Establishing Data-Derived Emissions Limitations

H. Gregor Rigo
Rigo & Rigo Associates, Inc.
1 Berea Commons, Suite 211
Berea, Ohio 44017

and

Gary Liberson
Environmental Risk Sciences, Inc.
1000 Thomas Jefferson St., NW, Suite 506
Washington, D.C.20007

INTRODUCTION

The emissions limitations found in regulations and permits have traditionally been set using engineering judgment, commercial considerations and public perception of potential harm and achievability. Given the restricted list of pollutants in Section 129 of the Clean Air Act (CAA) and EPA's decision to base municipal waste combustor (MWC) Maximum Achievable Control Technology (MACT) Floor determinations on limitations in enforceable permits, the need for determining these values from test data seems to be past. However, this does not stop local jurisdictions from imposing additional requirements. Local regulatory bodies and concerned citizens frequently want to regulate pollutants that are not addressable using EPA's permit limit procedure; too few permits include limitations on specific emissions. The need to correctly calculate achievable emissions remains. At the very least, it is imperative that correctly calculated limitations be placed in the administrative record supporting a permit so that if a problem occurs, a facility retains the ability to defend itself. Fortunately, emissions limitations can be calculated using statistical techniques that consider both regulatory constraints and source variability. These procedures provide a deterministic, objective link between test results and a lower limit that bounds achievable, data-derived regulatory and permit restrictions.

PHILOSOPHICAL FRAMEWORK

EPA set the MACT floor for MWCs by ranking individual pollutant limitations from existing permits from lowest to highest and selecting the 12 percent that were lowest based on the number of incinerators in a category. This group of permitted limits was then averaged. EPA's approach presumes that the permit limitations were being achieved and that the arithmetic average is routinely achievable. Averaged emissions limitations simply imply that some values are larger than the average and some are smaller. Consider, for example, averaging: 20, 20, 20, 20 and 10 to get 18. For this group, only one would be in compliance with the arithmetic average of the emissions limitations even when all five are operating at their permitted level. This points out an obviously unintended consequence of reading the adjective "average" (meaning "typical") in the CAA as the verb or noun meaning "arithmetic average". To establish an emissions limitation using the data average is clearly wrong.

To address this problem, a number of different approaches have been posited by the EPA and industry. The balance of this paper presents these approaches, along with a couple of additional conceptual approaches that could prove useful when faced with the problem of establishing permit limitations for previously unregulated pollutants.

STATISTICAL CONSIDERATIONS

Average emitted concentrations and variance estimates are derived from test data. Figures 1 and 2 show the theoretical statistical effect of uncertainty on the expected distribution of the sample average compared to the true value. The true distribution may be shifted to the left or right of the calculated average; it may also be flatter or more peaked because of the uncertainty with which the standard deviation is known. We can be reasonably certain, however, that the true mean is within the 95 percent statistical confidence level confidence limit for the average and the population standard deviation (the true variability of the sample) is similarly bounded by the 95 percent statistical confidence level confidence limit for the standard deviation. When these two sources of uncertainty are combined, the emissions limitation is correctly determined by the 95 percent statistical confidence level, percent coverage tolerance limits that contain a specified percentage of future tests at a given statistical confidence level result. The errors are incorporated by K-statistics[1,2].

Inherent Limitations of a Statistical Bound

Regardless of the quality of data and validity of the statistical procedures employed, there is a known chance that a well-operated plant will exceed the average. This is inherent in the statistical procedure. When working at the 95 percent statistical confidence level, there is a 5 percent chance that a valid measurement will be outside the calculated bound. Similarly, when limits are calculated to contain a specified percentage of future measurements, say 99 percent, there is still a one percent chance that a valid measurement will not be included when the plant operates as it did when tested to set the limits!

Consider the implications of these errors to a well-operated MWC with three units being tested for six reference method pollutants every year—a combined total of 18 tests per year must be passed. At the 95 statistical confidence level, one exceedance would be expected virtually every year when statistically perfect emissions limitations are promulgated since 1 out of 20 tests under the same conditions should result in a violation at the 95 percent confidence level. Similarly, if the limits are designed to contain 99 percent of the future test values at the 95 percent statistical confidence level, more than one violation of a set of consistently tight, but perfectly derived, emissions limitations can be expected.

Sources of Error

Characterizing stack emissions involves field sampling and recovery, sample handling, laboratory analysis, data entry, results calculation and interpretation. When three runs make up a performance test[3], the results are never exactly alike due to normal variability (random errors) introduced at each step of the characterization process. Analyzing emissions from different test conditions introduces another source of variability due to changes in the fundamental performance of the system. All the foregoing assumes that the same team and laboratory do all the testing. Different field teams, laboratories and chemical supply houses introduce additional sources of variability during retesting.

From a global perspective, available test data can be said to contain two broad types of error:

- within-test error—this characterizes the differences between the three runs or repeated measurements made close together in time to characterize a single test condition.
- between-test error—this characterizes the differences between the average obtained from a number of different tests taken under different conditions and at different times.

Each error source must be accommodated when emissions limitations are calculated. Just as performance is characterized by the average of the individual test runs, within-test error is characterized by the standard deviation of those same runs. Between-test variability is estimated as the standard deviation of the individual test condition averages for an individual plant. When these are combined, the overall source variability, including the effects of different test teams, laboratories, operating conditions and normal performance changes, are characterized. Combining either the between- or within-test variability from several test series or sources is known as pooling. This is a valid technique for maximizing the information.

Statistical Outliers

Even the best laboratories and field teams sometimes produce aberrant data. When there are enough replicates, a number of statistical tests can be employed to identify unusually large and small values. Leaving such values in the analysis bias the average and inflates the standard deviation. Removing such values leaves a facility exposed to exceedances for normal occurrences. This is an issue that requires careful attention and balance. Those interested in tight regulations will want outliers excluded. Those

interested in not being penalized for chance occurrences will want them included. Agreement on outlier identification and handling should be reached prior to applying any statistical treatment of the data.

Combining (Pooling) Data From Several Tests

Unless the overall variability (both the between- and within-test variability) are properly incorporated, the expected range of future test results cannot be characterized. Facilities that comprise the MACT Pool could routinely fail future tests even when tested under identical conditions if the emissions limitations are not properly set.

Emissions data sets typically contain results for a number of test conditions and units. Determination of the statistics for the pooled standard deviation and the effective number of runs for both the within- and between-test conditions can be accomplished with the same basic equations.

The data from individual test conditions are used to estimate within-test variance. Our review of the data indicates that the within-test standard deviation — estimated for each run using the natural logarithms of the data since the data are lognormally distributed[4] —can be treated as coming from an equal variance population.

Under this assumption, the pooled standard deviation (S_p) and effective number of runs (N_p) which characterize all test conditions (C), not just an individual test series, are defined by the following equations:

$$S_p = \sqrt{\frac{\sum (n-1)s_c^{\,2}}{\sum n - C}} \tag{1}$$

$$N_p = \sum n - C + 1 \tag{2}$$

Estimating the overall test condition standard deviation (S_o) and effective number of runs (N_o) by pooling the between- and within-test condition variability (S_b and S_w, respectively) is more complicated than simply pooling either of these two variability sources alone. Review of the available data indicates we cannot assume that the variances are equal. The pooling is done by using the following equations for combining two standard deviations when they cannot be assumed[5] equal:

$$S_o = \sqrt{\frac{(N_w - 1)S_w^2 + (N_b - 1)S_b^2}{(N_w + N_b - 2)}}, \tag{3}$$

$$N_o = 1 + \frac{(N_B\theta + N_w)^2}{\dfrac{(N_b\theta)^2}{N_w - 1} + \left(\dfrac{N_w}{N_b}\right)^2}, \text{ and} \tag{4}$$

$$\theta = \left(\frac{S_w}{S_b}\right)^2 \tag{5}$$

Ranking Test Series Averages

Before selecting best performance, the results of emissions testing have to be ranked. The arithmetic average of three runs complies with regulatory requirements for determining compliance, but it does not consider the precision with which a test result is known. In CETRED[6], EPA added twice the standard deviation to the test series average as an approximation of the value likely to contain the average with 95 percent confidence. Unfortunately, three times the standard deviation should have been added to account for the uncertainty associated with using the sample[7].

A better approach is to rank the facilities using the 95 percent statistical upper confidence limit for the average of lognormally distributed data as suggested by Land[8]. The following equations by Rigo[9] produce the same numeric estimates for the upper confidence limit, but have the advantage of being implementable using standard statistical tables and functions available in modern spreadsheets like Excel®:

$$Plus2 = \overline{X} + 2 \times S \tag{6}$$

$$Plus3 = \overline{X} + 3 \times S \tag{7}$$

$$UCL_{Land} = \mu + \tfrac{1}{2}\sigma^2 \left(t_{\nu,\alpha/2}\sigma^2 + \left[\tfrac{1}{2}\sigma^2 \left\{ \frac{N-1}{\chi^2_{N-1,1-\alpha}} - 1 \right\} \right]^2 \right)^{1/2} \tag{8}$$

Experience using the various formulations indicates that *Plus3* and UCL_{Land} produce generally similar rankings. *Plus2*, on the other hand, frequently differs materially from either the correct nonparametric limiting value and the upper confidence limit on the mean for lognormally distributed data.

Method Precision—The Accuracy of Reported Numbers

Test results that are below the analytical laboratory detection limit (ADL), and even those below the Reference Method Practical Quantitation Limit (RMPQL), contribute no meaningful information about the operation of a source since these results are either indistinguishable from zero or are too variable (error greater than ±30 percent for methods evaluated in accordance with 40 CFR 63, Method 301 Method Validation). These results must be eliminated from the data in order to calculate stable between- and within-test estimates of variability. When values below the ADL are eliminated, much of the "data" currently being used to establish variability are lost.

Correct Statistical Confidence Level and Interpretation of Data-Derived Limits

In the proposed Hazardous Waste Combustor Rule, there are five data-derived emissions limitations that must be met simultaneously (mercury, lead, semivolatile metals, low-volatile metals and total chlorine). For the statistics of the proposed rule to be at the 95 percent confidence level recommended by OSW for other RCRA applications and by the Office of Air Programs for New Source Performance Standards and Guidelines for Existing Sources (see 40 CFR 60 Appendix A, Method 19, for example), individual data-derived MACT Floor limits must be set at the 99 percent statistical confidence level.

Gastwirth[10] explains that "the probability that the *joint procedure* rejects" (the null hypothesis that a plant is in compliance) is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)^{11} \tag{9}$$

The above equation can be extended to cover the situation where 5 test conditions must be simultaneous met at the 95 percent statistical confidence level.

Using the Bonferroni approximation, the significance level for *k* comparisons to meet the intended α level is:

$$\alpha^* = \alpha/k \quad (10)$$

The intended overall statistical significance is achieved by using a statistical significance — α^* — of 0.01, which is the 99 percent statistical confidence level.

Data-derived emissions limitations provide floors below which no one should write or accept a limit in a regulation permit[12]. A data-derived emissions limitation based permit must include either additional margin or a provision that passage of a prompt retest deems the initial test a statistical aberration that is not subject to citizen suits and enforcement actions.

Pitfalls of Pollutant-by-Pollutant Calculations

A pollutant-by-pollutant (also know as HAP-by-HAP) approach to setting emissions limits precludes consideration of confounding and conflicting affects between and among the various proposed limitation. Potentially conflicting parameters must be considered when setting standards for various constituents. When pollutants are controlled by different techniques, it may be appropriate to analyze them separately. However, this decision must be re-verified for conflicts in performance of individual control technologies applied in series. For instance, the design of air pollution control devices [APCDs] to control particulates, semi-volatile metals and low volatile metals is the antithesis of designs needed to minimize the synthesis of dioxins and furans. Larger ESPs and baghouses reduce particulate and particulate related pollutant emissions. Unfortunately, these larger APCDs also increase the amount of time the flue gas is held at elevated temperature before being exhausted to the environment. More dioxins and furans are formed in ESPs and baghouses operating at the same temperature which have larger specific collector areas (SCAs) and lower air-to-cloth ratios (ACRs).

It must be recognized that several pollutants are controlled by the same device. Overlooking this point can quickly lead to the ridiculous conclusion that an individual source must be simultaneously equipped with ESPs (lower dioxins) and FFs (lower metals), but not both! This clear technical impossibility is easily avoided if a common pool of best performing facilities or at least a common set of characteristics for those plants is used.

ALTERNATIVE APPROACHES TO CALCULATING A DATA-DERIVED EMISSIONS LIMITATION

Over the past few years, a number of different methods of calculating data-derived emission limitations have been forwarded. These were originally developed to lend a scientific basis to a historically judgmental problem. Since the passage of the 1990 Clean Air Act Amendments with the requirement to establish Maximum Achievable Control Technology (MACT) floors for historically unregulated sources, the attention paid to this problem has increased. The following brings together the most recent versions of the concepts currently in use.

Combined Statistical—Technical Procedure

The combined statistical—technical procedure used by EPA in the Hazardous Waste Combustion Rule procedure is conceptually simple. All the data for a single pollutant and category are arrayed in ascending order. Additional test averages are included until 6% of the facilities for which complete data

are available are included. This means that several test conditions for some facilities will make it into the pool and some low emitting facilities will be excluded because of incomplete emissions characterization data. All other facilities having similar design and feed characteristics are identified and brought into the facility pool. These additional facilities with similar characteristics to the best performing group are called the expanded universe (EU).

The average emissions from the highest emitting test condition among this group of deemed equivalent facilities are used as a proxy for between-test variability. The data-derived MACT emissions limitations is set equal to the data-derived emissions limitations for that test condition using the pooled within-test variability for all included test conditions. Unfortunately, this specific formulation proffered by EPA did not recognize that the pooled within-test standard deviation is only an estimate of the population standard deviation and used the normal deviate ($\phi^{-1}(0.99)$) instead of the appropriate tolerance interval ($K_{0.055,\,0.99,\,N}$).

Equation (11) was used by EPA to determine data-derived emissions limitations and design values (annual average emissions) from the data used to characterize the pool facilities[13].

$$EL_{EPA} = \exp\left[\mu_3 + \phi^{-1}(0.99)\sigma_3\sqrt{1+\frac{3}{N}}\right] \tag{11}$$

where:

$$\mu_3 = \log(\bar{X}) - \frac{1}{2}\sigma_3^2 \tag{12}$$

$$\sigma_3^2 = \log\left[\frac{S^2}{3[\bar{X}]^2}+1\right] \tag{13}$$

$$\bar{X} = \exp\left(\mu_m + \frac{1}{2}\sigma_0^2\right) \tag{14}$$

$$S^2 = \left(\exp(\sigma_0^2)-1\right)\exp(2\mu+\sigma_0^2) \tag{15}$$

$$\bar{X} = \exp\left\{\log(EL_{EPA}) + \frac{1}{2}\log\left(\frac{1}{3}\exp(\sigma_0^2)+\frac{2}{3}\right) - \phi^{-1}(0.99)\left[\log\left(\frac{1}{3}\exp(\sigma_0^2)+\frac{2}{3}\right)\right]^{1/2}\right\} \tag{16}$$

In the published equation for S^2, σ_o instead of σ_o^2 was incorrectly used in two in places in the equation when compared to the published[14] derivation for S^2.

When $\bar{X}$ is calculated, the mean and standard deviation of the logarithms of the data for the highest emitting condition in the pool facilities are used. A more technically correct approach uses the highest test-condition average in the pool along with the pooled geometric standard deviation to estimate the internally consistent geometric mean for the highest emitting source. Of course, if the geometric standard deviation for the highest emitting source and the pool variance are similar, this refinement is not needed. Given the presence of outliers which can inflate variances, the refinement is prudent and turns out to be necessary in many cases.

As discussed previously, $\phi^{-1}(0.99)$ is a reasonable statistic to use when there is a very large emissions database that produces, say, more than 50 degrees of freedom. When fewer runs are available, the upper *two-sided tolerance limit K-Statistic for 99 percent coverage, 95 percent statistical confidence level and*

N runs should be used instead. Tabulated values can be found in previously referenced books by Natrella, and Hahn and Meeker.

An alternative to the combined statistical—technical procedure used for some pollutants was dubbed a breakpoint analysis. Here the average emitted concentrations, or preferably some related statistic that considers variability, is ranked from smallest to largest. Different equations[15] are used to estimate the slope (*SL*) and inflection (*I*) points:

$$SL = X_{i+1} - X_{i-1} \tag{17}$$

$$I = X_{i+1} + X_{i-1} - 2X_i \tag{18}$$

By plotting *SL* and *I* versus rank on the same graph, the slope of various segments of the line describing the data is visualized. Slope changes are marked by jumps in the value of *I* which mark inflection points in the curve. The breakpoint is marked by the largest value of *I* and the corresponding emitted concentration is X_i. Once X_i is identified, if it does not already explicitly incorporate uncertainty, the relevant concentration is described by previous Equation (11) except that $\bar{X}$ is replaced by X_i and α_o is the appropriate pooled value. Consideration should be given to using the median of the individual series standard deviations in this context.

12% MACT Approach

The 12% MACT approach differs from the HWC Rule approach in that progressively higher emitting test conditions are collected until 12% of the facilities for which data are available, but not less than 5 under Sections 112 and 129 of the CAA, are included. An expanded pool of similar facilities is then established and the data-derived emissions limitation is then calculated using the average of the EU emissions and within-test variability. Adjustments for the effect of below detection limits data using the δ-log approach is recommended by EPA. This approach obviously does not consider between-test variability, even via a proxy. Thus, it ignores good engineering and statistical practice and procedures. It is not considered viable by these authors and is not discussed further.

CETRED Approach

When the Agency began the regulatory process for HWCs, they published CETRED[16]. In that report, most of the particulate and PCDD/F data found in EPA's emissions database were analyzed. The methodology employed was similar to EPA's current 12% MACT approach except that the results of all test conditions for individual facilities were treated as a single characterization. The data-derived emissions limitations were done considering both between- and within-facility variance[17].

Most facilities have been tested several times and under various operating conditions. When the mean of the test series averages is calculated, it generally falls very close to the average developed by simply averaging each data point for all test conditions. This is an expected result since most test series involve the same number of runs and it doesn't matter how an overall average is calculated as long as each data point is included the same number of times.

The standard deviation, however, can and does explode (dramatically increase) when the results for individual test conditions are tightly clustered, but the various run conditions are widely separated. This is because the calculated standard deviation includes both the between- and within-test sources of variance.

In CETRED, the statistical methodology considered both the variability exhibited within each test series and between all test conditions contained in the best performing group using a tolerance limit formulation developed by Vangel. Vangel's method requires that the same number of runs have been conducted in each test. Since the majority of emissions test series use three runs, a statistically incorrect but practically inconsequential error is introduced by using the average number of runs in the formulas even though it is physically impossible to have a fractional number of runs. The emissions limitation is:

$$EL_{CETRED} = \bar{\bar{x}} + k^{*}\sigma_{x} \tag{19}$$

where: $\bar{X}_i = \dfrac{\sum X_{ij}}{iC}$ (20)

$S_1^2 = C\sum \dfrac{(\bar{x} - \bar{X}_i)^2}{i-1}$, the variance calculated between test series averages (21)

$S_2^2 = \sum\sum \dfrac{(X_{ij} - \bar{X}_i)^2}{i(C-1)}$, the pooled within test series variance (22)

$$Q = \frac{S_1^2}{S_2^2} \tag{23}$$

$$W = \left(1 + (C-1)/Q\right)^{-1/2} \tag{24}$$

$$k^{*} = MAX\left(Kc_{p,1-\sigma,ic}, \left[K_{p,1-\alpha,ij} - K_{p,1-\alpha,c}/\sqrt{C} + \left(K_{p,1-\alpha,C} - K_{p,1-\alpha,iC}\right)W\right] / \left(1 - 1/\sqrt{C}\right)\right) \tag{25}$$

CKRC Approach

EPA previously rejected the CETRED approach because no facility simultaneously met the data-derived emissions limitations when it was applied to hazardous waste combustors (cement kilns, incinerators, light weight aggregate kilns and boilers) in CETRED. An obvious touchstone, a valid calculated result, was failed. The approach, however, can be made workable. ERS[18] proposed a modification which implements the CETRED approach that considers both between-test condition variability and the use of a single group of facilities to establish simultaneously achievable emissions limitations for co-controlled pollutants. The results allow well-operated facilities to simultaneously meet the calculated emissions limitations.

Instead of using the highest emitting source in the EU as a proxy for between-test variability, the CETRED approach can be correctly applied by accounting for both between- and within-test variability and recognizing that the data are lognormally distributed. As done by EPA in CETRED, the Cement Kiln Recycling Coalition [CKRC] centered the distribution on the grand mean (average of the averages). All test conditions for the best performing 12 percent of the units in the hazardous waste burning cement kiln universe were incorporated. Of course, if more than 12 percent display below reference method practical quantitation limits (RMPQL) emissions, then they must all be included. Because some pollutants are co-controlled, a common pool for all pollutants should be defined so that internally consistent and technically compatible emissions limitations can be developed.

In the ERS report, the upper confidence limit for the average particulate emissions from each facility was used to characterize test series emissions. Consistent with EPA's methodology, these sources are ranked from lowest to highest. EPA, however, further subdivides the test results by test condition and only includes those that meet the Agency's criteria rather than all permitted plant operating conditions.

Any permitted operating condition for a facility must remain in the MACT analysis to properly characterize source emissions and not just the part of them which happen to occur under some test condition. That is, it is reasonable to statistically analyze each operating condition separately, but once a facility meets the conditions used to establish a pool facility for one constituent it should remain in for all constituents. All the operating conditions for a pool facility should also be included, unless it can be demonstrated that a specific controllable operating condition resulted in previously unacceptable emissions. All measured emissions are inherent characteristics of the source. You cannot simply declare an emissions set unrepresentative unless there is a specific action the operator can take to preclude such emissions. Put differently, when the pool is selected, EPA should not stop when the required number of facilities are included. They should keep adding facilities until all test conditions for the required number of facilities are included.

The MACT Floor should then be calculated using either the tolerance limit or prediction limit as described by the Office of Solid Waste[19]. The tolerance limit as discussed under the 6% MACT floor is a limit designed to contain a specified proportion of the population (e.g., 99 percent of the future test results from all plants). Prediction limits, on the other hand, are designed to contain the next specified number of sample values from the characterized facility assuming that the statistical characteristics of the plant do not change (e.g., the next performance and trial burns between recertifications).

As a result of inherent physical characteristics of the emission process, emissions data tends to be lognormally distributed[20]. Consequently, any procedure used to derive emission limitations must consider the lognormal characteristics of the emissions. A procedure for developing statistical limits for lognormally distributed data is outlined by Land[21]. Land's procedure was followed and the following equations estimate emissions limitations from lognormally distributed data including consideration of the uncertainty with which the standard deviation is known. Simulation studies comparing EPA's HWC rule formulation to that previously published by Rigo[22], demonstrated that EPA's equations for going from individual run data to 3-run averages is superior and included in the following equations:

$$EL_{CKRC-T} = \exp\left(\mu^3 + K_{N,P,1-\alpha}\sigma_3\left[1 + 3/N\right]^{1/2}\right), \text{ and} \quad (26)$$

$$EL_{CKRC-P} = \exp\left(\mu^3 + t_{N,1-\alpha_k}\sigma_3\left[1 + 3/N\right]^{1/2}\right) \quad (27)$$

Both EL_{CKRC-T} and EL_{CKRC-P} are conceptually equivalent to EPA's EL_{EPA} emissions limitation. The numerical values are different and the choice depends on whether the limit is to contain a percentage of all future tests or the next specified number of tests. The real question is whether the limitations should be based on the percentage of tests that will be in compliance from well-run, EU similar facilities or should the Agency be concerned about the number of statistical failures an individual facility is likely to face between permit renewals. Hahn and Meeker use the analogy of an astronaut who is not concerned about the amount of fuel consumed in an average trip to the moon or the percentage of trips that can be successfully completed with a given amount of fuel. Instead, like plant managers, astronauts are concerned about getting home on the fuel carried on their trip (the fines they will receive or jail time they will serve as a result of the number of times their plants will be tested during that manager's watch).

The $EL_{CKRC\text{-}P}$ is the correct limit to choose if a mechanism for dealing with statistically predictable exceedances is not included in a permit or regulation.

A Modified MWC MACT Approach

The MWC MACT floor was developed by taking the average of the lowest 12 percent of permitted emissions limitations. When establishing limits for previously unregulated pollutants, a conceptually consistent approach is to use the pooled within-test variability for all similarly equipped units in Equation (11) to estimate the 3-run average emissions limitation for each unit. The highest emitting source is selected to estimate the emissions limitation.

When examining the set of data-derived emissions limitations, it is important to understand the meaning of the adjective "average" in the phrase "average emissions limitation." Failure to do so can result in selecting a limit that is only achievable by a single facility — this can happen if one plant has aberrantly low emissions results and the balance are characterized by almost identical emissions profiles—and the arithmetic average is calculated. The arithmetic average will only be above one plant and not inclusive of the 12 percent specified in Sections 112 and 129 of the Clean Air Act. The HAP-by-HAP problem also remains. If a common set of facilities are not used to establish emissions limitations for co-controlled pollutants, the resulting ménage may not be routinely achievable by any facility, much less by the intended number.

Instead of averaging the emissions limits that characterize the best performing 12 percent; the correct approach is to use the highest data-derived emissions limitation for the pool facilities as the relevant limitation. This approach maintains consistency with EPA's previous practice in MWC regulations while making maximum use of the available data and avoiding the illogical result where only one of the facilities used to establish the emissions limitations is likely to meet the value.

COMPARATIVE RESULTS

Table 1 is a listing of 3-run average naphalene data obtained from tests conducted at MWCs. The concentrations are in ng/dsm^3 @ 7% O_2, the units used to quantify dioxins. To convert these values to ppm_{dv} @ 7% O_2, simply multiply by 1.88x10-9, 20,000 ng/dsm^3. Naphalene, for example, is 0.004 ppm_{dv} (3.8 ppb_{dv}) which can be compared to the OSHA PEL[23] of 10 ppm to provide an indication of the magnitude of these emissions.

The emissions limitations calculated by several of the approaches are:

- CKRC (next 5 tests) — 9,195 ng/dsm^3 @ 7% O_2
- CETRED —10,250ng/dsm^3 @ 7% O_2
- Modified MWC (average) — 12,426 ng/dsm^3 @ 7% O_2
- CKRC (99% Coverage) —12,705 ng/dsm^3 @ 7% O_2
- Combined Statistical—Technical Procedures — 13,054 ng/dsm^3 @ 7% O_2
- Breakpoint Analysis — 15,645 ng/dsm^3 @ 7% O_2
- Combined Statistical—Technical Procedures (Unequal variances) — 18,915 ng/dsm^3 @ 7% O_2
- Modified MWC (highest limitation) — 21,824 ng/dsm^3 @ 7% O_2

There are clear differences in the numerical results, a little more than a 2:1 range from lowest to highest. These differences may not seem significant, but the low end is where precautionary principle advocates

target. The high end, plus a prudent margin is where plant managers and people faced with $25,000 per day per violation fines and penalties focus their attention.

While professional judgment is involved, the methods finally selected must conform to the Clean Air Act definitions if a regulated HAP is involved or simply for public understandability. The CKRC and HWC rule approaches seem to be the best because they are scientifically defensible and correctly treat the law's grammatical construction. Setting an arbitrarily low limit is not useful if it doesn't affect plant design or operation. Regardless, the emissions are going to be what they are.

CONCLUSIONS

The problem of correctly using data to establish emissions limitations is not yet fully resolved. Great strides have been made over the past couple of years. Regulators have progressed from the rules of thumb that set emissions limitations 20 percent (or ¼ inch if a graph was being used) above the highest measured value. Now, data-derived limits—based on Congressional direction that existing facilities should all become as clean as the "emission control that is achieved in practice by the best controlled 12 percent of existing sources" and new facilities should be as clean as the best one already out there—are being calculated.

Unfortunately, the failure to recognize that data-derived emissions limitations are derived from data, has produced a number of strained interpretations of the Clean Air Act, Sections 112 and 129 designed to fulfill non-technical objectives. Clean proposals have been made including:

- EPA calculating statistical emissions limits for HWCs in CETRED that recognized source variability, even if the final produce was untenable because co-controlled pollutants were treated as if they were independent instead of being linked.
- EPA using the average of permit limitations in the MWC rule — while we can argue about the propriety of using the arithmetic average, the approach clearly recognizes that there is a necessary gap between the best performance ever measured and limits that can be achieved during routine testing.
- The HWC rule has taken this a step further by using data-derived emissions limitation based on the limit statistically achievable by the dirtiest plant in the clean pool. The clean pool is defined as all facilities with technology similar to the best performing group. This is essentially a proxy for considering between-test and plant variability in addition to within-test variability explicitly handled by the approach. Again, we can argue about the size of the pool (6% or 12%), but the concept of including all like facilities is sound.
- CKRC has extended EPA's original data-derived emissions limitation approach to recognize that the emissions characteristics of facilities for all co-controlled pollutants should be analyzed as a group using proper statistical techniques.
- A variant on the MWC and HWC rule and CKRC approaches that accounts for simultaneous control characteristics while preserving the basic frame-work of the MWC rule approach is suggested.

Each of these approaches produces different numerical results. Applied with intelligence, the results are generally comparable. Regardless, it must be recognized that data-derived emissions limitations have built-in probabilities of finding exceedances when a facility is operating exactly as it did when the limitation was established. Consequently, it is imperative that either margin be added or a violation declared only when a re-test is failed.

Needlessly low limits, however, do not affect a material improvement in the environment. They discourage innovation and waste resources. On the other hand, they do produce revenue from unavoidable fines and jobs from more testing. The method of calculating data-derived emissions limitations should provide an ample margin of safety against falsely finding violations. Public policy should not affect the way data-derived emission limitations are calculated, rather it should affect the margin of safety allowed and the decision to impose more stringent emissions limitations than the MACT floor.

NOMENCLATURE

σ is the standard deviation of the natural logarithms of the data.; $\sigma^2 = \ln\left(1 + \left[s/\ \bar{X}\right]^2\right)$

S is the standard deviation of the untransformed data; $S^2 = \exp^{(2\mu+\sigma^2)}\left(e^{\sigma^2} - 1\right)$

$\bar{X}$ is the arithmetic average of the concentrations; $\bar{X} = \exp^{(\mu+\frac{1}{2}\sigma^2}$

σ is the standard deviation of the log transformed data; $\sigma^2 = \ln\left(1 + \left[\frac{S}{\bar{X}}\right]^2\right)$

σ_3 is the standard deviation of the log transformed data associated with three-run averages derived from the pooled σ; see Equation (11)

EL_{EPA} is the data-derived emissions limitation using EPA's April 19, 1996 HWC procedure

EL_{CETRED} is the date-derived emissions limitation used by EPA in CETRED

$EL_{CKRC\text{-}T}$ is the data-derived tolerance limit (% coverage)

$EL_{CKRC\text{-}P}$ is the data-derived prediction limit (number of future tests covered)

EL_{MWC} is the average of the emissions limitations calculated using the pooled variance and each test series average then using the largest result

$\phi^{-1}(.99)$ is the 99th percentile of the normal probability distribution

N is the effective number of runs used to estimate σ

i is the number of runs conducted for each test condition

C is the number of test conditions in each average

k is the number of future test series to be contained

m is the number of runs to be averaged in a test series

P is the percentage of future test series to be contained

$t_{N-1,\alpha/k}$ is the t-statistic with the Bonferroni multiple approximation so it approximately equals the prediction limit for k future tests

$K_{N,P,1-\alpha}$ is the tolerance limit coefficient designed to include P percent of future occurrences at the 1-α statistical confidence level

X_i is the concentration for the ith ranked result

X_{i-1} is the next smaller concentration

X_{i+1} is the next larger concentration.

REFERENCES AND NOTES

1. N.G. Natrella, National Bureau of Standards Handbook No. 91, Experimental Statistics, USGPO, 1966.
2. Hahn & Meeker, Statistical Intervals, A Guide for Practitioners, McGraw-Hill, 1990.
3. Federal Regulations (40 CFR 60.8(f)) defines the results of a performance test as the arithmetic average of three individual test runs.
4. A more precise statement is that the assumption that the data are normally distributed can usually be rejected, but the assumption that the data were lognormally distributed cannot. It is impossible to prove what the underlying data distribution is; it is possible to demonstrate that an assumed distribution is not applicable.
5. Natrella, op cit., pp. 3-22 and 3-29.
6. USEPA/OSW, Combustion Emissions Technical Resource Document (CETRED), EPA530R-94-014, May 1994.
7. F. Pukelsheim, "The Three Sigma Rule," *The American Statistician*, Vol. 48, No. 2, pp. 88-91, May 1994.
8. C.E. Land, Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance, Selected Tables in Mathematical Statistics, Vol. III, 1975.
9. H.G. Rigo, "Selecting Statistically Meaningful Emission Rates," presented at Air & Waste Management Association's 86th Annual Meeting & Exhibition, Denver, CO, June 13-18, 1993.
10. J.L. Gastwirth, Statistical Reasoning in Law and Public Policy Volume 2 Tort Law, Evidence, and Health, Academic Press, 1988.
11. $\alpha = 0.05/5$; $0.05 = 0.01+0.01+0.01+0.01+0.01-0.01^5$
12. Rigo, op. cit.
13. USEPA, Draft Technical Support Document for HWC MACT Standards, Volume III: Selection of MACT Standards and Technologies, February 1996, pp. C-11, C-12.
14. J. Atchison, J.A.C. Brown, The Lognormal Distribution, Cambridge University Press, 1969, pp. 8, eq 2.7.
15. D.N. deG. Allen, Relaxation Methods, McGraw-Hill, 1954.
16. CETRED, op.cit.
17. M.G. Vangel, "Limits for a One-Way Balanced Random-Effects ANOVA Model," Technometrics, May 1992, V34, N2, pp. 176-185.
18. Environmental Risk Sciences, Inc., Environomics, Rigo & Rigo Associates, Inc., An Analysis of Technical Issues Pertaining to the Determination of MACT Standards for the Waste Recycling Segment of the Cement Industry, May 1995
19. EPA/OSW, Statistical Analysis of Ground Water Monitoring Data at RCRA Facilities, Interim Final Guidance, 1989 and Draft Addendum to Interim Final Guidance, 1992.
20. W.R. Ott, Environmental Statistics Analysis, Lewis Publishers, 1996.
21. Land, op. cit.

22. Rigo, op. cit.

23. U.S. Department of Labor, Air Contaminants—Permissible Exposure Limits, 29 CFR 1910.1000, 1989.

Table 1. Test series average naphthalene concentrations measured at MWCs.

FACILITY	TESTS	ARITHMETIC		LOGARITHMIC		Plus2	Plus3	UPPER CONFIDENCE LIMIT ON THE AVERAGE	TEST SERIES PREDICTION LIMIT FOR THE NEXT TEST	MWC APPROACH WITH CALCULATED LIMITS
		AVG.	STD.	AVG.	STD.					
151	3	340	16	5.828	0.046	372	388	369	446	2,546
24	2	1,741	6	7.462	0.003	1,752	1,758	1,765	1,872	13,047
24	2	1,746	91	7.464	0.052	1,928	2,019	2,635	6,957	13,072
24	3	2,772	47	7.927	0.017	2,865	2,912	2,852	3,048	20,773
175	3	1,016	371	6.880	0.359	1,758	2,129	3,940	16,536	7,288
24	4	1,713	598	7.388	0.416	2,909	3,508	3,979	11,127	12,112
175	3	1,487	444	7.271	0.330	2,375	2,818	4,787	18,266	10,774
175	12	2,356	2,941	7.254	1.020	8,238	11,179	5,864	19,502	10,598
168	9	3,742	3,302	7.977	0.706	10,347	13,649	8,320	19,464	21,824
136	6	2,005	1,116	7.397	0.815	4,238	5,354	8,376	29,833	12,225
175	3	2,114	790	7.611	0.369	3,694	4,483	8,787	38,064	15,133
67	3	5,888	1,224	8.667	0.203	8,336	9,560	9,878	23,442	43,514
168	7	5,511	4,338	8.436	0.584	14,186	18,523	10,207	24,785	34,540
433	3	6,688	1,291	8.796	0.189	9,271	10,562	10,633	23,834	49,516
136	4	1,602	1,020	7.213	0.677	3,642	4,662	11,379	51,657	10,166
91	3	5,974	1,386	8.675	0.250	8,747	10,133	12,368	35,440	43,886
24	3	4,141	1,318	8.303	0.324	6,777	8,095	12,934	48,378	30,234
166	4	10,468	2,402	9.233	0.257	15,272	17,674	15,645	30,023	76,662
85	3	441	324	5.874	0.655	1,090	1,414	27,070	221,319	2,665
24	2	2,807	385	7.935	0.138	3,577	3,962	33,401	289,563	20,935
168	9	7,963	12,222	8.471	1.013	32,407	44,629	36,365	101,991	35,790
24	2	2,017	320	7.603	0.160	2,658	2,978	54,777	583,494	15,019
24	2	1,756	295	7.464	0.169	2,346	2,641	69,932	802,491	13,068
91	3	8,866	4,381	8.976	0.478	17,629	22,010	85,017	499,153	59,259
24	3	30,264	11,862	10.271	0.364	53,987	65,849	121,494	518,520	216,450
4	3	29,446	12,453	10.243	0.436	54,351	66,803	207,987	1,096,749	210,558
91	3	25,636	13,920	10.056	0.531	53,476	67,396	421,886	2,793,124	174,573
83	5	629,971	257,853	13.294	0.376	1,145,676	1,403,529	1,035,528	2,165,717	4,447,403

Figure 1. Effect of location uncertainty on a normal distribution.

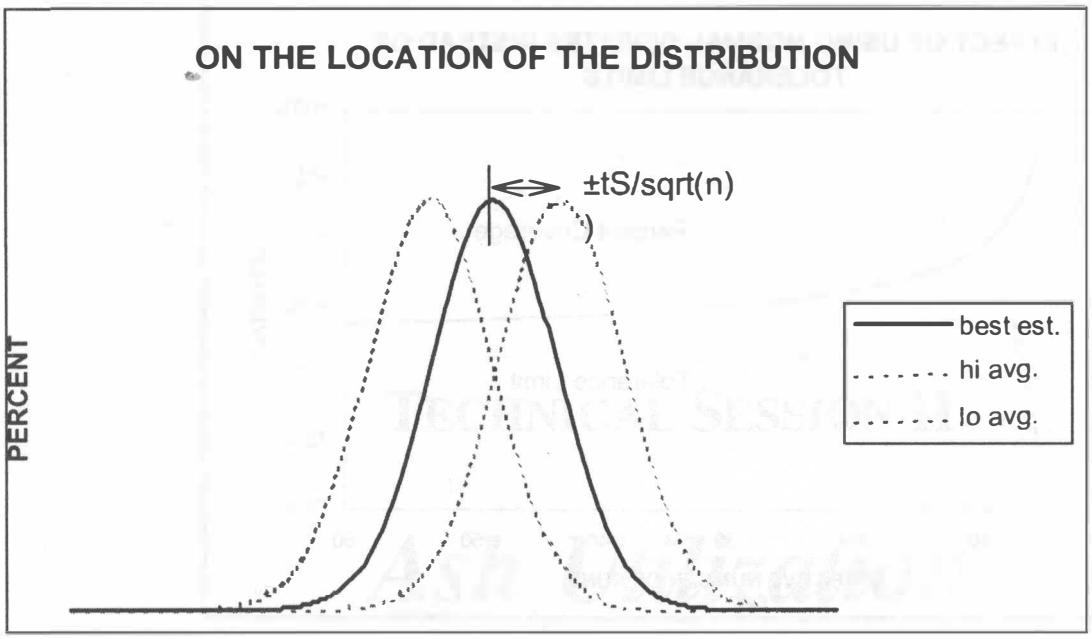

Figure 2. Effect of spread (standard deviation) uncertainty on a normal distribution.

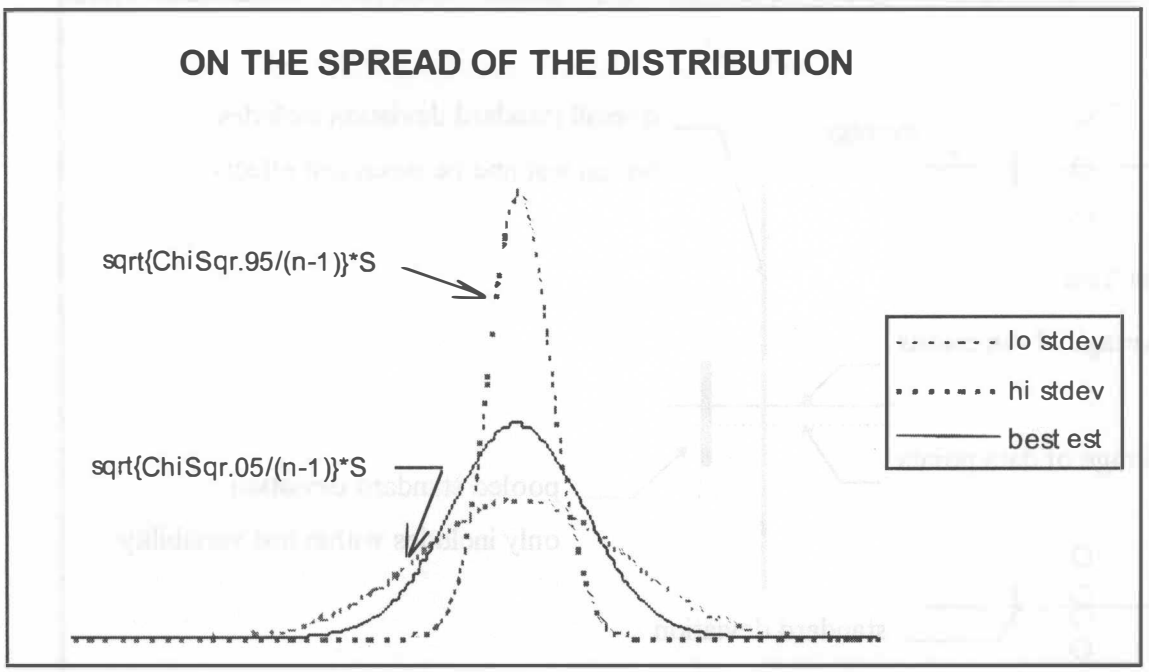

Figure 3. The difference between normal deviates and tolerance limits.

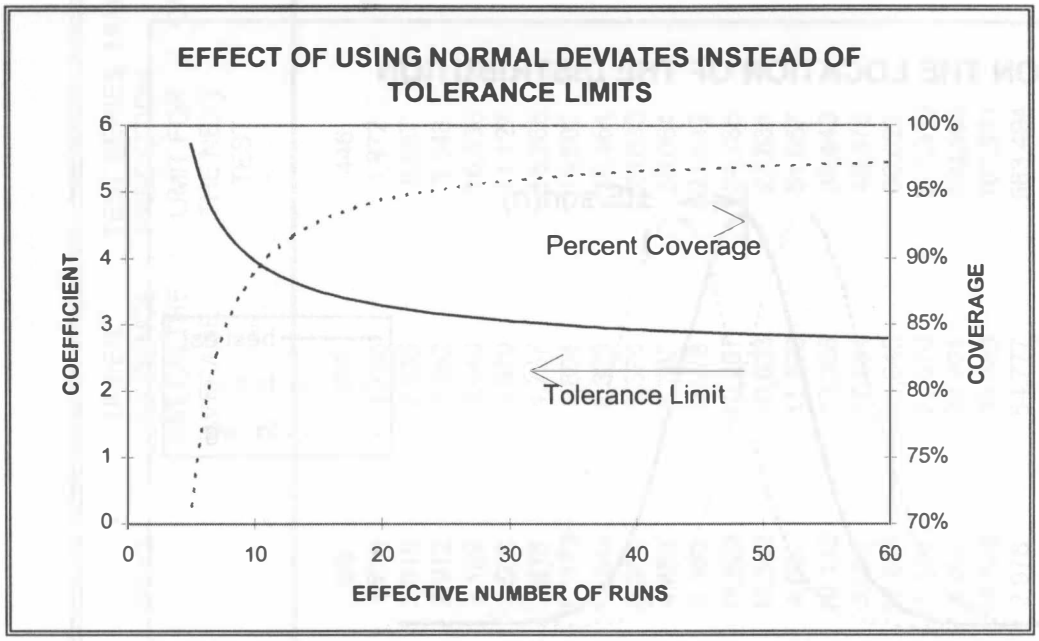

Figure 4. Conceptualization of the difference between the sources of variance.