

DIOXIN TEST METHOD PRECISION

H. Gregor Rigo

Rigo & Rigo Associates, Inc.
Strongsville, Ohio 44136

ABSTRACT

Sampling, sample recovery, laboratory preparation and analysis and performing calculations all introduce errors into a final test result. Even without any mistakes, random errors will remain. The precision of a method -- the size of likely random errors from all sources from sampling techniques through data reduction -- is determined by analyzing the differences between simultaneous measurements. For total dioxins, the error increases with concentration and at 26 ng/dsm³, the highest concentration with suitable data, 99 percent of individual measurements are likely to be within ± 75 percent of the true value. For ITEQ dioxins using a combination of USEPA Method 23 and European data, the precision of individual measurements also changes with concentration and is ± 110 percent at 0.1 ng/dsm³; when just the Method 23 data are used, precision does not change with concentration and it is ± 105 percent at 0.1 ng/dsm³.

Diluent correction introduces an additional source of error. The error contribution associated with the oxygen measurement is usually much less than 1 percent. So, it can be safely ignored and diluent corrected measurement imprecision can be determined by multiplying the uncorrected method precision by the same dilution correction factor used to correct the concentration.

CONCEPTUAL FRAMEWORK

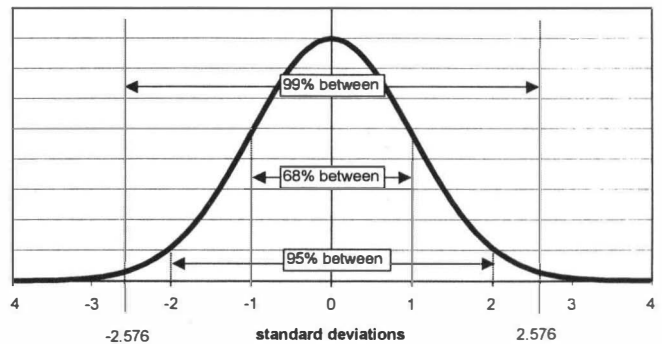
Errors are an inherent part of measurement. We are not talking about blunders and known mistakes that invalidate the results. Mistaken and biased measurements are not the subject of precision; precision applies to irreducible random errors.

For example, if we are measuring the length of a board and know that the tip of the ruler has been snapped off, this result is biased by however much of the ruler was lost. Precision, on the other hand, describes the spread in results we are likely to get measuring that board a number of times. No matter how carefully we align the ruler and read the scale, the answers will be slightly different. We cannot align the ruler exactly the same way each time. The board will not be perfectly square and true. We will interpolate slightly differently each time we read the scale. If we take many measurements, a distribution will result.

The spread of the distribution is affected by both the instrument resolution and measurement technique being used. Clearly, interpolation errors alone will be smaller when we are using a scale graduated to 1/32 of an inch instead of one marked every 1/4 inch. If we have a short ruler that has to be moved many times to measure the board, then we will have more error as well.

The error term associated with air pollution measurements is usually adequately represented by a normal distribution like the one shown in Figure 1 (Kennedy, et. al., 1995).

Figure 1. Normal distribution curve for individual measurements.

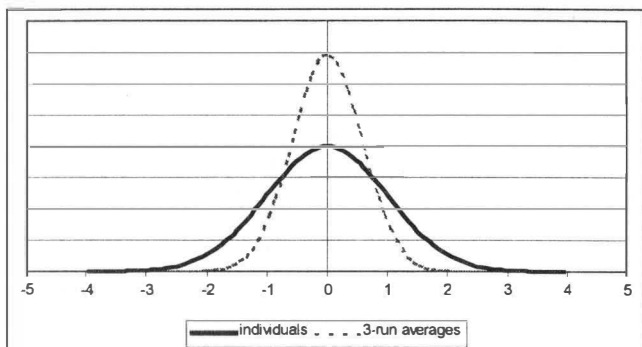


When errors are described by the normal distribution, then all we have to know is the standard deviation to determine the range of results we are likely to see. For example, 68 percent of the measurements will lie within ± 1 standard deviation of the true value; 95 percent will lie within ± 2 standard deviations and 99 percent will lie within ± 2.576 standard deviations.

When we average individual results, the averages are closer to the center than the individual measurements. This is because we are looking at the sampling distribution of the mean (average) rather than the sampling distribution for individual measurements. The standard deviation for averages is calculated by dividing the standard deviation for individual measurements by the square root of the number of individual measurements being averaged. For example, the standard deviation for 3-run averages is 58 percent of the standard deviation for individual measurements.

The effect of averaging data is shown in Figure 2. The dashed line describes the distribution of 3-run averages given the original, solid line, distribution. Of course, the number of standard deviations associated with including a given percentage of the results is the same, but the effect of averaging can be clearly seen. Virtually all the 3-run averages are bounded by two of the original standard deviations since this distance is almost 3.5 standard deviations for the 3-run averages.

Figure 2. Effect of averaging on the distribution.



Reduced to its simplest, the problem of describing measurement precision comes down to finding the population standard deviation (σ) for simultaneous measurements. Once we know the population standard deviation, we can look up how many standard deviations we need to go away from the average to make sure that a given fraction of all measurements are included.

COMPLICATIONS

Simultaneous Samplings

Measurement precision studies typically involve the repeated measurement of a standard. In the case of a chemical, standards are usually prepared and measured repeatedly. Unfortunately, this cannot be done for stack samples because it is practically impossible to keep the emissions from even the most well controlled source absolutely constant and the methods may be affected by unanalyzed gaseous constituents (matrix effects) so that synthetic gas mixtures can miss important sources of error. In fact, since the actual concentration at a source is not known, the best that can be done is to collect simultaneous samples and use those samples to estimate both the average concentration and standard deviation between measured concentrations.

EPA determined that when simultaneous samples are taken in close proximity to each other (say within 1 inch, 2.5 cm, for dual-trains that collect a pair of simultaneous samples and from four points within a 2.5 inch, 6 cm, square when a

quad-train to collect four samples simultaneously) the effects of spatial and temporal variability are minimized if not eliminated (Mitchell & Midgett, 1976; Midgett, 1997).

Sometimes data is not acquired using dual- or quad-trains. If the data is collected simultaneously from a portion of the facility where dilution and removal cannot occur (i.e., at different elevations in a stack or locations along a breaching), then the samples may be suitable.

This complication has a direct effect on estimating the precision of dioxin measurements. There are only 19 dual-train Method 23 dioxin measurements. (Rigo & Chandler, 1997). In addition, three sets of paired samples were collected during an EPA sponsored emissions assessment at a lightweight aggregate kiln when the source elected to sample in parallel to the EPA's sampling team (EER, 1997). These samples were collected on cross-traverses and analyzed by different laboratories. Consequently, they have additional sources of error, which might prove significant. Finally, as part of the preparation of a European Standard, 24 simultaneous dioxin measurements were made along a duct by multiple sampling teams, two analytical laboratories and three different methods (Dilution, Filter/Cooler and Cooled Probe) at three different facilities (European Committee for Standardization, 1996). Since European samples are typically 10 dnm³ and US samples are 4 dsm³, the European results were corrected for differing reference conditions and sample volumes. The similarity of simultaneous, but not collocated samples was verified during data analysis.

Small Sample Bias

As a practical matter, only a limited number of simultaneous, collocated samples can be collected at a time. This means that the average and standard deviation are based on small sample sizes.

As more data is acquired, the average converges to the population mean. The same can be said for the variance, the standard deviation squared.

Unfortunately, the sample standard deviation (S), what we use to estimate the population standard deviation, hence precision, does not converge to the population standard deviation, rather it converges to some fraction of that value (Natrella, 1966). Fortunately, that fraction is well known and all that has to be done is to multiply the sample standard deviation calculated using the small samples by the appropriate correction factor (Dixon & Massey, 1969) shown in Table 1.

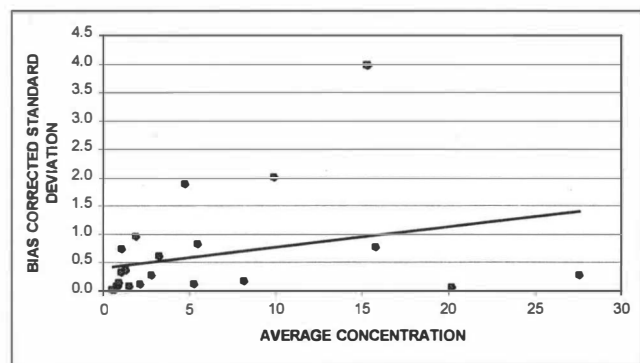
Table 1. Small sample size characteristics of the standard deviation for a normal distribution.

Sample Size, N	S is an unbiased estimator of:	Small Sample Bias Correction Factor
2	0.797 σ	1.253
3	0.886 σ	1.128
4	0.921 σ	1.085
5	0.940 σ	1.064
6	0.952 σ	1.051
7	0.959 σ	1.042
8	0.965 σ	1.036
10	0.973 σ	1.028
20	0.987 σ	1.013
30	0.991 σ	1.009

Transformations

Figure 3 is a plot of the bias-corrected sample standard deviations (S_{bc}) versus concentration (C) for total dioxins. Notice that the data looks like a wedge expanding to the right. This violates an assumption implicit in ordinary least squares (OLS) linear regression that the errors are distributed uniformly about the regression line.

Figure 3. Plot of S_{bc} versus C for total dioxins.



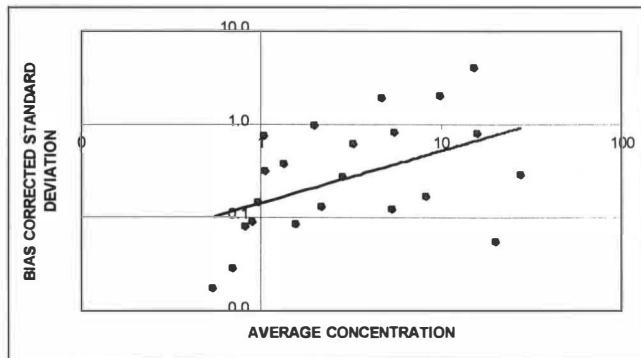
In Figure 4, the natural logarithms of the same data are plotted. Now, the data form a reasonably uniform band about the regression line. The transformed data are suitable for analysis using OLS, but the untransformed data are not.

Confidence bounds for σ

We use the population standard deviation (σ) to determine precision. If the standard deviation is constant over the range of concentrations measured, solving the problem begins by averaging either S_{bc} or the uncorrected variances and applying the small sample bias correction factor after taking the square root of the result. Then, the confidence interval likely to include σ is determined and the results are

multiplied by the appropriate number of standard deviations to generate the central estimate and upper and lower confidence bounds to characterize precision.

Figure 4. Plot of $\ln(S_{bc})$ versus $\ln(C)$ for total dioxins.



For data where S_{bc} is not constant, the relationship between S_{bc} and C is established and the error associated with fitting the curve is used to establish upper and lower confidence bounds for the regression line. These are multiplied by the appropriate number of standard deviations to generate the central estimate and upper and lower confidence bounds which characterize precision.

Finding out if S_{bc} is constant is easily done by assuming that it changes with concentration. If the regression coefficient for $\ln(C)$ is not statistically different than zero, then the S_{bc} is homogeneous (e.g. S_{bc} does not change with concentration). When the S_{bc} is heterogeneous; the regression coefficient changes with concentration.

If you only need precision at the average of the tested concentrations, then the following equation provides the appropriate bounds for σ given constant S_{bc} :

$$\sqrt{\frac{N-1}{\chi^2_{N-1, \alpha/2}}} S_{bc} < \sigma < \sqrt{\frac{N-1}{\chi^2_{N-1, 1-\alpha/2}}} S_{bc}$$

Where: N is the number of simultaneous measurements, α is the statistical significance level (0.05 corresponding to 95 percent statistical confidence for this work), and

$\chi^2_{N-1, \alpha}$ is the Chi-Squared statistic with N-1 degrees of freedom.

It is important to always remember that S_{bc} is an estimate of σ . If each of the tests were to be repeated, it is very unlikely that exactly the same answer would result. S_{bc} is an estimate of σ , not σ itself. σ is unknown and unknowable. We know, however, that σ is likely to be found

within the bounds calculated from an individual experiment a specified percentage of the time. Hence, it is necessary to describe the region likely to encompass the truth and not focus on an individual numerical result. The chance that the point estimate generated from any data set is σ is negligible; the chance that the confidence bounds include σ is high.

Calibration Curve or a Point on the Curve?

When σ is needed at more than one concentration, then an approach that captures the uncertainty associated with the regression line (even if the regression is for a constant with zero intercept – constant S_{bc} case) and not a single point on the line. This is the case for dioxin measurements since regulatory limits in the United States are expressed on a diluent corrected basis (S_{bc} is needed over a range of concentrations from roughly 50 to 150 percent of the regulatory value), emissions guideline values range between 15 and 125 ng/dsm³ corrected to 7% O₂ for total dioxins (sum tetra-through-octa homologue groups for dibenzo-p-dioxins and dibenzofurans) and there are several ITEQ concentrations found in current regulations including 0.2 and 0.4 ng/dsm³ corrected to 7%.

Least squares regression is used to determine the characteristics of the best fit line relating $\ln(S_{bc})$ and C . Once these characteristics are known, the following equation is used to determine $\ln(S_{bc})$ at any concentration:

$$\ln(S_{bc}) = a + b \ln(C)$$

The statistical confidence limits for the estimate are estimated by adding or subtracting:

$$\pm \sqrt{2F_{2,N-2,1-\alpha}} \cdot SER \cdot \sqrt{\frac{1}{N} + \frac{(\ln(C) - \langle \ln(C) \rangle)^2}{(N-1)S_{\ln(C)}^2}}$$

Where: a and b are the intercept and slope of the regression,
 $F_{2,N-2,1-\alpha}$ is the F statistic with 2 and $N-2$ degrees of freedom

SER is the standard error of the regression and is equal to the standard deviation of the residuals (predicted value minus actual value at each concentration) multiplied by $\sqrt{\frac{N-1}{N-2}}$,

$\langle \ln(C) \rangle$ is the average of the transformed concentrations, and

$S_{\ln(C)}^2$ is the standard deviation of the transformed concentrations.

If we were interested in using the results of this analysis to estimate only one point on the line, then the $\sqrt{2F_{2,N-2,1-\alpha}}$

term needs to be replaced by the Student's t-statistic ($t_{N-1,\alpha/2}$) (Natrella, 1966).

Retransformation Bias

When the S_s estimated at each concentration are retransformed by taking the exponential of the estimate and averaged, the result will not match the average of the original (untransformed) S_{bc} when the response variable (S_{bc}) has been nonlinearly transformed. Natural logarithms are a nonlinear transformation used to make the data suitable for analysis using regression, but the logarithmic transform introduces retransformation bias.

Values of \hat{S}_{bc} (S_{bc} determined for any C using the results of the regression analysis) are calculated using:

$$\hat{S}_{bc} = rT_{bcf} \exp^{\ln(S_{bc})}$$

Where: rT_{bcf} is the retransformation bias correction factor.

rT_{bcf} is determined by summing or averaging \hat{S}_{bc} calculated assuming that rT_{bcf} is 1 for all measured concentrations ($\ln(C)$) and dividing by the sum or average of S_{bc} (Duan, 1983).

Different numbers of simultaneous samples

As a final complication, different numbers of samples can make up each simultaneous set. When this happens, the S_{bc} is estimated using a different number of samples. The bias correction factor makes sure that all the results are pointing toward σ and not different fractions of σ ; however, the number of runs used to calculate each with S_{bc} determines the relative importance of each estimate.

If a weighted least squares program is available, the curve fit should be performed using the degrees of freedom (1 for a pair, 2 for a triplet and 3 for a quad-train) as weights.

The same thing can be accomplished by repeating each C - S_{bc} pair so that it appears in the data table the same number of times as it has degrees of freedom. After running the ordinary least squares regression on this data set, the standard error of the regression has to be multiplied by $\sqrt{\frac{N-1}{\sum \nu - 1}}$ to correct for the number of independent estimates of S_{bc} used in the analysis.

Where: $\sum \nu$ is the number of degrees of freedom in the original data set – also equal to the number of rows in the repeated data set – and N is the original number of pairs.

Further information on calculation procedures can be found in a forthcoming report by Lanier, W. S. and C. D. Hendrix (2000).

Diluent Correction

Emissions test results are diluent corrected to 7% O₂ in the United States by multiplying the directly measured result by $\frac{13.9}{20.9 - O_2}$ where O₂ is the oxygen concentration measured in the gas. Propagation of error techniques (ASME, 1998) were used to determine that:

$$S_{bc-7\%} = \frac{13.9}{20.9 - O_2} \sqrt{\exp(\hat{S}_{bc}^2) + \left(\frac{\ln(C)S_{O_2}}{20.9 - O_2}\right)^2}$$

Where: S_{O₂} is the standard deviation of oxygen measurements.

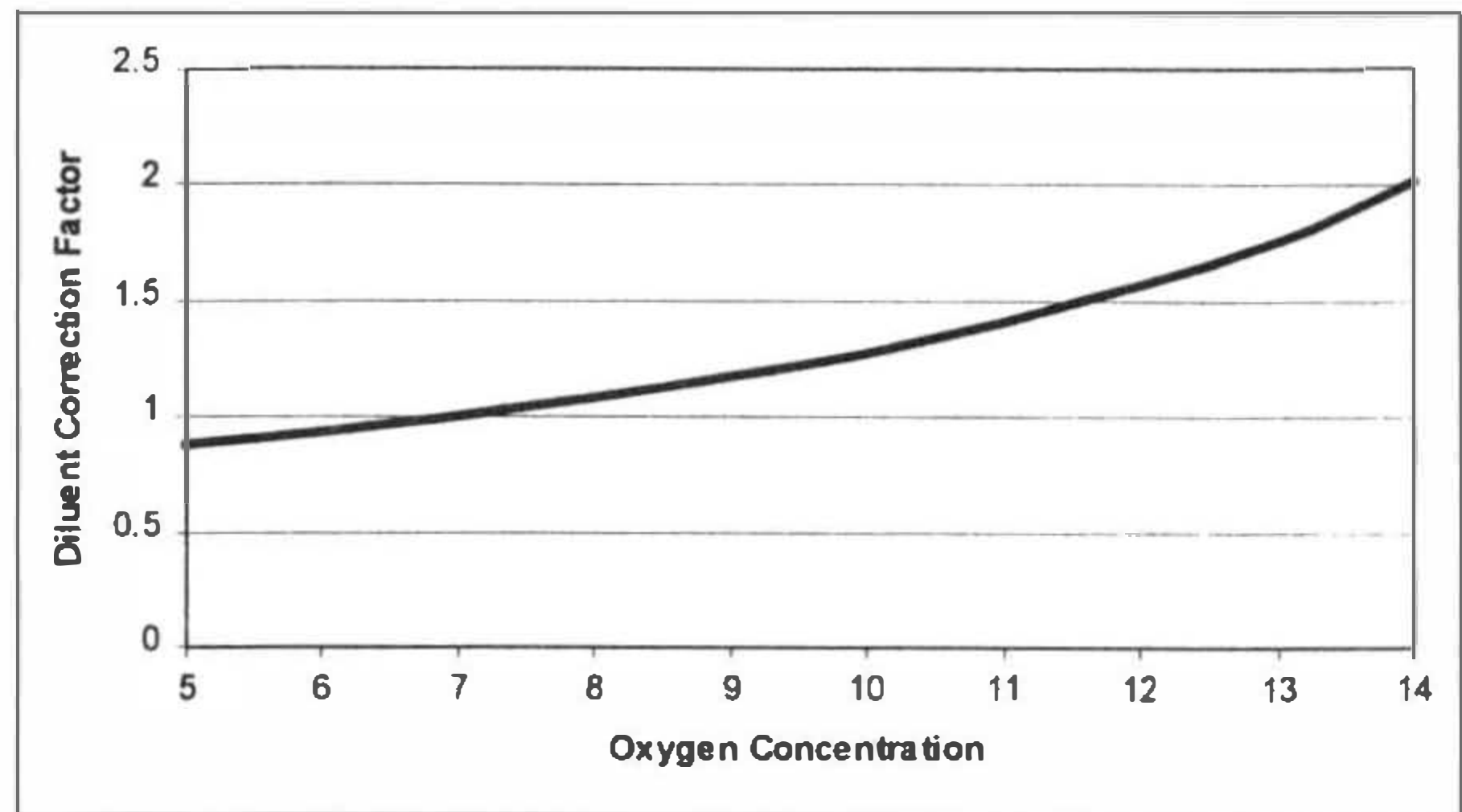
Analysis of several sets of oxygen analyzer Relative Accuracy Test Audit (RATA) data found that the standard deviation between data pairs is homogenous over a wide range of concentrations. Consequently, the standard deviation of the oxygen measurements was determined by averaging all the small sample bias corrected standard deviation estimates and dividing that result for ½ RATA hour measurements by the square root of 8 to get the standard deviation that applies to a 4-hour dioxin test. The precision of today's oxygen analyzers is about half that associated with Orsat analyzers (Midgett, 1977).

The effect of including the right-hand term in the uncertainty estimate was determined by calculating S_{bc-7%} for randomly selected dioxin concentrations. Oxygen concentrations were randomly selected between 5 to 13.5 percent, the range Rigo & Rigo Associates, Inc. proprietary emissions database indicates facilities operate. The precision of the dioxin measurement was calculated using the complete equation as well as just the term associated with left hand side of the square root term – the portion associated with the precision of the oxygen measurement. The ratio of the diluent corrected measurement method precision to the total precision was calculated. This process was repeated 10,000 times.

The measurement precision associated with oxygen measurements does not contribute significantly to the overall precision estimate. In fact, for Total dioxins and ITEQ dioxins estimated using all the available data, the largest difference was less than 0.02 percent. When just the Method 23 data was used to estimate ITEQ dioxins, 50 percent of the results were within 1 percent of correct answer ignoring the right-hand side of the diluent correction equation, 75 percent were within 2 percent of the correct answer and only 1 percent differed by more than 4 percent.

Consequently, it is reasonable to estimate the precision of a diluent corrected result by multiplying the precision estimate for the uncorrected result by the factor used to diluent correct the concentration. The values for this factor are shown in Figure 5.

Figure 5. Dilution correction factor as a function of percent oxygen in the gas to adjust results to 7% O₂.



EXPRESSING THE RESULTS

Measurement precision is expressed many ways. Analytical chemists frequently express the result as a coefficient of variance – the standard deviation divided by the concentration. Engineers tend to think in terms of how far a result may be removed from the real value, so they use a multiple of the standard deviation designed to include a specified percentage of the likely results. As a rule of thumb, 2S_{bc} provides 95 percent coverage and 3S_{bc} provides 99 percent coverage (ASME, 1998, ANSI, 1997).

For this analysis, σ is multiplied by 2.576 to estimate the bounds likely to contain 99% of the measurements made using the method at a specified concentration. This factor is used instead of the rule of thumb values because the uncertainty associated with S_{bc} is considered through the use of confidence bounds on S_{bc} rather than being included in the multiplier.

To make the presentation as simple as possible, upper and lower 95 percent confidence limits and the central point estimate are provided for the coefficient of variance. These bounds are re-expressed as bounds likely to encompass 99 percent of the individual runs and 3-run averages concentration units.

FINDINGS

Total Dioxins

Only dual-train and simultaneous cross-traverse Method 23 data exist for total dioxins because the European method validations studies only reported ITEQ dioxins.

Figure 6 through 8 present the results first in terms of coefficient of variances and then as the bounds likely to include

99 percent of the individual measurements and 3-run averages.

Figure 6. Coefficient of Variance for Total dioxins.

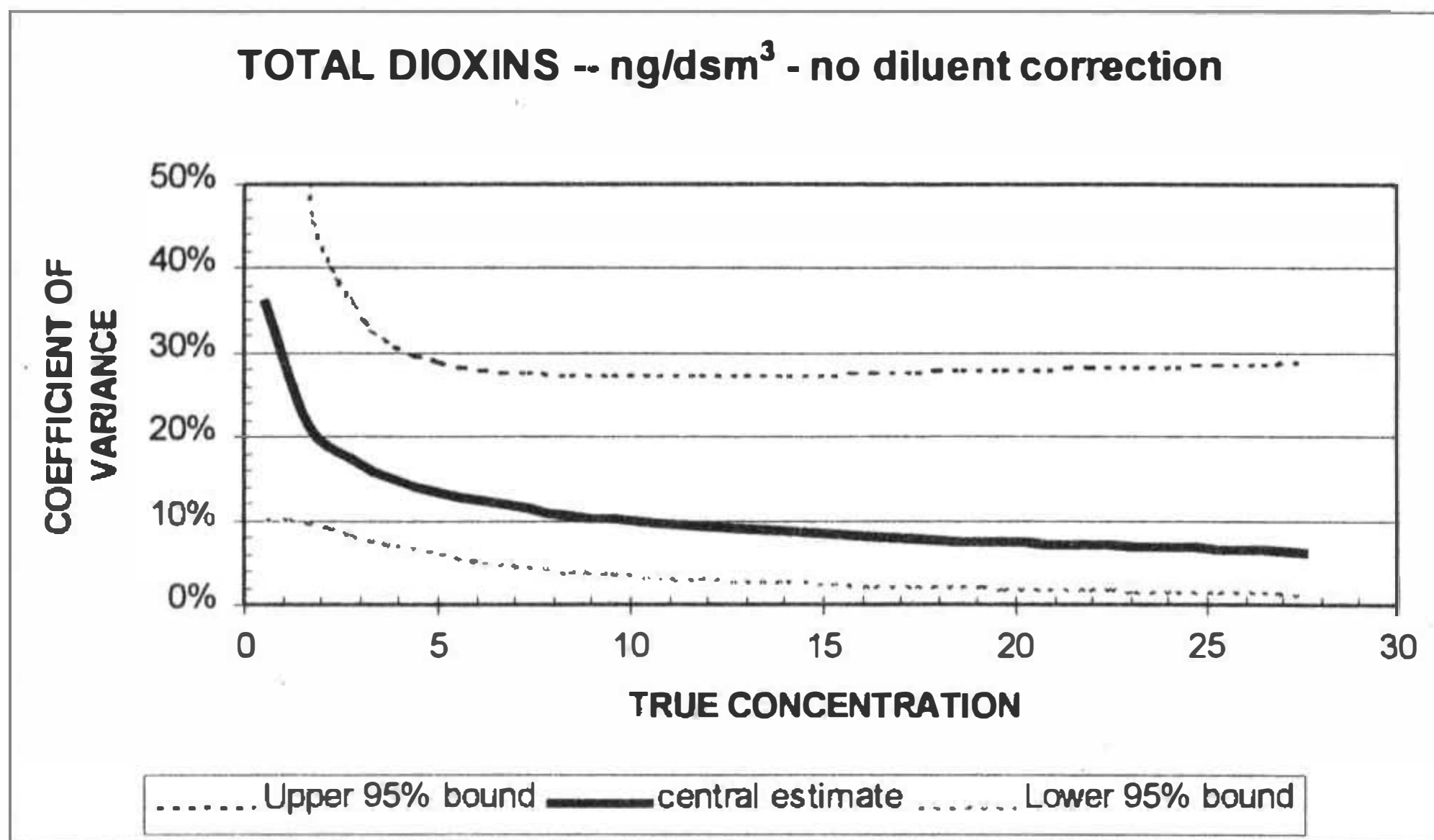


Figure 7. Individual measurement precision bounds.

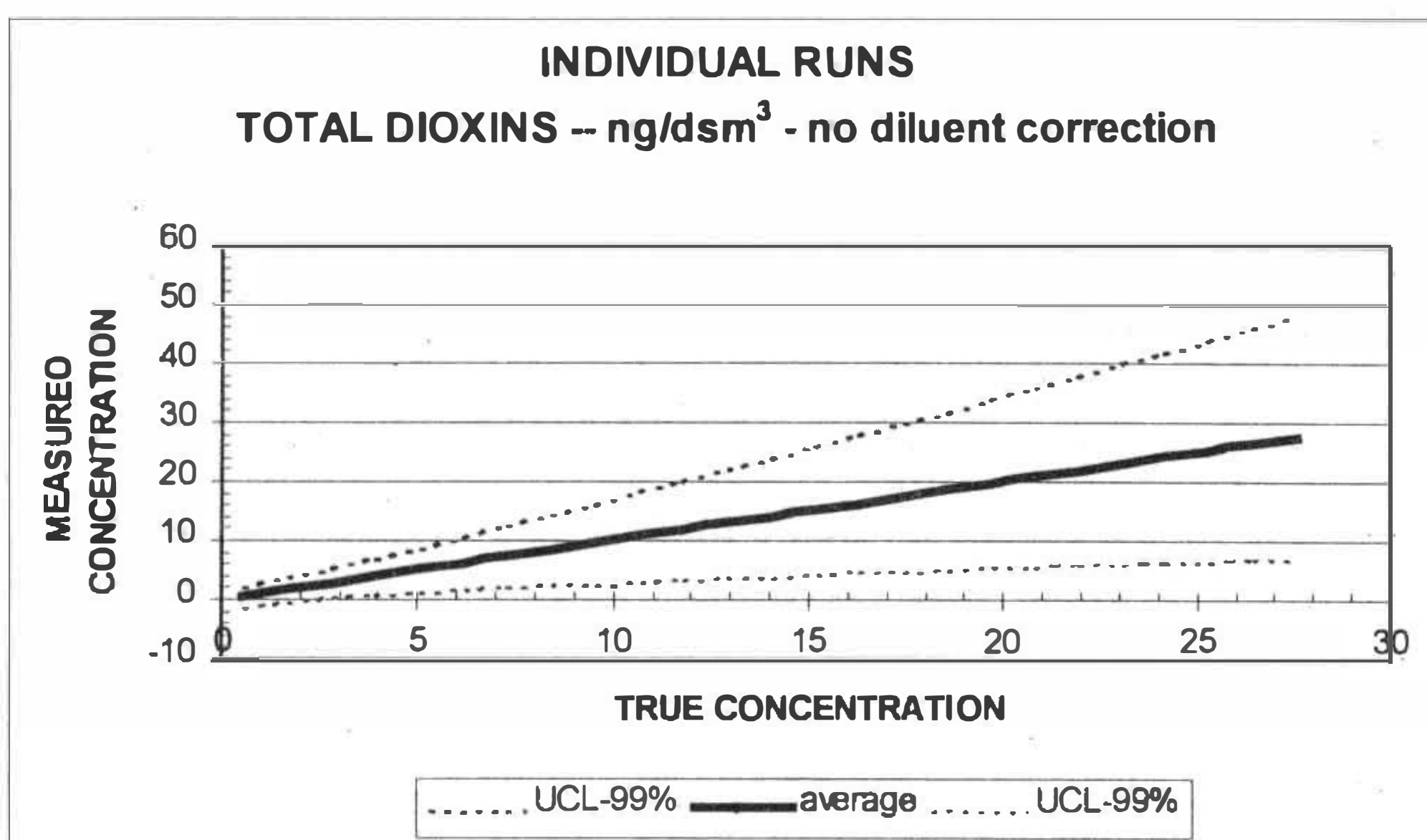
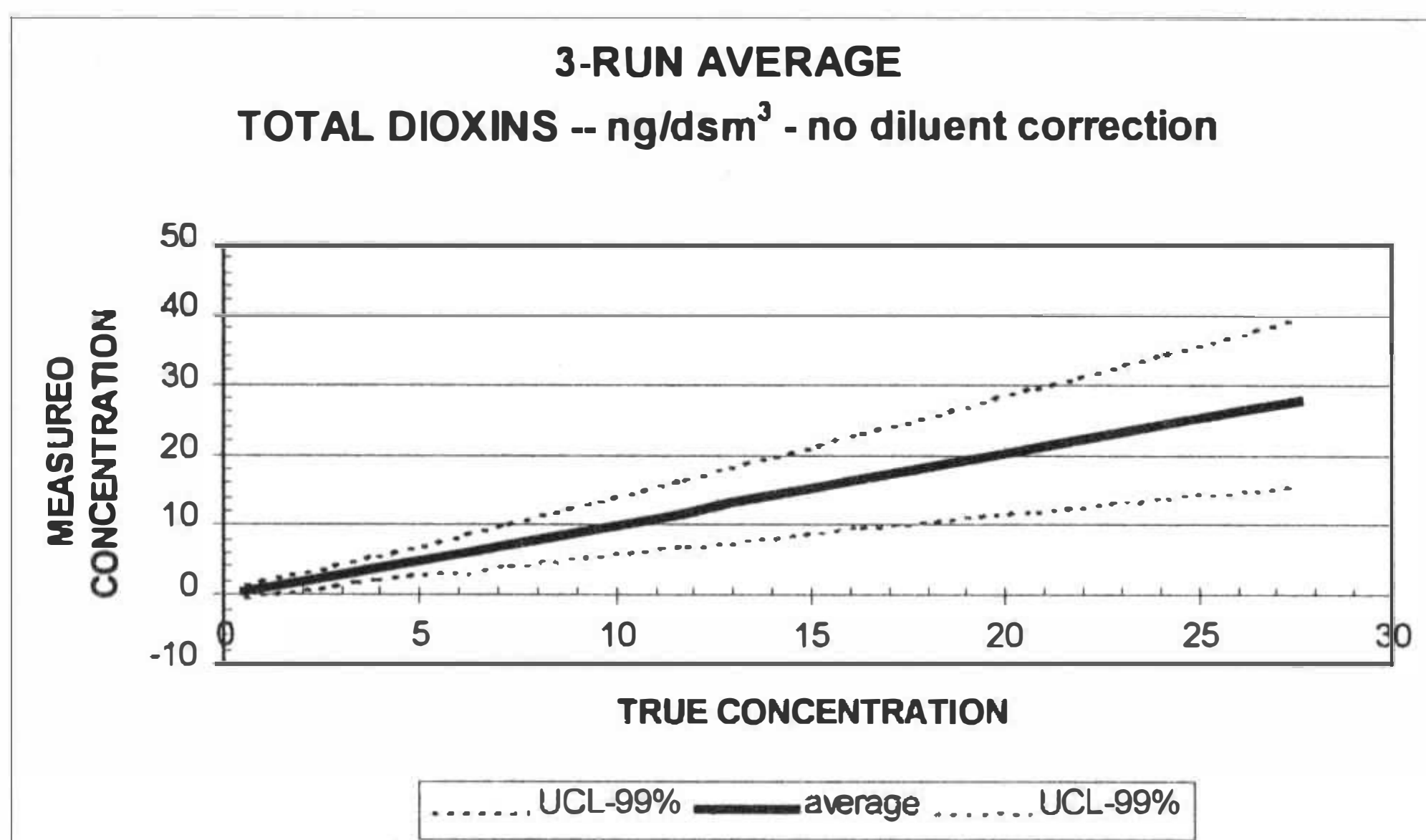


Figure 8. 3-run average precision bounds.



The most important observation is that the available total dioxin data does not cover most of the range of regulations in the United States. Since extrapolation is very dangerous

absent a validated fundamental model – we are using a statistical model, which describes the data over the available range, not a model built up from first principles that is validated and calibrated using the available data – extrapolation is not recommended.

The coefficient of variance for individual measurements could be as high as 30 percent. The uncertainty associated with 3-run averages is 30 percent as well. The implication is that someone trying to meet a dioxin standard needs to design their facility to emit roughly 2/3 of the standard if exceedances due to measurement precision alone are to be avoided. An even lower design point is needed to accommodate normal fluctuations in facility performance.

The statistical results of this analysis are summarized in Table 2. The slope coefficient is statistically significant, so precision changes with concentration.

Table 2. Statistical characteristics for the precision of total dioxin measurements.

	Coefficient	Std. Error	t	sig.
intercept	-1.939	0.372	-5.217	0.000
slope	0.559	0.23	2.43	0.025
Standard error of the estimate	1.2673			
N	22			
minimum	0.5426			
maximum	27.555			
average of ln(C)	1.1089			
standard deviation of ln(C)	1.2018			
retransformation bias correction factor	1.894			

The t-statistic for the intercept and slope are calculated by dividing the coefficient by its standard error. This t-statistic has a certain likelihood of being found by chance which is called its sig.(nificance). Both the intercept and slope are highly unlikely to be zero. The rest of the results are needed to calculate the retransformation bias corrected standard deviation at a specified concentration and the precision expressed as limits likely to include 99 percent of the measurements.

ITEQ Dioxins – Method 23 data only

Only the Method 23 results were used in this analysis of ITEQ dioxin precision. The standard deviation is homogeneous; it does not change with concentration. Consequently, after the data was analyzed using the natural logarithm transformed values, the analysis was repeated using untransformed values. The intercept is the average of the standard deviations and the slope has a value of zero. The Standard Error of the Regression is the standard deviation

of the bias corrected standard deviations. These statistical characteristics are needed to estimate precision along the line rather than at just the average concentration, which results when ordinary pooling of the standard deviations is used. The results of the analysis are summarized in Figures 9 through 11.

The most notable aspect of this analysis is that at a true concentration of 0.1 ng/dsm³, which roughly corresponds to 0.2 ng/dsm³ corrected to 7% Oxygen, individual run results as high as 0.2 and as low as 0 ng/dsm³ are likely. From a design-for-compliance point of view, it may be impossible to avoid occasional exceedances due to measurement error alone even at a facility emitting no dioxins. Providing prudent design margin to accommodate normal facility fluctuations may prove impossible.

Figure 9. Coefficient of Variance for ITEQ Dioxins – Method 23 data only.

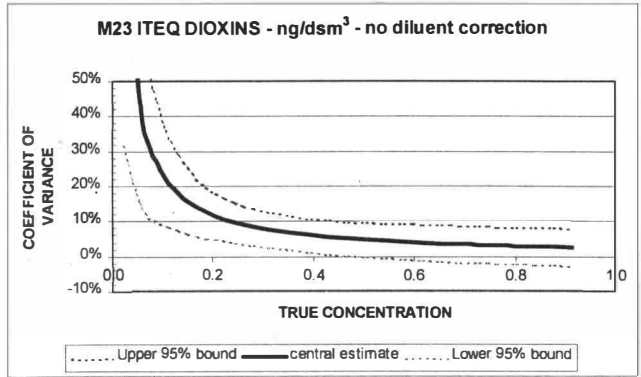


Figure 10. Individual run precision for ITEQ dioxins – Method 23 data only.

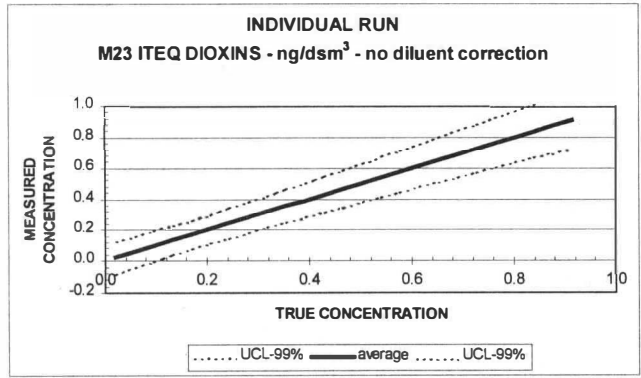
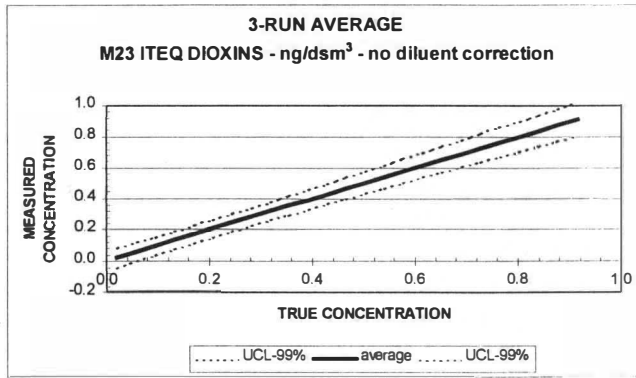


Figure 11. 3-run average precision for ITEQ dioxins – Method 23 data only.



The statistical characteristics of the data analysis are provided in Table 3.

Table 3. Statistical Characteristics of Method 23 ITEQ dioxin precision results.

	Coefficient	Std. Error	t	sig.
intercept	0.02373	0.564	-5.72	0.000
slope	0	0.211	2.334	0.024
Standard error of the estimate	0.02412			
N	22			
minimum	0.0209			
maximum	0.9128			
average of C	0.186565			
standard deviation of C	0.217279			
retransformation bias correction factor	1.000			

If the only desired result is at the precision at the average concentration (0.19 ng/dsm³), then the data can be analyzed by simply averaging the bias corrected standard deviations and determining the confidence bounds using the equations provided earlier in this paper. The results of this analysis are provided in Table 4.

The results in Table 4 for 0.19 ng/dsm³ dioxin are very similar to those in Figure 7. Figure 7's bands are slightly wider since they are designed to encompass the entire concentration range analyzed rather than just apply to the average concentration.

Table 4. Point estimate of Method 23 ITEQ dioxin precision.

	Coefficient
Average C	0.186565
Pooled S	0.02373
N	22
statistical significance	0.05
UCL-S	0.034
LCL-S	0.018
CV-upper bound	0.182
CV-mean	0.127
CV-lower bound	0.098
most uncertainty	0.274
mean	0.248
least uncertainty	0.234

ITEQ Dioxins – All available data

In addition to the dual-train data collected by Rigo & Chander (1997) and the cross-traverse results by EPA (1998), The European Committee for Standardization (EN) (1996 and Broker, 1998) conducted simultaneous testing at several locations in a single breaching at three different facilities. The EN testing involved three different test methods and generated a number of usable pairs and triplets.

The analysis of this data began by assuming that the three European methods are all different than Method 23. The initial assessment of the C-S_{bc} relationship used a suite of dummy variables, which allowed each method to have a different intercept and slope with Method 23 being the default. None of the coefficients were significant. This is not surprising since Hagenmaier (1987) found that the different dioxin measurement methods involved in the European validation effort produced the same total concentrations, but where the various dioxins are caught in the sampling train changed. That is, each of these methods provide a good dioxin estimate, but the distribution within the sampling train provides no meaningful information on the physical state of the dioxin before sampling.

Next, the analysis was repeated without the dummy variables and all the valid data for the individual dioxin measurements were combined into a single analysis. The standard deviation is a function of concentration.

Figures 12 through 14 present the results of the analysis for ITEQ dioxins using all the available data. The statistical characteristics are summarized in Table 5.

The most interesting observation is that combined data set covers the same range as the Method 23 only ITEQ dioxin analysis and the measurements are apparently less precise at

the upper end. This may be an artifact of using the log-log model in one case and a linear model in the other.

Figure 12. Coefficient of Variance for all the ITEQ Dioxin data.

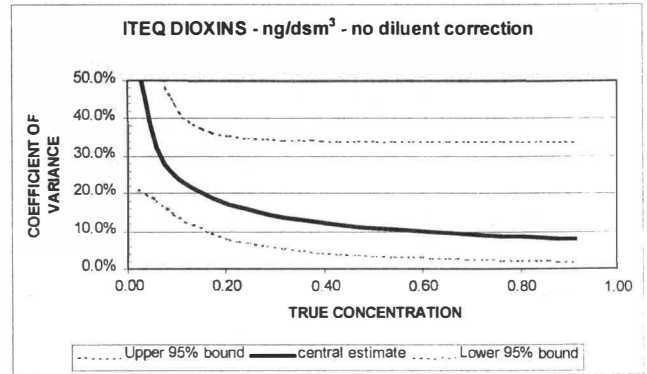


Figure 13. Individual measurement precision for all the ITEQ Dioxin data.

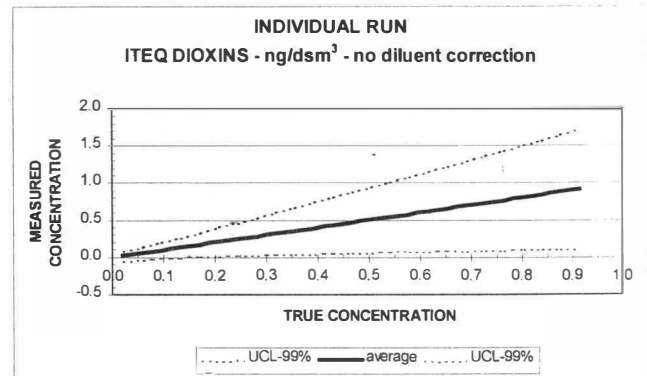


Figure 14. 3-run average measurement precision for all the ITEQ Dioxin data.

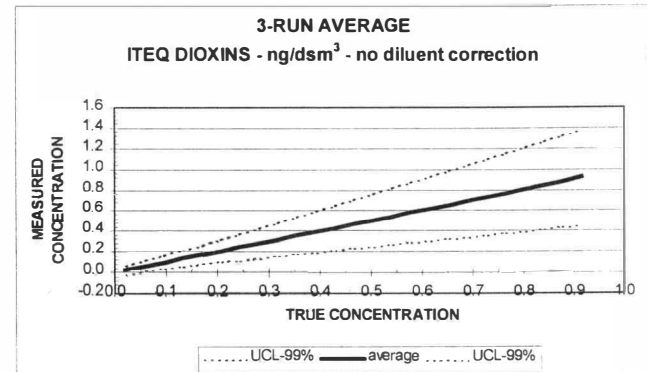


Table 5. Statistical Characteristics of Method 23, cooled probe, dilution probe and filter/condenser ITEQ dioxin precision measurements.

	coefficient	Std. Error	t	sig.
intercept	-3.228	0.564	-5.72	0.000
slope	0.492	0.211	2.334	0.024
Standard error of the estimate	1.4324			
N	46			
minimum	0.0209			
maximum	0.9128			
average of ln(C)	-2.4789			
standard deviation of ln(C)	0.9578			
retransformation bias correction factor	1.922			

HOW PRECISE IS PRECISE ENOUGH?

A measurement is precise enough as long as the results are suitable for their intended purpose. For example, a seamstress planning on a ¼ to a ½ inch seam is happy with measurements that are within 1/8 inch of the true value. Someone building a turbine bearing, however, wants the rotor within a thousandth of an inch of round.

The American Chemical Society's (ACS) Committee on Environmental Improvement began addressing this question by asking "whether a measured value is significantly different from that found for the sample blank" (Keith, et. al., 1983). The committee recommended that the Limit of Detection (LOD) for a sample be taken as 3σ where σ is the measured at the blank level. The ACS committee also recommends establishing the Limit of Quantitation (LOQ), the lower limit of the useful range for a measurement method, as 10σ because this concentration the "correspond(s) to an uncertainty of $\pm 30\%$ in the measured value ($10\sigma \pm 3\sigma$) at the 99% confidence level" (Keith, et. al., 1983). Unfortunately, the ACS definition of LOQ is only completely correct and internally consistent when σ is constant. When σ changes with concentration, then the $\pm 30\%$ criteria is only met when σ is measured at the concentration corresponding to 10σ . That is, the criteria are met when the coefficient of variance is 10%.

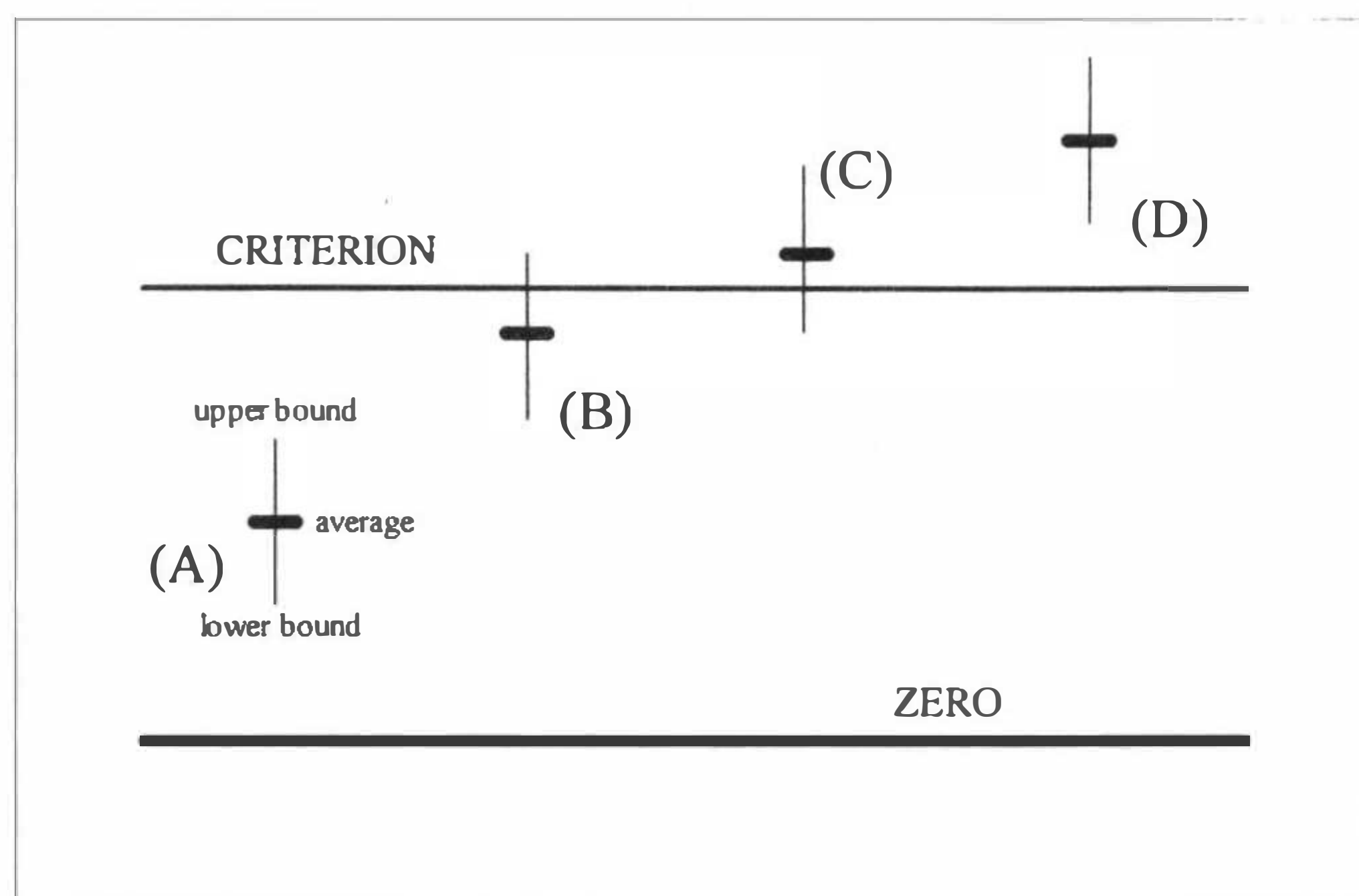
In the preamble to the publication of Method 301, *Field Validation of Pollution Measurement Methods from Various Waste Media*, the USEPA explained that "the field validation procedure of Method 301 is appropriate for validating emissions data from the source and at the levels only for which the method was validated." EPA goes on to explain that: "the practical limit of Quantitation (PLQ) is the lowest level above which quantitative results may be obtained with

an acceptable degree of confidence. For this protocol, the PLQ is defined as 10 times the standard deviation, S_0 , at the blank level. This PLQ corresponds to an uncertainty of $\pm 30\%$ at the 99-percent confidence level. The PLQ will be used to establish the lower limit of the test method" (EPA, 1992).

Method 301 includes a procedure to use when S_0 changes with concentration. Unfortunately, this procedure breaks down when S_0 increases with concentration – the case with most emissions measurements. Instead, EPA declared that when 99 percent of the individual results are within $\pm 30\%$ percent of the true concentration the precision is acceptable (EPA, 1999). This interpretation maintains the integrity of the definition found in Method 301. It is also consistent with the approach to acceptable measurement method precision taken by the National Institute for Occupational Safety and Health (NIOSH). NIOSH finds that "over a concentration range of 0.1 to 2 times the exposure limit, the method can provide a result that is within $\pm 25\%$ percent of the true concentration 95 percent of the time" (NIOSH, 1998). However, EPA's recent interpretation of Method 301 criteria for method applicability can create a regulatory gap when the precision limits for a test method is larger than $\pm 30\%$ percent or, alternatively, the coefficient of variance is greater than 10 percent at regulatory concentration levels.

Another way to look at an acceptable level of precision is to consider how the results are going to be used. Figure 15, adapted from the NIOSH Manual of Analytical Methods (1994), shows that precision can be used to determine if a facility is in compliance (A), out of compliance (D) or if the results of a test are indeterminate (B and C).

Figure 15. Compliance status using NIOSH criteria.



The EPA's approach to determining compliance does not allow an indeterminate category. Instead, EPA requires that the average of three runs be below the criterion (EPA,

1971). This approach allows regulators and the public to conclude at the end of each test whether or not the facility was in or out of compliance with its permit limits. Unfortunately, if precision is not known or properly considered when a limit is set, compliance becomes a statistical game of chance rather than the result of facility good design, maintenance and operation.

Returning to Figure 15, test results for D will always exceed the criterion and be considered violations of a not-to-exceed permit limit set at the line labeled "criterion". Facility C might pass some tests since the lower bound on measurement precision is above the criterion, but would fail more than half of the time since the average is above the criterion. Facility B, on the other hand, would pass most tests since the average is below the criteria, but fail some since the upper bound is above the criteria. Only facility A, a facility with upper bound emissions characteristic below the criterion can be expected to be below the criterion whenever it is tested.

Historically, EPA accommodated measurement uncertainty and process variability by adding margin to the emissions they expect from well-controlled sources when establishing New Source Performance Standards (Ajax, 1995). For data-derived MACT standards, no explicit consideration has been made.

This all leads to an obvious point. Measurement precision must be known and fully considered when limits are set if those limits are not going to be exceeded due to random measurement error – chance – alone.

For dioxin measurements specifically, measurement precision is not known over the entire regulatory range and it cannot be safely assumed that (im)precision is negligible at the levels which have not yet been characterized.

NEXT STEPS

The database of available multi-train measurements needs to be expanded to cover the entire regulatory range. Otherwise, it is impossible to determine how much margin is needed between a regulatory limit and the emissions control system design point to allow for measurement precision.

While matrix effects have not yet been encountered, the available data is too limited to conclude that precision data collected at one type of source is applicable to all others. Consequently, multi-train data should be collected from a variety of source categories to verify that precision is only affected by the targeted pollutants and not other flue gas constituents that might affect the analytical result.

REFERENCES

Ajax, R. L., "The effect of compliance test frequency on the stringency of technology based standards," Attachment 4 to Cement Kiln Recycling Coalition (CKRC) comments, Hazardous Waste Combustion MACT Proposed Rule, RCRA Docket No. F-96-RCSP-FFFF (Document Number 00170).

ANSI, American national standard for expressing uncertainty – U.S. guide to the expression of uncertainty in measurement, ANSI, October 9, 1997.

ASME, Test Uncertainty – Instruments and Apparatus – Supplement to ASME Performance Test Codes, ASME, PTC-19.1-1998.

Broker, G., et. al., 'Standardisation of dioxin measurement procedures for incinerator waste gases according to the European Standard EN 1948,' *Gefährstoffe – Reinhaltung der Luft*, V58, 1998, pp 141-147.

Dixon, W. J., and F. J. Massey, Introduction to Statistical Analysis, 3^d ed., McGraw-Hill, New York, 1969, p 136.

Duan, N., "Smearing estimate: a nonparametric retransformation method," *JASA*, V78, N383, September 1983, pp 605-10.

Comite Europeen de Normalisation, Stationary source emissions – determinatin of the mass concentratin of PCDDs/PCDFs – Part 3: Identification and quantification, EN 1948-3, December 1996.

Energy and Environmental Research Corporation (EER), Dioxins/Furans, HCl, Cl₂, and related testing at a hazardous waste burning light-weight aggregate kiln, USEPA contract No. 68-D2-0164, October 10, 1997.

Hagenmaier, J. P., "Validation of Sampling Equipment for Dioxin," in *Proceedings Municipal Waste Incineration*, October 1-2, 1987, Montreal, Quebec.

Keith, L. H. et. al., "Principles of Environmental Analysis," *Analytical Chemistry*, v55, 1983, pp 2210-2218.

Kennedy, E. R., et. al., Guidelines for air sampling and analytical method development and evaluation, NIOSH, Cincinnati, May 1995.

Lanier, W. S., and C. D. Hendrix, Reference method accuracy and precision (ReMAP) -- Phase I – precision manual air emission measurements, ASME, in process.

Midgett, M. R., "How EPA validates NSPS Methodology," *ES&T*, V11, N7, 1977, pp 655-659.

Mitchell, W. J. and M. R. Midgett, "Means to evaluate performance of stationary source test methods," ES&T, V10, N1, January 1976, pp 85-88.

National Institute for Occupational Safety and Health, NIOSH Manual of Analytical Methods, 4th ed. DHHS (NIOSH) Publication 94-114, August 1994.

Natrella, M. G., Experimental Statistics, NBS Handbook 91, USGPO, 1966, pp 1-10 and 5-13 thru -27.

Rigo H. G. and A. J. Chandler, Retrofit of waste-to-energy facilities equipped with electrostatic precipitators – sensible heat exchanger cooling – Phase II, ASME, CRTD-Vol. 44, 1997.

USEPA, EPA regulations on standards of performance for new stationary sources – Subparts A-Z, Code of Federal Regulations, V40, Part 60.8(f).

USEPA, National emissions standards for hazardous air pollutants: compliance extensions for early reductions, Federal Register, V57, N250, December 29, 1992, 61989-91.

USEPA, Method 301 – Field validation of pollutant measurement methods from various waste media, Code of Federal Regulations, V40, Part 63, Appendix A, 1992.

USEPA, Final Technical Support Document for HWC MACT Standards, Vol VIII, 1999.