ORAL ARGUMENT NOT YET SCHEDULED

UNITED STATES COURT OF APPEALS
FOR THE DISTRICT OF COLUMBIA CIRCUIT

UNITED STATES SUGAR CORPORATION, Petitioner, v. UNITED STATES ENVIRONMENTAL PROTECTION AGENCY, et al., Respondents.	Docket No. 11-1108 (and consolidated cases)

NOTICE OF FILING RESPONSE TO REMAND OF THE RECORD

Respondents United States Environmental Protection Agency, et al., (collectively "EPA") hereby gives notice of filing the attached Response to the remand of the record ordered by the Court in its Order of May 15, 2014, ECF No. 1493171.

Respectfully submitted,

SAM HIRSCH
Acting Assistant Attorney General

/S/ Norman L. Rave, Jr.
NORMAN L. RAVE, JR.
PERRY ROSEN
Environmental Defense Section
Environment & Natural Resources Division
United States Department of Justice
P.O. Box 7611
Washington, D.C. 20044
(202) 616-7568
Counsel for Respondents

July 14, 2014

CERTIFICATE OF SERVICE

I hereby certify that on this 14th day of July, 2014, I caused a copy of the foregoing document to be served by the Court's CM/ECF system on all counsel of record in this matter.

/S/ Norman L. Rave, Jr.
Norman L. Rave, Jr.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

JUL 14 2014

OFFICE OF
AIR QUALITY PLANNING
AND STANDARDS

MEMORANDUM

SUBJECT: EPA's Response to Remand of the Record for Major Source Boilers

FROM: Stephen D. Page, Director
Office of Air Quality Planning and Standards

TO: Docket ID No. EPA-HQ-OAR-2002-0058

This document responds to the court's remand of the record to allow EPA to provide further explanation of how its use of the Upper Prediction Limit (UPL) in setting emissions standards pursuant to section 112 and 129 of the Clean Air Act (CAA) meets the requirements of section 112(d)(3) and section 129(a)(2). Specifically, this document explains EPA's use of the UPL, which is the statistical methodology EPA uses as the primary tool to account for emissions variability when setting emissions standards under these sections, and explains how the UPL is used to calculate the average emissions limitation achieved over time by the best performing source or sources.

I. Emissions Standards under Section 112 and 129

For major stationary sources of hazardous air pollutants (HAP) under section 112 and for sources regulated under section 129,[1] the CAA requires emission standards that represent the maximum achievable control technology, or "MACT" level of control. The MACT standard setting process is a two-step process that starts with the agency determining the level of control that is currently achieved by the best-controlled similar source (for new source standards) or the average of the best-performing sources (for existing source standards). That level of control is the minimum level of control required by the CAA regardless of cost or feasibility, and it is commonly called the "MACT floor." This document focuses on that first step in the MACT standard setting process, as it is in that step that the agency uses the UPL to both evaluate average emissions performance and estimate variability to determine the level of control sources currently achieve.[2]

For new sources located at facilities that are major sources of HAP, section 112 of the CAA requires EPA to establish standards that are no less stringent than the emissions control that is

[1] A "major source" is a stationary source or group of stationary sources at a single location and under common control that emits or has the potential to emit ten tons per year or more of any hazardous air pollutant or 25 tons per year or more of any combination of hazardous air pollutants. Any stationary source that is not a major source is an "area source." CAA § 112(a). For area sources of HAP, EPA has the discretion to set emissions standards under section 112(d)(5) based on generally available control technology in lieu of MACT standards.

[2] The second step in the MACT standard-setting process is the evaluation of whether beyond-the-floor controls are appropriate, as provided in section 112(d)(2) of the Act. This second step is not addressed in this document.

Internet Address (URL) • http://www.epa.gov
Recycled/Recyclable • Printed with Vegetable Oil Based Inks on Recycled Paper (Minimum 25% Postconsumer)

achieved in practice by the best controlled similar source. The statute does not define "achieved in practice" nor does it dictate the manner in which the agency determines which source is the best controlled similar source, instead leaving it to the agency's discretion to make those determinations. For existing sources located at facilities that are major sources of HAP, standards must be no less stringent than the average emissions limitation achieved by the best performing sources for which EPA has emissions information. Again, the CAA leaves to EPA's discretion the manner in which to calculate "the average emissions limitation achieved" by the best performing sources. Under section 112, the CAA recognizes that source categories and subcategories have differing numbers of sources and provides category size-specific instructions on determining standard stringency for existing sources. Specifically, standards for categories or subcategories with fewer than 30 sources must be based on the average emissions limitation achieved by the best performing 5 sources, and standards for categories or subcategories with 30 or more sources must be based on the average emission limitation achieved by the best performing 12 percent of sources for which EPA has emissions information. Section 129 requires EPA to establish standards for new sources that are no less stringent than the emissions control that is achieved in practice by the best controlled similar source for nine specified criteria and hazardous air pollutants, and standards for existing sources that are no less stringent than the average emission limitation achieved by the best performing 12 percent of sources in the category or subcategory for the same nine pollutants.[3]

II. Remand of the Record

EPA promulgated the major source industrial, commercial, and institutional boilers rules and the commercial and industrial solid waste incinerators (CISWI) rule on March 21, 2011, and subsequently made certain revisions to the rules on January 31, 2012. On August 20, 2013, the D.C. Circuit Court of Appeals issued its decision in National Assn. of Clean Water Agencies v. EPA, which involved challenges to EPA's MACT standards for sewage sludge incinerators, issued under section 129 of the Act. See 734 F.3d 1115. In the NACWA decision, the D.C. Circuit remanded certain aspects of the rule for further explanation, including the question of how the UPL represents the MACT floor for new and existing units, as required by the CAA. The sewage sludge incinerator rule was issued on the same day as the boilers and CISWI rules, and used the same general methodology for calculating the MACT floors. For this reason, EPA requested an opportunity to supplement the record in pending challenges to the boilers and CISWI rules, to provide the explanation of the agency's analysis of variability in setting the MACT floor standards that the court believed was needed in the record for the sewage sludge incinerator rule. The court granted EPA's motion for a remand of the record on May 15, 2014. This document provides that additional explanation.

There are several key points, addressed in more detail below, that underlie EPA's methodology for calculating MACT floor standards through the use of the UPL. First, the floor standards reasonably account for variability in the emissions of the sources used to calculate the standards. This variability occurs due to a number of factors, including measurement variability (both sampling and analysis) and short term fluctuations in the emission levels that result from short-

[3] Unlike section 112(d)(3)(A), section 129(a)(2) does not direct the agency to base the standards on 12% of the sources for which the agency has data, and the EPA has interpreted this to mean that the incinerator standards require the standards to be based on 12% of the sources in the category whether or not the agency has data on all the sources in the category.

term changes in fuels, processes, combustion conditions, and controls. Second, because the emissions data available to EPA is in the form of short-term stack tests and the standards must be complied with at all times, the agency uses the UPL to estimate the average emissions performance of the units used to establish the MACT floor standards at times other than when the stack tests were conducted. Thus, the UPL results in a limit that represents the average emissions limitation achieved by the best performing sources over time, accounting for variability in emissions performance.

III. Overview of MACT Floor Methodology

In establishing MACT floors, we use the available information to determine the average performance of the best performing sources (for existing source floors) and the average performance of the best-controlled similar source (for new source floors). Each MACT standard is based on limited data from sources whose emissions are expected to vary over their long term performance. For this reason, and because sources must comply with the MACT standards at all times, consideration of variability is a key factor in establishing these standards. This variability in emissions is due to numerous factors, including operation of control technologies, variation in combustion materials and combustion conditions, variation in operation of the unit itself, and variation associated with the emission measurement techniques. In order to account for variability that is reflected in the available data that we use to calculate MACT floors, we use the UPL.

The NACWA court recognized that EPA has significant discretion in establishing the MACT floors, including how to account for variability. See NACWA v. EPA, 734 F.3d at 1142-43 ("EPA can decide what value the MACT floors are supposed to represent, as long as that decision is a reasonable interpretation of the statute.") However, the court expressed concern about the interpretation it believed EPA was taking of the "average" emissions performance of the best performing sources. Specifically, the court asked for further clarification of how "'average' means the average of a future 3-run compliance test." Id. at 1143. As explained below, EPA does not interpret the term "average" as used in section 112(d)(3) in this manner. Rather, EPA interprets the average to mean the average emissions over time, based on both the calculated average of all emissions test data from the best performing source or sources and the available information regarding the variability of emissions. The UPL, which EPA uses to account for variability among the best performing sources, reflects an upper limit for the emissions of those sources at times other than when the emissions tests occurred. This not only is a prediction of the emissions performance of those sources in tests conducted in the future, but is also an indication of the range of current average emissions performance of those units.

In calculating MACT floors, EPA analyzes all available emissions information and then uses the emissions information for the best performing source or sources to calculate a value that represents the level achieved in practice by the best controlled similar source for the new source MACT floor and a value that represents the average emission limitation achieved by the best performing twelve percent of sources for the existing source MACT floor. EPA must establish the MACT floor standards based on the emissions information before the agency, and the standards must be met at all times by sources subject to them. The available emissions data are generally in the form of short term, three-run stack tests, with each test run lasting for between 1 and 4 hours. Those short term tests cannot encompass the emissions performance of a source over time. EPA interprets "emissions performance," as used in section 112(d)(3) to mean the

emissions of a source over the long term, rather than just during a short-term stack test. This is a reasonable interpretation of the statutory directive that the MACT standard be no less stringent than the average emissions limitation "achieved by" the best performing sources, since the emissions level of a source is not expected to always be the same as it is during stack testing.

Therefore, EPA applies a methodology that predicts the actual emissions levels the source is achieving at times other than when stack testing was conducted. This enables EPA to calculate the average emissions performance of the best performing sources over time. Therefore, the MACT floor calculation takes into account the inherent variability in emissions performance to more accurately reflect the range of the best performing sources' emissions over time. The remainder of this discussion focuses on why the statistical methodology that we used - the UPL - provides results that reasonably represent the level achieved in practice by the best-controlled similar source for new sources, and the average emissions limitation achieved by the best performing emissions units for existing sources.

IV. Use of the Upper Prediction Limit

The UPL is a value derived from widely accepted and commonly used statistical principles, and represents the upper end of a prediction interval.[4] The UPL is used in a wide variety of industries and professions, such as manufacturing, finance, healthcare and others.[5] The UPL is calculated using one of several equations that must be selected based on certain characteristics of each dataset. The most important parameters in the UPL calculation are the mean (average) performance of the unit or units, the distribution of the dataset, the variance in the measured emission values (that is, the individual data points), and the sample size. The equations that are used to calculate the UPL are explained in more detail below.

In the context of development of MACT floors, the UPL is a value, calculated from a dataset, that identifies the average emissions level that a source or group of sources is meeting and would be expected to meet a specified percent of the time that the source is operating. That percent of time is based on the confidence level used in the UPL equation, as explained in more detail below. For a single best controlled emission source that serves as the basis for a new source MACT floor, the 99 percent UPL is the emissions level that the source would be predicted to be below during 99 out of 100 performance tests, including emissions tests conducted in the past, present, and future, based on the short term stack test data available for that source. It is important to note that the UPL does not represent the *worst* emissions performance of the best performing units at any time, but rather the *average* level expected to have been achieved over time, as the first element of the UPL equation is the average of the short-term emissions test data from the best performing units. In other words, the 99 percent UPL is the level of emissions that we are 99 percent confident is achieved by the average source represented in a dataset over a long-term period based on its previous, measured performance history as reflected in short term stack test data.

---

[4] National Institute of Standards and Technology, *NIST/SEMATECH e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/predlimi.htm>.

[5] See, e.g., Gibbons, Robert D. and Coleman, David E. (2001). *Statistical Methods for Detection and Quantification of Environmental Contamination*, New York: Wiley, pp 21-31, and Luko, S.N. and D.V. Neubauer, Statistical Intervals Part 2. *Standardization News*, October/November 2011. <http:www.astm.org/standardization-news/data-points/statistical-intervals-part-2-so11.html>.

In sum, the UPL predicts the level of emissions that the sources upon which the floor is based are expected to meet over time, considering both the average emissions level achieved as well as emissions variability and the uncertainty that exists in the determination of emissions variability given the available, short-term data. EPA did not have emissions information from sources at all times each source was operating, and therefore determined it was necessary to apply a methodology that addressed the fact that the data were not complete. Statistical tools are commonly used in science and industry because they permit predictions to be made based on the available data. As explained in more detail below, the UPL equations include a term that adjusts the average emission level of the units to account for both the emissions variability observed in the data and the uncertainty that exists in that estimate of variability for a sample of the given size, to estimate the performance of the units over time. As shown by the t- and z-statistics incorporated into the UPL equations, the uncertainty regarding emissions variability is higher in small samples and is lower in large samples, but it is always accounted for in UPL.

The 99 percent UPL value reasonably estimates the level currently achieved in practice by the best controlled similar source and the average emission limitation currently achieved by the best performing 12 percent of sources by estimating the level achieved by the best source (for new sources) or by the average of the best sources (for existing sources) ninety-nine percent of the time the sources are operating. It is reasonable to interpret the statutory requirement that the MACT floor level reflect the "emission control that is achieved in practice" by the best controlled similar source and the "average emission limitation achieved by" the best performing 12 percent of sources as a level that the average level the best performers are meeting on a consistent basis over time, not just at the single point in time during which emissions test data were collected. If EPA were to simply take the average of the stack test data from the best performing source or sources, the MACT floor would not represent the levels of emissions achieved in practice over time because the MACT floors would not reflect variations in material inputs (e.g., fuels), control device performance, operating unit performance (e.g., changes in combustion conditions that impact emissions), and other factors that affect unit performance over time. This is demonstrated by the fact that even single three run tests, which are performed over a short period of time, typically show different emissions levels during each individual test run. It would not be reasonable to establish a MACT floor standard that the best performing sources would exceed routinely. But neither would it generally be reasonable to establish a standard that all the best performing 12 percent of existing sources can meet without any modification because the statute requires the agency to establish the standard at the *average* level of performance of the best 12 percent of sources. If EPA simply calculated the average (mean) of the emissions data from the best performing 12 percent of sources, that would only lead to a limit that can be met by all the existing sources in the floor at all times in the unlikely hypothetical situation where all the sources have identical emissions test data and controls and operate in exactly the same manner– i.e. the sources would have to be essentially identical and use identical fuels or other material inputs. Since that situation does not represent real-world source categories, the agency takes the available data and applies a statistical formula to account for variability in emissions of the best performing source or sources.

V. Establishment of MACT Floors

At first glance, developing a HAP emissions limit in accordance with the statutory requirements seems a straightforward task, requiring simply that EPA collect all available valid data from all the sources in a specific category or subcategory, rank the data to select the best performing twelve percent of the category or subcategory, and calculate the average emission level achieved by that group of sources for a particular pollutant. However, multiple factors complicate these calculations. First, emission test results taken during a three-run stack test, as is the case with virtually all the HAP emission data available to EPA, represent three "snapshots" of a source's operations during the test period and those results generally will not represent a source's full range of normal operations or emissions levels. However, because there are three separate runs in a test, a single test will in most cases show some of a particular source's variability over the short period of time during which testing was conducted. Further, EPA generally has limited data with which to establish MACT standards, primarily because the data needed are not generally collected by sources prior to regulation (particularly for HAP, but also for other pollutants such as nitrogen oxides, sulfur dioxide, and carbon monoxide regulated under section 129). The collection of such data is a time- and resource-intensive effort, as emissions tests can take more than a week to perform, analyses of samples from the tests can take more than a month, preparation of test reports typically takes several months, and costs range up to $40,000 per test per pollutant. For example, during our data collection efforts for the boiler and CISWI rules we required hundreds of facilities to conduct emissions testing, and we estimated an average cost of about $90,000 for each facility that was required to conduct testing.

As explained above, EPA establishes MACT floor standards under sections 112 and 129 as limits that reflect the level achieved in practice by the best controlled similar source, or the average emissions limitation achieved by the best performing 12 percent of sources over time, but we necessarily must do that based on the emissions information that is available. In this case, that information consisted of emissions test data that represents the specific instances in the sources' operations when testing was conducted. The UPL is a common statistical technique that allows EPA to use emissions test data and the data characteristics, such as the distribution and sample size, along with the intrinsic variability associated with those data,[6] to calculate an emissions limit based on a specified level of confidence such that an average best performing existing source would not be expected to exceed the limit a specified number of times. For instance, if an emission limit were set at the 50 percent UPL, we would expect the average unit to have emissions above the level of the standard 50 percent of the time.

This section explains how EPA applied the UPL to the emissions data from the best performing source or sources to calculate MACT floor limits that represent the average emissions performance of those sources, while accounting for variation in those sources' emissions over time.

A. Ranking Sources and Establishing the Distribution of Data

After collecting emissions data, EPA analyzes the data both to identify and rank the best performing sources and to evaluate the data distribution. This evaluation is necessary because the

[6] Gibbons, p 30.

UPL equation that is used to account for variability and calculate the MACT floor standard depends on the distribution of the data. Typical data distributions are normal and lognormal, as shown in Figures 1 and 2. Therefore, the first step in applying the UPL is to rank sources (based on lowest emissions) to identify the best performing sources and then analyze the distribution of the emissions data from those best performing sources. The ranking is used to establish which set of units make up the top 12 percent of performers for each pollutant and subcategory and which unit EPA considers the single best-controlled for each pollutant and subcategory. Once these units are identified, EPA determines the data distribution as discussed below.

In order to select statistical equations (to calculate the UPL) that are appropriate for each individual dataset, EPA evaluates each individual dataset to determine the data distribution, as noted above. The individual datasets that we evaluate include the datasets for the best performing unit for each pollutant and subcategory for the new source MACT floor standards and the datasets for the best performing 12 percent of units for each pollutant and subcategory for the existing source MACT floor standards. To evaluate the distribution of the data, EPA checks each dataset for distribution characteristics by applying skewness and kurtosis tests. Skewness is a measure of symmetry (or the lack of symmetry) of a distribution, and is useful in determining the distribution because normal distributions are symmetric while lognormal distributions are not. Kurtosis is a measure of whether a dataset has a similar, more distinct, or flatter peak than a standard normal distribution.[7] Figures 1 and 2 show simple examples of normal and lognormal distributions. In each figure for the purpose of EPA's analyses, the x-axis is the level of emissions and the y-axis represents the probability of a particular value occurring (established based on the number of occurances in the dataset at each level of emissions). If both the skewness and kurtosis tests suggest the data exhibit a normal distribution, EPA evaluates the emissions data for the source category using statistical equations for normally distributed data. If the test results are inconsistent regarding normality, or if both test results suggest non-normality, EPA evaluates the emissions data for the source category using statistical equations for lognormally distributed data.

EPA uses this approach because, absent evidence of a normal distribution, air emissions data have historically been shown to exhibit lognormal distribution characteristics such as 1) run values always above zero; 2) a majority of run values will be low; and 3) a few run values will be high but will have a low probability of occurring.[8,9,10] These characteristics yield a left-skewed distribution curve, meaning the distribution has a long tail to the right.

---

[7] National Institute of Standards and Technology, *NIST/SEMATECH e-Handbook of Statistical Methods*, <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>.

[8] Ott, Wayne R. (1990) *A Physical Explanation of the Lognormality of Pollutant Concentrations*, Journal of the Air & Waste Management Association, 40:10, 1378-383, DOI: 10.1080/10473289.1990.10466789.

[9] Gilbert, Richard O. (1987), *Statistical Methods for Environmental Pollution Monitoring*, New York: Van Nostrand Reinhold, p 152.

[10] Gibbons, Robert D. and Coleman, David E. (2001), *Statistical Methods for Detection and Quantification of Environmental Contamination*, New York: Wiley, p 369.

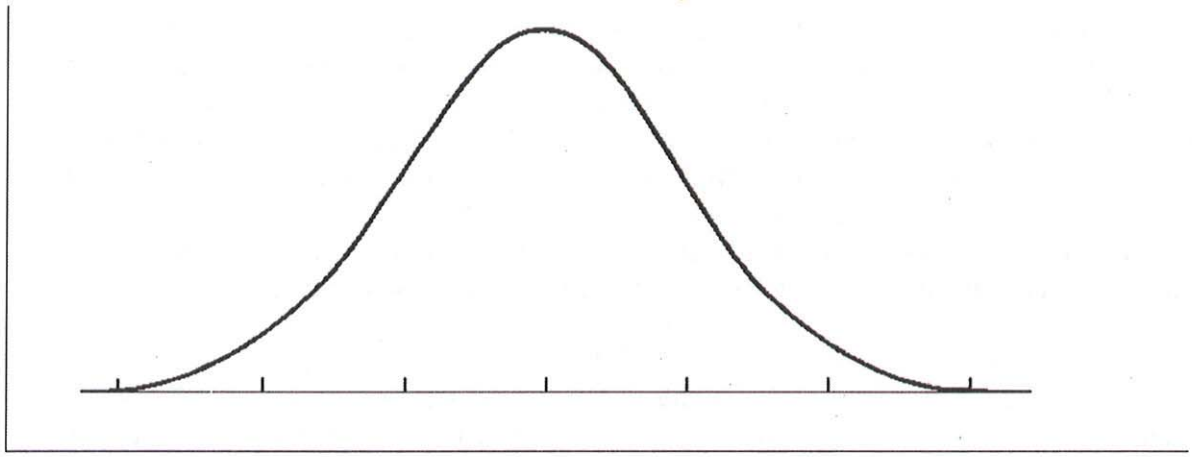

Figure 1. Normal Distribution[11]

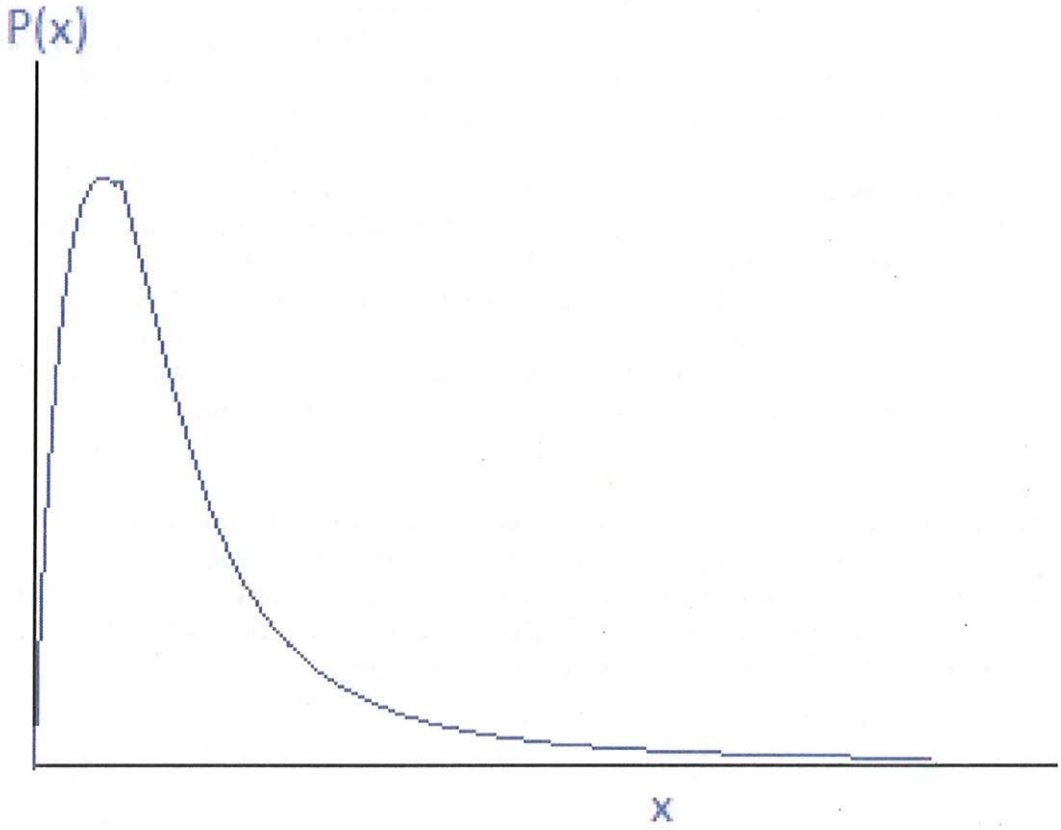

Figure 2. Lognormal Distribution[12]

[11] http://www.psychstat.missouristate.edu/introbook/sbk10m.htm.
[12] http://math.tutorvista.com/statistics/normal-distribution.html.

The following example demonstrates the characteristics of a lognormal distribution curve. Consider a dataset obtained from three filterable particulate matter (PM) emissions tests, each consisting of three test runs, from three independent sources. The nine runs produced the following values measured in pounds PM per million British Thermal Units (lb PM / MMBTU): 0.00134, 0.00139, 0.00153, 0.00319, 0.00387, 0.00417, 0.00970, 0.0128, and 0.0243. This dataset, referred to as Case A, yields both skewness and kurtosis consistent with a lognormal distribution. Figure 3 shows Case A data as solid green squares plotted on the lognormal distribution determined from the characteristics of the run data. Based on those characteristics, one would expect 50 percent of all filterable PM values to be less than 0.0039 lb PM / MMBTU. In addition, one would expect about 95% of all filterable PM values to be less than 0.0188 lb PM / MMBTU. The long tail to the right, as shown in Figure 1, is consistent with the expectation that high values can exist, but they have a low probability of occurring – below one percent for values greater than 0.0255 lb PM / MMBTU.

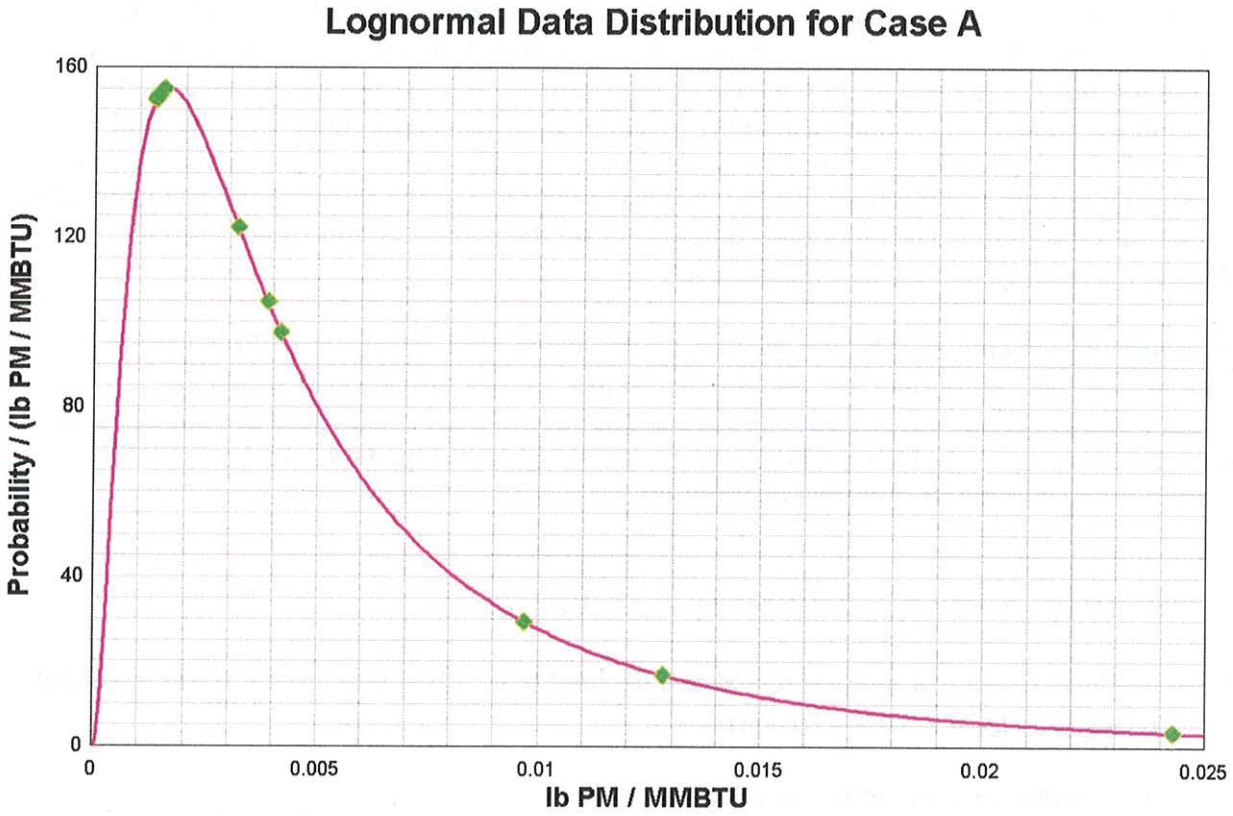

FIGURE 3. Lognormal Data Distribution Example. The smooth curve represents the density function of the lognormal distribution determined from the characteristics of the filterable PM emissions test data (shown as squares) from Case A.

B. Calculating the MACT Floor

Once EPA identifies the best performing single source or best performing twelve percent of sources and the distribution of each dataset, we calculate the UPL, as described below, to identify an emissions limit that represents the level the average unit among the best performing

units has met, is meeting, and would be expected to meet in practice with a specific confidence level.[13] The confidence level, in this case 99 percent, is the percentage of measurements (past, present, and future) that are predicted to fall at or below the UPL value. Stated differently, the 99 percent UPL is the level of emissions that we are 99 percent confident is achieved by the average source represented in a dataset over a long-term period based on its previous, measured performance history as reflected in short term stack test data. The UPL value is based on the available emissions data from the best performing single source (for new source standards) or the best performing 12 percent of sources (for existing source standards). The UPL incorporates the intrinsic variability of the individual run measurements used to calculate an emission test average.[14] While the UPL can be calculated at different confidence levels, we selected the 99 percent level in order to provide reasonable assurance that the limit can be met at all times by a source with emissions at the average level achieved by the best performing source or sources.

C. Applying the UPL Equations

Once the distribution is established and the confidence level selected, a UPL value is calculated using six components. Four of these components are derived from the emissions dataset: the average (or mean), the variance of the dataset, the t-statistic or z-statistic (depending on the distribution characteristics), and the number of samples, which is associated with the total number of emissions test runs. For example, if there are results from six emissions tests, there typically are at least 18 emissions test run values (or samples), since an emissions test generally consists of 3 test runs.[15] The other two components are the number of run values used to calculate the average and the confidence level.

UPL equations are data-distribution specific, so that the normal distribution UPL equation is different from the lognormal UPL equation. Even though they differ due to separate mathematical properties associated with each distribution, the UPL equations share a common format that is composed of two parts which are added together. Equation 1 shows the UPL equation for a dataset with a normal distribution.

Equation 1. $$UPL_{100-(\alpha\times100)} = \bar{x} + t_{(n-1),(1-\alpha)}\sqrt{s^2\left(\frac{1}{m}+\frac{1}{n}\right)}$$

where α = level of significance expressed as a decimal (e.g., 1% significance = 0.01), note that confidence level = $100 - (\alpha \times 100)$;

$\overline{x}$ = average or mean of test run data;

[13] Gibbons, p 28.
[14] Gibbons, pp 30-31.
[15] EPA test methods typically require 3 test runs. In cases where a test included fewer than 3 valid test runs, we typically do not include the test as a valid test for the purpose of MACT floor calculations. However, if limited or no other data are available for a particular source and pollutant, we consider the use of such data on a case-by-case basis.

t = t score, the one-tailed t value of the Student's t distribution for a specific degree of freedom and level of significance;

n = number of test runs;

s^2 = variance of test run data; and

m = number of test run values used to calculate the test average (generally 3).

The first part of the UPL equation, $\bar{x}$ is the average of the best performing sources' emission measurements. Thus, the average of the emission measurements from the best performing source or sources is a fundamental element, and is in fact the starting point, of the UPL calculation. The second part of the equation, i.e., everything to the right of $\bar{x}$, is associated with the variance of the run data, s^2; the total number of runs in the dataset, n; the confidence level, calculated from α; and the number of run values used to calculate the emissions average for each test result, m.

As shown in Equation 1, the UPL is directly related to the sources' emissions average. In other words, as the emissions average goes up or down, so does the UPL value. The UPL also is directly related to the confidence level and to the variance, meaning that as either of these values go up or down, so does the UPL value. However, the UPL is inversely related to the number of runs in the dataset and to the number of runs averaged for each emissions test. Therefore, as either of these values increases, the UPL value decreases and as either of these values decreases, the UPL value increases. A primary reason for this result is that the amount of uncertainty associated with the UPL is higher when sample sizes are smaller. Larger datasets provide better estimates of the average, the standard deviation and the distribution of the data. A lower boundary for the UPL exists at every significance level. This lower boundary corresponds to what the UPL would be if an infinite number of samples was available because the uncertainty component of the UPL becomes negligible as the sample size increases. In practical terms, this lower boundary represents a value that includes almost no uncertainty when the dataset is large. In other words, even if we had an infinite number of samples, the 99 UPL would not be below this lower-bound value. Therefore, the first half of the UPL equation takes the average emissions value of the sources in the MACT pool and the second half of the equation addresses the variability of that average emissions level based on the factors identified in the variables in that portion of the equation.

Equation 2 shows the UPL equation for a dataset with a lognormal distribution.

Equation 2.

$$UPL_{100-(\alpha\times100)} = e^{\hat{\mu}+\frac{\hat{\sigma}^2}{2}} + \frac{Z_{(1-\alpha)}}{m}\sqrt{me^{2\hat{\mu}+\hat{\sigma}^2}\left(e^{\hat{\sigma}^2}-1\right)+m^2e^{2\hat{\mu}+\hat{\sigma}^2}\left(\frac{\hat{\sigma}^2}{n}+\frac{\hat{\sigma}^4}{2(n-1)}\right)}$$

where α = level of significance expressed as a decimal (e.g., 1% significance = 0.01); note that confidence level = $100 - (\alpha \times 100)$;

e = base of the natural logarithm (≈ 2.718282);

$\hat{\mu}$ = mean of the log transformed run data ($= \frac{1}{n}\sum_{i=1}^{n} \ln(x)_i$);

$\hat{\sigma}^2$= variance of the log transformed run data ($= \frac{1}{n}\sum_{i=1}^{n} (\ln(x)_i - \hat{\mu})^2$);

z = z score, the one-tailed z value of the z distribution for a specific level of significance;

m = number of run values used to calculate the test average (generally 3); and

n = number of runs.

The widely-accepted statistical procedure for calculating a UPL for log-normally distributed data involves performing a log-transformation of the data. As with Equation 1, the first part of the UPL equation, $e^{\hat{\mu}+\frac{\hat{\sigma}^2}{2}}$, is associated with the average of the best performing sources' emissions. The second part of the equation, i.e., everything to the right of $e^{\hat{\mu}+\frac{\hat{\sigma}^2}{2}}$, is associated with the variance of the run data, $\hat{\sigma}^2$; the number of runs in the dataset, n; the confidence level, calculated from α; and the number of run values used to calculate the test average, m.

UPL Equation 2 operates in a manner similar to Equation 1. The first part of UPL Equation 2 is directly related to the emissions average and corresponds to an estimate of the average of the lognormal data. In other words, as the emissions average goes up or down, so does the UPL value. The second part of the UPL equation is directly related to confidence level and to the variance of the average of the lognormal data, meaning that as either of these values go up or down, so does the UPL value. As in the normal case, the second part of the UPL equation is inversely related to the number of runs in the dataset and to the number of runs averaged for an emissions test. Therefore, as either of these values increase, the UPL value decreases and as either of these values decrease, the UPL value increases.

Consistent with Equation 1, a lower boundary for the UPL exists for Equation 2 at every significance level when an infinite number of samples is provided. In practical terms, this lower boundary represents the minimum amount of variability associated with a future, past, or present emissions test average.[16] Furthermore, the selection of a higher confidence level for a UPL calculation widens the interval of results, meaning that a higher confidence level increases the UPL value. As with Equation 1, the first half of the UPL equation takes the average emissions value of the sources in the MACT pool and the second half of the equation addresses the variability of that average emissions level based on the factors identified in the variables in that portion of the equation.

To provide an example of how the UPL calculation is performed, the application of the UPL to a specific set of data from a set of best performing sources in EPA's sewage sludge incinerator rule is explained below (Case B).

Case B shows how one takes data from the best performing sources and translates that into an emissions limit. There are 150 existing units in the multiple hearth subcategory for sewage

[16] Gibbons, p 31.

sludge incinerators, so there are 18 units that are best performing sources for Case B since 12 percent of 150 sources is 18 sources. Charts 1 and 2 show PM emissions data for each run of testing for the 18 best performing sources (note that we do not always have the same number of test runs for each source).

	Unit PM Emissions, mg/dscm								
Run	1	2	3	4	5	6	7	8	9
1	2.08	1.99	0.82	2.87	16.02	15.25	25.95	41.72	38.90
2	0.50	0.55	5.57	15.42	12.02	18.31	26.90	39.64	30.40
3	1.05	2.73	5.41	8.31	15.84	19.08	20.05	20.40	32.24
4				18.55			28.59	40.18	
5				18.82			23.66	40.46	
6				2.70			29.56	39.24	
7								19.10	
8								20.60	
9								21.40	

Chart 1. PM emissions data, mg/dscm, for Units 1 – 9 by individual run.

	Unit PM Emissions, mg/dscm								
Run	10	11	12	13	14	15	16	17	18
1	47.11	33.76	49.65	42.53	48.90	45.99	72.26	80.51	70.31
2	43.86	41.11	34.07	40.52	46.64	124.20	60.89	80.46	88.68
3	27.71	44.49	28.54	37.87	70.58	45.33	78.61	69.49	83.01
4	27.91		56.75			89.06	63.62	73.99	74.90
5	52.90		48.47			28.92	94.64	79.78	69.80
6	37.16		24.03			55.09	63.06	75.36	85.83
7						45.03		81.62	
8						32.13		77.11	
9						55.30		67.15	

Chart 2. PM emissions data, mg/dscm, for Units 10 – 18 by individual run.

Using Excel software, the characteristics of the best performers from this source category are as follows:

$\overline{x}$ = mean of test run data = 41.33;

n = number of test runs = 90;

s^2 = variance of test run data = 736.24;

kurtosis = -0.31; and

skewness = 0.47.

As the kurtosis and skewness results are both near zero, we conclude that the data exhibit a normal distribution. Given a normal distribution, we use Equation 1 to calculate the UPL. Because we are using a confidence level of 99, the significance level, α, calculated using the confidence level is 0.01 (1-0.99). With 90 run values, the t value as determined from Excel software is 2.37 ($t_{(90-1),(1-0.01)} = 2.37$). Since emission test averages, such as those used to demonstrate compliance with the UPL-based limits, include 3 individual test runs, the value of m is set at 3. The 99 percent UPL is calculated from the best performers' dataset for a normal distribution as follows:

$$UPL_{100-(\alpha\times100)} = \bar{x} + t_{(n-1),(1-\alpha)}\sqrt{s^2\left(\frac{1}{m}+\frac{1}{n}\right)}$$

$$UPL_{100-(0.01\times100)} = 41.33 + t_{(90-1),(1-0.01)}\sqrt{736.24\left(\frac{1}{3}+\frac{1}{90}\right)}$$

$$UPL_{99} = 41.33 + 2.37\sqrt{253.59}$$

$$UPL_{99} = 41.33 + 37.74 = 79.07.$$

As mentioned earlier, this 99 percent UPL is composed of two parts: the first part which is directly related to the emissions average of the dataset (41.33) and the second part which is related to the variance, level of confidence, number of runs, and number of runs that constitute an emissions test (37.74). Thus, the equation includes both the average emissions level achieved by the best performing sources, and also a consideration of the variability of those emissions, including the variability associated with conducting an emissions test at other times. The calculation shows that the best performing sources would be expected to have an emissions test value that would not exceed 79.07 mg PM/dscm, based on the average of the best performing 12 percent of sources' emissions data and the consideration of variability within the data.

Most, but not all, of the emissions test data from the best performing twelve percent of sources is below this value. As shown in Figure 4, values of individual test run data from these sources vary, with 10 values above the 99 percent UPL. The range of individual unit values is related to an individual unit's variability. Individual run values above the 99 percent UPL demonstrate that even units among the best performers may need to apply some combination of additional control devices, modified process operations, or modified raw materials in order to be able to meet an emissions limit set at the 99 percent UPL. This is to be expected, and further demonstrates that EPA's use of the UPL to calculate the MACT floor is reasonable, since the MACT floor represents the *average* emission level achieved by the best performing sources, not the *worst* emission level achieved by those sources. For instance, units 15, 16, 17, and 18 (the "bottom 4" out of 18 units in the top 12 percent in this example) all have at least one test run value above the 99 percent UPL for the entire dataset, but no source among the best performing 12 percent of sources had an overall average above the 99 percent UPL, although several sources had an average that is close to the UPL and one source had a 3-run test average above the UPL.

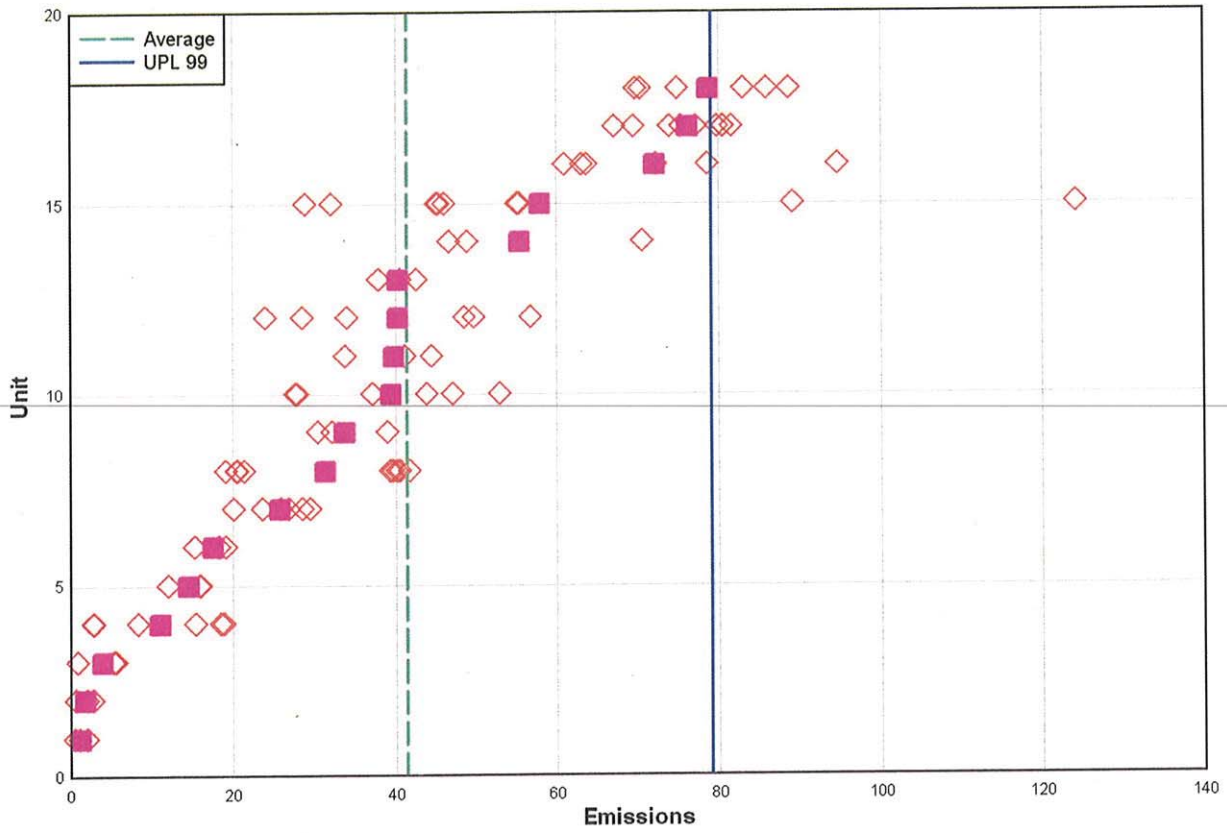

FIGURE 3. Filterable PM data from 18 multiple hearth sewage sludge incineration units. Individual run data shown as open red diamonds, unit average emissions test data shown as solid pink squares, the average of all runs shown as the dashed green line, and the 99 percent UPL is shown as the solid blue line.

VI. Conclusion

In conclusion, EPA's methodology for establishing the MACT floors is reasonable and represents the average emissions achieved by the best performing sources with consideration of the variability in the emissions of those sources. The methodology first identifies the best performing sources by ranking sources from lowest to highest emissions. Next, the distribution of the data is evaluated in order to identify which UPL equation to apply. After that, the average of all the short-term stack test data from all the best performing sources (or the single best-controlled source, for new sources) is calculated. Finally, the UPL is applied to that calculated average value to account for variability such that the MACT floor standard can be met by a unit with emissions at the average level of the best performing source or sources.