables or programs). Therefore, at 95 % confidence in the evaluation, the unknown M is bracketed by:

$$m - 1.960(\upsilon/\chi^2_{\upsilon\ 0.05})^{1/2} \times m \cdot s < M < m + 1.960(\upsilon/\chi^2_{\upsilon\ 0.05})^{1/2} \times m \cdot s \tag{11}$$

for greater than 95 % of measurements.

7.2.2 In this case, the combined (relative) uncertainty u_c is:

$$u_c = s \tag{12}$$

but if the meaning of Eq 3 is sought, the coverage k factor in Eq 11 is now:

$$k = 1.960(\upsilon/\chi^2_{\upsilon\ 0.05})^{1/2} \tag{13}$$

7.2.3 In Fig. 1 the coverage factor k of Eq 13 is plotted versus degrees of freedom υ and is seen to approach 1.960 as $\upsilon \rightarrow \infty$ corresponding to 7.1. However, Fig. 1 indicates that over a wide range of degrees of freedom adopted in practical method evaluations, k is of the order of 3 in order to achieve 95 % evaluation confidence.

NOTE 3—Specification of an evaluation confidence level together with coverage probability (both taken here to equal 95 %) relates to the statistical theory of tolerance or prediction intervals (**5**).

7.3 *Continual Method Evaluation (Type A Uncertainty):*

7.3.1 Preferred, though often not practical in air quality measurements, is an n-measurement calibration giving an estimate s for σ with $\upsilon = n - 1$ degrees of freedom *every time* a practical method is applied. Then it is possible to show that the true value M is bracketed by:

$$m - t_{\upsilon\ 0.975} \times m \cdot s < M < m + t_{\upsilon\ 0.975} \times m \cdot s \quad \text{at probability} = 95\ \% \tag{14}$$

where $t_{\upsilon\ 0.975}$ is the student-t 97.5 % quantile at υ (also found in statistical sources).

7.3.2 Therefore the coverage factor k is now given by:

$$k = t_{\upsilon\ 0.975} \tag{15}$$

7.3.3 In Fig. 1 this coverage factor is plotted versus the number υ of degrees of freedom in the evaluation. As can be seen from the figure, with continual method evaluation, the coverage factor is close to 2 over a range of values for υ. In fact, this is the reason behind the now nearly traditional use of the value 2 for the coverage factor.

7.3.4 The use of the traditional coverage factor = 2 simply gives intervals bracketing the unknown measurand with interpretation specific to the measurement circumstances. Of course, as alluded to in Section 3, if U is actually reported with the traditional coverage factor 2, then, if needed, an expanded uncertainty with 95 % evaluation confidence is easily obtained by multiplication (by about 3/2).

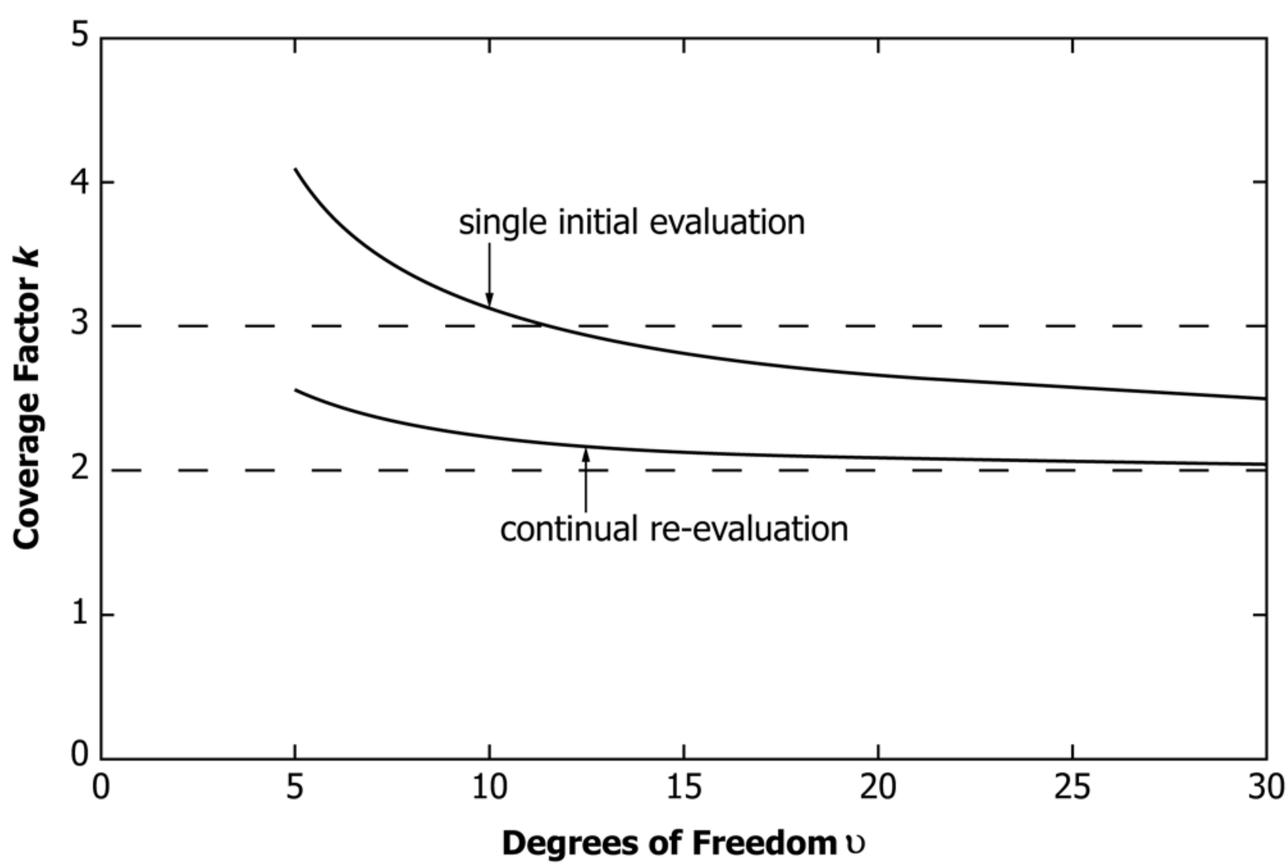

FIG. 1 Comparison of Coverage Factors *k* for Single Initial Method Evaluation versus Continual Evaluation with υ Degrees of Freedom

7.4 *Uncertainty Characterization of Unknown Bias or Systematic Error (Type B Uncertainty):*

7.4.1 Unknown systematic error or bias in a measurement may originate in several ways. For example, if a method is not re-calibrated at each application, then error from the finiteness of an initial calibration may be present as a non-random variable in subsequent applications. Even if re-calibrated, bias may result from repeated use of a reference material or method, itself with unknown bias. In either case, the uncertainty component corresponding to uncertain method bias may be taken as the uncertainty in the bias itself.

7.4.2 *Reference Uncertainty*—As an example, suppose a method is repeatedly calibrated by a reference method that is itself biased, though is negligibly variable (as example). Then the estimated mass in measuring unknown M may be represented as:

$$m = M(1+\Delta_{ref})+M\cdot\varepsilon \tag{16}$$

where the standard deviation estimate s for the normally distributed random variable ε may be obtained from the calibration experiment, and where Δ_{ref} is the unknown bias of the reference method. (See Appendix X2 for details on a similar situation, including finite-calibration bias.)

7.4.2.1 Suppose that all that is known about the reference bias Δ_{ref} is that it is bounded by a constant positive quantity Δ_{max}, often a matter of judgment, so that:

$$|\Delta_{ref}| < \Delta_{max} \tag{17}$$

7.4.2.2 ISO GUM generally suggests handling this situation by approximating (evaluation to evaluation) Δ_{ref} as uniformly distributed between $\pm\Delta_{max}$. Then it is simple to compute an inter-evaluation variance as:

$$Var[\Delta_{ref}] = \frac{1}{3}\Delta_{max}^2 \tag{18}$$

7.4.2.3 Δ_{max} characterizes a shortcoming in the method evaluation, as does an imperfect initial determination of σ, the standard deviation of ε (see 7.2). Thus, assuring confidence (for example, 95 %) in the calibration with the same (prediction or tolerance) sense as in 7.2, a coverage factor k can be selected so that an expanded uncertainty given by:

$$U = k\sqrt{\frac{1}{3}\Delta_{max}^2+s^2} \tag{19}$$

brackets the unknown M for a high fraction (for example, 95 %) of measurements (see Appendix X2 for details).

7.4.2.4 In other words, the uncertainty component u_{Δ} for the bias is:

$$u_{\Delta} = \sqrt{\frac{1}{3}\Delta_{max}^2} \tag{20}$$

and again the random uncertainty component u_{random} is:

$$u_{random} = s \tag{21}$$

with combined uncertainty u_c given by:

$$u_c = \sqrt{u_{\Delta}^2+u_{random}^2} \tag{22}$$

and:

$$U = k\,u_c \tag{23}$$

7.4.3 *Finite-Calibration Uncertainty*—Similarly to 7.4.2, correcting bias by a *single n*-mean estimate m_{ref} of a reference mass M_{ref} (again with estimated corrected standard deviation s) and then calibrating subsequent application measurements by a calibration factor M_{ref}/m_{ref} leads to an uncertainty component u_n given by:

$$u_n = s/n^{1/2} \tag{24}$$

7.4.3.1 In this case the two components u_n and u_{random} are not independent if u_{random} is given by Eq 21.

NOTE 4—If an *n*-measurement calibration is effected at *each* application measurement, then the value in Eq 24 still appears as part of the calibration uncertainty, but now refers to a random rather than systematic variation.

7.4.4 *Large Bias Magnitude of Unknown Sign*—There are examples in air quality measurement where the range of unknown bias may be large relative to the variable components of uncertainty. For example, aerosol samplers used for measuring dust concentrations according to one of the international sampling conventions (ISO 7708), for example, respirable, thoracic or inhalable, generally differ in particle-size acceptance from convention. Therefore, in sampling a particular site with aerosol of unknown particle size distribution range, an unknown and sometimes large bias relative to convention is possible.

7.4.4.1 With large bias magnitude, the ISO GUM approach of 7.4.2 of combining uncertainty components squared may be replaced by a linear combination of bias magnitude uncertainty and variability uncertainty. On the other hand, the usual ISO GUM approach (with coverage factor k = 2) gives similar uncertainty values. For an example, see Section X2.3.

7.5 *Analysis of a Round Robin Evaluation:*

7.5.1 Analysis of a specific round robin evaluation of a measurement method as applied by several independent labs is presented here, illustrating features of uncertainty characterization. Suppose that the overall method bias magnitude, though unknown, is likely smaller (as can be decided if necessary through a student-t test) than correctible by the round robin itself because of the size of the inter-lab variation and the small number of labs tested. Appendix X1 indicates that in this case, the bias uncertainty component can be taken conservatively as the bias estimate itself. Uncertainty components estimated include u_{intra} characterizing within-lab method variability, u_{inter} for the variability between labs, and u_{bias} for overall bias of the labs (averaged together) relative to spiked or reference values.

7.5.2 Following the ISO GUM principle of analysis of uncertainty into elemental sources, the value u_{intra} obtained for the within-lab uncertainty may sometimes then be expressed in terms of its own individual components as exemplified in (7.1 – 7.4). Often, however, such a breakdown may not be entirely understood. In this case, a comparison of u_{intra} to an elemental analysis may yet be useful.

7.5.3 Calculation of the various uncertainty components is presented informally here, since depiction of the data graphically leads to intuitive interpretation. More information is given in Appendix X3 for those interested. The details of the round robin are somewhat simplified for illustration, but may be modified for other designs. For example: a larger number of labs may take part than considered here, resulting possibly in a useful method bias correction; variation between labs' internal method uncertainty may be significant and characterized; and uncertainty in reference material may be accounted for.

7.5.4 *Assumptions:*

7.5.4.1 L (for example, 6) labs take part in the round robin.

7.5.4.2 S (for example, 6) spiked samples are sent to each lab.

7.5.4.3 The true relative standard deviation σ_{ref} of the spiked values is assumed negligible.

7.5.4.4 Similarly, bias in the preparation of the spiked values is assumed negligible.

7.5.4.5 The six labs are assumed to have similar (within-lab) variability.

7.5.5 *Data and Analysis:*

7.5.5.1 Suppose each of $L = 6$ labs is presented with samples prepared with spiked amounts of an analyte as shown in Table 2.

7.5.5.2 Measurements from the six labs corresponding to the six samples are shown in Table 3.

7.5.5.3 The error of each of the measurements in Table 3 relative to the reference values of Table 2 is easily computed and is depicted in Table 4.

7.5.5.4 The variability within each lab can be estimated with $S - 1$ degrees of freedom by computing the estimated variance (the standard deviation squared) within each row of Table 4. The result is shown in Table 5.

7.5.5.5 Also shown in Table 5 is the mean, with $L \times (S - 1)$ degrees of freedom, of the six lab variances, whose square root represents the within-lab uncertainty component u_{intra}.

7.5.5.6 Returning now to Table 4, the bias of each of the labs relative to reference can be estimated by averaging within each row. The result is presented in Table 6.

7.5.5.7 The standard deviation, with $L - 1$ degrees of freedom, of the six lab values in Table 6 is also shown together with the lab biases averaged together. These two numbers represent the interlab uncertainty u_{inter} (neglecting the averaged out intra-lab variation) and bias uncertainty u_{bias} components, respectively.

7.5.5.8 Finally, the combined (relative) uncertainty u_c may be computed as:

$$u_c = \sqrt{u_{bias}^2 + u_{inter}^2 + u_{intra}^2} = 13.5\ \% \tag{25}$$

7.5.5.9 Adopting the traditional coverage factor $k = 2$ (though 95 % evaluation confidence $\Rightarrow k = 3.4$), the expanded uncertainty U is:

$$U = k \cdot u_c = 27\ \% \tag{26}$$

7.5.5.10 These results may be summarized as in Table 7.

8. Reporting Uncertainty

8.1 The following should appear in reports documenting method uncertainty. As examples, see ISO 16702 regarding diffusive sampling or **(8)** concerning specifically stack monitoring.

8.1.1 List of sources corresponding to the uncertainty components estimated.

8.1.2 Values of the uncertainty components together with the number of degrees of freedom involved in their measurement.

8.1.3 Statement of estimate sensitivity to and distribution of significant influence quantities.

8.1.4 Classification of components in terms of Type A or Type B uncertainty estimates.

8.1.5 Estimation of bias or a statement of its assumed negligibility.

8.1.6 Statement as to how bias was minimized if a correction has been effected.

8.1.7 Value of the combined standard uncertainty u_c.

8.1.8 Value of the expanded uncertainty U, together with the value taken for the coverage factor k.

9. Keywords

9.1 air quality; characterizing uncertainty; measurements

TABLE 2 Spiked Samples (µg)

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
1.00	1.00	2.50	2.50	5.00	5.00

TABLE 3 Mass Measurements (µg) from 6 Participating Labs

Lab	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
1	1.044	1.132	2.785	2.653	5.663	6.057
2	0.932	1.062	2.463	2.659	4.844	5.154
3	0.944	0.904	2.152	2.068	4.369	4.436
4	0.954	0.959	2.370	2.227	4.254	4.901
5	1.298	1.196	2.974	3.168	6.046	5.781
6	1.016	1.063	2.616	2.444	5.105	5.176

TABLE 4 Error (Fractional Discrepancy) of Measurements Relative to Spiked Reference Samples

Lab	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
1	0.044	0.132	0.114	0.061	0.133	0.211
2	–0.068	0.062	–0.015	0.063	–0.031	0.031
3	–0.056	–0.096	–0.139	–0.173	–0.126	–0.113
4	–0.046	–0.041	–0.052	–0.109	–0.149	–0.020
5	0.298	0.196	0.189	0.267	0.209	0.156
6	0.016	0.063	0.047	–0.022	0.021	0.035

TABLE 5 Values of the Variance Within Each Lab and Resulting Intra-Lab Uncertainty Component

Lab	Variance Within Lab
1	0.0036
2	0.0029
3	0.0016
4	0.0024
5	0.0028
6	0.0009
Mean variance within:	0.00235
√ → Std. Dev. within:	u_{intra} = 0.048

TABLE 6 Mean Lab Error (from Table 4) whose Standard Deviation Represents the Inter-Lab Uncertainty Component and whose Mean Overall Bias Gives the Bias Uncertainty Component

Lab	Mean Bias for Each Lab
1	0.116
2	0.007
3	–0.117
4	–0.069
5	0.219
6	0.026
Std. Dev. between labs:	u_{inter} = 0.123
Mean overall bias:	u_{bias} = 0.030

TABLE 7 Measurement Uncertainty Summary

Source	Uncertainty Component	Degrees of Freedom	Type
inter-lab variation	12.3%	5	A
intra-lab variation	4.8%	30	A
bias	3.0%	...	A
Combined standard uncertainty u_c = 13.5 %			
Coverage factor k = 2.0			
Expanded uncertainty U = 27.0 %			

APPENDIXES

(Nonmandatory Information)

X1. UNCERTAINTY AND THE SYMMETRIC ACCURACY RANGE *A*

X1.1 Definition

X1.1.1 ISO GUM defines *accuracy* qualitatively (see Section 3) in terms of the closeness between measurement and (true) measurand value. Accuracy is therefore broader than uncertainty in that accuracy may reflect a *known* bias between measurement and measurand value. In contrast, uncertainty focuses on what is *unknown*.

X1.1.2 As intimated within ISO GUM, there is no uniquely useful way to quantify accuracy. However, several specific quantifications are in use. One of these, the *symmetric accuracy range* A, is closely tied mathematically to uncertainty, and has been applied (**2**) for evaluating candidate workplace atmospheric concentration measurement methods and for documenting the uncertainty of methods in their application. A is defined as the fractional range, symmetric about (for example) the true mass M, within which 95 % of sampler measurements m are to be found.

X1.1.3 Explicitly:

$$M \times (1 - A) < m < M \times (1 + A) \quad \text{for 95\% of measurements } m \tag{X1.1}$$

X1.1.3.1 This definition implies that the accuracy range function A must increase with both random effects described by σ_m (the true relative standard deviation) and the estimate's mean *bias*, both expressed relative to the (unknown) true value M. These feature can be seen in Fig. X1.1, where A, in the case of normally distributed estimates, is plotted as a set of contours.

X1.1.4 With estimates m approximately normally distributed, the accuracy range A is accurately given (**2**) simply by:

$$A = 1.960 \times \sqrt{bias^2 + \sigma_m^2} \quad \text{if } |bias| \text{ is small (that is, } |bias| < \sigma_m / 1.645) \tag{X1.2}$$

$$A = |bias| + 1.645 \times \sigma_m \text{ if } |bias| \text{ is large} \tag{X1.3}$$

X1.2 Collapse of the Accuracy Measure to the Expanded Uncertainty *U*

X1.2.1 If only a (preferably small) unknown measurement bias exists, then measurement uncertainty may be specified in terms of the accuracy range confidence limit $A_{95\ \%}$, accounting for method evaluation error. Suppose the confidence limit $A_{95\ \%}$ << 100 %, then Eq X1.1 can be reexpressed so that for 95 % of all method validations,

$$m - (m \times A_{95\%}) < M < m + (m \times A_{95\%}) \text{ for} > 95\% \text{ of the estimates } m \tag{X1.4}$$

X1.2.2 Thus, $A_{95\ \%}$ provides intervals bracketing the true mass M—with the *double confidence* sense of a guaranteed-tolerance prediction interval (**5**), that is, confidence (95 %) in the method validation and confidence (also taken here equal to 95 %) in the ultimate application. Therefore from the point of view of bracketing (Eq 3) the measurand value M, the expanded uncertainty U (relative to the estimate m) may be taken to equal:

$$U = A_{95\ \%} \tag{X1.5}$$

if 95 % evaluation confidence is required of intervals bracketing true measurand values.

X1.3 Applications of the Accuracy Range *A* Beyond Uncertainty

X1.3.1 If *bias* is not corrected, then the accuracy range A goes beyond what is uncertain. Nevertheless, A still has

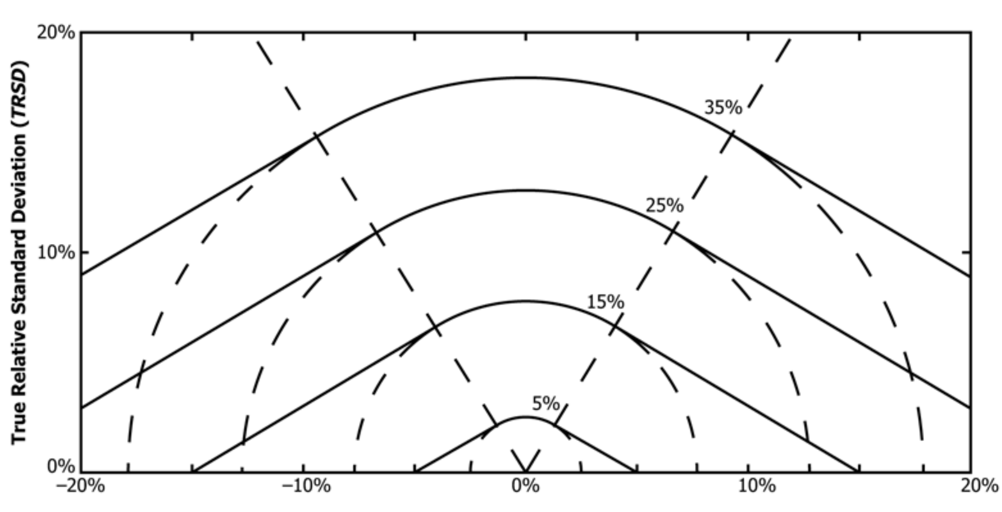

FIG. X1.1 Contours of Constant Accuracy Range *A* versus Bias and (True) Relative Standard Deviation

applications. It can be used to set up performance criteria for deciding if a given method can be used without further correction.

X1.3.2 For example, for a given aerosol assessment application, there may exist a number of different types of samplers. With a range of aerosol sizes to be sampled, the samplers are generally biased relative to each other and to any ideal sampling convention (ISO 7708). Bias relative to a sampling convention, in fact, may dominate the errors in the use of a particular sampler type. Furthermore, correction of this bias in the mean may be unacceptable because of historical or other reasons. Use of the function A makes it possible to state limits on the accuracy of a measurement procedure, even though one cannot know the actual accuracy if the aerosol size distribution is not determined.

X2. UNCERTAIN BIAS

X2.1 Details are presented here on determining or setting a limit of confidence in an initial method calibration with two independent sources of bias. Suppose a method is calibrated by a finite comparison to a possibly biased but otherwise accurate reference-method mass m_{ref} in measuring an unknown constant mass M_{cal}:

$$m_{ref} = M_{cal}\left(1+\Delta_{ref}\right) \quad \text{(X2.1)}$$

where the bias magnitude $|\Delta_{ref}| << 1$.

X2.1.1 During calibration, the method, biased by unknown Δ_{raw}, yields a number n of uncorrected (raw) estimates m_{raw} so that:

$$m_{raw} = M_{cal}\left(1+\Delta_{raw}\right)+M_{cal}\cdot\varepsilon_{raw} \quad \text{(X2.2)}$$

X2.1.2 The random variable ε_{raw} is normally distributed about 0 with unknown variance $\sigma_{raw}{}^2$, but is estimated by $s_{raw}{}^2$ (with $n - 1$ degrees of freedom).

X2.1.3 Therefore, an estimate for Δ_{raw} is given in terms of the n-sample mean $\bar{m}_{raw}$ by:

$$_{est}\Delta_{raw} = \left(\bar{m}_{raw} - m_{ref}\right)/m_{ref} \quad \text{(X2.3)}$$

$$\approx\Delta_{raw} - \Delta_{ref}\left(1+\Delta_{raw}\right)+\bar{\varepsilon}_{raw}$$

after expanding in Δ_{ref} (but not in Δ_{raw}, whose magnitude is not necessarily small relative to 1.0).

X2.1.4 Suppose that all that is known about the reference method bias Δ_{ref} is that its magnitude is bounded by Δ_{max}, and, as suggested in 7.4, that Δ_{ref} may be taken as uniformly distributed (calibration-to-calibration) between its limits. Then the expected value of $\Delta_{ref}{}^2$, the (intercalibration) variance $\sigma_{\Delta_{ref}}{}^2$ of Δ_{ref}, is given by:

$$\sigma^2_{\Delta_{ref}} = E\left[\Delta^2_{ref}\right] \quad \text{(X2.4)}$$

$$=\frac{1}{3}\Delta^2_{max}$$

X2.1.5 Following an initial calibration suppose the method is applied as in 7.2 without re-calibration, but with bias partially eliminated by correcting raw measurements using a factor $m_{ref}/\bar{m}_{raw}$ equal to $(1 + {}_{est}\Delta_{raw})^{-1}$.

X2.1.6 Then in measuring unknown mass M, the corrected measurement value m is given by:

$$m = \frac{M\left(1+\Delta_{raw}\right)+M\cdot\varepsilon_{raw}}{1 + {}_{est}\Delta_{raw}} \quad \text{(X2.5)}$$

$$=M\left(1+\Delta\right)+M\cdot\varepsilon$$

where the corrected bias Δ and true relative standard deviation are given by:

$$\Delta = \frac{\Delta_{raw} - {}_{est}\Delta_{raw}}{1 + {}_{est}\Delta_{raw}} \quad \text{(X2.6)}$$

$$=\Delta_{ref} - \bar{\varepsilon}_{raw}/\left(1 + {}_{est}\Delta_{raw}\right) \quad \text{(from Eq. X2.3)}$$

$$\equiv\Delta_{ref}+\Delta_n$$

and:

$$\sigma = \frac{\sigma_{raw}}{1 + {}_{est}\Delta_{raw}} \quad \text{(X2.7)}$$

X2.1.7 Then the small-bias limit (Eq X1.2) for the symmetric accuracy range A is given by:

$$A^2/1.960^2 = \left(\Delta_{ref}+\Delta_n\right)^2+\sigma^2 \quad \text{(X2.8)}$$

with two independent sources of bias, Δ_{ref} and Δ_n, the latter normally distributed with variance σ^2/n. Eq X2.4 and Eq X2.6 indicate that A may be estimated from:

$$A^2_{est}/1.960^2 = \frac{1}{3}\Delta^2_{max}+\frac{1}{n}s^2+s^2 \quad \text{(X2.9)}$$

$$=u^2_c, \quad \text{the combined standard uncertainty}$$

X2.1.8 Determining a coverage factor k by demanding 95 % confidence in the calibration then requires a 95 %-confidence limit $A_{95\ \%}$, which may be determined by a chi-square approximation (similar to the Smith/Satterthwaite/Welch procedure in Refs 9-8):

$$\upsilon_{eff}\frac{A^2_{est}}{A^2} = \upsilon_{eff}\frac{\frac{1}{3}\Delta^2_{max}+\left(\frac{1}{n}+1\right)s^2}{\left(\Delta_{ref}+\Delta_n\right)^2+\sigma^2} \quad \text{(X2.10)}$$

$$\approx\chi^2_{\upsilon_{eff}}$$

X2.1.9 The effective number υ_{eff} of degrees of freedom is calculated by means of propagation of errors by requiring that the variances of the two lines in Eq X2.10 agree, noting that both numerator and denominator (requiring estimates of the kurtoses of Δ_{ref} and Δ_n) are variable in this case. The result is:

$$\upsilon^{-1}_{eff} = u_c^{-4}\left[\frac{2}{45}\Delta^4_{max}+\frac{2}{3}\Delta^2_{max}s^2+\left[\upsilon^{-1}\left(1+\frac{1}{n}\right)^2+\frac{1}{n^2}\right]s^4\right] \quad \text{(X2.11)}$$

X2.1.10 Finally, $A_{95\ \%}$ or the expanded uncertainty U is given by:

$$U = k\cdot u_c \quad \text{(X2.12)}$$

$$k = 1.960 \cdot \sqrt{\upsilon_{eff}/\chi^2_{\upsilon_{eff}\,0.05}} \quad (X2.13)$$

X2.2 In the case of continual calibration giving a running value s with υ degrees of freedom, yet with bias uncertainty residual from an initial evaluation, again the combined and expanded uncertainties are given as in Eq 22 and 23, but now with coverage factor k given in terms of an effective number υ_{eff} of degrees of freedom by:

$$k = t_{\upsilon\ 0.975} \cdot \sqrt{\upsilon_{eff}/\chi^2_{\upsilon_{eff}\,0.05}} \quad (X2.14)$$

for expressing 95 % confidence in the initial evaluation.

X2.2.1 Note that requiring only *mean* (rather than 95 %) confidence in the initial evaluation eliminates the square root factor from Eq X2.14, resulting in Eq 15.

X2.3 In the case of large bias magnitude, linearity replaces the root square combination of components (see Appendix X1 and particularly, Eq X1.3) for providing estimates of coverage. However, for consistency with ISO GUM, the small-|*bias*| branch of Eq X1.2 may be taken even in the case of large |*bias*|, in which case the expanded uncertainty U proportional to sums of uncertainty components squared becomes an ordering parameter, rather than providing intervals quantitatively bracketing measurand values in a simple way. An example will show the slight differences. Suppose as in 7.4.1 the bias Δ is only known to be uniformly distributed (for example, assessment site-to-site) between $\pm\Delta_{max}$. Suppose the true relative standard deviation (*TRSD*) is well-known and small (as example). Then the symmetric range accuracy is given by:

$A = |\Delta| + 1.645\ TRSD$
$A_{est} = \Delta_{max}/2 + 1.645\ TRSD$
$A_{95\ \%} = 0.95\ \Delta_{max} + 1.645\ TRSD$

X2.3.1 Therefore, taking $A_{95\ \%}$ as the combined uncertainty U:

$U = 0.95\ \Delta_{max} + 1.645\ TRSD$
linear in Δ_{max} and *TRSD*

X2.3.2 This compares to the now common ISO GUM approach:

$u_c = [⅓\ \Delta_{max}{}^2 + TRSD^2]^{1/2}$
$U = k\ u_c$
$k = 2$

X2.3.3 As a specific example, suppose Δ_{max} = 50 % and *TRSD* = 7.5 %, then:

U = 59.8 % (linear)
U = 59.6 % (root sum of squares)

X2.3.4 The results are fortuitously close. However, it should be remembered if necessary that the meaning of the more rigorous (linear) result is that for 95 % of sites visited, the expanded uncertainty covers true values of the aerosol concentration at greater than 95 % of measurements, whereas the rote use of k = 2 does not have any specific meaning in this case.

NOTE X2.1—With U so large as in this example, the expansion leading (for example) from Eq 5 to Eq 6 does not converge very well, and asymmetric intervals about true measurand values may be more useful in the case that $A_{95\ \%}$ is known well:

$$M \times (1 - A_{95\ \%}) < m < M \times (1 + A_{95\ \%})$$
$$\Downarrow$$
$$m / (1 + A_{95\ \%}) < M < m / (1 - A_{95\ \%})$$

X3. EXAMPLE: ROUND-ROBIN UNCERTAINTY ANALYSIS

X3.1 Model

X3.1.1 Mathematical details are given here for slightly more general conditions than given in (7.5). Assume that estimated and unknown reference (that is, spiked) masses m and $_{ref}m$ (independent quantities biased relative to nominal spiked values $_{nom}m$) can be described by the following:

$$m_{sl} / {}_{ref}m_{sl} = (1 + bias) + {}_{intra}\varepsilon_{sl} + {}_{inter}\varepsilon_{l} \quad (X3.1)$$

$$_{nom}m_{sl} / {}_{ref}m_{sl} = (1 - bias_{ref}) - {}_{ref}\varepsilon_{sl}$$

where the εs are normal, 0-expection value, random variables with intra- and inter-lab variances $\sigma_{intra}{}^2$ (approximated here as lab-independent), $\sigma_{inter}{}^2$, and $\sigma_{ref}{}^2$ (the latter assumed known), respectively.

X3.1.2 The quantity *bias* represents the bias of the candidate method relative to (unknown) measurand values. All the σs are *total relative standard deviations* (*TRSD*s relative to measurand values). The samples are labeled by $s = 1, \ldots, S$, and $l = 1, \ldots, L$, where L is the number of participating labs. Thus, the relative discrepancy *error* between estimates m and $_{nom}m$ and with variance denoted as σ^2 is:

$$error_{sl} \equiv (m_{sl} - {}_{nom}m_{sl}) / {}_{nom}m_{sl} \quad (X3.2)$$

$$= bias + (bias_{ref} + {}_{ref}\varepsilon_{sl}) \cdot (1 + bias) + {}_{intra}\varepsilon_{sl} + {}_{inter}\varepsilon_{l}$$

retaining only first-order terms, yet allowing arbitrarily large *bias*, which may be corrected as below (and always neglecting Cauchy-distribution-like effects).

X3.1.3 The two unknown variances $\sigma_{intra}{}^2$ and $\sigma_{inter}{}^2$ are estimated here for the case that the number of samples S is large enough and $\sigma_{intra}{}^2$ small enough that $\sigma_{intra}{}^2/S$ can be neglected relative to the inter-lab variance $\sigma_{intra}{}^2$. In this case, the statistical analysis is very simple to carry out. However, if this approximation is not valid, then the variances may be obtained by the method of analysis of variance.

X3.2 Intra-Lab Uncertainty

X3.2.1 Eq X3.2, for each lab l, provides an estimate $s_l{}^2$ for σ^2 (as calculated on the hand-held calculator):

$$s_l^2 = \frac{1}{s-1} \sum_s (error_{sl} - error_{.l})^2 \quad (X3.3)$$

with $(S - 1)$ degrees of freedom

$$\equiv \sigma_{ref}^2 (1 + bias)^2 + s_{intra\ l}^2$$

X3.2.2 Averaging over the L labs then gives an estimate $s_{intra}{}^2$:

$$s_{intra}^2 = s_{.}^2 - \sigma_{ref}^2 (1 + bias)^2 \text{ with } L \cdot (S-1) \text{ degrees of freedom} \quad (X3.4)$$

where a dotted index signifies an index average, and where *bias* will be replaced by its estimate below (Eq X3.6).

X3.3 Inter-Lab Uncertainty

X3.3.1 Furthermore, an intra-lab mean $error_{.l}$ focuses in on the inter-lab variation. Thus, $\sigma_{inter}{}^2$ is estimated (neglecting the averaged-out intra-lab uncertainty as described above) by:

$$s_{inter}^2 = \frac{1}{L-1}\sum_l (error_{.l} - error..)^2 \quad \text{with } (L-1) \text{ degrees of freedom} \tag{X3.5}$$

X3.4 Bias Estimation

X3.4.1 The bias may be estimated as the mean of *error*:

$$bias_{Est} \equiv Mean[error] \tag{X3.6}$$

$$= bias + (bias_{ref} + {}_{ref}\varepsilon_{..}) \cdot (1 + bias) + {}_{intra}\varepsilon_{..} + {}_{inter}\varepsilon_{.}$$

where dotted indices again signify index averages.

X3.4.2 Only the first term is constant, evaluation to evaluation, if the reference bias varies.

X3.5 Bias Correction

X3.5.1 Now suppose the number L of labs participating in the round-robin experiment is so large that a future bias-corrected mass estimate m_{Corr}, following the single round-robin evaluation may reasonably be determined by dividing the raw estimate m by (1 + $bias_{Est}$). The result, for a lab drawn at random in estimating unknown mass M, is:

$$m_{Corr}/M = (1 + bias_{Corr}) + ({}_{intra}\varepsilon + {}_{inter}\varepsilon)/(1 + bias_{Est}) \tag{X3.7}$$

where the corrected but unknown $bias_{Corr}$ may be shown equal to:

$$bias_{Corr} = (bias - bias_{Est})/(1 + bias_{Est}) \tag{X3.8}$$

$$\approx -bias_{ref} - {}_{ref}\varepsilon_{..} - ({}_{intra}\varepsilon_{..} + {}_{inter}\varepsilon_{.})/(1 + bias_{Est})$$

where again the dots refer to averages over the (unknown) round-robin values.

X3.6 The Uncertainty Components

X3.6.1 The inter- and intra-lab uncertainty components can be immediately read from Eq X3.7 for the corrected mass estimates:

$$u_{inter} = s_{inter}/(1 + bias_{Est}) \tag{X3.9}$$

$$u_{intra} = s_{intra}/(1 + bias_{Est})$$

X3.7 Bias Correction Uncertainty

X3.7.1 As the bias is only known in approximation, the uncertainty in the bias correction must have an effect on the method uncertainty. Detailed analysis **(5)** indicates that the distribution of $bias_{Corr}$ (Eq 2-8) determines the uncertainty component u_{bias} and coverage factor k with the sense of prediction. (See also Appendix X2.)

X3.7.2 The only unknown is the distribution of the first term $bias_{ref}$ in Eq X3.8. Suppose that all that is known of $bias_{ref}$ is that $|bias_{ref}| < \Delta_{max}$. Then the variance of $bias_{ref}$ (evaluation to evaluation) is given (see 7.4, Appendix X3) by:

$$\text{Variance}[bias_{ref}] = \frac{1}{3}\Delta_{max}^2 \tag{X3.10}$$

X3.7.3 Adopting Eq X3.10, the uncertainty component u_{bias} associated with the bias correction uncertainty is given approximately (neglecting the averaged intra-lab and reference uncertainty and assuming that the analytical lab is drawn at random) by:

$$u_{bias}^2 = \frac{1}{3}\Delta_{max}^2 + \frac{1}{L}u_{inter}^2 \tag{X3.11}$$

X3.8 The Combined and Expanded Uncertainties

X3.8.1 All the uncertainty components can now be estimated (including a Type B estimate of the sampling pump uncertainty) and are pooled together to give the combined standard uncertainty u_c:

$$u_c^2 = u_{inter}^2 + u_{intra}^2 + u_{bias}^2 + u_{pump}^2 \tag{X3.12}$$

X3.8.2 The expanded uncertainty then is U given by:

$$U = k \times u_c \tag{X3.13}$$

where k may be calculated to provide 95 % confidence in the round robin as in Appendix X2.

X4. UNDETECTABLE OUTLIERS

X4.1 An example of how to deal with a particular type of non-normal measurement results along the lines of ISO GUM is presented here. Suppose a candidate method is evaluated relative to a well-characterized reference method. Suppose further that it is found that the candidate method rarely, but ever-so-often, gives anomalous results or outliers which are entirely useless as measurement. Otherwise the measurement results are found to be approximately normally distributed. As the reference method will not be available during subsequent practical applications of the candidate method, the outliers cannot then be detected and eliminated.

X4.2 Nevertheless, a symmetric accuracy range A can still be constructed giving 95 % coverage, by finding coverage A_f at tighter frequency f of the non-anomalous points than 95 % to compensate for the existence of the outliers. If r is the outlier rate, then the fraction of all the points, including outliers, covered by A_f is $f \times (1 - r)$. Therefore, if f is determined so that:

$$f \times (1 - r) = 95\% \tag{X4.1}$$

then A_f gives the 95 %-coverage as the usual symmetric range accuracy.

X4.2.1 Similar to Eq X4.2, $A = A_f$ is determined by:

$$A^2 = u^2 \times [bias^2 + TRSD^2 - TRSD^2_{ref}] \quad (X4.2)$$

in terms of estimates $bias^2$, variance $TRSD^2$, confounded with reference variance $TRSD^2_{ref}$, hence the final term $TRSD^2_{ref}$, assumed known.

X4.3 However, now u is not 1.960 but is the normal quantile of ½ [100 % + f], which, according to Eq X1.1 must equal ½ [100 % + 95 %/(1− r)] to compensate for the outliers at rate r. At outlier rate r = 0, u_f = 1.960. The value of u at given r can be read from Fig. X4.1. Naturally, as is evident in the figure, the outlier rate must be less than 5 % in order to have an overall accuracy range including outliers to specify coverage at 95 % confidence. Nevertheless, as intuitive from the figure, the curve u is so flat, that A_f is forgiving of a large uncertainty in the initial estimate of the anomaly rate r.

X4.4 The 95 %-confidence limit $A_{95\ \%}$ may be approximated by:

$$A_{95\%} = q \times \hat{A} \quad (X4.3)$$

where the estimate Â is the function of Eq X4.2 in terms of estimates $bi\hat{a}s$, $TR\hat{S}D$, and $\hat{r}$.

X4.4.1 Outliers are identified as such and an estimated outlier rate $\hat{r}$ estimated if their occurrence is unusual assuming a normal distribution determined from data without the outliers. For example, a point X may be considered an outlier if Prob(X > 3 std from mean) < 0.001.

X4.4.2 The factor q is chosen to approximate $\hat{A}^2/A^2$ as proportional to a chi-square variable with effective degrees of freedom υ_{eff} chosen as with **(11-10)** to reproduce the variance of $\hat{A}^2/A^2$, given the variances of $bi\hat{a}s$, $TR\hat{S}D$, and $\hat{r}$. The result, applying lowest order propagation of errors, is that the factor q is given by:

$$q^2 = \upsilon_{eff}/\chi^2_{0.05\ \upsilon_{eff}} \quad (X4.4)$$

defined in terms of a chi-square 5 %-quantile at effective number of degrees of freedom υ_{eff} :

$$\upsilon_{eff}^{-1} = \upsilon^{-1}\hat{u}^4 TR\hat{S}D^2 \frac{TR\hat{S}D^2 + 2 \cdot bi\hat{a}s^2}{\hat{A}^4} + \upsilon_{outlier}^{-1} \quad (X4.5)$$

$$\upsilon_{outlier}^{-1} = \pi\hat{r}(1-\hat{r})^{-3} n^{-1}\hat{u}^{-2} Exp[u^2] \times 0.95^2 \quad (X4.6)$$

X4.4.3 The numerator in Eq X4.5 is simplified if $bias$ is negligible (or if corrected by means of the evaluation, leading to:

$$E[bi\hat{a}s^2] \approx (TRSD^2 / n)$$

X4.4.4 At r = 4 %, u increases by 30 % over 1.960, and more importantly, the uncertainty in the estimate $\hat{r}$ when the number n of data points in the evaluation = 50 (for example) brings the confidence limit on r close to the wall at r = 5 %, with the effective number $\upsilon_{outlier}$ of degrees of freedom dropping to about 4, with marked effect on $A_{95\ \%}$. However, $\upsilon_{outlier}$ extremely rapidly increases with decreasing $\hat{r}$, equaling about 25 at $\hat{r}$ = 2 %.

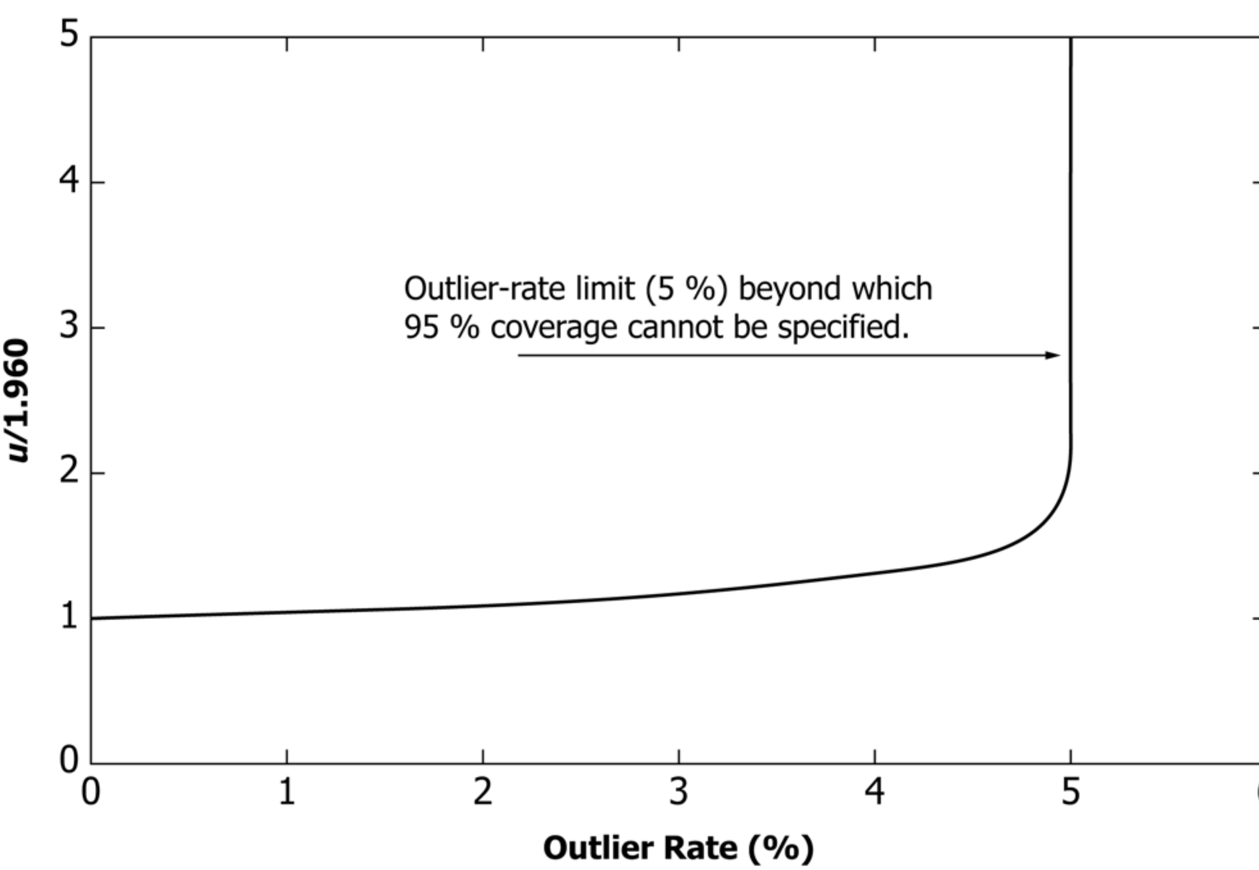

FIG. X4.1 Factor *u* Giving 95 %-Coverage Symmetric Range Accuracy in Terms of the Outlier Rate

X4.5 The usual result is obtained then, taking the expanded uncertainty $U = A_{95\ \%}$:

$$U = k\,u_c \qquad (X4.7)$$

$$u_c = \sqrt{bi\hat{a}s^2 + TR\hat{S}D^2 - TRSD_{ref}^2}$$

$$k = u_f[\hat{r}] \times \sqrt{\upsilon_{eff}/\chi^2_{\upsilon_{eff}\ 0.05}}$$

REFERENCES

(1) Taylor, B. N., and Kuyatt, C. E., Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, *NIST Technical Note 1297*, 1994.

(2) NIOSH: "Measurement Uncertainty and NIOSH Method Accuracy Range," Chapter P, *NIOSH Manual of Analytical Methods*, eds. Schlecht, P. C., and O'Connor, P. F., 4th ed., 3rd Suppl., DHHS (NIOSH) Publication No. 2003-154, 2003, and references within.

(3) Bartley, D. L., Slaven, J. E., Andrew, M., Rose, M. C., and Harper, M., "Uncertainty Determination for Non-destructive Chemical Analytical Methods Using Field Data," *Journal of Occupational and Environmental Hygiene*, 2007.

(4) Vaughan, N. P., Chalmers, C. P., and Botham, R. A., "Field Comparison of Personal Samplers for Inhalable Dust," *The Annals of Occupational Hygiene*, Vol 34, 1990, p. 553.

(5) Aichison J., and Dunsmore, I. R., *Statistical Prediction Analysis*, Cambridge University Press, Cambridge, England, 1975.

(6) EURACHEM/CITAC GUIDE—Measurement Uncertainty Arising from Sampling, 1st Ed., 2007.

(7) EURACHEM/CITAC GUIDE—Use of Uncertainty Information in Compliance Assessment, 1st Ed., 2007.

(8) Environment Agency, Technical Guidance Note M2—Monitoring of Stack Emissions to Air (and references therein), United Kingdom, 2007.

(9) Welch, B. L., *Journal of the Royal Statistical Society*, Supplement 3, 1936, p. 29; *Biometrika*, Vol 29, 1938, p. 350; *ibid.*, Vol 34, 1947, p. 28.

(10) Satterthwaite, F. E., "An Approximate Distribution of Estimates of Variance Components," *Biometrics*, Vol 2, 1946, p. 110.

(11) Smith, H. F., "The Problem of Comparing the Results of Two Experiments with Unequal Errors," *Journal for the Council of Scientific and Industrial Research*, Vol 9, 1936, p. 211.