

Estimation of Laboratory Analytical Uncertainty Using Laboratory Control Samples

By Thomas Georgian, Ph.D.

A simple method to estimate laboratory analytical error.

International Standardization Organization (ISO) 17025, "General Requirements for the Competence of Testing and Calibration Laboratories," requires testing laboratories to "apply procedures for estimating uncertainties of measurement." In instances in which a rigorous assessment of measurement error is not possible, testing laboratories are directed to "attempt to identify all the components of uncertainty and make the best possible estimation." When reporting test results, laboratories are also required to enclose "a statement on the estimated uncertainty of measurement...when it is relevant to the validity or application of the test results."

Unfortunately, there appears to be little or no guidance on how to estimate measurement uncertainty in an environmental production laboratory setting. A large number of replicates are typically required to adequately characterize uncertainty. However, it would probably be impractical for environmental testing laboratories to assess uncertainty by performing replicate analyses for each environmental sample, for example, because of higher production costs. Duplicates (e.g., matrix spike duplicates) are usually analyzed on a per batch basis at only a 5% frequency, and typically provide only a qualitative assessment of measurement uncertainty. A simple procedure for estimating laboratory measurement uncertainty from laboratory control samples (LCSs) is proposed.

PROPOSED APPROACH FOR ESTIMATING UNCERTAINTY

LCSs are routinely analyzed with environmental samples to evaluate overall method performance. The LCS is used to assess the ability of a method to successfully recover the analytes of interest from a homogeneous matrix of known composition and typically consists of a "clean" matrix-reagent water, Ottawa sand, or some other purified material that is spiked with known concentrations of analytes. The LCS is typically processed with a batch of 20 or fewer environmental samples using the same preparatory and determinative procedures for the environmental

samples. The percent recovery, R , for each analyte in a LCS is calculated from the equation:

$$R = \frac{C_m}{C_s} \times 100$$

where C_m is the measured analyte concentration and C_s is the "known" or "theoretical" spike concentration; for example, from some certified reference standard. For brevity, the percent recovery for an analyte in a LCS will be referred to as the "LCS recovery" or as the "recovery." The spike concentration for a LCS is typically near the mid-calibration range. For an ideal test method, R is 100%.

Control and warning limits for LCS recoveries are often obtained from "X-bar" or "mean" control charts when at least 20 or 30 data points have been collected. Given a set of n LCS recoveries, $\{R_i | i = 1, 2, \dots, n\}$, the upper and lower control limits are typically set at three standard deviations above and below the mean percent recovery, respectively:

$$UCL = \bar{R} + 3 s_R, \quad LCL = \bar{R} - 3 s_R$$

where \bar{R} and s_R are the mean and standard deviation for a set of LCS percent recoveries:

$$\bar{R} = \frac{\sum_{i=1}^n R_i}{n}, \quad s_R = \sqrt{\frac{\sum_{i=1}^n (R_i - \bar{R})^2}{n-1}}$$

Similarly, upper and lower warning limits are set at plus and minus two standard deviations from the mean percent recovery, respectively:

$$UWL = \bar{R} + 2 s_R, \quad LWL = \bar{R} - 2 s_R$$

If the LCS recoveries are normally distributed, then the probability that a recovery will fall between the upper and lower control limits is 99%, and the probability that a recovery will fall between the upper and lower warning limits is 95%. (More accurately, about 99.7% of the area of a normal curve is within three standard deviations

of the population mean and about 95.4% of the area is within two standard deviations of the mean). A recovery will fall outside of the control range due to random error or "chance variation" at a frequency of less than 1%. Hence, the method is viewed to be "in statistical control" when the LCS recovery falls within three standard deviations of the mean and no "abnormal" patterns are observed, such as cyclic trends or six consecutive increasing or decreasing recoveries.

If matrix interference and sample heterogeneity are not significant components of the total measurement uncertainty, or if an estimate of minimum measurement uncertainty is desired, then laboratory control sample data can be used to estimate the uncertainty for sample test results. The mean LCS recovery provides a measure of method bias and the standard deviation, $s_{\bar{R}}$, provides a measure of method precision (e.g., the variability that arises because of random error). The standard deviation (or the laboratory's warning and control limits) can be used to calculate 95% or 99% confidence intervals for sample test results. However, the confidence intervals will also be dependent upon whether or not significant method bias exists. Hence, both biased and unbiased test methods will be addressed.

It should be noted that, in theory, matrix spike (MS) recovery data would provide a more realistic estimate of measurement uncertainty. However, in practice, environmental testing laboratories rarely maintain matrix-specific statistical control and warning limits. Laboratories typically establish statistical limits by method of analysis rather than by matrix. For example, for the purposes of establishing MS control limits, groundwater, drinking water, surface water and wastewater analyzed by the same analytical method are routinely considered to be a single matrix, "water." In addition, the data used to calculate control limits for matrix spikes might include recoveries that have been impacted by significant matrix interferences. These practices tend to give rise to inappropriately wide statistical limits. In order to establish MS control limits from MS recovery data, the matrix of interest must be well-defined and free of gross interferences. Otherwise, this approach will not be viable.

In order to estimate analytical uncertainty using LCS recovery data, the presence or absence of method bias must first be established. If the recoveries are normally distributed, a two-sided significance test using the t -distribution can be per-

formed to determine whether or not \bar{R} is statistically different from 100%:

$$\bar{R} = 100\%: |\bar{R} - 100| / s_{\bar{R}} \leq t$$

$$\bar{R} \neq 100\%: |\bar{R} - 100| / s_{\bar{R}} > t$$

The symbol t represents the critical value of the Student's t -distribution, $t_{n-1, \alpha/2}$, for $n - 1$ degrees of freedom and the $(1 - \alpha)$ 100% confidence level. If n is large (e.g., there are at least 20 or 30 data points), then t will be approximately equal to two and three for the 95% and 99% levels of confidence, respectively. The term $s_{\bar{R}}$ in the denominator is the standard deviation of the mean recovery.

$$s_{\bar{R}} = s_R / \sqrt{n}$$

The "null hypothesis" that \bar{R} is not significantly different from 100% is accepted if the first inequality is true; the "alternative hypothesis" that \bar{R} is significantly different from 100% is accepted if the second inequality is true.

Assume that there is a known bias for an analytical method; in other words, the percent recovery for the LCS is statistically different from 100%. A measured result, c , for an environmental sample may be corrected for bias by dividing the result by the fraction of analyte recovered in one or more laboratory control samples:

$$\text{Bias corrected result} = c / \hat{r}$$

where

$$\hat{r} = R/100 \text{ or } \bar{R}/100$$

\bar{R} is the mean LCS recovery and R is a LCS recovery associated with the environmental sample (e.g., the recovery for the LCS processed with the environmental sample). However, it should be noted that correction for bias using the mean LCS recovery rather a single recovery generally results in a more reliable estimate. This is especially true when extremely low bias (e.g., $R < 10\%$) or high method variability exists (i.e., when precision is poor). Under these circumstances, it is recommended that bias correction be performed using the mean LCS recovery.

In accordance with ISO terminology for reporting uncertainty, assume that some "measurand" Y (i.e., the quantity being measured), is a function of N independent parameters X_1, X_2, \dots, X_N :

$$Y = f(X_1, X_2, \dots, X_N)$$

An estimate y of the measurand (Y) is determined by substituting the "input estimates" x_1, x_2, \dots, x_N for the values of

the N parameters, or "input quantities," X_1, X_2, \dots, X_N :

$$y = f(x_1, x_2, \dots, x_N)$$

The "combined standard uncertainty" for y , $u_c(y)$, is multiplied by a factor k_p , called the "coverage factor," so that the interval

$$y \pm k_p u_c(y) \quad (1)$$

has some specified high probability p , called the "coverage probability" or "level of confidence," of containing the actual value of the measurand. The product $k u_c(y)$, which is often represented by the symbol U , is called the "expanded uncertainty." The "combined standard uncertainty" is determined from the "law of propagation of uncertainty" or "uncertainty propagation formula":^{2,3}

$$u_c(y) = \sqrt{\sum_{i=1}^N (\partial f / \partial x_i)^2 u(x_i)^2} \quad (2)$$

Each partial derivative $\partial f / \partial x_i$, called a "sensitivity coefficient," is equal to $\partial f / \partial X$ evaluated at $X_i = x_i$. The "standard uncertainties" $u(x_i)$ are the estimated standard deviations (i.e., the square roots of the estimated variances) for the "input estimates." Note that if the probability distribution for y is normal and the uncertainty for $u_c(y)$ is negligible—in other words, the estimated standard deviation $u_c(y)$ is essentially equal to the actual standard deviation for the population (σ_y), then the probability that the interval $y - k u_c(y)$ to $y + k u_c(y)$ about the measured result y will contain the value of the measurand Y (i.e., the population mean μ_y) will be approximately 95% and 99% when $k = 2$ and $k = 3$, respectively. Typically, a value of 2 is used for k .

Therefore, if we let

$$y = f(c, \hat{r}) = c / \hat{r}$$

then it follows from Equation 1 that the actual analyte concentration for the sample will fall within the interval

$$c / \hat{r} \pm k_p u(c / \hat{r}) \quad (3)$$

at the level of confidence p . Assuming that c and \hat{r} are independent, the combined standard uncertainty may be determined using the uncertainty propagation formula (Equation 2):

$$\begin{aligned} u(c/\hat{r}) &= \sqrt{[\partial f(c, \hat{r}) / \partial c]^2 u(c)^2 + [\partial f(c, \hat{r}) / \partial \hat{r}]^2 u(\hat{r})^2} \\ &= (c/\hat{r}) \sqrt{(u(c) / c)^2 + (u(\hat{r}) / \hat{r})^2} \quad (4) \end{aligned}$$

The quantity $u(x)/x$ (e.g., $x = c$ or \hat{r}) will be referred to as the "relative uncertainty of x ."

Since the fraction of recovered analyte is related to the percent recovery by a constant factor (i.e., 1/100):

$$u(\hat{r} / \bar{r}) = u(r) / r$$

where $r = R$ or \bar{R} . The above equality and substitution of Equation 4 into Equation 3 gives the following interval for the bias corrected result:

$$(c / \hat{r}) \left[1 \pm k \sqrt{(u(r) / r)^2 + (u(d) / d)^2} \right] \quad (5)$$

For the purpose of conceptualization, the term in the brackets (Equation 5) may be viewed as a "precision term" that takes random measurement error into account and the factor $1 / \hat{r}$ may be viewed as a "bias term" that takes systematic measurement errors into account. The "precision term" accounts for random measurement uncertainty associated with the measured sample concentration c and the calculated percent recovery r . Note that if there were no method bias, then a bias correction would not be performed and would not contribute to the total uncertainty for the sample test result. Therefore, the standard uncertainty $u(r)$ for the recovery would be set equal to zero and the recovery \hat{r} would be set equal to unity in Equation 5, giving the interval:

$$c [1 + k u(d) / c] = c \pm k u(c) \quad (6)$$

A comparison of bracketed term in Equation 5 to that in Equation 6 indicates that bias correction increases the uncertainty interval for the test result.

If it is assumed that the relative uncertainty is approximately constant over the quantitative range of the analytical method, then

$$u(d) / c \approx u(R) / R = s_r / R \quad (7)$$

In addition, assume that when bias correction is performed using the mean recovery, the method is in statistical control, so that

$$|R - \bar{R}| \leq 3 s_r$$

This essentially means that the recovery R (e.g., associated with a batch of samples) is not significantly different than the mean recovery \bar{R} (i.e., $\bar{R} \approx R$). Therefore, when $r = \bar{R}$ in Equation 5 (i.e., when bias correction is performed using the mean LCS recovery) and n is large, the following approximation may be made using Equation 7:

$$(u(d) / d)^2 + (u(\bar{R}) / \bar{R})^2 \approx (s_r / R)^2 + (s_r / \bar{R})^2 / n \approx (s_r / R)^2 \quad (8)$$

Using the above approximation, the

interval for the bias corrected result may be approximated as follows:

$$c / (\bar{R}/100) [1 \pm k s_r / \bar{R}] \quad (9)$$

Note that in deriving Equation 9, it was assumed that the LCS is in control in order to substitute \bar{R} for R . The assumption that the relative uncertainty (standard deviation) is constant will be valid for a sample concentration c sufficiently near the spiking concentration (C_s) for the LCS and will be appropriate when the standard deviation is approximately a linear (increasing) function of concentration. Uncertainty is often approximately proportional to analyte concentration for measurements that are well above the detection limit.⁴⁶ The assumption that the relative standard deviation is constant will probably be reasonably accurate within the calibration range of the method and for analyte concentrations near the LCS spiking concentration.

If $r = R$ in Equation 5 (i.e., when bias correction is performed using a single LCS recovery value), it follows from Equation 7, that the interval for the bias corrected result is approximately:

$$c / (R/100) [1 \pm k \sqrt{2} s_r / R] \quad (10)$$

Assuming a normal distribution for the bias corrected result, $k \approx 2$ and $k \approx 3$ for confidence levels of 95% and 99%, respectively. Therefore, intervals for the bias correct result can be estimated from the statistical warning and control limits for the LCS recoveries:

$$100 (c / \bar{R}) (1 \pm L / \bar{R}) \quad (11A)$$

$$100 (c / R) (1 \pm \sqrt{2} L / R) \quad (11B)$$

The symbol L represents the half width of the warning range or control range, respectively (e.g., determined from the laboratory's in-house control charts):

$$L_{99\%} = (UCL - LCL) / 2 = 3 s_r, \\ L_{95\%} = (UWL - LWL) / 2 = 2 s_r$$

Equation 11B generally results in wider intervals for bias corrected sample concentrations than Equation 11A because the mean recovery constitutes a less uncertain estimate for the bias correction term than a single recovery. Equation 11A is especially convenient to use because the "precision term" $(1 \pm L / \bar{R})$ and the bias correction term $(100 / \bar{R})$ for the measured sample concentration will be fixed for all environmental samples analyzed by a test method when the mean recovery is associated with a fixed set of LCS warning or

control limits. When Equation 11B is used, the "precision term" and bias correction will be constant for only the batch of environmental samples that were processed with the LCS possessing the recovery of R .

When there is no method bias, the interval for sample result may be estimated from the LCS warning or control limits from Equation 11B by setting the mean recovery equal to 100%:

$$c (1 \pm L / 100) \quad (12)$$

It should be noted that Equations 10 and 11 are not applicable for concentrations that are less than the method quantitation limit (e.g., test results at or near the method detection limit). The estimates are derived from LCS spike concentrations that are well within the quantitative range of the method.

Furthermore, by definition, the quantitation limit is the lowest concentration for which quantitatively reliable values exist or are desired for sample test results. Uncertainty is inherently large and is usually unknown for concentrations less than the quantitation limits (e.g., the initial calibration curve does not generally model instrumental response to the method detection limit). Therefore, it is assumed that, at least with respect to the approach being presented here, error intervals (e.g., Equation 3) are not applicable for concentrations below the method quantitation limit.

Finally, it should be noted that both Equations 10A and 10B were derived assuming nonzero recoveries. Since extremely low laboratory control sample recoveries can result in wide confidence intervals that may not be quantitatively reliable, it is recommended that bias correction not be performed (especially when using Equation 10B) unless the LCS percent recovery is significantly greater than zero.

Because low bias is probably more common than high bias (e.g., for extractable organic analyses) and is more likely to impact data quality, for the purposes of illustration, assume that low method bias exists and a nonzero mean recovery is available to perform the bias correction. For example, assume that the mean LCS recovery is 50%, the LCS recovery associated with the test results is in control, the control limits are 20% to 80%, and $c = 10$ parts per billion (ppb) for some sample test result. Equation 13A may be

used to construct an interval for the test result:

$$100 (10 \text{ ppb}/50) (1 \pm 30/50) = 20 \pm 12 \text{ ppb}$$

In other words, for the reported value of 10 ppb, the actual concentration of the analyte is estimated, at a 99% level of confidence, to be between 8 ppb and 32 ppb. This estimate will be reasonably accurate if the LCS control limits reflect actual method performance (e.g., were generated using a sufficient number of data points), and method variability is not strongly concentration dependent (i.e., for concentrations equal to or greater than the method quantitation limit). However, if matrix effects and sample heterogeneity significantly contribute to the total uncertainty, then the LCS warning and control limits will provide only lower bound estimates of the uncertainty. Nevertheless, since environmental laboratories typically maintain in-house control chart limits for laboratory control samples, this approach appears constitute at least one possible cost-effective strategy to comply with the ISO 17025 requirement to "apply procedures for estimating uncertainties of measurement."

Furthermore, it is noted that the estimation of uncertainty from LCS warning and control limits can also be used to compare environmental test results with project decision limits (e.g., cleanup goals or risk-based action levels) and to help qualify test results during data validation. For example, when a low bias exists and the primary objective of the remedial effort is to determine whether or not contamination is above or below some decision limit, then upper confidence limits calculated from Equation 11A or Equation 11B could be compared with the decision limit. In particular, assume that the LCS recovery, R , for a batch of samples falls outside the acceptance range and low bias is indicated. The upper confidence limit could be calculated for each test result using Equation 11B and subsequently compared with the project decision limit. If the upper confidence limit is less than the project decision limit, then the test result could be qualified as estimated with low bias (e.g., using the "J-flag") and it could be concluded that the contamination is present at a concentration below the decision limit.

However, if the test result is less than the decision limit but the upper confidence limit is greater than the project

decision limit, then the test result does not demonstrate that the analyte is present at a concentration less than the decision limit. Hence, it might be appropriate to reject the test result because of the unacceptable (low) LCS recovery (e.g., to qualify the result with the "R flag").

Finally, it should be noted that, in general, bias correction is not currently an accepted practice for the reporting of environmental test results. Batch QC samples (e.g., matrix spike and LCS recoveries) are routinely reported with environmental test results but are typically used to perform only a qualitative evaluation of the data. For example, during data validation, test results associated with a low LCS recovery may be qualified as estimated. However, estimated or "J-qualified" test results associated with an unacceptable LCS would typically be used, such as for human or ecological risk assessments, without quantitative corrections for bias. Therefore, when a low bias exists, this approach will result in an under estimation of contamination in a study area. From a technical perspective, failure to correct for bias does not appear to be consistent with the ISO 17025 objective "to identify all the components of uncertainty and make the best possible estimation." Both random and systematic errors (e.g., bias) need to be taken into account to establish the "best" estimate for a test result.⁷

CONCLUSION

A simple, cost-effective strategy to estimate method uncertainty has been presented as a possible approach to help comply with the ISO 17025 requirement to "apply procedures for estimating uncertainties of measurement." Confidence limits for test results (equal to or greater than the method quantitation limits) are estimated using the statistical warning and control limits environmental laboratories routinely maintain for laboratory control samples. Uncertainty is estimated in a quantitative manner by taking both method precision and bias into account.

The approach will be valid if the in-house LCS warning and control limits reflect actual method performance and relative method error (e.g., as measured by the relative standard deviation) is not strongly concentration dependent. Furthermore, since LCSs are typically "clean" homogeneous matrices, in general, the approach will provide only a minimum estimate of method uncertainty for actual

environmental samples.

The proposed strategy can also be used to support data assessment activities such as data validation. In particular, the resulting quantitative assessment of the uncertainty could be used to help evaluate the impact of unacceptable laboratory control samples and to determine the assignment of data qualifiers for the associated sample test results. 

Thomas Georgian, Ph.D., is a chemist with the U.S. Army Corps of Engineers (USACE), Hazardous, Toxic, and Radioactive Waste, Center of Expertise in Omaha, NE. As a member of the Chemical Data Quality Management Branch, he performs on-site laboratory inspections for the USACE Laboratory Validation Program, provides technical assistance to various USACE field offices, and generates QA/QC guidance for environmental investigations and remedial efforts. Prior to joining the Army Corps of Engineers, he was the director of a drinking water laboratory in New York City. Georgian can be contacted at thomas.georgian@usace.army.mil.

REFERENCES

1. International Organization for Standardization (ISO). *Guide to the Expression of Uncertainty in Measurement*. ISO. Geneva, Switzerland. 1993.
2. Taylor, Barry N., and Chris E. Kuyatt. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. National Institute of Standards and Technology (NIST) Technical Note 1297. Gaithersburg, MD. September 1994.
3. Thompson, M., S.R. Ellison, A. Fajgelj, A.P. Willetts, and R. Wood. Harmonised guidelines for the use of recovery information in analytical measurement. IUPAC/ISO/AOAC International/EURACHEM Technical Report. Presented at the *Symposium on Harmonisation of Quality Assurance Systems for Analytical Laboratories*, Orlando, FL. September 4-5, 1996.
4. Kahn, H.D., C.E. White, K. Stralka, and R. Kuznetsovski. Alternative estimates of detection. Presented at the *20th Annual Conference of Analysis of Pollutants in the Environment*, Norfolk, VA. May 7-8, 1997.
5. Stewart II, R.E. Analytical variability. Presented at the *20th Annual Conference of Analysis of Pollutants in the Environment*, Norfolk, VA. May 7-8, 1997.
6. Taylor, John K. *Quality Assurance of Chemical Measurements*, pp. 79/82. Lewis Publishers. 1989.
7. EURACHEM Measurement Uncertainty Working Group. *EURACHEM/CITAC Guide: Quantifying Uncertainty in Analytical Measurement, Second Edition* (draft). June 1999.