August 28, 2002

MEMORANDUM

To: William Maxwell, EPA/OAQPS/ESD/CG

From: Jeffrey Cole, RTI

Subject: Statistical Analysis of Mercury Test Data Variability in Support of a Determination of the MACT Floor for the Regulation of Mercury Air Emissions from Coal-Fired Electric Utility Plants

Background

In 1999, the U.S. Environmental Protection Agency (EPA) conducted the Electric Utility Steam Generating Unit Mercury Emissions Information Collection Effort (EU/ICE) to gather information about mercury emissions from the coal-fired electric utility industry. This effort led to the collection of stack test reports on 80 furnace/boiler units. The EPA is currently using the results of these tests to determine the maximum achievable control technology (MACT) floor for the regulation of mercury emissions. The EPA is now seeking to quantify the uncertainty component that should be added to the mean values of the best 12 percent of the units chosen for MACT floor.

Objective

The objective of this analysis is to evaluate the variability in the determination of the average performance of the best 12 percent of 80 units that were tested under the 1999 EU/ICE (Scenario 1). After this, RTI will apply the statistical method used in Scenario 1 to the best units in two subcategorization scenarios: Scenario 2 - subcategorization by coal type with fluidized-bed combustors (FBC) included, and Scenario 3 - subcategorization by coal type without FBC. Note: this is only one method of addressing the variability or uncertainty associated with emission performance testing. Future memoranda may discuss other possible methods.

PROCEDURE

Basis for MACT Floor (Scenario 1)

In the determination of the MACT floor for the existing coal-fired power plant units, RTI used the results of the emissions tests for a set of 80 power plant units (Scenario 1). The air emissions of mercury (lb Hg/trillion Btu, derived using F-factors) were evaluated for each of the 80 tests, and the best 10 units (i.e., top 12 percent of 80 units) were identified. The best ten units are identified in Table 1, together with their average (mean) air emissions of mercury.

Types of Uncertainty

In Scenario 1, the top 12 percent of the data are the best units. When we identify the top units in MACT floor analysis, the average of those best 12 percent of the units is the first estimate of the MACT floor. In Table 1, this average mercury emission rate is 0.175 lb Hg/TBtu. This average number has two elements of uncertainty associated with the value of the number. One uncertainty is the actual value of the long-term average of the best units, since there are only a few tests available that represent the best 12 percent of the units and there is measurement uncertainty associated with each test set of 3[a] measurements.

The second uncertainty is the variability of emissions for these best units under the worst foreseeable circumstances under normal operating conditions. This second uncertainty includes operational variability. There is no direct measurement of operational variability, although there was some operational variability included in the measurement uncertainty associated with each test set of three measurements.

Evaluating Uncertainty

RTI was required to identify and use a method to demonstrate a reasonable characterization of the top 12 percent of sources that the Act defines as the basis of the MACT floor. One key component of this methodology is the use of conventional statistical analysis of the two uncertainties that are identified here.

RTI has used a statistical model to evaluate the uncertainty in the data base and two components of the uncertainty have been evaluated:

1. the uncertainty due to measurement error (includes very limited operational variability), and
2. the uncertainty due to the evaluation of the air emissions at different locations.

[a] Three of the 80 units had only 2 usable data points because of apparent errors in transporting samples or in laboratory sample analysis. However none, of these 3 units are in the Scenario 1, top 12 percent (top 10 units).

The uncertainty due to measurement error (component 1 above) is presented in Table A-1 for each of the 10 best tests.

Table 1. Mercury Test Unit Emission Means for Top 10 Performing Plants (Scenario 1, Top 12 percent of 80 Units Tested) *

Rank	Test unit	Coal	Controls	ICR plant mercury emissions (lb/TBtu, based on F-factor)
1	Kline	Waste Bituminous	FF	0.0816
2	Scrubgrass	Waste Bituminous	FF	0.0936
3	Mecklenburg	Bituminous	SDA/FF	0.1062
4	Dwayne Collier Battle Cogeneration Facility	Bituminous	SDA/FF	0.1074
5	Valmont	Bituminous	FF	0.1268
6	Stockton	Bituminous - PetCoke	FF	0.1316
7	SEI - Birchwood	Bituminous	SDA/FF	0.2379
8	Intermountain	Bituminous	FF/FGD	0.2466
9	Logan	Bituminous	SDA/FF	0.2801
10	Salem	Bituminous	ESP-CS	0.3348
			Average	**0.175**

* Three of the 80 units had only 2 usable data points because of apparent errors in transporting samples or in laboratory sample analysis. However, none of these 3 units are in the top 12 percent (top 10 units). Two of the 80 stack tests were done on the same unit at different times (Gibson Generating Station).

FF = Baghouse (fabric filter)
SDA = Dry lime/spray dryer adsorber
FGD = A flue gas desulfurization wet scrubber (lime or limestone)
ESP-CS = Electrostatic precipitator (cold-side, meaning this ESP is installed at a location downstream of the air preheater)

Estimating Measurement Uncertainty

When only a few test sets are available in the entire data pool for the evaluation of the performance of the best 12 percent of the units due to industry subcategorization or other factors, it is not possible to evaluate the uncertainty accurately using only the smaller data pool. As it appears that the measurement uncertainty is a function of the mercury level and the entire data set exhibits a wide variation in mercury levels, the entire data set rather than only 12 percent of the data base will provide a more reasonable estimate of the measurement uncertainty of the best 12 percent of the units. Consideration of a larger data base tends to reduce the statistical uncertainty. Equation 1 describes the statistical model that was obtained to characterize the measurement uncertainty for any test set of three measurements:

Equation **(1)**

$$\sigma = \sqrt{-0.00066 + 0.0239x^{1.4302}}$$

where F is the standard deviation of a normal distribution that describes the uncertainty that is associated with the measurement error, and x is the average of the mercury air emissions that were measured in the three tests.

The variability of emissions for these best units under the worst foreseeable circumstances (component 2 above) is estimated by an analysis of the entire data set. The uncertainty is found for all data, then is added as a portion of the uncertainty for the best 12 percent of units. The combined uncertainty components (measurement error and long-term operation) are added to the averages (means) of the best 12 percent and used to establish upper values for the 90 percent, 95 percent, and 99 percent confidence limits for the distribution of the best 12 percent of units.

Estimating Operational Uncertainty

In this statistical analysis, it is assumed that the long-term performance of the MACT unit will be better than or equal to the performance of the average of the best 12 percent of the tested 80 units that the Act defines as the basis of the MACT floor, and confidence limits due to unit performance variability will be added to this average through conventional statistical techniques. The operational uncertainty for the same plant operating at the average performance of the best units is assumed to be less than the unit-to-unit uncertainty of the entire set of best units. If the emissions of a unit are measured at a value greater than the upper 95 percent value, then there is less than 5 percent confidence that the unit is achieving the performance of the MACT unit. Application of the two uncertainty components to subcategorization scenarios are described below.

Results for Scenario 1

In the first scenario, RTI analyzed the MACT floor for all the existing coal fired power plant units (no subcategorization). In this analysis, the results from all EU/ICE emissions tests (80 emissions test reports) were used. Two of the 80 stack tests were done on the same unit at different times (Gibson Generating Station). The air emissions of mercury (lb Hg/trillion Btu, derived using F-factors) are evaluated for each of the 80 tests, and the best 10 units are identified. The best 10 units (top 12 percent) are identified in Table 1, together with their average (mean) air emissions of mercury.

The estimates of the upper limits for the average value of the performance of the top 10 units (upper 12 percent of 80 tests) are listed in Table 2. The values of these emission limits are based upon an average of 0.175 lb Hg/TBtu. Because of the measurement uncertainty in the average of the emission factors, these upper limits are somewhat greater than the average value, and depend on the confidence limit.

Results for Scenario 2 and 3

RTI then applied the same emissions variability methodology to two subcategorization scenarios: Scenario 2 - subcategorization by coal type with FBC included; and Scenario 3 - subcategorization by coal type without FBC.

In the second scenario, the MACT floors for the existing coal-fired power plant units subcategorized by fuel type, RTI used the results of the emissions tests for a set of 78 power plant units (34 bituminous-fired, 32 subbituminous-fired, and 12 lignite-fired). The two waste fuel-fired units are not included in this scenario. The air emissions of mercury (lb Hg/TBtu, derived using F-factors) are evaluated for each of the 34, 32, or 12 tests, respectively, and the average of the best units in each subcategory is identified. This averaging is done in one of two ways. If there are 30 or more units in the industry that match the subcategory, the average of the top 12 percent of the data sets available are used. If there are 29 or fewer units in the industry that match the subcategory, the average of the top 5 sets of available data is used. There are well over 30 bituminous- and subbituminous-fired units in the electric utility industry (1999), respectively. Thus, the bituminous (34 x 0.12 = 4.08 or 4 sets) and the subbituminous (32 x 0.12 = 3.84 or 4 sets) subcategories are based on 4 sets of 3 data points each. There are only 29 lignite-fired units in the electric utility industry (1999); therefore, the lignite subcategory is based on 5 sets of 3 data points (except Leland Olds with 2 data points, 14 data points total).

When the top 12 percent of the units in MACT floor analysis are identified (by subcategory), the average of the best 12 percent of the units is the first estimate of the MACT floor. In Table 3, the average mercury emission rates for bituminous-, subbituminous-, and lignite-fired (with FBCs) are 0.118, 0.764, 5.032 lb Hg/TBtu, respectively.

In the third scenario, the MACT floors for the existing coal-fired power plant units subcategorized by fuel type (without FBCs), the results of the emissions tests for a set of 74 power plant units (33 bituminous-fired, 31 subbituminous-fired, and 10 lignite-fired) were used. The air emissions of mercury (lb Hg/TBtu, derived using F-factors) are evaluated for each of the 33, 31, or 10 tests, respectively, and the average of the best units in each subcategory is identified. This averaging is done one of two ways. If there are 30 or more units in the industry that match the subcategory, the average of the top 12 percent of the data sets available are used. If there are 29 or fewer units in the industry that match the subcategory, the average of the top 5 sets of the data available is used. There are well over 30 bituminous- and subbituminous-fired units in the electric utility industry (1999), respectively. Thus, the bituminous (33 x 0.12 = 3.72 or 4 sets) and the subbituminous (31 x 0.12 = 3.84 or 4 sets) subcategories are based on 4 sets of 3 data points each. There are only 29 lignite-fired units in the electric utility industry (1999), therefore, the lignite subcategory is based on 5 sets of 3 data points (except Leland Olds with 2 data points, 14 data points total).

Table 2. Resulting potential MACT floor levels for all data that incorporate variability at various confidence limits (lb/TBtu) *

Mean of the best 12%	90% limit	95% limit	99% limit
0.175	0.232	0.251	0.292

* Based on 238 data points.

When the top 12 percent of the units in MACT floor analysis are identified (by subcategory), the average of the best 12 percent of the units is the first estimate of the MACT floor. In Table 4, the average mercury emission rates for bituminous-, subbituminous-, and lignite-fired (without FBCs) are 0.145, 1.048, 6.403 lb Hg/TBtu, respectively.

The estimates of the upper limits for the average value of performance for the top best performing units (upper 12 percent or equivalent of appropriate tests) in scenarios 2 and 3 are shown in Table 5 and 6. Because of the combined measurement and operational uncertainty in the averages of the emission factors, these upper limits range from about 12 percent to 146 percent greater than their average values, depending on the confidence limit and scenario.

The details of the associated calculations are presented in the attachment to this technical memorandum and the results for Scenario 1 are presented in Table A-5 of that attachment. This methodology is described as Concept A, Approach 2, and Model 2. Scenarios 2 and 3 are also based on the same methodology (equations) but were calculated by spreadsheet.

Possible Future Actions

If the EPA elects to provide emission-averaging procedures, statistical techniques may be used to further adjust the upper values of 90 percent, 95 percent, and 99 percent confidence limits on the distribution of the best 12 percent of units to reflect the effect of emission averaging on the upper values. Other approaches to establish MACT floors may include technical analyses of emissions reduction performance based on elements such as feed composition, operational characteristics of single or combined control systems, combustion effects, and data transformation.

Table 3. Mercury Test Unit Emission Means of Top Performing Plants for Subcategorization by Coal Type, Including FBC

Rank	Test unit	Air pollution control(s) and/or furnace type	ICR Plant mercury emissions means (lb/TBtu, based on F-factor)
Bituminous-fired Units			
1	Mecklenburg	SDA/FF	0.1062
2	Dwayne Collier Battle Cogeneration Facility	SDA/FF	0.1074
3	Valmont	FF	0.1268
4	Stockton	FBC/FF	0.1316
		Average	**0.118**
Subbituminous-fired Units			
1	AES Hawaii	FBC/FF	0.4606
2	Clay Boswell 2	FF	0.6633
3	Craig 3	SDA/FF	0.7248
4	Cholla 3	ESP-HS	1.2066
		Average	**0.764**
Lignite-fired Units			
1	R.M. Heskett	FBC/ESP-CS	3.9768
2	Antelope Valley	SDA/FF	4.0042
3	Leland Olds	ESP-CS	4.0233
4	Stanton Station 10	SDA/FF	6.2517
5	Stanton Station 1	ESP-CS	6.9024
		Average	**5.032**

FF = Baghouse (fabric filter)
SDA = Dry lime/spray dryer adsorber
FGD = A flue gas desulfurization wet scrubber (lime or limestone)
ESP-CS = Electrostatic precipitator (cold-side, meaning this ESP is installed at a location downstream of the air preheater)
ESP-HS = Electrostatic precipitator (hot-side, meaning this ESP is installed at a location upstream of the air preheater)

Table 4. Mercury Test Unit Emission Means of Top Performing Plants for Subcategorization by Coal Type, Excluding FBC

Rank	Test unit	Air pollution control(s) and/or furnace type	ICR Plant mercury emissions (lb/TBtu, based on F-factor)
Bituminous-fired Units			
1	Mecklenburg	SDA/FF	0.1062
2	Dwayne Collier Battle Cogeneration Facility	SDA/FF	0.1074
3	Valmont	FF	0.1268
4	SEI-Birchwood	SDA/FF	0.2379
		Average	**0.145**
Subbituminous-fired Units			
1	Clay Boswell 2	FF	0.6633
2	Craig 3	SDA/FF	0.7248
3	Cholla 3	ESP-HS	1.2066
4	Craig 1	ESP-HS/FGD	1.5955
		Average	**1.048**
Lignite-fired Units			
1	Antelope Valley	SDA/FF	4.0042
2	Leland Olds	ESP-CS	4.0233
3	Stanton Station 10	SDA/FF	6.2517
4	Stanton Station 1	ESP-CS	6.9024
5	Lewis & Clark	Particulate Scrubber (not FGD, no lime or limestone)	10.8315
		Average	**6.403**

FF = Baghouse (fabric filter)
SDA = Dry lime/spray dryer adsorber
FGD = A flue gas desulfurization wet scrubber (lime or limestone)
ESP-CS = Electrostatic precipitator (cold-side, meaning this ESP is installed at a location downstream of the air preheater)
ESP-HS = Electrostatic precipitator (hot-side, meaning this ESP is installed at a location upstream of the air preheater)

Table 5. Resulting potential MACT floor levels by fuel that incorporate variability at various confidence limits w/FBC (lb/TBtu)

Fuel	Mean of the best 12%	90% limit	95% limit	99% limit
Bituminous *	0.118	0.132	0.138	0.157
Subbituminous *	0.764	1.102	1.250	1.703
Lignite **	5.032	6.379	6.905	8.324

* Based on 4 sets of 3 data points.
** Based on 5 sets of 3 data points (except Leland Olds with 2 data points, 14 data points total).

Table 6. Resulting potential MACT floor levels by fuel that incorporate variability at various confidence limits w/o FBC (lb/TBtu)

Fuel	Mean of the best 12%	90% limit	95% limit	99% limit
Bituminous *	0.145	0.221	0.255	0.357
Subbituminous *	1.048	1.459	1.638	2.188
Lignite **	6.403	8.528	9.358	11.597

* Based on 4 sets of 3 data points.
** Based on 5 sets of 3 data points (except Leland Olds with 2 data points, 14 data points total).

ATTACHMENT: Accounting for Variability in MACT-based Limits

C. Andrew Clayton, RTI
G. Gordon Brown, RTI

OBJECTIVE AND OVERVIEW

The objective of the work described in this document is to develop statistically-based threshold values that can serve as MACT-based limits. Several statistical approaches are considered and compared. In addition, three potential concepts for defining limits are considered. These three concepts relate to alternative ways of accounting for uncertainty:

Concept A: Define a limit as the mean emission of the top 12 percent of the units plus a term that accounts for the uncertainty in the estimate of that mean.

Concept B: Define the limit as in A, but also add a component that accounts for the variation in a three-run mean for a (hypothetical) unit that is operating at the mean of the top 12 percent.

Concept C: Define the limit as an upper percentile of the distribution of units within the top 12 percent.

Concept A is the most conservative in that it will produce the smallest limits while Concept C is the least conservative and will produce the largest limits.

The statistical approaches considered are as follows:

Approach 1: Using only the emissions data for the units within the top 12 percent, estimate variance components for between-unit and within-unit sources of variation. Use these estimates to account for uncertainty, using each of the concepts above (A, B, and C). This approach assumes that within-unit variation, for the top performing units, is constant. Limits are based on normality assumptions for both between-unit and within-unit distributions.

Approach 2: Using all emissions data, estimate a relationship between within-unit (i.e., run-to-run) variances and unit means. (We estimated parameters for three different models.) Apply this relationship to estimate within-unit variances for the top 12 percent of the units and to derive the estimated variance component associated with between-unit variation. Use these variance estimates to account for uncertainty, using each of the concepts above (A, B, and C). This approach assumes that within-unit variation depends only on the performance level of the unit and that the chosen model adequately approximates the relationship between within-unit variance and the performance level. Limits are based on normality assumptions for both between-unit and within-unit distributions..

Approach 3: Using all emissions data, estimate a relationship between within-unit variances and unit means, as for Approach 2. Apply this relationship to estimate within-unit variances for the top 12 percent of the units; to account for uncertainty in each of these unit means, add the estimated amount of within-unit variation to produce a series of unit-specific limits. Select an upper percentile of these limits as the overall limit. This approach is appropriate for Concept C. It relies on normality of the within-unit distribution, but makes no assumption about the form of the between-unit distribution.

In this document, the approaches are applied to the top performing units without regard to unit subcategorizations; however, the approaches could also be applied to subcategories of units.

DATA

The data consisted of stack emission factor measurements (lb Hg/TBtu of fuel burnt, calculated via the F-factor method) from 80 units. In general, three replicate runs were performed at each unit. For 3 of the units, only 2 of the runs yielded usable data; hence, there were 237 runs overall. The top-performing 12 percent of the units consisted of those 10 units with the smallest means. The means and the observed within-unit variances for these units are listed in Table A-1. Figure A-1 shows a plot of the within-unit variances versus the unit means for all 80 units.

It should be noted that the within-unit variances represent measurement errors and *short-term* variation in a unit's performance, since this component is based on run-to-run variation within a unit. On the other hand, a between-unit variance component, among some group of units, encompasses both unit differences and longer-term temporal variability resulting from temporal variation in feed stock, operating conditions, etc. within a unit. This occurs because each unit is observed only over a short period of time and, thus, these effects cannot be separated from one another.

DETAILS OF STATISTICAL APPROACHES

<u>Notation</u>

M = the number of units
m = the number of top-performing units
n_i = the number of measurements for the i^{th} unit
n = the total number of measurements among the top-performing units
X_{ij} = the j^{th}-run measurement of the emission factor at the i^{th} unit (j=1,2,3)

$\overline{X}_i$ = mean for unit i

Table A-1. Means and Within-Unit Variances for Top Performing Units

Rank	Unit	Unit mean	Observed within-unit variance
1	Kline	0.08164	0.000007
2	Scrubgrass	0.09360	0.000125
3	Mecklenburg	0.10619	0.000263
4	Collier	0.10742	0.000110
5	Valmont	0.12683	0.001968
6	Stockton	0.13165	0.000100
7	SEI	0.23791	0.022967
8	Intermountain	0.24664	0.009698
9	Logan	0.28015	0.075198
10	Salem	0.33482	0.025756

$\overline{X}$ = overall mean of top performing units
s_i = the within-unit standard deviation for unit i.

Approach 1.

One approach to the problem is the following. It uses only the data for the top performing units; it assumes that the within-unit variability among these units is the same. The steps are outlined below.

Step 1. For the top-performing units, compute the unit means, $\overline{X}_i$, the overall mean, $\overline{X}$, and s_i, the within-unit standard deviations.

Step 2. Perform an analysis of variance (ANOVA) to determine estimates of the between-unit component of variance, s_P^2, and the within-unit component of variance, s_W^2. The ANOVA table is as follows:

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Expected values of mean squares
Between units	m-1	$PSS = \sum_{i=1}^{m} n_i (\overline{X}_i - \overline{X})^2$	$PMS = PSS / (m-1)$	$s_W^2 + K s_P^2$
Within units	m-n	$WSS = \sum_{i=1}^{m} (n_i - 1) s_i^2$	$WMS = WSS / (n-m)$	s_W^2
Total	n-1	$TSS = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (X_{ij} - \overline{X})^2$		

In the above, $K = \sum_{i=1}^{m} n_i^2 [(1/n_i) - (1/n)] / (m-1)$. K will have the value 3 if all units have 3 replicate measurements. The variance components are estimated as

(1) $\hat{s}_W^2 = WMS$

(2) $\hat{s}_P^2 = (PMS - \hat{s}_W^2) / K = [(TSS - (n-m)WMS) / (m-1) - WMS] / K$

Step 3. Determine V, the estimated variance of 3-run means as

(3) $V = \hat{s}_P^2 + \hat{s}_W^2 / 3 = (TSS - (n-m)WMS) / (m-1) + WMS / 3$

It should be noted that V=(1/3)PMS if K=3, and that there are d=m-1 degrees of freedom associated with V. (In this case it is not actually necessary to partition the V into components; it can be calculated directly from the variance of the unit means.) If K is different from 3, then the degrees of freedom can be approximated by Satterthwaite's formula.

Step 4A. For concept A, determine an upper 100" percent confidence limit (e.g., for "=0.95, a 95 percent confidence limit) for the overall mean of top performing units as $L_A = \overline{X} + t_{d,a} \sqrt{V/m}$, where $t_{d,"}$ is the 100" percent percentage point of the t distribution with d degrees of freedom. Use L_A as the MACT limit (Concept A).

Step 4B. For concept B, determine an upper 100" percent confidence limit for a 3-run average for a unit performing at the overall mean of the top performing units as $L_B = \overline{X} + t_{f,a} \sqrt{(V/m) + (\hat{s}_W^2 / 3)}$, where $t_{f,"}$ is the 100" percent percentage point of the t distribution with f degrees of freedom. Determine the degrees of freedom, f, via the formula

$$f = \frac{[(V/m) + (\hat{s}_W^2/3)]^2}{\frac{(V/m)^2}{m-1} + \frac{(\hat{s}_W^2/3)^2}{n-m}}.$$

Use L_B as the MACT limit (Concept B).

Step 4C. For Concept C, determine an upper 100" percent confidence limit for 3-run means as $L_C = \overline{X} + t_{d,a}\sqrt{V}$. Use L_C as the MACT limit (Concept C).

Approach 2.

As noted above, Approach 1 assumes there is a common within-unit variance, s_W^2. The data for the top 10 performers have observed variances ranging from 0.000007 to 0.075198, suggesting that it is unlikely that such an assumption is valid. A plot of the s_i^2 versus the unit means (for all units) also reveals that within-unit variances tend to increase with increasing level of unit emissions (see Figure A-1). Hence an alternative approach is to model the within-unit variances (or standard deviations) as a function of the 3-run unit means and to use those modeled variances to derive threshold limits similar to the L values above (Step 4). This approach is expected to work well if the modeling can be applied to a reasonably large data set that covers a fairly large range of variation.

The steps are as follows:

Step 1. Select a class of models relating s_i to the unit means, and estimate the parameters of the model. (We used several models and applied them to all the data and to several subsets of the data; the models and results are presented below.). Let s[x] denote the estimated standard deviation when the level of the unit mean is x.

Step 2. Calculate the estimated within-unit mean square (for the top-performing units) based on the model as

(4) $$WMS = \sum_{i=1}^{m} (n_i - 1)(s[\overline{X}_i])^2 \; / \; (n-m)$$

Step 3. Calculate estimates of the within-unit and between-unit variance components by substituting the model-based WMS value for the WMS used in equations (1) and (2).

Step 4. Calculate

(5) $$V = \hat{s}_P^2 + (s[\overline{X}])^2 / 3 \; .$$

Step 5A. For Concept A, determine an upper 100" percent confidence limit for the overall mean of top performing units as $L_B = \overline{X} + t_{d,a}\sqrt{V/m}$, where V is from equation (5) and $t_{d,"}$ is the 100" percent percentage point of the t distribution with d degrees of freedom. Use L_A as the MACT limit (Concept A).

Step 5B. For concept B, determine an upper 100" percent confidence limit for a 3-run average for a unit performing at the overall mean of the top performing units as

$$L_B = \overline{X} + t_{d,f}\sqrt{(V/m) + (s[\overline{X}])^2/3} \qquad ,$$

where V is from equation (5) and $t_{f,"}$ is the 100" percent percentage point of the t distribution with f degrees of freedom. Determine the degrees of freedom, f, via the formula

$$f = \frac{[(V/m) + ((s[\overline{X}])^2/3)]^2}{\dfrac{(V/m)^2}{m-1} + \dfrac{((s[\overline{X}])^2/3)^2}{M-q}}$$

where M is the total number of units (i.e., the number of units used in the modeling of the variance-versus-mean relationship) and q is the number of model parameters that were estimated. Use L_B as the MACT limit (Concept B).

For Concept C, perform the following additional steps:

Step 5C. Calculate an upper 95th percentile of the between-unit distribution as

(1) $$U_{0.95} = \overline{X} + Z_{0.95}\hat{s}_P$$

where $Z_{0.95}$ denotes the 95th percentage point of a standard normal distribution. (Note: Other percentage points could be chosen; however, more extreme [i.e., higher] percentage points do not seem warranted when m is small.)

Step 6C. Determine V, the estimated variance of 3-run means at that point of the between-unit distribution as:

(2) $$V = \hat{s}_P^2 + (s[U_{0.95}])^2/3$$

Step 7C. Calculate an upper 100" percent confidence limit for 3-run means as L_C (as in Step 4C of Approach 1) using the V from equation (7): $L_C = \overline{X} + t_{d,a}\sqrt{V}$. Use L_C as the MACT limit (Concept

C). The appropriate degrees of freedom to associate with V is debatable. We have chosen to be conservative and have used d=m-1 as the degrees of freedom, as was done in Approach 1.

<u>Approach 3.</u>

Approach 3 makes use of the same strategy of modeling the within-unit variances as was done for Approach 2. It uses the results of the modeling in a different way, however. As noted above, the resultant limit from this approach is compatible only with Concept C.

Step 1. Use the same modeling approach as described in Step 1 of Approach 2.

Step 2. Determine the upper 100"th percentile for a 3-run average for each of the top-performing units, using the model-based estimate of variance:

(3) $$U_i = \bar{X}_i + Z_a s[\bar{X}_i] / \sqrt{3}$$

where $Z_{"}$ is the upper 100"th percentile of the standard normal distribution. A normal-distribution critical value is used, rather than a value from a t distribution, since the number of within-run variances that are modeled is assumed to be large.

Step 3. Choose an upper percentile of the distribution of these U values (or the maximum U_i if the number of the top performers is small) as the MACT limit L_C.

RESULTS

<u>Approach 1.</u>

Step 1. The overall mean for the 10 top performing units is 0.17468 lb/Trillion Btu. Table A-1 shows the estimated unit means and variances.

Step 2. The ANOVA table is as follows:

Source of variation	Degrees of freedom	Sum of squares	Mean squares
Between units	9	0.22359	0.02484
Within units	20	0.27238	0.01362
Total	29	0.49597	

Estimates of the within- and between-unit variance components are thus 0.01362 and 0.00374, respectively (from equations (1) and (2)).

Step 3. The estimate of V is 0.00828. Its square root is 0.091. (As noted above, for this data set, V could be calculated directly from the variance of the unit means, since all of the top performing units have 3 runs.)

Step 4. The limits for three different confidence levels for each of the concepts are presented in Table A-2 below.

Table A-2. Limits Based on Approach 1

Concept	**Limit**	**Degrees of freedom for t**	**Confidence level**		
			▪ = 0.90	**▪ = 0.95**	**▪ = 0.99**
A	L_A	9	0.214	0.227	0.256
B	L_B	26	0.271	0.300	0.356
C	L_C	9	0.301	0.341	0.431

<u>Approach 2.</u>

Step 1. Three different models were used to relate the within-unit variances and means:

Model 1: $s \approx a + bx$

Model 2: $s \approx \sqrt{a + bx^p}$

Model 3: $s \approx \sqrt{bx^p}$

Note that the parameters have different meanings in the different models. These models were fit on the log scale (e.g., for Model 1, the log of s was modeled as a function of log(a+bx)) using non-linear least squares (SAS PROC NLIN). The log scale was used since standard deviations tend to have a variance that increases with their magnitude. Each of these models was applied to 4 data sets: the top 10 performing units (top 12 percent), the top 40 performing units (top 50 percent), all 80 units, all units except for one (Colstrip) which appeared to be an outlier. Results are summarized in Table A-3. The left hand part of the table shows the modeling results – namely, the estimated parameters, their approximate standard errors, 95 percent confidence interval estimates, the correlation matrix of the parameter estimates, and the mean square error (MSE) from the model. For each data set, the most general model, Model 2, appeared to perform best. The parameters are poorly estimated if only the top 10 performers are used. The results for the other three data sets are very similar. This is demonstrated by the results presented in the rightmost columns of Table A-3, which are described below:

- the estimated WMS values (Step 2, equation (4)),
- the estimated between-unit variance component (Step 3)
- the estimated value of the 95th percentile of the between-unit distribution of the top performing units (Step 5C, equation (6)), and
- the estimated limit L_C associated with Concept C (Step 7C).

Except for the cases where the variance estimation was based on only the top 10 performers, the limits appear to be insensitive to both model specification and to the particular data set used.

The remaining steps of Approach 2 were carried out for each of the model forms but only the model estimation results from the "all data" case are included. The predicted within-unit variances for the top performing units are shown in Table A-4. The resulting limits are presented in Table A-5, by Concept and model form.

<u>Approach 3.</u>

Table A-6 illustrates Approach 3. It shows the estimated unit-specific U values, as defined in equation (8), using "=0.95 and the estimated within-unit variances (see Table A-4). These are given for the three model forms, as applied to the 80-unit data set. The maximum values (last row of Table A-6) correspond approximately to the 95 percent percentile of the distribution of the Us; these values (approximately 0.38 to 0.40) are somewhat smaller than the 95 percent limits derived for L_C via Approach 2 (~0.43, see Table A-5). This is for two reasons. First and probably foremost is the fact that a conservative number of degrees of freedom was used in Approach 2 (t=1.833) as compared to the Z value used in Approach 3 (Z=1.645). Second, for Approach 3, the maximum U does not necessarily correspond exactly to the 95th percentile of the distribution of top performing units (depending on how one defines such a percentile) while Approach 2 produces an estimated limit for the 95th percentile of the distribution under the assumption of approximate normality. Both Approaches 2 and 3 yield limits greater than the corresponding L_C limit from Approach 1 (equal to 0.341), which assumed constant within-unit variances. The reason for this is that the model-based within-unit variances are smaller than the ANOVA-based estimate of Approach 1. This yields a higher between-unit component of variance for Approach 2 than for Approach 1. The limits for Approach C for the other " values are provided in Table A-7.

Table A-3. Summary of Results for Approach 2, for Three Models Applied to Four Data Sets

Data set	Model	Parm	Estimate	Approx S.E.	Lower 95 percent Limit	Upper 95 percent Limit	Corr with b	Corr with p	MSE	$\hat{s}_W^2$	$\hat{s}_P^2$	$U_{0.95}$	$(s[U_{0.95}])^2$	L_C
Top 12 percent	1	a	-0.0475	0.0112	-0.0734	-0.0216	-0.993		0.3415	0.00630	0.01160	0.3518	0.07769	0.5297
		b	0.6114	0.1328	0.3051	0.9178								
	2	a	-0.00007	0.000061	-0.00021	0.000078	0.748	0.878	0.2999	0.01155	0.00596	0.3016	0.03732	0.4233
		b	11.1683	17.9833	-31.3557	53.6924		0.973						
		p	4.7555	0.9109	2.6015	6.9094								
	3	b	39.5130	58.3655	-95.0793	174.1		0.968	0.3458	0.01409	0.00324	0.2683	0.01354	0.3612
		p	5.6384	0.7668	3.8701	7.4067								
Top 50 percent	1	a	0.00454	0.00769	-0.0110	0.0201	-0.548		0.9823	0.00080	0.01751	0.3923	0.00741	0.4338
		b	0.1242	0.0249	0.0738	0.1745								
	2	a	-0.00156	0.00136	-0.00431	0.00119	-0.579	0.946	0.7725	0.00214	0.01607	0.3832	0.00718	0.4238
		b	0.0254	0.00747	0.0103	0.0405		-0.283						
		p	1.1128	0.2946	0.5159	1.7098								
	3	b	0.0178	0.00543	0.00684	0.0288		0.014	0.9267	0.00121	0.01705	0.3896	0.00734	0.4308
		p	1.6076	0.2472	1.1073	2.1080								
All	1	a	0.0126	0.00843	-0.00415	0.0294	-0.396		0.9771	0.00092	0.01738	0.3915	0.00559	0.4290
		b	0.0952	0.0125	0.0703	0.1201								
	2	a	-0.00066	0.000440	-0.00153	0.000220	-0.878	0.946	0.8952	0.00145	0.01680	0.3879	0.00551	0.4250
		b	0.0239	0.00693	0.0101	0.0377		-0.678						

Data set	Model	Parm	Estimate	Approx S.E.	Lower 95 percent Limit	Upper 95 percent Limit	Corr with b	Corr with p	MSE	$\hat{s}_W^2$	$\hat{s}_P^2$	$U_{0.95}$	$(s[U_{0.95}])^2$	L_C
		p	1.4302	0.1719	1.0880	1.7724								
	3	b	0.0166	0.00444	0.00773	0.0254		-0.585	0.9440	0.00105	0.01724	0.3907	0.00558	0.4280
		p	1.6496	0.1538	1.3435	1.9557								

Table A-3. Summary of Results for Approach 2, for Three Models Applied to Four Data Sets (Cont'd)

Data set	Model	Parm	Estimate	Approx S.E.	Lower 95 percent Limit	Upper 95 percent Limit	Corr with b	Corr with p	MSE	$\hat{s}_W^2$	$\hat{s}_P^2$	$U_{0.95}$	$(s[U_{0.95}])^2$	L_C
All but one	1	a	0.0137	0.00848	-0.00320	0.0306	-0.398		0.9465	0.00095	0.01735	0.3914	0.00552	.42863
		b	0.0920	0.0121	0.0679	0.1160								
	2	a	-0.00067	0.000442	-0.00155	0.000209	-0.877	0.946	0.8614	0.00145	0.01681	0.3880	0.00545	.42488
		b	0.0233	0.00664	0.0101	0.0365		-0.675						
		p	1.4116	0.1692	1.0747	1.7485								
	3	b	0.0161	0.00424	0.00764	0.0245		-0.580	0.9117	0.00105	0.01725	0.3907	0.00551	0.4279
		p	1.6326	0.1514	1.3312	1.9340								

Table A-4. Observed and Model-Based Within-Unit Variances, for Three Models Applied to All Data

Rank	Unit	Unit mean	Observed within-unit variance	Model 1 predicted variance	Model 2 predicted variance	Model 3 predicted variance
1	Kline	0.08164	0.000007	.000416329	.000006628	.000265663
2	Scrubgrass	0.09360	0.000125	.000464095	.000150047	.000332867
3	Mecklenburg	0.10619	0.000263	.000517166	.000309742	.000409882
4	Collier	0.10742	0.000110	.000522506	.000325806	.000417743
5	Valmont	0.12683	0.001968	.000610424	.000589536	.000549449
6	Stockton	0.13165	0.000100	.000633312	.000657867	.000584317
7	SEI	0.23791	0.022967	.001244885	.002408714	.001550881
8	Intermountain	0.24664	0.009698	.001304231	.002570895	.001645871
9	Logan	0.28015	0.075198	.001544834	.003216053	.002030739
10	Salem	0.33482	0.025756	.001981107	.004341002	.002725007

Table A-5. Limits Based on Approach 2

Concept	Limit	Model for relating variance and mean *	Degrees of freedom for t	Confidence level		
				- = 0.90	- = 0.95	- = 0.99
A	L_A	1	9	0.233	0.252	0.293
		2	9	0.232	0.251	0.292
		3	9	0.233	0.251	0.293
B	L_B	1	12	0.236	0.255	0.296
		2	14	0.235	0.254	0.293
		3	12	0.236	0.256	0.297
C	L_C	1	9	0.367	0.429	0.566
		2	9	0.363	0.425	0.560
		3	9	0.366	0.428	0.565

* Model 1: $s \approx a + bx$

Model 2: $s \approx \sqrt{a + bx^p}$

Model 3: $s \approx \sqrt{bx^p}$

The results are based on model parameter estimates from the full data set of 80 units.

Table A-6. Model-Based Upper 95 Percent Limits, for Three Models Applied to All Data

Rank	Unit	Unit mean	Observed within-unit variance	Model 1 predicted upper 95 percent limit	Model 2 predicted upper 95 percent limit	Model 3 predicted upper 95 percent limit
1	Kline	0.08164	0.000007	0.10102	0.08409	0.09712
2	Scrubgrass	0.09360	0.000125	0.11406	0.10523	0.11093
3	Mecklenburg	0.10619	0.000263	0.12778	0.12290	0.12541
4	Collier	0.10742	0.000110	0.12913	0.12456	0.12683
5	Valmont	0.12683	0.001968	0.15030	0.14989	0.14909
6	Stockton	0.13165	0.000100	0.15555	0.15601	0.15461
7	SEI	0.23791	0.022967	0.27142	0.28452	0.27531
8	Intermountain	0.24664	0.009698	0.28094	0.29480	0.28517
9	Logan	0.28015	0.075198	0.31748	0.33401	0.32295
10	Salem	0.33482	0.025756	0.37709	0.39739	0.38439

Table A-7. Limits Based on Approach 3*

Model	Confidence Level		
	▬ = 0.90	▬ = 0.95	▬ = 0.99
1	0.368	0.378	0.396
2	0.384	0.398	0.425
3	0.374	0.385	0.406

* Limits are compatible with Concept C and are taken as the worst-case U_i (Salem) among the top performing units (since there were 10 such units).

Figure A-1. Plot of Within-Unit Variances Versus Unit Means