ATTACHMENT 3

RMB Consulting & Research, Inc.

5104 Bur Oak Circle
Raleigh, North Carolina 27612

Phone (919) 510-5102
FAX (919) 510-5104

MEMORANDUM

TO: Utility Air Regulatory Group

FROM: Ralph L. Roberson, P.E.

DATE: August 3, 2011

SUBJECT: Technical Comments on EPA's Proposed Electric Generating Unit Rule

I. INTRODUCTION

On May 3, 2011 EPA proposed its National Emission Standards for Hazardous Air Pollutants from Coal- and Oil-Fired Electric Utility Steam Generating Units.[1] Because the emission standards set forth in this NESHAPs are based on emission reductions assuming application of maximum achievable control technology (MACT), such rules are often referred to as "MACT rules" or "MACT standards." The Utility Air Regulatory Group (UARG) asked me, Senior Consultant with RMB Consulting & Research, Inc. (RMB) to review and to provide technical comments on EPA's proposed EGU MACT Rule. UARG asked me to focus its review on the key components of any MACT rulemaking such as choice of surrogates; appropriate treatment of emission variability; calculation of MACT floors and compliance determinate requirements.

II. EPA'S APROACH TO PM SHOULD BE REVISED

Total PM Is Not Appropriate Surrogate

EPA proposes to regulate total PM, which is defined as the sum of filterable PM and condensable PM, solely on the basis of the behavior of selenium (Se). I disagree with EPA's decision for several reasons. First, there is overwhelming data (both historical and the 2010 EGU ICR) that support using filterable PM as the surrogate for antimony, arsenic, beryllium, cadmium, chromium, cobalt, lead, manganese, and nickel. While there is variability in the Se results, EPA's own data show exceptionally high removal percentages for all of the metals for all coals and all control technology configurations.[2] EPA states that the results for Se removal were less consistent. However, when we examine EPA's results closely, it appears that EPA is trying to distinguish Se where there is very little real difference. For example, EPA states that the results for Se control were consistently very good when subbituminous coal was fired. EPA also states that when a fabric filter was the primary control device, Se control was consistently good. Thus, the only questionable configuration for Se control appears to be when bituminous coal is fired and an electrostatic precipitator (ESP) is the only control technology.

[1] 76 Fed. Reg., 24,976 (May 3, 2011).
[2] 76 Fed. Reg. 25,038, col. 3 (May 3, 2011).

I believe EPA has unnecessarily complicated the control and regulation of non-Hg HAP metals based on shaky technical grounds. My analysis of the ICR data leads us to conclude that a unit cannot comply with the emission limits in the proposed rule while burning bituminous coal and only having ESP control technology. EPA's own analysis projects the installation of fabric filters for 166 GW of capacity.[3] Likewise, an owner/operator cannot burn bituminous coal and expect to comply with the proposed acid gas limitation without installing either wet or dry scrubbing. A simple calculation demonstrates that bituminous coal with a nominal chlorine content equal to 750 ppm will require approximately 97 percent removal to comply with the proposed 0.002 lb/10^6 Btu limit. This is a significant scrubbing requirement and will almost certainly require wet scrubbing.

$$HCl\ inlet = \frac{750\ ppm}{12,000\ Btu/lb} \times \frac{36.5\ lb\ HCl}{35.5\ lb\ Cl} = 0.064\ {}^{lb}/_{10^6\ Btu}$$

$$Removal = \frac{In - Out}{In} = \frac{0.064 - 0.002}{0.064} \times 100\% = 97\%$$

I believe all stakeholders would be better served with a filterable PM limit as the single non-Hg HAP metals surrogate. EPA should be aware that to the extent Se is neither controlled nor adequately characterized by filterable PM, the Agency certainly has not demonstrated that Se is collected in the condensible PM fraction. EPA's concluded that Se removal percentages were not consistent when burning bituminous coal with only ESP control.[4] But this conclusion is based on the observation that some Se was collected in the Method 29 impingers and not on the Method 29 filters. However, EPA Method 29 impingers are not the same solutions as Method 202 impingers, which is the EPA reference method for condensable PM. In other words, EPA has failed to provide any data that demonstrate Se is present in condensable PM.

Condensable PM Measurement Issues

As noted above, total PM consists of two components, filterable PM and condensable PM. Since no single EPA method measures both filterable and condensable PM, a minimum of two different EPA sampling methods must be utilized to determine total PM emissions. For the ICR, EPA specified OTM-28 for condensable PM measurement. Since the section 114 ICR letters were mailed by EPA to EGUs (December 2009), the requirements of OTM-28 have been incorporated into EPA Method 202, which is one of the proposed compliance methods. Method 202 has been flawed since it was issued by the Agency 20 years ago. Despite recent cosmetic changes to Method 202 by the Agency, the method remains flawed and yielded very inconsistent ICR test results. As EPA is aware, the Electric Power Research Institute (EPRI) has conducted numerous analyses on the EGU ICR data, and EPRI will be submitting detailed comments under its own cover. Among the EPRI results I am privy to are a series of regression analyses of the individual metals versus the various PM fractions (i.e., filterable, condensable and total). The PM component with clearly the least explanatory power was condensable PM. The reason for

[3] *Regulatory Impact Analysis of the Proposed Toxics Rule*, U.S. Environmental Protection Agency, p. 8-14, March 2011.

[4] Note that EPA's did not evaluate Se control with an ESP followed by a wet scrubbing system, which is quickly becoming the most prevalent configuration for burning bituminous coal.

lack of correlation is likely not only due to the fact that non-Hg metals are predominately solids (filterable PM) at stack temperatures, but also because of the poor quality of condensable PM data collected with EPA Method 202.

Suggested Approaches for Establishing Filterable PM Emission Limit

One alternative would be to follow the approach contained in EPA's PM Floor Spreadsheet,[5] but focus on filterable PM instead of total PM. Starting with the above-referenced EPA spreadsheet file, I identified all of the filterable PM data and sorted the units based on each unit's minimum reported filterable PM test average.[6] This is consistent with the procedure EPA has used for Hg and HCl for all of the proposed subcategories. I constructed a new filterable MACT pool using the 131units with the lowest reported filterable PM test averages. I computed $\bar{x}$ based on the average filterable PM emission value for each of the 131 units in the MACT pool, and I computed the variance of all the test values (i.e., 221 averages) in the pool. This approach yielded a upper prediction limit (UPL) = 0.013 lb/10^6 Btu.[7] As documented in Section III of this memorandum and also discussed in Allen Analytics Report to UARG, EPA's UPL approach fails to properly account for variability. Using an "R" multiplier value equal to 1.69, as recommended in the Allen Analytics Report,[8] yields a recommended filterable PM limit equal to 0.022 lb/10^6 Btu.[9]

As a second approach to developing a filterable PM limit, I wanted to more closely follow the approach EPA used for its Hg floor analysis. The key to this approach is to consider all of the Part II and Part III filterable data in the variance calculation, rather than just the minimum values that were used above and are included in the EPA PM floor calculations. By including all of the Part II filterable PM data along with the Part III filterable PM data, we may characterize enough hours of plant operation to eliminate the need for variability adjustment factor described in the Allen Analytics Report. Starting with the 131 best performing units in the MACT pool (determined by minimum test values), I went into EPA Microsoft® Access Database[10] to retrieve all of the additional filterable PM data for the MACT pool units. I added all of available Part II filterable PM data to the 131 MACT pool units, which were defined by the lowest reported filterable PM test averages. First, I computed $\bar{x}$ based on the average filterable PM emission value for each of the 131 units in the MACT pool, and I computed the variance of all the test

[5] See the EPA file, *floor_analysis_coal_pm_031611.xlsx.*
[6] It is difficult to divine exactly the procedure used by EPA in computing total PM from the filterable and condensable PM measurements provided in the above-referenced spreadsheet file. EPA may have inadvertently omitted the filterable PM data contained in Column "CS" of Tab "PM_coal_MMBtu." Regardless, I used all available filterable PM data from EPA's spreadsheet file for my analysis.
[7] Details of EPA's UPL approach are discussed in Section III of the memorandum.
[8] *Evaluation of Proposed EGU MACT Floor Emission Limits for Hg, PM, and HCl*, prepared by Jonathan Allen, Sc.D., P.E., Allen Analytics LLC, Tucson, AZ, July 30, 2011.
[9] The Allen report also contains a maximum recommended "R" values of 3.89, which appears to be driven by a PM CEMS dataset that contains an inordinately long start-up event. Given that most of the other R values are in the 1.3 to 1.7 range plus the fact that EPA's proposed rule provides for a diluent cap during start-up and shutdown events, I believe the "R" = 1.69 recommendation is reasonable, albeit it perhaps conservative.
[10] See the EPA file, *eu_partiii.mdb.*

values (i.e., 460 averages) in the pool. This approach yielded a UPL = 0.033 lb/10^6 Btu.[11] Second, I computed $\bar{x}$ based on the minimum filterable PM emission value for each of the 131 units in the MACT pool, and I computed the variance of all the test values (i.e., 460 averages) in the pool. This approach yielded a UPL = 0.029 lb/10^6 Btu. These two UPL results are driven by the variance term, which may be unrealistically large. I did not screen any of the Part II filterable results based on either outlier or data quality concerns. Visual inspection of the filterable PM results for some of the units revealed more variability than one would expect for "normal" operating conditions. Some of the PM test results may have been for specialized situations such as fuel tests or testing to develop a compliance assurance monitoring (CAM) plan. I simply did not have time to review individual test reports in effort to ferret out unrepresentative test results, which would serve to reduce the variance term as well as the calculated UPL.

I believe both of the two different computational approaches described and implemented above illustrate that EPA's UPL approach based on limited datasets fails to capture long-term variability. Both approaches indicate that EPA's UPL approach does not yield emission limits that are truly achievable on a continuous basis. This conclusion is confirmed by the results obtained from using the variability adjustment factor suggested in the Allen Analytics Report as well as from using EPA's UPL approach, but only when used with adequate datasets to sufficiently characterize variability.

III. CRITIQUE OF EPA'S VARIABILITY ANALYSIS

Overview of EPA's Variability Analysis

EPA's attempt to address emission variability through the use of an upper prediction limit (UPL) is fundamentally flawed. The UPL approach does not accomplish what the Agency purports it to accomplish. Failing to address variability correctly means EPA's proposed rule is technically deficient and also at odds with several rulings by the D.C. Circuit Court of Appeals.[12]

It is not obvious where within our comments is the optimum location to introduce this issue, but I believe all of the EPA floor calculations are fatally flawed because of EPA's decision to use the lowest measured value for each unit, when multiple test values were available. This is particularly an egregious error in the total PM calculations because EPA often had simultaneous or near simultaneous measurements of filterable PM employing different EPA methods and the Agency ignore variability and simply combine the lowest filterable PM concentration with a condensable PM concentration to arrive at a total PM concentration. (I address the PM measurement issue of multiple methods in Section IV of this memorandum.)

---

[11] This UPL result is driven by the variance term, which may be unrealistically large. I did not screen any of the Part II filterable results based on either outlier or data quality concerns.

[12] See, for example, *National Lime Association v. EPA*, 627 F.2d 416 (DC Cir 1980) (holding that EPA failed to show how the standard proposed was achievable under the range of operating conditions that might affect the emission that was being regulated).

I decided to introduce this topic in the variability section to illustrate how the Agency's decision biases the floor results and completely fails to address within unit variability. Based on one of the EPA spreadsheet files, apparently eight complete Hg test series were submitted for the BL England facility.[13] The test set averages ranged from a low of 0.0296 lb/10^{12} Btu to a high of 2.43 lb/10^{12} Btu – almost two orders of magnitude. Yet, EPA used only the lowest value (0.0296) to compute the mean of the best performing units for mercury. Based on the eight sets of BL England Hg data, the best or central estimate of performance is given by the mean, which is equal to 0.578 lb/10^{12} Btu.

What UPL Does and Does Not Accomplish

It would be perhaps gratuitous to quibble with some of EPA's statements leading up to the Agency's UPL analysis. For example, EPA incorrectly applies the Central Limit Theorem and concludes that the Agency can assume normal distributions because the sample size (ICR datasets, in this case) is 15 or larger. In point of fact, when we compute the kurtosis and skewness statistics for any of the ICR datasets (PM, HCl and Hg), it becomes abundantly clear that the data distributions are not normal. Also, as mentioned above, EPA does not actually calculate the correct average ($\bar{x}$) performance of the MACT pool because the Agency elects to use the minimum value of multiple test results. But such quibbles are not at the heart of EPA's misdoings.

Issues With EPA's Variability Analysis

EPA's attempt to address emissions variability through the use of an upper prediction limit (UPL) is fundamentally flawed. The UPL approach does not accomplish what the Agency purports it to accomplish. Failing to address variability correctly means EPA's proposed rule is technically deficient. EPA used the following formula to estimate the UPL for the best performing unit:

$$UPL = \bar{x} + t(0.99, n-1) \times \sqrt{s^2\left(\frac{1}{n}+\frac{1}{m}\right)}$$

Where:

n = the number of test runs for best performing source
m = the number of test runs in the compliance average
$\bar{x}$ = mean of the data for top performing unit
t(0.99, n – 1) = 99th percentile of the T-Student distribution with n – 1 degrees of freedom
s^2 = variance of the data from the top performing source.

The problem with EPA's approach is that the Agency is applying the UPL formula to very incomplete data, especially for the new unit analysis. For each HAP, EPA typically used three sampling runs that were conducted under relatively constant operating conditions and performed very close in time (i.e., at a maximum, over 3 consecutive days) for the single, best performing

[13] See, *floor_analysis_coal_hg_051811.xlsx.*

unit. The variance (s^2) that EPA calculates using the formula above is only representative of a very limited set of operating conditions and probably little, if any, fuel variability. Thus, EPA is only predicting the 99th percentile of a very limited range of operation and not necessarily a level that can be complied with at all times and under all operating conditions.

For total PM and HCl, EPA determined that there should be 131 units in the MACT pool for the category of coal-fired units designed for coal > 8,300 Btu/lb. Because EPA preselected among the best performing units for ICR stack testing, I agree with EPA's decision regarding the size (12% of 1,091 = 131) of the MACT pool for PM and HCl. The problem is that EPA's approach, at best, estimates the 99th percentile of 131 individual data snapshots. EPA has no idea how representative any single ICR snapshot is or what frequency of time a given unit can achieve the level at which operated during the ICR test.

Our concern with EPA's incomplete treatment of emission variability includes both the new unit and existing unit analysis. It may be easier to comprehend our criticism in the context of EPA's new unit analysis because we are only working with the emissions from a single, best performing unit.

EPA's UPL approach is flawed because the Agency's analysis fails to address how representative any perceived best performing ICR stack testing results are relative to long term operation and performance. The ICR instructions specifically directed (1) the stack tests to be representative of the fuel that is routinely burned and (2) all pollution control equipment to be operated in accordance with manufacturers' specifications for proper operation during emissions testing.[14] What the ICR test results do not tell us is how representative are the fuel and operating characteristics for each unit in the database of longer term performance. Is the ICR data snapshot indicative of : (1) a really good day(s) of fuel and operating parameter characteristics; (2) an average day(s); or (3) performance that can be achieved a relatively high percentage of the time? The plain truth is that EPA idea has no idea how representative any give ICR data snapshot is of long term performance. Since EPA cannot determine where along a unit's performance "curve" any given ICR test result resides, the Agency's claim that the proposed emission limits are based upon a 99th percent UPL is simply speculation and not supported by an information or analysis in the rulemaking docket.

My analysis of CEMS data, which will be presented later in this memorandum, demonstrates just how inadequate EPA's variability approach actually is in the context of the proposed limits being continuously achievable. As we noted earlier, some of the issues are easier explained in the context of proposed emission limits for new units because only a single unit is involved. EPA's proposed HCl limit for new units is 0.30 lb/GWh. The proposed limit is supported by the Agency's UPL analysis of HCl emission data from the Logan Generating Plant.[15] However, the same EPA spreadsheet lists a Part II HCl test result for the Logan Generating Plant equal to 0.66

---

[14] *Supporting Statement for OMB Review of ICR No. 2362.01 (OMB Control Number 2060-0631): Information Collection Request for National Emission Standards (NESHAP) for Coal- and Oil-Fired Electric Utility Steam Generation Units*, Part B, p. 7,U.S. Environmental Protection Agency, Research Triangle Park, NC, December 24, 2009 (cited hereafter as "2010 ICR Supporting Statement").

[15] See EPA Spreadsheet, *floor_analysis_coal_hcl_031611.xlxs*, Tab = HCl_New_MW, Cell = C102.

lb/GWh, which is over twice the proposed limit. In other words, the unit upon which EPA based the proposed HCl emission cannot meet that limit based on a single additional stack test. One can only imagine how the Logan Generating Plant would fare under the proposed HCl limit using CEMS data, including periods of start-up and shutdown.

Analysis of PM CEMS Data

In anticipation of EPA's proposed EGU MACT rule, I solicited PM CEMS data from a number of utility companies. However, in requesting PM CEMS data, I limited my requests to those units selected by EPA for mercury, non-Hg metals and PM testing in Part III of the 2010 ICR. RMB's request was very basic; we asked for a minimum of one calendar year of all quality assured, hourly PM emission averages.

I conducted several different analyses of the PM CEMS data. Since we received and began analyzing the PM datasets before Administrator Jackson signed the proposed EGU MACT rule (March 16, 2011), I decided to examine several different averaging times, ranging from 3-hour rolling averages to 30-day rolling averages. This type of analysis illustrates how the length of averaging time impacts the ability of a unit to comply with a given numerical limit. My results are summarized in a series of tables presented in Appendix A to this memorandum.
The values presented in the cells of Tables A1 – A6 represent the percentage of time the unit's emissions were below the hypothetical numerical limit shown as the each header. The reader should be aware that the number of compliance periods in a calendar year is dependent on the averaging time. I enumerated the number compliance periods in a year, using the unrealistic operating scenario[16] that a unit operates 24 hours per day for 350 days, which equals 8,400 operating hours.

Averaging Time	Number of Compliance Periods
8-hour rolling	8,393
24-hour rolling	8,377
Daily (24-hour discrete)	350
Weekly (168-hour discrete)	50
30-day rolling	321

With reference to the tables, if a unit complies with a numerical limit 90 percent of the time, then it would have 838 (0.10 x 8,377) exceedances based on 24-hour rolling averages but only 32 (0.10 x 321) exceedances based on 30-day rolling averages. The results shown in the tables are somewhat counter-intuitive results because *conventional wisdom* is that as the averaging time increases, so does the compliance rate. I believe the results shown in the tables may be partially the result of the vagaries of the computations (i.e., the number of potential exceedances in a year is a function of the averaging period examined). However, I believe there is an "operational" explanation for the observed results. That is, when an excursion is high enough to cause an exceedance of a longer-term average (e.g., 30-day rolling) that event remains in future

[16] It is unrealistic to assume a unit never experiences a forced outage during the year, and thus operates a full 24-hour cycle each day. Nonetheless, this example illustrates how the number of compliance period in a year vary as a function of assumed averaging time

calculation and results in additional exceedances. Thus, notwithstanding conventional wisdom regarding the lengthening of averaging time, my analysis shows that the percent compliance does not increase significantly as the averaging time increases. Thus, while the number of exceedances decrease with increasing averaging time, the percentage of time does not decrease very significantly.

PM CEMS Data Illustrates Emission Variability is Not Accounted for Properly

One of the compliance options for non-Hg metallic HAPs in the proposed rule is to achieve a total (filterable plus condensable) PM emission limit of 0.03 lb/10^6 Btu based on periodic stack tests. The owner/operator may then choose to demonstrate continuous compliance by installing and operating a PM CEMS in accordance with proposed §63.10010(g).[17] [In another section of comments, I have explained that proposed approach to establishing the filterable PM limit is unworkable and therefore must be corrected.] Since EPA's proposed approach is based on using 30-day rolling averages, I used the hourly PM CEMS data to calculate a series of 30-day rolling averages. Next, I prepared a cumulative distribution frequency (CDF) plots of each PM CEMS dataset. The CDF plots are presented as Figures B-1 through B-6 in Appendix B. The x-axis of the CDF plots represents filterable PM emissions in units of the standards, lb/10^6 Btu. The y-axis shows the cumulative or combined percentage of time that the 30-day rolling averages are less than the corresponding PM emission limits, shown on the x-axis.
Figures B-1 through B-6 show that an ICR stack test can be conducted at any point along a unit's emission distribution. EPA's UPL calculation fails to include or address this fact. For example, Figure B-1 indicates that the ICR test was conducted at an emission rate that Unit 7 operates at infrequently. Conversely, Figure B-2 shows that the ICR test was conducted at an emission rate that Unit 8 can achieve almost all the time. Interestingly enough, Units 7 and 8 are "sister" units at the Oak Creek Power Plant (OCPP), receive coal from the same coal pile and have similar PM control technologies. Again, EPA's UPL analysis is completely indifferent to this situation (i.e, one unit can achieve its ICR emission rate infrequently while a sister unit achieves its ICR rate almost all the time). As we stated earlier, EPA's UPL analysis focuses on identifying the 99th percentile of snapshot tests, but has no comprehension of how representative any single ICR snapshot test is of long-term performance.

The CDF plots in Appendix B clearly show, EPA's UPL analysis fails to account for how representative each ICF data snapshot is of longer term emissions. Moreover, an acceptable response is not, "we (EPA) proposed 30-day rolling averages and that solves the short-term data issue/problem." The CDF plots prove that EPA's UPL analysis has not demonstrated that the proposed emission limits are achievable, as required by the Clean Air Act and subsequent case law.

IV. EPA's PROPOSED TOTAL PM LIMIT MAY BE BASED ON BIASED DATA

Although not elaborated on in any of the technical support document, EPA's approach for determining total PM emission appears to bias the results low. Total PM consists of two components, filterable PM and condensable PM. Since no single EPA method measures both

[17] 76 Fed. Reg. 25.112, col. 2 (May 3,2011).

filterable and condensable PM, a minimum of two different EPA sampling methods must be utilized to determine total PM emissions. For the ICR, EPA specified OTM-28 for condensable PM measurement.[18] However, because all filterable methods may not be applicable to all types of stacks, the ICR included several filterable PM measurement methods (i.e., Method 5, OTM-27 and Method 29).

For the Part III ICR data, I concluded EPA selected the lowest filterable PM value among up to three options and combined this result with the condensable PM result from OTM-28. EPA implemented the same approach in processing any Part II PM data to arrive at total PM. Next, EPA selected the lowest total PM value between the Part III and Part II ICR results, and placed this value in a spreadsheet column titled, *Total_Particulate_Calc-min.*[19] EPA then sorted the total PM data, identified the best performing (lowest) 131 units, conducted its UPL calculation on these 131 values and arrived at 0.026 lb/10^6 Btu. At best, EPA has determined the 99th UPL of a group of minimum values.

Comparing ICR PM Test Results to Historical Test Results

As I began to review EPA's initial releases of the ICR Part III data (October 2010), I noted some of the lowest filterable PM results I have ever observed. The following antidotal experience caused me to research this issue in greater detail. There are two units in the database for which my company (RMB) has been working on a Pollution Control Upgrade Analysis for 2 to 3 years. RMB staff reviewed numerous stack tests for these units and conducted a number of computer simulation runs – all which resulted in us concluding that EPA's consent decree limit of 0.03 lb/10^6 Btu was not achievable other than by replacing the existing electrostatic precipitators (ESPs). I was surprised to find in EPA's ICR database a filterable PM result for one of the units equal to 0.013 lb/10^6 Btu. As a result of this experience, I began to contact utility companies and EPRI to inquire if others were observing unusually low PM concentrations.

The response to our calls was prompt and convincing. Following are two graphs, Figures 1 and 2, prepared by Southern Company Services. Figure 1 is for Plant Scholz, a facility with two coal-fired EGUs located in Florida. As Figure 1 shows, the two units have been individually tested a number of times over the last 5 years. The mean PM emission rate for Units 1 and 2, based on 16 tests (48 runs) is 0.018 lb/10^6 Btu. The probability of obtaining the Method 29 result of 0.004 lb/10^6 Btu is about 2.5 percent; the probability of obtaining the OTM-27 result of 0.002 lb/10^6 Btu is less than 1.5 percent. In other words, if the Scholz stack were tested 100 times, you would expect to obtain the ICR-reported results 1 to 3 times. The statistics for Plant Miller (Figure 2) are not as impressive as for Plant Scholz because the number of tests is much smaller. Nonetheless, Figure 2 shows that the ICR results are distinctly lower than previous PM test results.

---

[18] Since the section 114 ICR letters were mailed by EPA to EGUs (December 2009), the requirements of OTM-28 have been incorporated into EPA Method 202. Similarly, the requirements of OTM-27 have been incorporated into EPA Method 201A. See, 75 Fed. Reg. 80,118 (December 21, 2011).

[19] See EPA Spreadsheet, *floor_analysis_coal_pm_031611.xlxs*, Tab = PM_Coal_MMBtu.

Figure 1. PM Emissions Test Results for Plant Scholtz

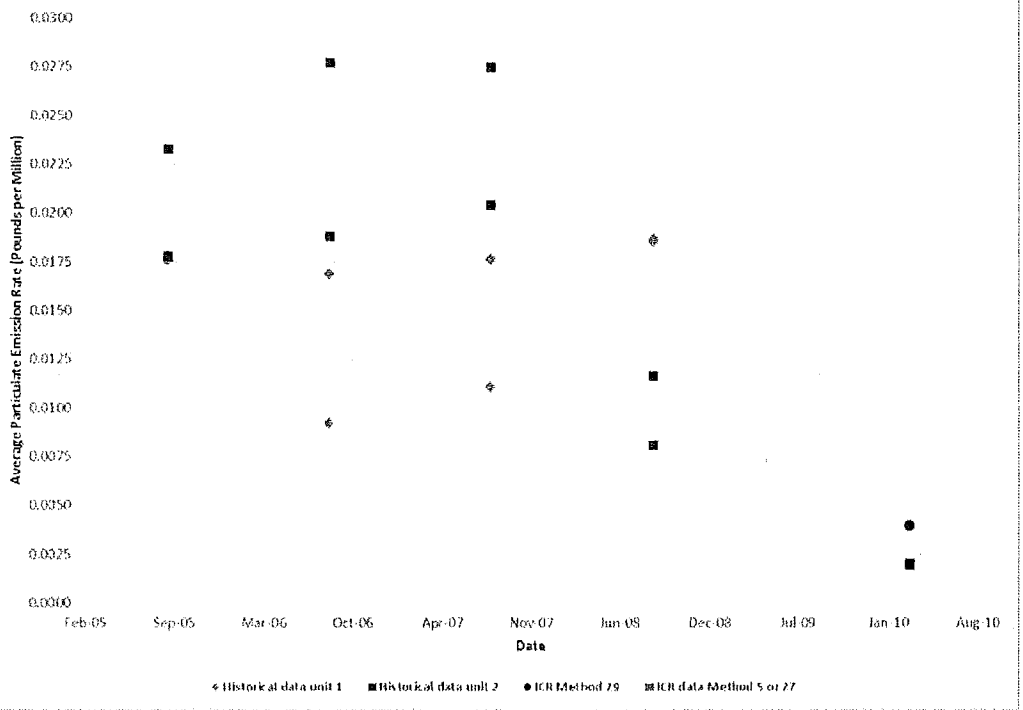
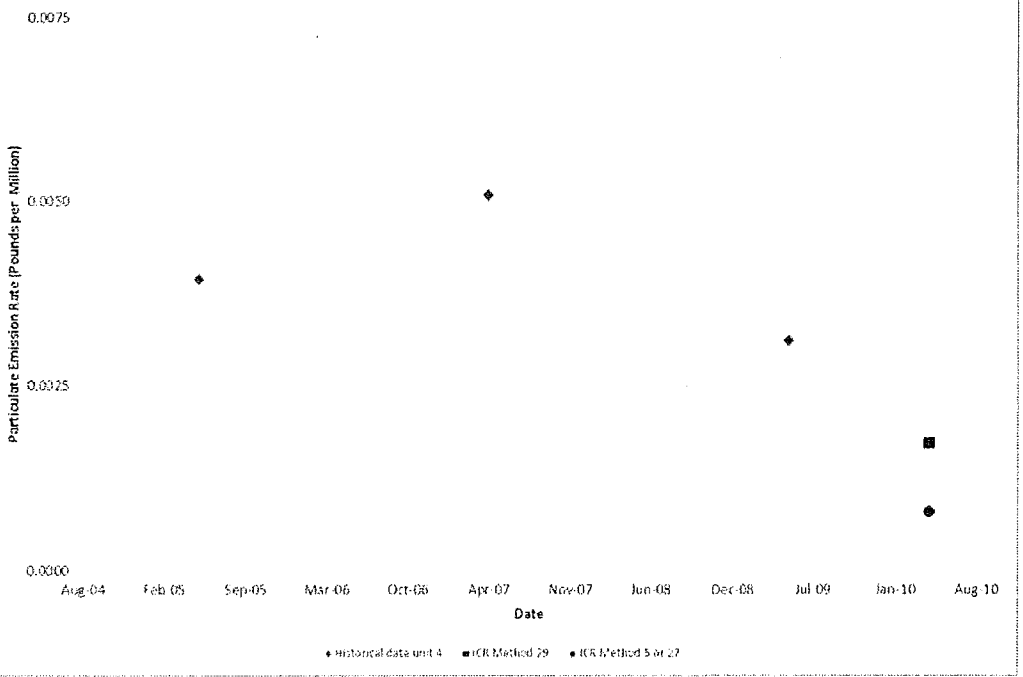


Figure 2. PM Emissions Test Resutls for Plant Miller



Developing and Testing Hypothesis

My thought process was to identify the differences between the required ICR testing and historical PM stack testing on EGUs. There were two obvious differences: (1) the ICR sampling runs were of longer duration (i.e., 4 hours versus 1- or 2-hour runs) and (2) the ICR specified Method 29 as the primary filterable PM method whereas historical tests relied on Method 5.

I know EPRI and the American Electric Power (AEP) undertook a field project to test the above-posited explanations. That work is reported in a standalone report, which will be submitted to EPA under separate cover. I understand that the AEP project was able to replicate the situation where Method 29 produced significantly lower PM measurements than did concurrent Method 5 results. Of course, the primary different between the two methods with respect to PM determinations is that Method 29 recommends a high purity quartz filter without organic binder whereas Method 5 specifies a glass-fiber filter without organic binder.

Additional Tests Designed to Specifically Evaluate Potential PM Bias

Earlier in 2011, I reviewed some of the above-discussed observations with representatives from Sunflower Electric Cooperative (Sunflower). Sunflower has a second unit permitted for its Holcomb station, but did not commence construction prior to EPA's May 3, 2011 EGU MACT proposal date. Presumably, Holcomb 2 would eventually be subject to whatever final MACT limits EPA may issue for new coal-fired units. Therefore, Sunflower is interested in any potential low biases in the ICR data that could unfairly lower the new source PM limits.

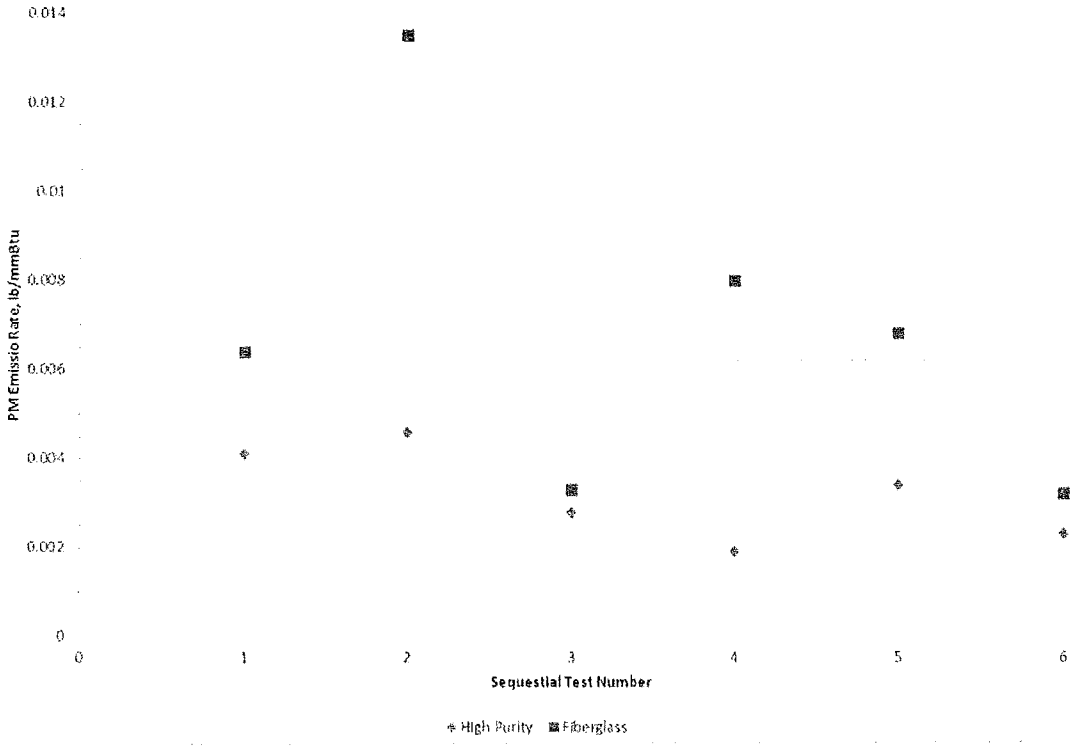
The following test matrix was developed and implemented at Holcomb 1 in April 2011. Although Holcomb 1 has been in commercial service 28 years, it is a well-controlled unit being one of the first units built subject to EPA's Subpart Da new source performance standards. The test matrix was implemented in duplicate on consecutive days. The Method 29 (M29) runs used high purity quartz filter without organic binder, and the Method 5 (M5) runs used a glass-fiber filter without organic binder. The results of these tests along with some interesting Method 29 findings (method detection limits versus some of EPA's proposed emission limits for new sources) are documented in a final test report.[20]

Run	Duration	Method	Temperature	Duration	Method	Temperature
1	4-Hr	M29	250 °F	4-Hr	M5	250 °F
2	2-Hr	M29	250 °F	2-Hr	M5	250 °F
3	2-Hr	M29	320 °F	2-Hr	M29	250 °F

I extracted filterable PM results from the above-referenced report and prepared Figure 3. Regardless of test duration (i.e., 2-hr versus 4-hr), the Holcomb results indicate that Method 29 results with a quartz filter consistently yields lower PM concentrations. While EPA may not be able to resolve this issue during the comment period, at a minimum EPA should revise Table 5 of the proposed rule to allow either Method 5 or Method 29 for PM testing.

[20] *Filterable Particulate and Trace Metals Emission Study Test Report,* prepared by Platt Environmental Services, Inc., Oak Brook, IL, prepared for Sunflower Electric Power, July 2011.

Figure 3. Comparative PM Measurements for Holcomb Unit 1



V. EPA's MERCURY MACT POOL IS NOT JUSTIFIED

EPA explains that the Agency targeted the best performing 15 percent of the coal-fired EGUs for ICR testing in the three HAP groups. Therefore, EPA concluded it had emissions data or information on the best performing 12 percent of all existing coal-fired EGUs. In the proposed rule, EPA used data from 12 percent of all units in the category for the two HAP groups. For Hg, EPA decided to only use "*the top 12 percent of the data obtained because, even though we required Hg testing for the units testing for the non-Hg metallic HAP, we did not believe those units represented the top performing 12 percent of sources for Hg in the category at the time we issued the ICR.*"[21]

This statement is inconsistent with the previous EPA statements and suggests some revisionist thinking. First, in responding to a comment received during the ICR notice and comment process regarding the selection of coal- and oil-fired units for testing. EPA states:

> *For the Hg and other non-mercury metallic HAP group, EPA believes that units with the newest PM controls installed represent those units meeting the lowest PM emission limits and, thus, are believed to be among the top performers with respect to Hg and other non-*

[21] 76 Fed. Reg. 25,023, col. 1 (May 3, 2011).

> *mercury metallic HAP emissions. Therefore, EPA has selected 175 units with the newest PM controls installed; of these 175, the newest 170 operating units will be required to conduct Hg and other non-mercury metallic HAP testing. This number includes units with ACI installed and other control configurations are also included (e.g., for SO_2 and NO_X control).*[22]

Then, again in the Agency's final Supporting Statement to the Office of Management and Budget (OMB), EPA discusses selecting units with activated carbon inject (ACI) and being identified as top performing units. *"The units selected also include a number with ACI installed. As units have been identified as meeting the criterion of being a "top performer" unit, substitution of units will not be permitted."*[23] Why would EPA have included "a number of units with ACI installed," if not trying to quantify the best performing units for mercury control? Referring to Attachment 11 of the 2010 ICR Supporting Statement, of the 170 units EPA selected for Hg and non-Hg metallic HAP testing, approximately 120 were equipped with fabric filters. As EPA undoubtedly knows from past experience, when a flue gas contains any measureable amount of unburned carbon, then a fabric filter becomes a highly efficient Hg control technology. The ICR data confirms the above statement. When I examine the best performing Hg units, as tabulated in one of EPA's spreadsheets,[24] I counted a grand total of <u>one</u> unit that does <u>not</u> have a fabric filter among the <u>top 60</u> units.

It is possible that when EPA initially examined a summary of the mercury ICR test results, the Agency may have questioned its success in truly identifying the best performing units. In fact, when I examined EPA's initially-posted spreadsheet, I was very surprised to observe how many of the apparent top performing units were not among the Part III ICR units. Quickly, I determined that EPA's initially-posted spreadsheet[25] contained a data conversion error that resulted in about 20 units having their results underreported by a factor equal to 1,000. This was communicated to EPA, and EPA subsequently revised its analysis.[26,27]

My review of EPA's revised Hg spreadsheet indicates that once EPA fixed its conversion error, the Hg data appear very similar to the other ICR data. That is, EPA did select the best performing units for Hg control so the Agency should have used 12 percent (0.12 x 1,061 = 127) of the coal-fired fleet in the Hg MACT pool.[28]

UARG asked me to recreate EPA's Hg analysis except to base the analysis on 12 percent of the entire subcategory of units designed to burn coal >8,300 Btu/lb, which we believe is 127 units.

---

[22] *Response to Comments Received on Proposed Information Collection Request (Published on July 2; 74 FR 31725), p.27,* U.S. Environmental Protection Agency, Research Triangle Park, NC, November 5, 2009 (emphasis added).

[23] See, 2010 ICR Supporting Statement.

[24] EPA's revised spreadsheet titled, *floor_analysis_coal_hg_051811.xlxs.*

[25] EPA's spreadsheet titled, *floor_analysis_coal_hg_031611.xlxs.*

[26] Letter to EPA Administrator Lisa P. Jackson from Lee B. Zeugin, Counsel to the Utility Air Regulatory Group, dated May 6, 2011.

[27] Letter to Lee B. Zeugin, Counsel to the Utility Air Regulatory Group from Gina McCarthy, EPA Assistant Administrator for Air, dated May 18, 2011.

[28] Since EPA proposed Subcategory 2 for Hg for 30 units designed for coal < 8,300 Btu/lb, the number of units in Subcategory 1 for Hg is equal to 1,091 – 30 = 1,061.

I began my analysis with EPA's latest (revised) spreadsheet, *floor_analysis_coal_051811.xlxs*. Using another spreadsheet file recently made available by EPA, *PartII_III_Hg.xlxs*, I was able to add units that EPA ranked 41st through 127th to the above-reference EPA spreadsheet. In terms of "numbers," I expanded the Hg MACT pool from 40 to 127 units[29] and the number of test averages from 80 to 263. Next, I replicated EPA's UPL approach exactly (i.e., computed $\bar{x}$ as the average of each unit's minimum test average and computed the variance of 263 test averages). I obtained a UPL = 2.10 lb/10^{12} Btu.

There are probably many variants to EPA's approach. A rather obvious approach would be to compute $\bar{x}$ based on the average emission for each of the 127 units in the MACT pool and to compute the variance of all the test values in the pool. This approach yields a higher UPL = 2.24 lb/10^{12} Btu. I have never been entirely comfortable with EPA's ranking and selecting the MACT pool units based on minimum test values. I believe most statisticians would say that extreme values (i.e., minimums or maximums) are the most uncertain measurements in a dataset. Accordingly, I computed the mean of all available Hg test averages, resorted the units based on mean Hg emissions and selected the best performing 127 units. Next, I computed $\bar{x}$ as the average of each unit's mean test average and computed the variance of 266 test averages). I obtained a UPL = 1.42 lb/10^{12} Btu. I believe this is the single, most statistically defensible results that can be obtained from the Hg ICR data.

It is interesting to observe that the last UPL result presented (1.42) does not have the minimum value for $\bar{x}$, but it does contain the minimum variance component. By taking the average before the units are ranked, I believe one is able to minimize the impact of some unusually low Hg values, which may not be able to be replicated. A unit with an unusual and perhaps randomly low Hg measurement gets placed in the pool, but subsequent tests are higher and more variable. I believe this is the fundamental reason why a pool based on minimum measured values does not guarantee or necessarily produce the minimum UPL result.

Lastly, UARG asked if I could provide any alternative results for EPA's Hg MACT pool, which only contains the lowest 40 emitting units. One obvious alternative would be to compute $\bar{x}$ from the mean of each unit's test averages (as opposed to the minimum test average) and to compute the variance from the same 80 test averages that EPA used. Using this approach, I calculate a UPL = 1.30 lb/10^{12} Btu.

VI. EPA'S APROACH TO NEW UNITS SHOULD BE OVERHAULED

For the reasons provided in the memorandum, I believe that the new unit emission limits in EPA's proposed MACT rule will mean that no new coal-fired EGUs can be constructed in the United States. There are several reasons that lead to this conclusion. First, as discussed in detail below, EPA's reliance on a "Franken-Plant" approach has produced a suite of emission limits that no existing unit either meets or can meet. Second, EPA's reliance on its UPL approach for new units fails miserably to account for variability. For new units, EPA has selected the unit with the lowest test average for each HAP and then used only three test runs to account for

---

size equal 127 is determined by taking 12 percent of the coal-fired fleet of 1,091 after subtracting the 30 ed for burning coal with a heating value < 8,300 Btu/lb; that is, 0 .12(1,091 – 30) = 127.

variability. Three test runs, conducted at most over three consecutive days under basically identical operating conditions and probably lasting no more than 4 hours per test run. Does EPA seriously believe a facility that will operate for at least 40 years over a myriad of operating variables (e.g., seasonal challenges, varying loads, upset conditions, etc.) and probably receive coal from a numerous suppliers can adequately be characterized with three 4-hour snapshot tests taken over at most a 72-hour period?

Franken-Plant Approach

Just as Dr. Frankenstein's fictitious monster bore no resemblance to an actual human being, EPA's "Franken-Plant" bears no resemblance to the actual, best performing EGU. EPA's approach for setting emission limits for existing sources is equally flawed to the Agency's approach for setting emission limits for new sources. However, it is easier to demonstrate and comprehend the Agency's error for new sources. The relevant statutory provision is, *the maximum degree of reduction in emissions that is deemed achievable for new sources in a category or subcategory shall not be less stringent than the emission control that is achieved in practice by the best controlled similar source.*[30] Note that the statue refers to a single source – not multiple sources. If Congress had intended for EPA to set emission limits based on a "Franken-Plant" approach, the statue would read -- *best controlled similar sources.*
Using three EPA spreadsheets that the Agency posted on one of its web sites,[31] it is fairly straightforward to determine which individual unit EPA used to set the MACT floor for new units. Those units are listed in Table 1. From Table 1, it should be obvious that no existing unit meets all of the proposed emission limits for a new EGU.

Table 1. EPA's Franken Plant Approach For New Units.

Pollutant	Facility	99% UPL (lb/MWh)	Total Metal Ranking
Total PM	AES Hawaii	0.049	11th
Total Metals	Cedar Bay Unit A	3.3×10^{-5}	--1st --
Antimony (Sb)	AES Hawaii Unit 2	7.6×10^{-8}	11th
Arsenic (As)	Oak Grove Unit 1	1.6×10^{-7}	104th
Beryllium (Be)	Chamber Cogen Unit 2	2.2×10^{-8}	7th
Cadmium (Cd)	Walter Scott Unit 4	3.7×10^{-7}	3rd
Chromium (Cr)	PSEG Mercer Unit 1	1.7×10^{-5}	56th
Cobalt (Co)	Cholla Unit 3	7.2×10^{-7}	62nd
Lead (Pb)	Oak Grove Unit 1	8.8×10^{-7}	104th
Manganese (Mn)	Weston Unit 4	3.1×10^{-6}	3rd
Nickel (Ni)	Weston Unit 4	3.2×10^{-6}	3rd
Selenium (Se)	PSEG Mercer Unit 1	2.5×10^{-5}	56th
Hydrogen Chloride (HCl)	Logan Unit 1	2.6×10^{-4}	n/a

[30] 42 U.S.C. §7412(d)(3), emphasis added.
[31] See, *floor_analysis_coal_pm_031611.xlsx, floor_analysis_coal_hcl_031611.xlsx, and floor_analysis_coal_hg_051811.xlsx.*

Mercury (Hg)	Nucla Unit 1	1.7×10^{-5}	n/a

In the above table, we included dotted horizontal lines to differentiate among the optional emission limits for non-Hg metallic HAPs. The mercury and hydrogen chloride limits must be complied with individually, independent of the option chosen for non-mercury metallic HAPs. Table 1 clearly demonstrates the fallacy in the EPA's "Franken-Plant" approach. Consider the following EPA statements:

> *For the non-Hg metallic HAP, we chose to use PM as a surrogate. Most, if not all, non-Hg metallic HAP emitted from combustion sources will appear on the flue gas fly-ash. Therefore, the same control techniques that would be used to control the fly-ash PM will control non-Hg metallic HAP.*[32]

Oak Grove Unit 1 is EPA's basis for two individual metallic HAP emission limits, arsenic and lead. The key question is how or what control technology could the Oak Grove owners add to meet the other metallic HAP limits given (1) EPA's statement that the same control techniques that work for fly-ash PM also work for non-Hg metallic HAPs and (2) Oak Grove is already the best performing unit for not one but two non-Hg metallic HAPs. EPA's rejoinder to my comment may very well be that compliance with the individual non-Hg metallic HAP limits is an option and not a requirement. This is an inadequate response and misses the point. EPA should not be permitted to base a portion of a suite of emission limits upon the performance of a single unit when that same unit cannot comply with the other enforceable components of that same suite of emission limits. Moreover, the unit that formed the basis for one of the regulated HAPs (e.g., total PM) may not meet one of the other mandatory limits (e.g., HCl). AES Hawaii Unit 1 is the basis of the new unit total PM limit, but the lowest HCl emission results reported in the ICR data are approximately 50 times higher than the proposed new unit HCl limit.[33] Nucla Unit 1 is the basis of the new unit total Hg limit, but the total PM results reported in the ICR data are almost an order of magnitude higher than the proposed new unit total PM limit.

Speaking of AES Hawaii, it is the only coal-fired plant in Hawaii, and the plant has a generating capacity of 180 MW. Unit 1 is only capable of supplying one-half of the steam required by the 180 MW turbine/generator, so Unit 1 in effect has a capacity of 90 MW. The AES Hawaii unit burns coal, which is imported from Indonesia. To supplement the imported coal, the unit also burns old tires, used motor oil, and carbon from Board of Water Supply filters. While EPA is mandated to set limits for new sources based on the *maximum degree of reduction in emissions that is deemed achievable*, such degree of emission control must be *achieved in practice by the best performing similar source*. It is quite clear there is not or most probably will not be another similar source to AES Hawaii in the continental United States. Beyond the *similar source* issue, I believe EPA made a computational error in converting the AES Unit 1 total PM results from input units (lb/10^6 Btu) to output-based units (lb/MWh). EPA mistakenly assumed that both AES units have a capacity of 180 MW; in point of fact, the capacity of the two-unit plant is 180 MW. This error is easily verified in EPA's spreadsheet, because it shows Unit 1 to have a heat

[32] 76 Fed. Reg. 25,039, col. 3 (May 3, 2011).
[33] EPA's proposed HCl limit for new units is 0.0003 lb/MWh; the minimum ICR test result for AES Hawaii Unit 1 is equal to 0.0146 lb/MWh.

rate of 5.03 million Btu per MWh, when the correct value is exactly twice that or 10.06 million Btu per MWh. When I correct the heat rate or conversion error in three individual total PM runs and simply repeat EPA's UPL calculation, I obtain 0.10 lb/MWh. Even as an unrepresentative unit that AES Hawaii may be, it does not support a total PM emission limit of 0.05 lb/MWh.

Alternative Approach for New Units

Over the last several years, I have had the opportunity to consult for several potential EGU owner/operators that now find themselves in a quandary. I am referring to projects that have worked through the regulatory maze (e.g., obtained the necessary permits to construct such as prevention of significant deterioration (PSD) permits) and often withstood legal challenges (e.g., administrative law judge hearings) but now are faced with proposed MACT emission limits that cannot be achieved. Two projects with which I am familiar (i.e., Sunflower's Holcomb 2 and LP Power's Longleaf Plant) are both permitted as non-major (area) sources of HAPs. Since these two projects represent significant generating capacity (895 and 1,200 MW, respectively), we know the emission limits contained in the respective permits must be very low.

Before finalizing the EGU MACT limits for new units, I believe EPA should review the emission limits contained in these various state-of-the-art power projects. I understand that legal opinions exist that permitted emission limits do not equate to "achievable" emission limits. Likewise, I do not believe emission limits like those contained in EPA's proposed EGU MACT rule for new units, which control equipment suppliers cannot and will not guarantee, satisfy the definition "achievable" emission limits either. The simple fact is that after years of negotiation and/or litigation, the emission limits contained in these permits reflect the best level of performance that existing control technology can achieve. In other words, if lower emission limit guarantees were available from the equipment suppliers, the permit limits would be lower.

VII. WORK PRACTICE STANDARDS

During the approximate 6-month ICR Notice and Comment period, I was frequently critical of EPA's approach for dioxin/furan (D/F) testing and also for the non-D/F organic testing. Based on previous experience with the EPRI Power Plant Integrated Systems Chemical Emission Studies (PISCES), I believed that organic emissions from EGUs are predominantly below the method detection limits. Our experience also indicates that organic compounds that are measured at or above detection limits are often the result of contamination. Most of the EPA's stack sampling methods use copious quantities of organic solvents (e.g., acetone, methylene chloride) for recovering samples and/or sample train clean-up; therefore, it can be very difficult to avoid sample contamination when non-organic tests are being conducted concurrent with organic sampling. The ICR Data Review report prepared by EPRI and submitted to EPA concluded that many of the organics measured above the detection limits were likely the result of sample contamination.[34]

[34] *Data Quality Evaluation of Hazardous Air Pollutants Measurements for the US Environmental Protection Agency's Electric Utility Steam Generating Units Information Collection Request.* EPRI, Palo Alto, CA: 2010. 1021216.

All of the above said, I commend EPA on what I believe is the only appropriate means to deal with all of the non-detected organic results in the ICR database. EPA correctly concludes that both the D/F compounds and the non-D/F organic HAPs are routinely below the detection limits of the EPA test methods, even when extended sampling times and volumes were implemented. Work practice standards proposed pursuant to section 112(h) is the most effective and efficient approach for addressing organic emissions from EGUs. I should note that I am not providing comments on the specifics requirements of the work practice standards proposed in this rule as that would be outside my area of expertise; however the work practice concept is clearly superior to specific numerical limits.

VIII. CONFUSING COMPLIANCE REQUIREMENTS

The proposed compliance and testing obligations in EPA's proposed EGU MACT rule are among the most confusing I have ever attempted to review. I believe the most likely coal-fired configurations to survive EPA's proposed rule will (1) opt for the total PM limit and utilize PM CEMS for continuous filterable PM measurements; (2) employ wet or dry scrubbing for acid gas control and elect to use already-installed SO_2 CEMS; and (3) directly monitor mercury using either Hg CEMS or sorbent trap monitoring system. For the above-described scenario in which either all of the regulated HAPs or their surrogates are continuously monitored, **no** additional burdens (e.g., coal sampling, operating parameter limits, etc.) should be imposed. I would like to believe EPA understands this logic, but the manner in which the proposed rule is drafted is very unclear. For example, proposed §63.10007 (c) states:

> (c) *You must conduct each performance test under the specific conditions listed in Tables 5 and 7 to this subpart. You must conduct performance tests at the maximum normal operating load while burning the type of fuel or mixture of fuels that has the highest content of chlorine, fluorine, non-Hg HAP metals, and Hg, and you must demonstrate initial compliance and establish your operating limits based on these tests.*[35]

While proposed §63.10007 (c) continues beyond the above-quoted passage, there is no indication but what every affected facility is subject to the requirements. How any owner/operator is supposed to able to divine that "fuel with the highest content of chlorine, fluorine, non-Hg HAP metals, and Hg" is being burned during the performance test is not explained. However, a more important point is that such clairvoyance is not needed when either all of the regulated HAPs or their surrogates are continuously monitored.

Proposed §63.10008 lays out, in gory detail, a fuel sampling and analysis regime that requires two full columns in the Federal Register. Granted, proposed §63.10008(a) contains an *as applicable clause*, but its meaning is never explained. However, using the same logic as in my previous paragraph, the fuel information is neither meaningful nor relevant when either all of the regulated HAPs or their surrogates are continuously monitored.

---

[35] 76 Fed. Reg. 25,107, col. 3 (May 3, 2011).

IX. PM CEMS DILEMMA

EPA's proposed rule would require the use of PM CEMS for continuous compliance if the owner/operator elects the total PM option instead of complying with either the total or individual non-Hg HAP metal limits. I term this a dilemma because I believe the PM CEMS approach is much less burdensome than frequent stack testing coupled with rigid operating parameter limits that are mandated for either of the other two options (i.e., total or non-Hg individual HAP metals). The dilemma results from the status of PM CEMS technology. That is, commercially available continuous PM monitors, especially those that are based on the principal of light scattering, do not provide a direct measure of PM emissions. By direct measure, I mean that the instrument should measure the mass of PM and the volume of flue gas from which that mass of PM was sampled. Instead, EPA Performance Specification 11 (PS-11) requires that PM CEMS be correlated to a series of manual stack testing results, and the permissible "tolerance" between PM CEMS readings and corresponding stack test values allowed by PS-11 is significant. I have experience with some 20 or so PM CEMS installations on coal-fired EGUs. In general, complying with the initial PM CEMS certification (PS-11) requirements has not proven overly difficult. However, unlike SO_2 and NO_X monitoring systems, the utility industry has little, if any, experience with PM CEMS as continuous compliance devices. Because of this inexperience, it is difficult to forecast the level training and quality assurance support that any successful PM CEMS compliance monitoring program may require.

Measurement Issues During Start-up and Shutdown

Of course, this PM CEMS technology limitation poses another major problem with respect to start-up and shutdown emissions. That is, PM CEMS are not correlated during such times because, among other things, EPA typically requires normal, full-load operation for CEMS certification testing. Moreover, EPA manual test methods like the ones used for PM CEMS correlation testing (e.g., Method 5) were never designed to be used during transient plant operations. Such methods were designed and, if validated, were done so only during steady-state plant operation.

EPA did take a positive step by proposing the use of a 10 percent O_2 diluent cap during start-up and shutdown periods.[36] The proposed diluent cap will prevent some hourly PM values, when calculated in the unit of standard ($lb/10^6$ Btu), from being recorded at extremely high and unrepresentative levels. Unfortunately, EPA's proposed diluent cap does not address the underlying uncertainty and potential inaccuracy. Another constructive step EPA could consider in dealing with start-up and shutdown events (especially since the rulemaking record is silent with respect to any Agency data analysis pertaining to impact of such events) is to weight the hourly $lb/10^6$ Btu PM values by heat input. In other words, each hourly PM value in the 30-day rolling average would be multiplied times the heat input for that hour and the hourly summation

[36] 76 Fed. Reg. 25,106, col. 2 (May 3, 2011). Note that the overwhelming majority of gaseous CEMS currently installed and operating on EGUs measure CO_2 instead of O_2. Most, if not all, EGUs utilize a dilution/extractive technology, which has proven to be reliable and accurate under EPA's Part 75 Acid Rain Monitoring Program. However, it is not feasible to monitor O_2 with a dilution/extractive system because the diluent air contains 20.9 percent O_2 and would overwhelm the amount of O_2 in the sample extracted from the stack.

divided by the total heat input during the 30-day rolling average period. Weighting by heat input would help to normalize the impact of any potentially elevated PM concentrations during start-up events because heat inputs are necessarily lower during start-up events, and the actual mass of PM emitted is also lower because of a reduced firing rate. Emission weighting as suggested above is consistent with the rationale EPA provided in the preamble to proposed revisions to Subpart Da new source performance standards to address that very same start-up and shutdown issues.[37]

Continuous Compliance Issues

To my knowledge, no organization (e.g., EPA, EPRI, ASTM, etc.) have completed any studies that would inform a decision relative to using PM CEMS for determining compliance. However, EPRI has initiated a Tailored Collaboration (TC) project titled, "Comprehensive Continuous PM Monitoring Field Study." In November 2010, the EPRI project team installed five individual PM CEMS on a coal-fired power plant stack in the upper Midwest. The project scope of work includes both initial and final PM correlation testing, perhaps 6 to 8 months apart. The initial correlation test was conducted the first week in February 2011. In between the two correlation test periods, the EPRI team will monitor and document the operability and reliability of each PM CEMS and also evaluate potential alternative calibration techniques. The output of this project will certainly update and inform our knowledge of PM CEMS. Unfortunately, I do not expect those results will be available until the end of 2011.

Conflicting Requirements

Of course, for PM CEMS to be a viable option, EPA must straighten out the disarray of the PM CEMS approach contained in the proposed rule. We begin this discussion with contradictory language, which is probably a drafting error resulting from a hurried rulemaking schedule. Proposed §63.10011(d) states, *the operating limit will be the average of the PM filterable results of the three Method 5 performance test runs.* Meanwhile, Table 3, Line 8 states, *maintain the PM concentration (mg/dscm) at or below the highest 1-hour average measured during the most recent performance test demonstrating compliance with the total PM emission limitation.* This inconsistency can be resolved quickly; we assume EPA meant for the operating limit to be based on the highest 1-hour value recorded by the PM CEMS during the most recent performance test. However, resolving the inconsistency only solves a minor part of the PM CEMS problem. As noted above, PS-11 requires that the PM CEMS output be correlated to a series of manual stack testing runs (minimum of 15 is required). Two of the principal "performance" criteria are the widths of the confidence interval and the tolerance intervals about the correlation line that is drawn. The widths of both acceptable intervals are tied to the PM emission limit; the confidence interval (95%) must be within 10% of the PM emission limit and the tolerance interval must have 95% confidence that 75% of all possible values are within 25% of the PM emission limit. It is not clear how one would demonstrate compliance with PS-11 in the absence of a defined filterable PM emission limit. Thus, the PM CEMS compliance approach as proposed by the Agency is simply unworkable and internally incongruent with existing EPA requirements (i.e., PS-11).

[37] 76 Fed. Reg. 25,061, col. 3 (May 3, 2011).

X. CONCLUSIONS

Given all of the data collected and submitted to EPA by the utility industry, it is unfortunate that EPA management committed to a politically expedient rulemaking schedule and failed to allow EPA staff sufficient time to develop a quality proposed EGU MACT rule. The hurried and committee-like approach required to get a proposed rule signed by 3/16/2011 produced a document with numerous errors and inconsistencies – many of which have been discussed in this memorandum.

EPA's proposed approach to regulate total PM and then establish "custom" filterable PM limits based on performance testing is nonsensical and must be fixed. While I remain cautiously optimistic that EPA will significantly revise its PM approach between the proposed and the final rule, it is difficult to guess what the Agency may ultimately decide. Likewise, EPA must either significantly revise its approach to setting limits for new units or accept that this Administration has brought an end to constructing new, coal-fired EGU capacity in the United States. It has been frustrating trying to preparing meaningful comments on this proposed rule, because numerous revisions and/or corrections are warranted and there is perhaps limited opportunity to respond to those changes.

Table A-1. PM CEMS Emission Summary for Oak Creek Unit 7
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.004	<0.006	<0.008	<0.010	<0.012	<0.014
3-Hr Rolling	60.2%	90.2%	96.9%	97.8%	98.1%	98.6%
8-Hr Rolling	59.6%	89.9%	96.7%	97.6%	98.3%	98.8%
24-Hr Rolling	57.1%	89.8%	96.4%	97.4%	97.9%	99.2%
Daily	56.8%	89.5%	96.5%	97.5%	98.1%	99.4%
Weekly	52.9%	88.2%	94.1%	98.0%	98.4%	98.0%
30-Day Rolling	53.3%	85.1%	90.5%	96.4%	99.8%	99.8%
Long-Term Hourly Mean = 0.004 lb/10^6 Btu						

Table A-2. PM CEMS Emission Summary for Oak Creek Unit 8
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.007	<0.009	<0.011	<0.013
3-Hr Rolling	66.0%	82.7%	91.2%	95.5%
8-Hr Rolling	66.0%	83.9%	92.4%	96.2%
24-Hr Rolling	66.6%	86.0%	93.2%	97.2%
Daily	66.0%	85.3%	92.8%	96.8%
Weekly	66.7%	86.1%	95.8%	100%
30-Day Rolling	65.9%	92.7%	100%	100%
Long-Term Hourly Mean = 0.007 lb/10^6 Btu				

Table A-3. PM CEMS Emission Summary for Cross Unit 1
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.012	<0.015	<0.018	<0.020	<0.022	<0.025
3-Hr Rolling	68.5%	91.1%	96.7%	97.7%	98.1%	98.5%
8-Hr Rolling	68.1%	90.8%	96.3%	97.5%	97.9%	98.2%
24-Hr Rolling	66.6%	89.7%	95.7%	96.4%	96.8%	97.6%
Daily	66.8%	90.9%	96.2%	96.5%	96.9%	97.9%
Weekly	65.2%	87.0%	93.5%	93.4%	93.5%	93.5%
30-Day Rolling	58.5%	82.0%	82.0%	90.0%	90.3%	93.4%
Long-Term Hourly Mean = 0.012 lb/10^6 Btu						

Table A-4. PM CEMS Emission Summary for Clover Unit 1
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.002	<0.003	<0.004	<0.005	<0.006
3-Hr Rolling	26.8%	89.3%	96.0%	99.3%	99.6%
8-Hr Rolling	25.7%	89.8%	95.8%	99.1%	99.4%
24-Hr Rolling	24.0%	90.7%	95.4%	99.1%	99.2%
Daily	24.3%	90.8%	95.7%	99.3%	99.7%
Weekly	21.7%	89.1%	95.7%	100%	100%
30-Day Rolling	16.2%	94.0%	100%	100%	100%
Long-Term Hourly Mean = 0.002 lb/10^6 Btu					

Table A-5. PM CEMS Emission Summary for Clover Unit 2
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.002	<0.003	<0.004	<0.005
3-Hr Rolling	89.1%	99.6%	99.9%	99.9%
8-Hr Rolling	89.7%	99.7%	99.8%	99.8%
24-Hr Rolling	90.0%	99.6%	99.8%	99.8%
Daily	89.2%	99.4%	99.7%	100%
Weekly	94.3%	100%	100%	100%
30-Day Rolling	98.9%	100%	100%	100%
Long-Term Mean = 0.002 lb/10^6 Btu				

Table A-6. PM CEMS Emission Summary for Spurlock Unit 1
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.003	<0.004	<0.005	<0.006
3-Hr Rolling	90.3%	99.6%	99.6%	99.7%
8-Hr Rolling	96.4%	99.3%	99.3%	99.4%
24-Hr Rolling	97.7%	98.7%	98.8%	99.0%
Daily	97.6%	98.8%	99.1%	99.1%
Weekly	98.0%	100%	100%	100%
30-Day Rolling	100%	100%	100%	100%
Long-Term Mean = 0.002 lb/10^6 Btu				

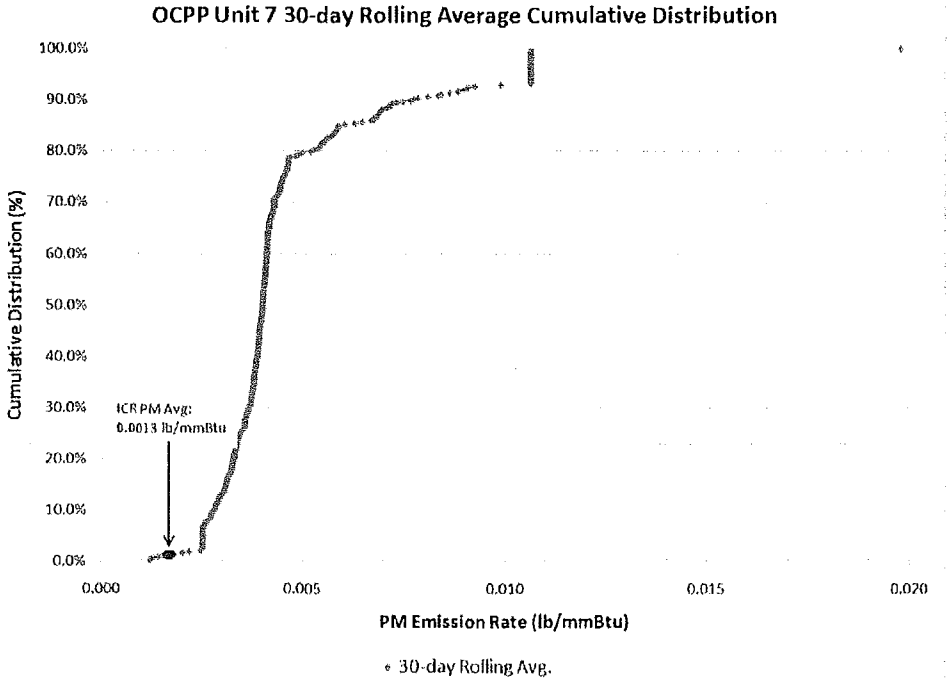


Figure B-1: Oak Creek Unit 7 30-Day Rolling Average Cumulative Distribution Plot

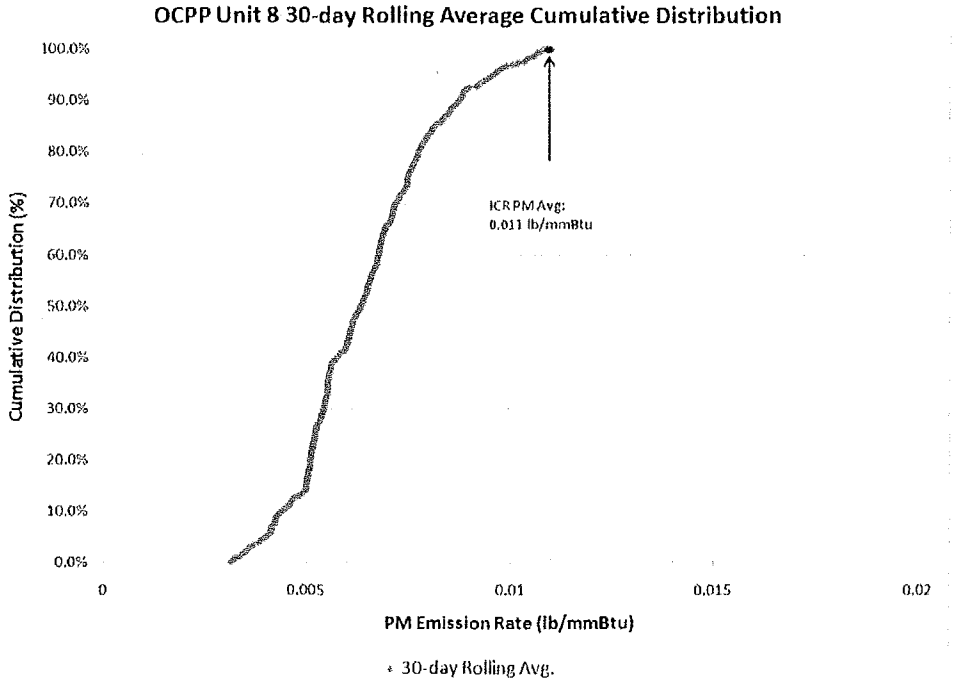


Figure B-2: Oak Creek Unit 8 30-Day Rolling Average Cumulative Distribution Plot

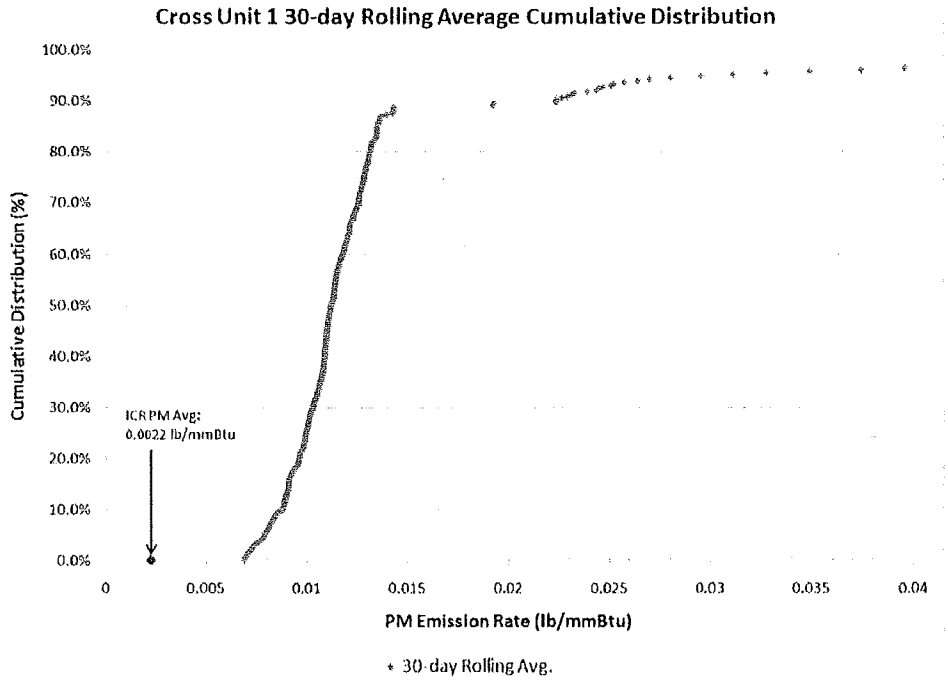


Figure B-3: Cross Unit 1 30-Day Rolling Average Cumulative Distribution Plot

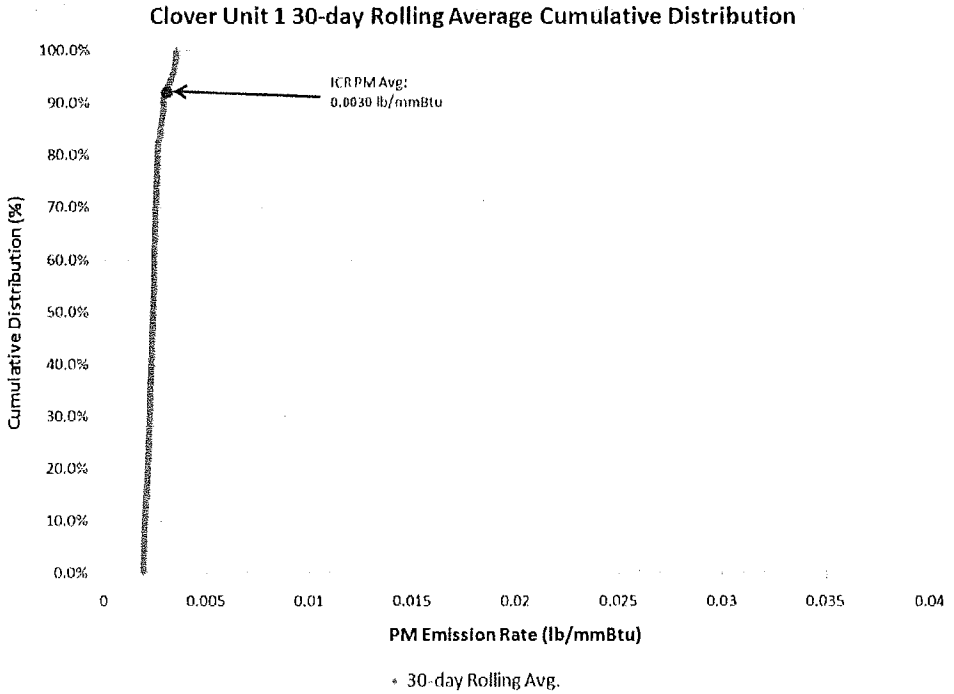


Figure B-4: Clover Unit 1 30-Day Rolling Average Cumulative Distribution Plot

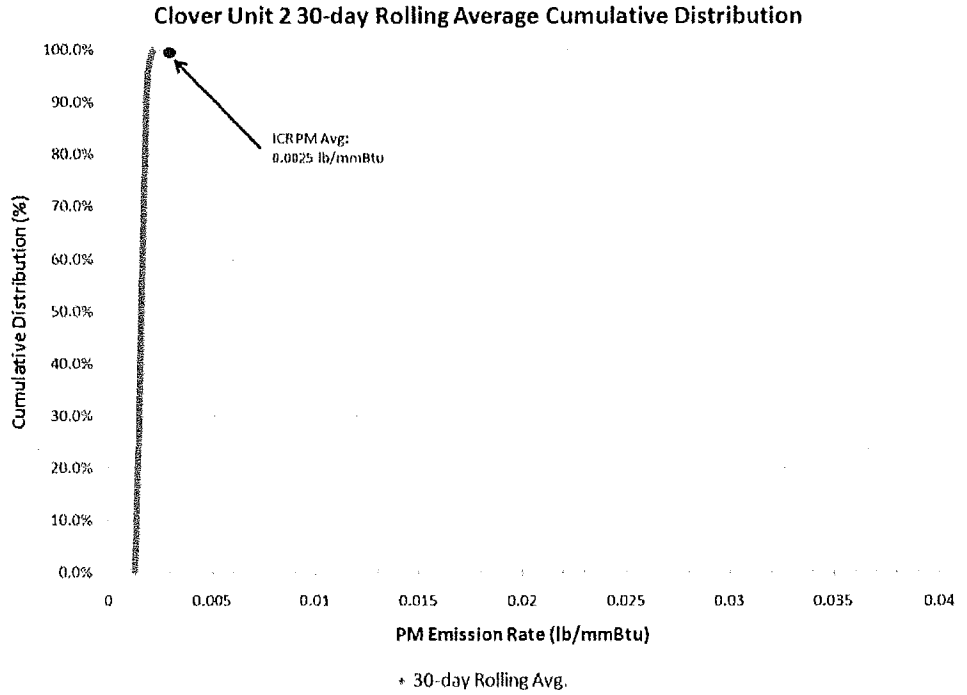


Figure B-5: Clover Unit 2 30-Day Rolling Average Cumulative Distribution Plot

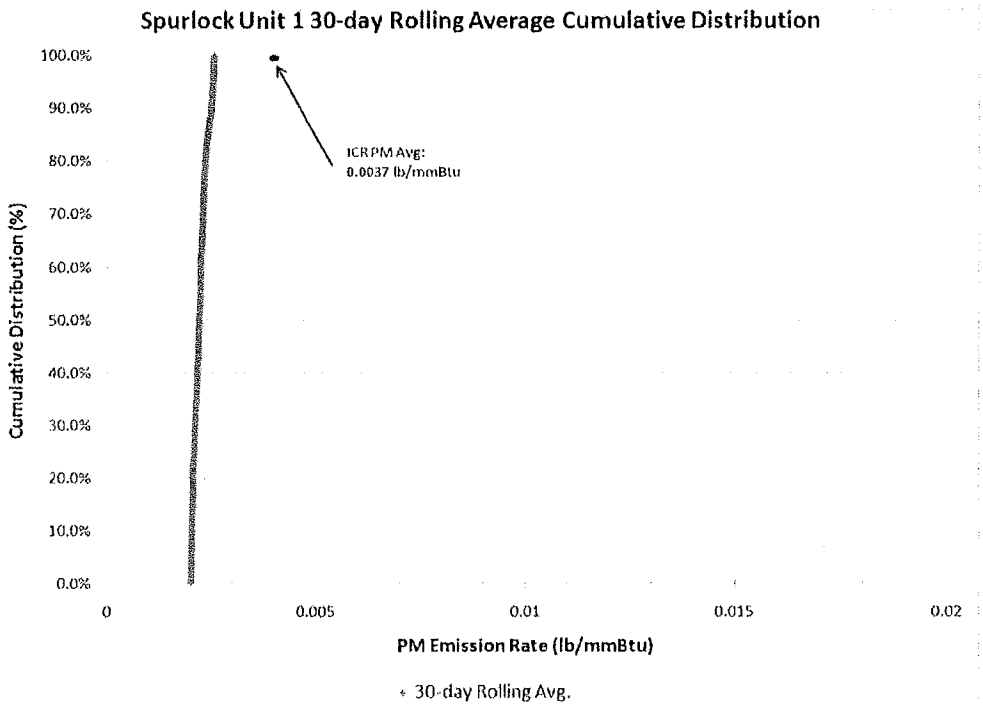


Figure B-6: Spurlock Unit 1 30-Day Rolling Average Cumulative Distribution Plot

ATTACHMENT 4

Upper Predicted Limits for HAP Emissions from Coal-Fired EGUs Using Stack Test and CEMS Data

Final Report for Utility Air Regulatory Group

Jonathan O. Allen, Sc.D., P.E.

Allen Analytics LLC

3444 N. Country Club Rd., Tucson, AZ 85716-1200
jon@allen-analytics.com

3 August 2011

This report (A2-2011-04) was prepared under UARG Agreement AA/HAP/1. The Technical Project Managers were Corey A. Tyree (Southern Company) and J. Michael Geers (Duke Energy).

Executive Summary

EPA has proposed continuous limits for Hazardous Air Pollutants (HAPs) emitted by coal-fired electric utility steam generating units (EGUs). Ideally these emission limits would have been based on long-term continuous data from the entire fleet of EGUs. However, HAP emission data to characterize emissions from the EGU fleet were mainly available as short-term stack tests, and EPA based its proposed HAP emission limits on these data. Stack tests measure emissions on a single day, and so include no information about correlated emissions or emissions during infrequent events. We have developed a method to correct long-term emission limits calculated from stack test data using the available Continuous Emission Monitoring System (CEMS) data. In this method an empirical correction factor, R, is determined as the ratio of the 99th percentile of historical 30-day emission averages to emission limits that would have been calculated from continual stack tests of individual top performing units. These ratios were in the range 1.07 to 3.89 and indicate that, because of correlated emissions and startups in the case of PM, long term emission averages from these units were higher than would have been calculated from stack tests alone.

In order to estimate 30-day rolling average emission limits from short-term stack test results, EPA calculated Upper Prediction Limits (UPLs) for each HAP. The UPL equation was derived from t-statistics and is correct for data that are independent and normally distributed; however the CEMS data demonstrate that HAP emissions are not independent. Further, EPA incorrectly implements the t-statistics UPL equation by 1) calculating mean emissions from the lowest stack tests, not all available stack tests, for the lowest emitting units, and 2) basing compliance on a single stack test instead of 30-day emission averages. Thus the t-statistics UPL equation as implemented by EPA has no theoretical justification and should be viewed as a purely empirical equation. Such an empirical approach may yield reasonable long term emission limits for top performing units; however, if it does not, empirical modifications are justified.

We calculated UPL values from stack test data using t-statistics and statistical simulations. We used the most complete stack test data sets available to us; these were Hg stack tests from the top 127 units compiled from ICR data by RMB Consulting, and ICR Part ii and Part iii stack test data for filterable PM compiled by RMB Consulting. Note that only plant minimum ICR Part ii stack test data were available for filterable PM. The t-statistics UPLs were calculated using correct implementations of the UPL equation presented by EPA. The simulation UPLs were calculated as the 99th percentile of a large number (10^7) of averages of randomly selected stack test results. UPL values calculated using t-statistics and statistical simulations agreed to within a few percent.

We then examined Hg and PM Continuous Emission Monitoring System (CEMS) data from top performing units to determine historical long-term emission averages, and to assess whether temporal correlations and infrequent operating conditions significantly affect emission limits attainable by these units. One year of PM CEMS data were available for each of four units. These data comprise 30286 hours of filterable PM measurements; for comparison the EPA PM MACT floor calculation is based on 131 6-h measurements. Between 3 and 41 months of Hg CEMS were available for five units. These data comprise 57907 hours of Hg measurements; for comparison the EPA Hg MACT floor calculation is based on 40 6-h measurements.

Long-term emission limits attained by the top performing units were calculated from CEMS data as the 99th percentile of historical 30-day emission averages. In order to calculate t-statistics UPLs comparable with the CEMS data, we calculated "synthetic stack test" (SST) results for each unit. SST were calculated as the average of 6 consecutive hours of CEMS emission data when the unit was at or above a full load threshold. The SST data were then used to calculate t-statistics UPL values for each unit. Ratios of 99th percentile historical 30-day emission averages to SST-based t-statistics UPLs were 3.89 for the PM CEMS data which included startup and averaged 1.80 for Hg CEMS data. Ratios greater than 1.0 indicate that these units had

higher long term emission rates than one would expect based on stack tests alone. This ratio, R, was then applied to UPLs calculated from stack test data in order to correct UPLs to account for correlated emissions and emissions during infrequent events (see Table 1).

Table 1: HAP Upper Prediction Limits Calculated from Stack Test and CEMS Data

HAP	t-Statistics UPL (lb/MBtu)	Correction Factor, R	Corrected UPL, UPL * R (lb/MBtu)
Hg (127 Units)	0.816×10^{-6}	1.80	1.47×10^{-6}
Filterable PM	0.00441	3.89	0.0172

In summary, we have developed a method to estimate long-term emission limits based on the available stack test and CEMS data. Stack test data were used to calculate UPLs assuming independent and normally distributed emissions; UPLs were then scaled by a single empirical correction factor, R, in order to include the effects of correlated emissions and emissions during infrequent events observed in the CEMS data. The present analyses were limited by the available data. Specifically, no HCl CEMS data were available, Hg CEMS data were available for five units, and PM CEMS data including startup were available for one unit. In addition filterable PM stack test data included only plant minimum data from ICR Part ii data; it is likely that the t-statistics UPL for filterable PM would be greater if based on the complete ICR Part ii PM stack test data after quality assurance. Improved estimates of long term emission limits may be calculated using our method and more extensive data.

Introduction

EGU MACT Limits

The U.S. Environmental Protection Agency (EPA) recently has proposed emissions standards for hazardous air pollutants (HAPs) emitted by coal-fired electric generating units (EGUs) (76 FR 24976). The proposed rule includes emission limits based on maximum achievable control technology (MACT). These limits are for long-term averaged emissions, e.g. 30-day emission averages; however the "MACT Floor" calculations on which the emission limits are based use short-term stack test results. In order to extrapolate short term measurements to long term limits, EPA assumed that the emission data were independent and normally distributed then calculated Upper Prediction Limits (UPLs) using a t-statistics approach. EPA has made available on the rule docket website the data and equations used for the MACT Floor calculation in Microsoft Excel™ spreadsheets and Microsoft Access™ databases. In this report we enumerate the assumptions and critique EPA's implementation of the MACT Floor calculations for mercury (Hg), particulate matter (PM), and hydrochloric acid (HCl). We also examine large continuous emission systems (CEMS) data sets to independently evaluate whether the calculated MACT floors are achievable by top performing units.

Prior Work on Hg Emission Control

In collaboration with Southern Company we have compiled over 200 plant-months of operational and Hg emissions data from ten plants which control Hg using co-benefit or activated carbon injection (ACI) technologies (Allen, Looney, and Tyree, 2011). These data demonstrate that both ACI and "co-benefit" control technologies remove 80-90% of Hg. However, Hg emissions were neither independent nor normally distributed. Hg emissions from co-benefit controlled units had seasonal variations in emissions, with generally higher emissions during the third quarter associated with high summer load. This is consistent with the observation that Hg emissions were positively correlated with load likely due to a combination of reduced Hg oxidation in the SCR and greater Hg re-emissions from the FGD liquor (Tyree and Allen, 2010). Hg emissions from ACI-controlled units were affected by ESP performance, with decreased Hg emissions associated with "de-tuning" of the ESPs. Once the ESPs were de-tuned, the baghouses were more effective due to higher ash content of the flue gas. Higher than average Hg emissions were observed at one ACI-controlled unit coincident with changes in the control settings of the new equipment.

In this work we also used the t-statistics UPL equation to calculate 99th percentile emissions over 30-day averaging periods from short-term measurements. A large number of stack test results were estimated from continuous Hg emissions data for periods when operating conditions were like those used during stack tests, i.e. boilers at full load and air quality control systems operating effectively. Thirty-day average emissions were calculated. The actual 99th percentile 30-day average emission levels were compared with the t-statistics UPL results for the same unit. Seven of ten units had actual 30-day emission averages that were up to 76% percent greater than the t-statistics UPL results. These results suggest that emissions limits calculated based on the assumption of independent and normally distributed emissions significantly underestimate actual long-term Hg emissions from units with effective Hg controls.

Approach

We have analyzed Hg, and PM stack test and CEMS data collected for EPA's rulemaking in order to:

- Quantitatively evaluate the assumptions of independent and normally distributed emissions inherent in EPA's MACT floor calculations.
- Critically evaluate application of the t-statistics UPL equation used to by EPA to establish MACT floor levels.
- Calculate actual 99th percentile 30-day average emission levels of Hg and PM from top performing units using CEMS data.
- Propose scaling factors for the t-statistics UPL equation which account for correlated emissions observed in CEMS measurements.

Emission Limits Based on Stack Test Results

EPA has proposed 30-day rolling average emission limits for HAPs based short-term stack test results. In order to estimate long-term emission limits from short-term stack test results, EPA applied textbook statistical methods. Student's t-statistics were used to set long-term emission limits which matched stack test results of the "top performing" units with a 99% confidence interval. In this section we review this statistical approach, discuss the assumptions which underlie this approach, critique EPA's implementation of the statistical method, and present statistical simulations of 30-day rolling average emissions based stack test data. Here "top performing" units are those whose stack test results were used in EPA's MACT floor calculations.

Statistical Basis

Student's t-statistic may be used to calculate the likelihood that measurements from two normally-distributed populations have the same mean. The t-statistic in this case is (Casella and Berger, 2002):

$$t_{m+n-2} = \frac{\bar{y} - \bar{x}}{\sqrt{s^2\left(\frac{1}{m} + \frac{1}{n}\right)}} \tag{1}$$

Here t_{m+n-2} is the t-statistic with $m + n - 2$ degrees of freedom, $\bar{y}$ and $\bar{x}$ are the sample means from two populations; s is the sample standard deviation; and m and n are the number of measurements from the populations. The resulting t-statistic may then be compared with tabulated values for $m + n - 2$ degrees of freedom to determine the likelihood that the populations have the same mean. The standard deviations for the two populations are assumed to be equal. The sample standard deviation would be calculated as:

$$s^2 = \frac{\Sigma(y - \bar{y})^2 + \Sigma(x - \bar{x})^2}{m + n - 2} \tag{2}$$

An example application of Student's t-statistic would be to determine whether adult men from two cities have the same average height; $\bar{x}$ would be the average of n height measurements from one city and $\bar{y}$ would be the average of m height measurements from the second city. Equation 1 may be rearranged to:

$$\bar{y} = \bar{x} + t_{m+n-2}\sqrt{s^2\left(\frac{1}{m} + \frac{1}{n}\right)} \tag{3}$$

An Upper Prediction Limit (UPL) for $\bar{y}$ can be calculated for a specific confidence interval. The UPL represents the highest value of $\bar{y}$ which is consistent with two populations drawn from normal distributions with the same mean:

$$\text{UPL} = \bar{x} + t_{m+n-2,99}\sqrt{s^2\left(\frac{1}{m} + \frac{1}{n}\right)} \tag{4}$$

Here $t_{m+n-2,99}$ is the t-statistic for $m + n - 2$ degrees of freedom and 99% confidence interval for a one-sided t-test. The sample standard deviation may be estimated from first population as:

$$s^2 = \frac{(n-1)\sum(x-\bar{x})^2}{m+n-2} \tag{5}$$

The ratio $\sqrt{\frac{n-1}{m+n-2}}$ scales the sample standard deviation from the x population to the pooled variance of the x and y populations, which are assumed to have equal standard deviations.

For the example comparing adult heights from two cities, the UPL would be the highest average height from the second city that is consistent with equal average heights at the selected confidence interval. This could be calculated using only data from the first city.

EPA has presented Equation 4 as the basis for its MACT floor calculations. These calculations appear to have been designed so that the long-term emissions less than the limits would match stack test results of the top performing units with a 99% confidence interval. However, EPA has made a number of errors in the application and implementation of the UPL equation.

Normal Distribution Assumption

Equation 4 is based on the assumption that the data, here HAP emissions, are independent and normally distributed. In the MACT rule EPA invokes the Central Limit Theorem as a justification for assuming that emissions data are normally distributed (76 FR 25041):

> When the sample size is 15 or larger, one can assume based on the Central Limit theorem, that the sampling distribution of the average or sampling mean of emission data is approximately normal, regardless of the parent distribution of the data. This assumption justifies selecting the normal-distribution based UPL equation for calculating the floor.

This is a fundamental misstatement of the Central Limit Theorem. The theorem states that means of *independent* and *identical* random variables approach a normal distribution (Casella and Berger, 2002). It is incorrect to assume, based on sample size alone, that a data are normally distributed. In fact, we have shown that Hg emissions controlled using either co-benefit or activated carbon injection are correlated with load, and so are not independent and identical (Tyree and Allen, 2010; Allen, Looney, and Tyree, 2011). Thus EPA has omitted an essential step from their statistical analysis by neither evaluating the accuracy of the assumed distribution, nor the effect of a mismatch between the data and assumed distribution on the resulting emission limits.

The HAP stack test data used in EPA's MACT floor analyses were examined to determine whether they are, in fact, normally distributed. The stack test data were taken directly from EPA spreadsheets posted on the rule docket; floor_analysis_coal_hcl_031611.xlsx for HCl, floor_analysis_coal_hg_051811.xlsx for Hg, and floor_analysis_coal_pm_031611.xlsx for PM (see Table 2). Additional stack test data for Hg were prepared by RMB Consulting using ICR Part ii and Part iii data from the EPA spreadsheet partii_iii_hg.xls. Additional stack test data for filterable PM were prepared by RMB Consulting using ICR Part ii.

Table 2: HAP Stack Test Data Inventory

Data Set Name	HAP	Spreadsheet Name	Number of Units	Number of Stack Tests
HCl	HCl	floor_analysis_coal_hcl_031611.xlsx	131	172

Hg40	Hg	floor_analysis_coal_hg_051811.xlsx	40	80
Hg127	Hg	partii_iii_hg.xls	127	265
PM-T	Total PM	floor_analysis_coal_pm_031611.xlsx	131	131
PM-F	Filterable PM	Filterable PM data Jon.xlsm	131	221

The goodness of fit between data and a parametric distribution may be evaluated graphically. In EPA's MACT floor calculations, emission means were determined from the minimum stack tests for each top performing unit while the emission variances were determined from all stack test results for the top performing units. Plant minimum and all stack tests results from each data set were compared graphically with the fitted normal distributions (see Figures 1-7). The normal distribution curves were scaled to have the same area as the bar graph, and not all of the normal distributions are shown on the plots. Normal distributions were poor matches for every collection of stack test results.

A number of approaches may be used to evaluate quantitatively whether data fit a normal distribution. Jarque and Bera (1987) presented an efficient test of normality. The Jarque and Bera statistic (JB) is calculated as

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{5}$$

Here S is skewness and K is kurtosis. Skewness is a measure of the asymmetry of the data around the sample mean. The skewness of the normal distribution is zero.

$$S = \frac{\frac{1}{n}\Sigma(x_i - \bar{x})^3}{\sqrt{\left(\frac{1}{n}\Sigma(x_i - \bar{x})^2\right)^3}} \tag{5}$$

Kurtosis is a measure of the fraction of outliers in a distribution. The kurtosis of the normal distribution is 3.

$$K = \frac{\frac{1}{n}\Sigma(x_i - \bar{x})^4}{\left(\frac{1}{n}\Sigma(x_i - \bar{x})^2\right)^2} \tag{5}$$

Small values of JB are consistent with a normal distribution. The probabilities that data were normally distributed were calculated using Monte-Carlo simulation. JB and the associated probabilities were calculated using the jbstat function in the Matlab Statistics Toolbox version 7.4 (www.mathworks.com).

Using Jarque and Bera statistics the goodness of fit between data and a parametric distribution may be evaluated quantitatively (see Table 3). The probabilities that data sets which included all the stack tests were normally distributed were all negligible, less than 0.1%. The probabilities that data sets which included only plant minimum stack tests were normally distributed were all less than 12%. However, these data sets appear to match uniform distributions more closely than normal distributions (see Figures 2, 4, 6, and 7). With the exception of the Hg40 plant minimum data set, one can reject the hypothesis that the data are normally distributed at the 95% confidence level.

Table 3: Quantitative Tests of Normality for Stack Test Data Sets

Data Set	Data	Jarque Bera	Probability of Normal

		statistic	Distribution
HCl	All	11455.1	< 0.1 %
HCl	Plant minimum	9.3	1.8 %
Hg40	All	1291.1	< 0.1 %
Hg40	Plant minimum	2.7	11.4 %
Hg127	All	4503.9	< 0.1 %
Hg127	Plant minimum	18.1	0.4 %
PM-T	Plant minimum	11.6	1.1 %
PM-F	All	1904	< 0.1 %
PM-F	Plant minimum	9.2	1.9 %

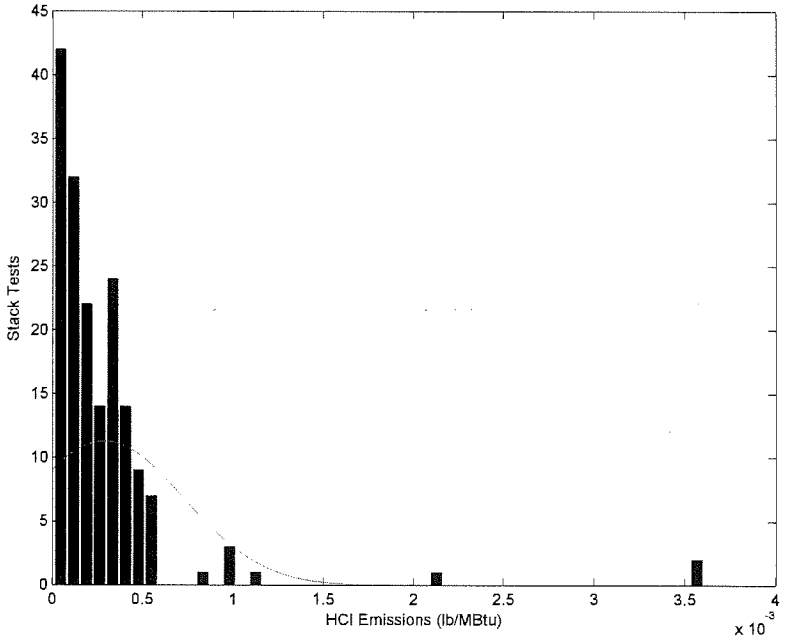


Figure 1: Distribution of all HCl stack test results.

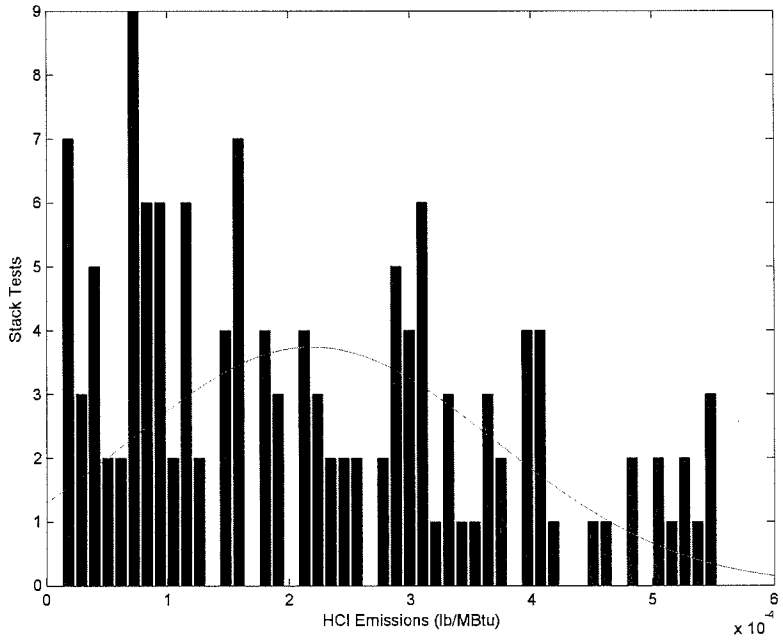


Figure 2: Distribution of plant minimum HCl stack test results.

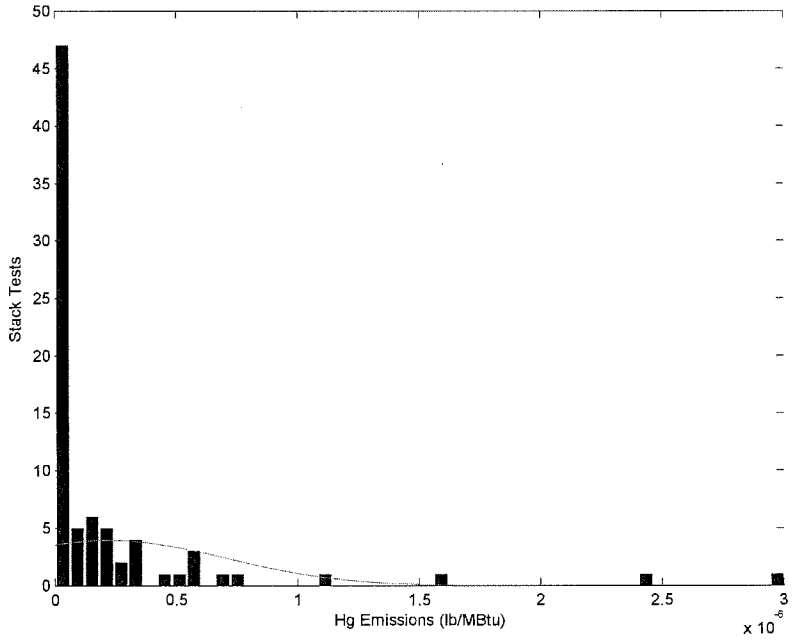


Figure 3: Distribution of all Hg stack test results for top 40 units.

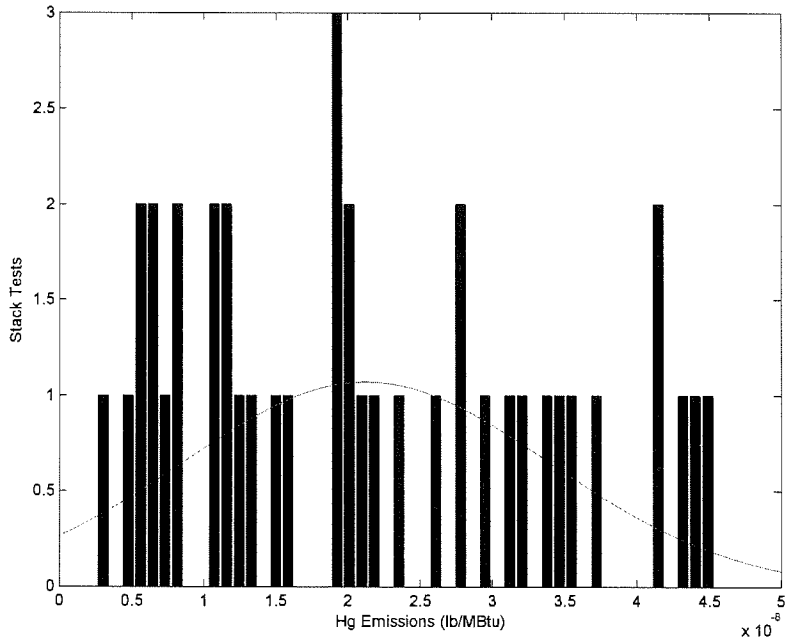


Figure 4: Distribution of plant minimum Hg stack test results for top 40 units.

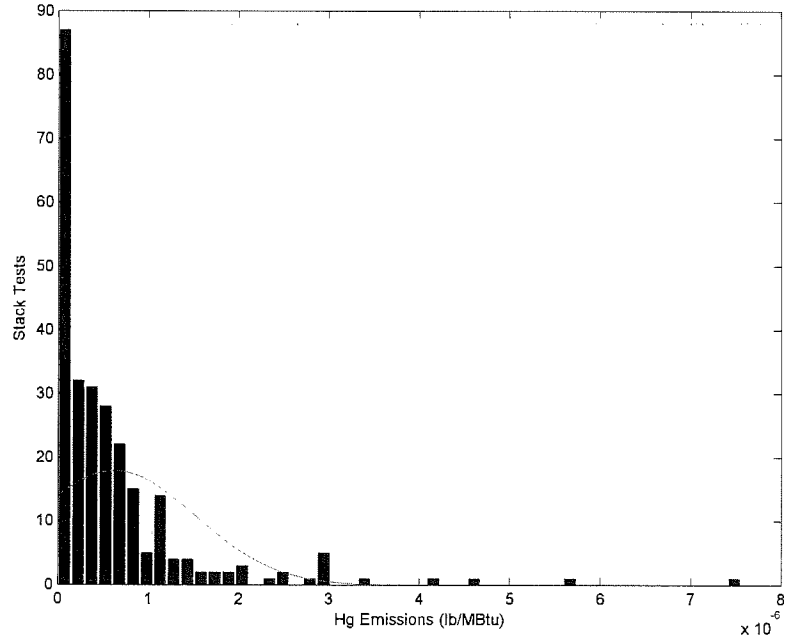


Figure 5: Distribution of all Hg stack test results for top 127 units.

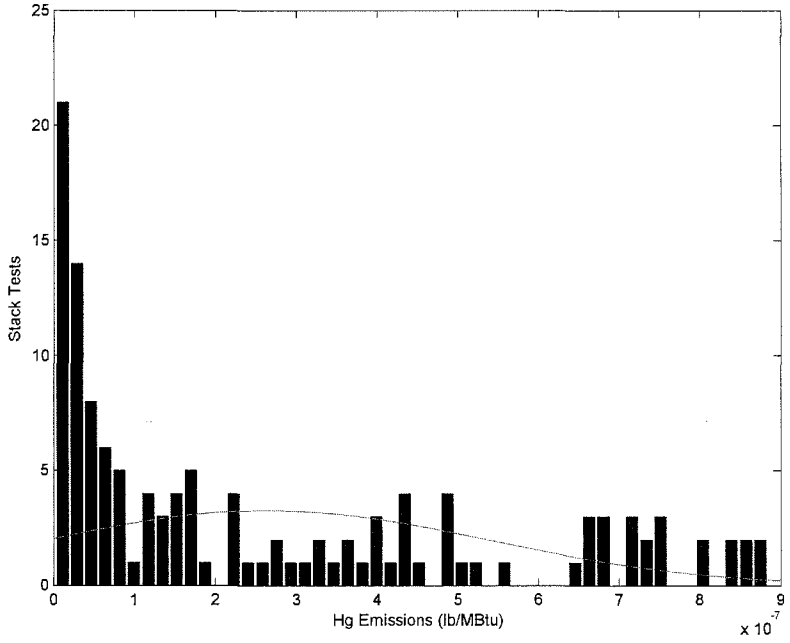


Figure 6: Distribution of plant minimum Hg stack test results for top 127 units.

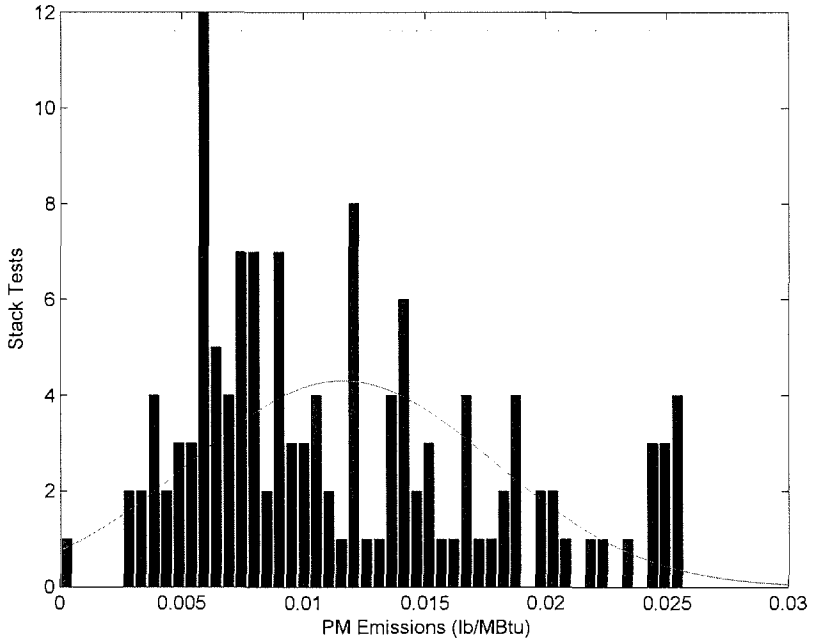


Figure 7: Distribution of plant minimum total PM stack test results.

UPL Equation Implementation

EPA has made a number of errors in the implementation of the UPL equation; specifically, EPA used values for the equation variables which are not consistent with textbook statistics. Each variable is discussed separately below. These comments are based on analyses of EPA spreadsheets posted on the rule docket: floor_analysis_coal_hcl_031611.xlsx for HCl, floor_analysis_coal_hg_051811.xlsx for Hg, and floor_analysis_coal_pm_031611.xlsx for PM.

EPA inconsistently calculated the mean and sample standard deviation from the top performing unit stack tests in Equation 4. The $\bar{x}$ value was calculated as the mean of the plant *minimum* results for the top performing units while s^2 was calculated as the sample variance for *all* the results. The use of different sample populations to calculate the mean and variance is arbitrary and not justified by the statistical approach. Further, the use of minimum emission tests for each unit is not an accurate characterization of that unit's performance. The MACT floor values are therefore based on *minimum emission tests*, not best performing units. This error significantly reduced the calculated UPL values for HCl and Hg. Note that only plant minimum stack tests were included in the PM MACT floor calculation spreadsheet.

EPA's calculations use $n-1$ as the number of degrees of freedom. The correct value is $m+n-2$ which also includes the m compliance measurements used to calculated $\bar{y}$. Applying this correction changes the results slightly; e.g., for $n = 40$ and $m = 120$, t should be 2.35, not 2.43.

EPA incorrectly calculated the sample variance used in Equation 4. Sample variances were calculated as if the data population included only the stack tests used in the MACT floor calculations; however the populations includes both these stack tests and the compliance samples. As explained above, the correct sample variance is that of the pooled stack tests and compliance samples (see Equation 5). The pooled variance will be smaller than calculated by EPA. The variability portion of the UPL equation, $t_{m+n-2}\sqrt{s^2\left(\frac{1}{m}+\frac{1}{n}\right)}$, will be less than that calculated by EPA by a factor of $\sqrt{\frac{n-1}{m+n-2}}$. For the case n is approximately equal to m and both are much greater than one, this factor is approximately $\frac{1}{\sqrt{2}} = 0.707$.

EPA used $m = 1$ in the UPL calculations. This would be correct if the compliance test were a single stack test. However, the proposed HAP emission limits are 30-day averages. The value of m should be approximately 120; this is the number of hours in 30 days (720) divided by the number of hours in a stack test (6). The value of m would be smaller than 120 if a unit did not operate for 720 hours in 30-day period. This change reduces the UPL from what was calculated by EPA for Hg emissions by a up to a factor of 5. The proposed regulatory limits for HCl and PM are a combination of 30-day rolling averages for CEMS data and individual stack tests. In these cases, two separate UPLs should have been calculated, one with $m = 1$ and one with $m = 120$.

Corrected and Simulation UPL Calculations

UPL values can be calculated using correct inputs to Equation 4 and stack test data from the spreadsheets. The $\bar{x}$ and s^2 values were calculated using all the stack test data for the top performing units. The degrees of freedom for the t-statistic were $m+n-2$. Pooled sample

variances were calculated using Equation 5. The value of m was 120. These results are referred to as "t-Statistics UPL" (see Table 4). Note that these results correct implementation issues identified above, but continue to be based on the assumption that HAP emissions are independent and normally distributed.

UPL values may be calculated from the stack test data without assuming normally distributed data using statistical simulations. The simulation approach is an example of non-parametric statistics because it is not based on parameters determined for an assumed distribution, e.g. a normal distribution. In this approach a large number (10^7) of 30-day emission averages were calculated by averaging 120 randomly selected stack test results. The 99th percentile emission average may be used to estimate the UPL ("Simulation UPL" in Table 4). The simulation program was tested by comparing simulation results with theoretical results for model distributions. The simulation UPL for input uniform and normally-distributed random variables were compared with normal and t-statistic distributions. Simulation results matched the theoretical values to at least 3 significant digits. Figures 8-11 show distributions of the emission averages as blue bars and the fitted normal distributions as green lines. The simulation and t-statistics UPLs are shown as solid and dashed red lines, respectively.

The t-statistic and simulation UPLs agree to within a few percent. The main difference between these two approaches is that the t-statistic approach assumes that the data are normally distributed and the simulation approach does not. These results suggest that for the present data and large m, the assumption of normally distributed data does not have a large effect on the calculated emission limits. This is likely because sums of large numbers of independent data are normally distributed (Central Limit Theorem) and the t-statistic distribution is very similar to the normal distribution.

As discussed above the value of m is only known approximately; m is the number of operating hours in 30 days of operation (720 or less) divided by the number of hours in a stack test (6). Thirty-day emission averages were calculated by simulation for a range of m values using the Hg127 data set (see Figure 12). The simulation UPLs decrease as m increases (see Table 5). The simulation UPLs are always greater than the t-statistics UPLs, although the difference decreases as m increases. Because only days with some hours of operation are included in the 30-day average, a minimum value for m is 30; this represents a unit operated for only 6 hours a day for 30 days.

Table 4: Upper Prediction Limits Calculated Using t-Statistics and Simulation

Data Set Name	HAP	Number of Units	Number of Stack Tests	t-Statistic UPL (lb/MBtu)	Simulation UPL (lb/MBtu)
HCl	HCl	131	172	3.83×10^{-4}	3.97×10^{-4}
Hg40	Hg	40	80	0.323×10^{-6}	0.332×10^{-6}
Hg127	Hg	127	265	0.816×10^{-6}	0.834×10^{-6}
PM-T	Total PM	131	131	0.0129	0.0130
PM-F	Filterable PM	131	221	0.00441	0.00449

Table 5: Simulation and t-statistics UPL for a range of m values.

Number of Periods in Compliance Average (m)	t-Statistic UPL (lb/MBtu)	Simulation UPL (lb/MBtu)
30	1.01×10^{-6}	1.08×10^{-6}
60	0.895×10^{-6}	0.932×10^{-6}
90	0.845×10^{-6}	0.870×10^{-6}
120	0.816×10^{-6}	0.834×10^{-6}

Both the t-statistics and simulation UPL calculations are based on the assumption that the data are independent, i.e. emissions are not correlated in time. However, we have shown that Hg emissions using either co-benefit or ACI control are indeed correlated in time (Tyree and Allen 2010; Allen, Looney, and Tyree 2011). One might expect that HCl and PM emissions will be similarly correlated with process events, for example startup and extended periods of high load. Stack tests, which measure a unit's emission at optimal conditions on a single day, include no information about correlated emissions or emissions during infrequent events. In the next sections we examine CEMS data to determine historical long-term emission averages, and to assess whether temporal correlations and infrequent operating conditions significantly affect emission limits attainable by top performing units.

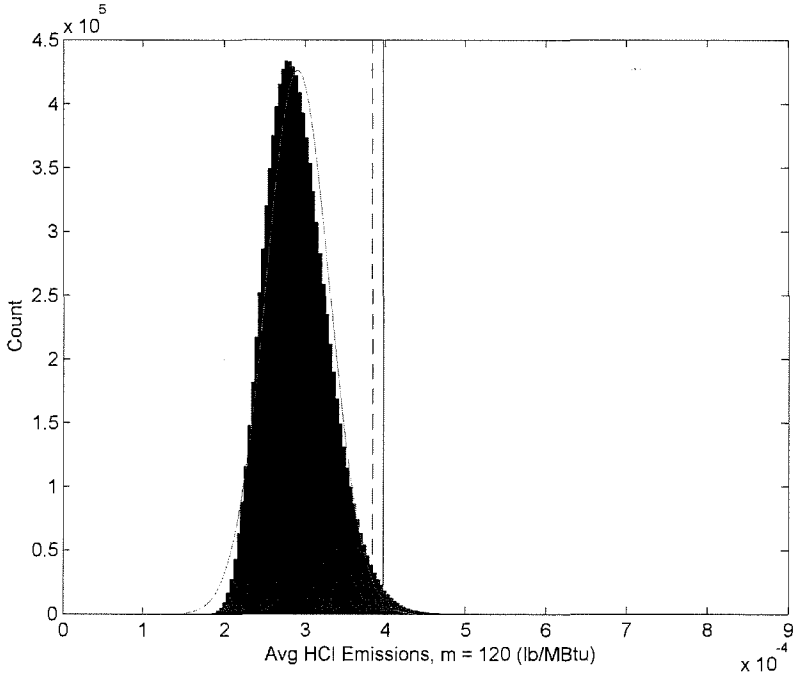


Figure 8: Distribution of simulated 30-day HCl emission averages using stack test data.

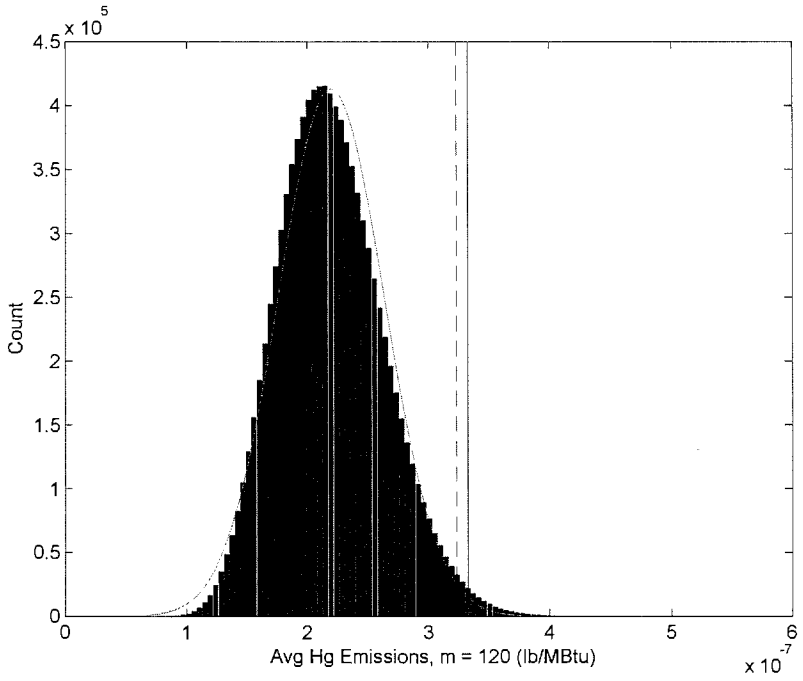


Figure 9: Distribution of simulated 30-day Hg emission averages using Hg40 stack test data.

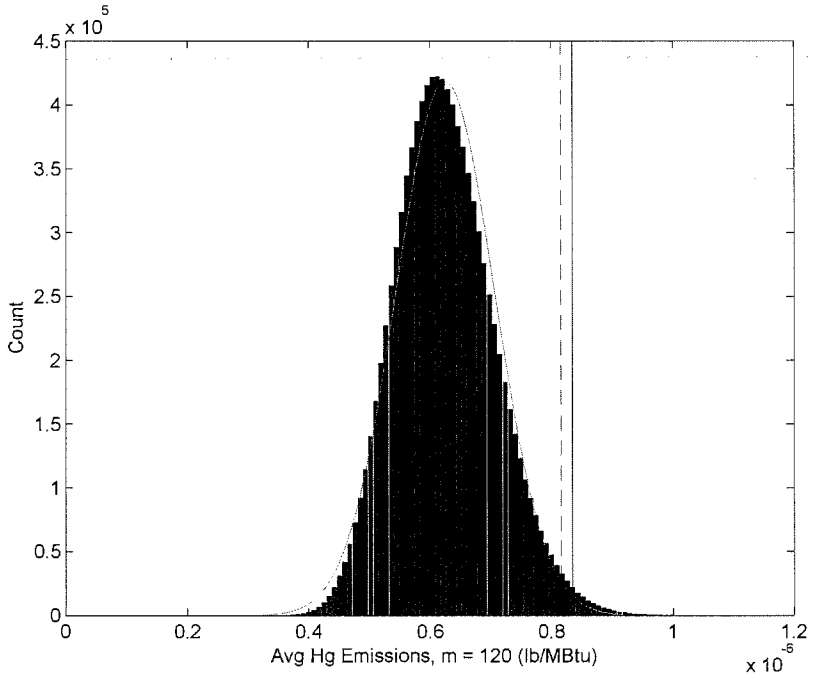


Figure 10: Distribution of simulated 30-day Hg emission averages using Hg127 stack test data.

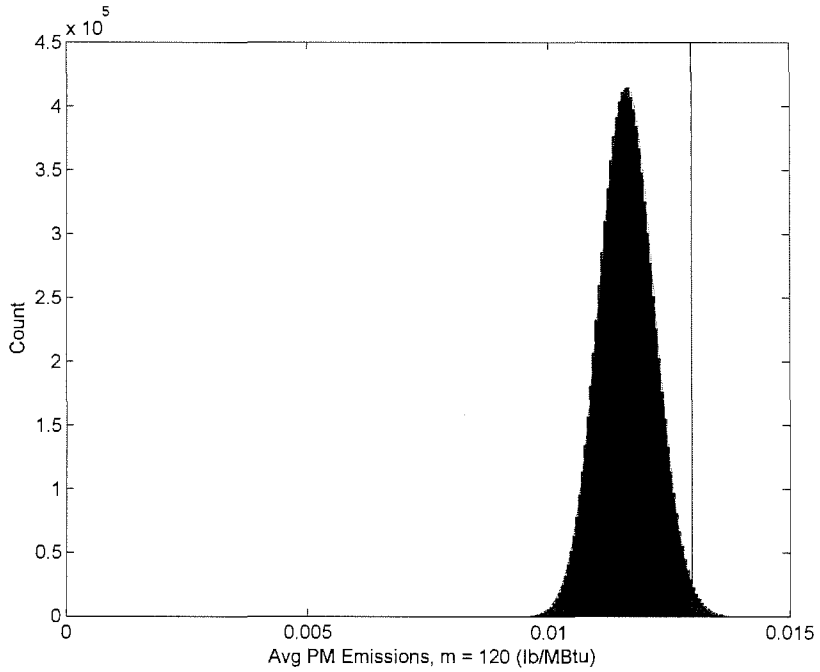


Figure 11: Distribution of simulated 30-day total PM emission averages using stack test data.

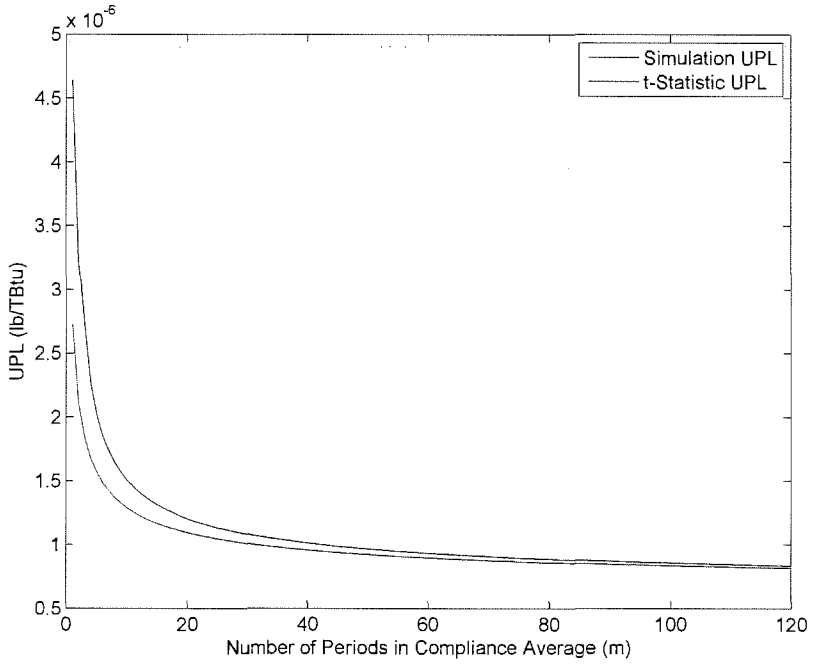


Figure 12: Simulation and t-statistics UPL as a function of m for Hg127 stack test data.

Particulate Matter CEMS Data

As discussed above stack tests are limited to short term measurements of HAP emissions at optimal conditions on a single day. PM CEMS data were available for a few top performing units listed on EPA's MACT floor calculation spreadsheets. These data were used to determine historical long-term emission averages while making no assumptions regarding the distribution of emissions data. Thus long-term data were used to evaluate long-term emission standards using measurements like those which would likely be used to determine compliance. These historical long-term emission averages may in turn be compared to those predicted from stack tests in order to assess the accuracy of calculated UPLs.

Data Source and Quality Assurance

Continuous PM emission data were provided by RMB Consulting for 7 units (see Table 6). These data cover 8 plant-years. Top performing units are those included in EPA's PM MACT floor spreadsheet. Hourly averaged PM emission and supporting data were imported directly from the original Excel spreadsheets into Matlab for analysis. The supporting data for each plant included the gross megawatt generation and PM CEMS quality assurance flags. Heat input data were also available for Cross Unit 1.

PM emission data were replaced with NaN (not-a-number) values and excluded from the analysis if the data were flagged as "invalid", "calibration", "maintenance", or "out of control". Data flagged as "suspect" or "exceedence" were included if they were not otherwise flagged. PM emissions from OCPP Unit 8 at 1100 on 17 April 2009 was also deemed to be invalid; the high PM concentration reported for this hour was comparable to those reported for other hours in the same day which had been flagged as invalid.

Table 6: PM CEMS Data Inventory

Plant Name	Top Performing Unit	Data Period		Number of Avg. Periods		
		First Day	Last Day	Hours	Days	30-Day
Clover Unit 1	X	1 Oct 2009	30 Sep 2010	7455	365	288
Clover Unit 2		1 Oct 2009	30 Sep 2010	8306	365	323
Cross Unit 1	X	1 Nov 2009	31 Oct 2010	6894	365	260
OCPP Unit 7		1 Jan 2009	30 Jun 2010	7735	546	288
OCPP Unit 8		1 Jan 2009	30 Jun 2010	11177	546	432
Spurlock Unit 1	X	31 Dec 2009	30 Dec 2010	7968	365	308
Spurlock Unit 4	X	31 Dec 2009	30 Dec 2010	7969	365	307

These data comprise a total of 30286 hours of valid PM measurements during the operation of top performing units. For comparison the EPA PM MACT floor calculation is based on 131 6-h measurements (786 hours of data). PM CEMS data were not included in EPA's MACT floor calculations. PM CEMS instruments generally optical measurements which are scaled to match

filterable PM, thus the PM CEMS data are not directly comparable to the MACT floor total PM stack tests which represent filterable plus condensable PM.

PM Emission Averages

Daily emission averages were calculated as the mean of hourly PM emissions including only operating hours. Thirty-day average emissions were calculated as the mean of hourly PM emissions during operating hours during 30 days which included only days with at least one operating hour. Operating hours were those for which heat input was positive, or if heat input data were not available, gross megawatt generation was positive.

PM emission averages are shown for the four top performing units (see Figures 13-16). Hourly emission averages are shown as small green dots. Daily emission averages are shown as blue dots. Thirty-day emission averages are shown as red dots plotted at the end of each period. The date labels mark the start of each period; e.g. 'Jul10' marks the start of 1 July 2010. The y-axis scales have been selected to include the maximum daily emission average; hourly emissions greater than this value may not be displayed. Effects of digitization can be seen in the Spurlock Unit 1 and 4 data; these the measurements were reported with a precision of 0.001 lb/MBtu.

The PM emission averages demonstrate that top performing units' PM emissions were correlated in time and strongly affected by startup events. Periods of correlated relatively high emissions were apparent for Clover Unit 1 in December 2010, Spurlock Unit 1 in September 2010, and Spurlock Unit 4 in March-April 2010 and October-November 2010. Startup events also significantly affected 30-day emission averages for Cross Unit 1 in April 2010, Spurlock Unit 1 in November 2010, and Spurlock Unit 4 in June 2010.

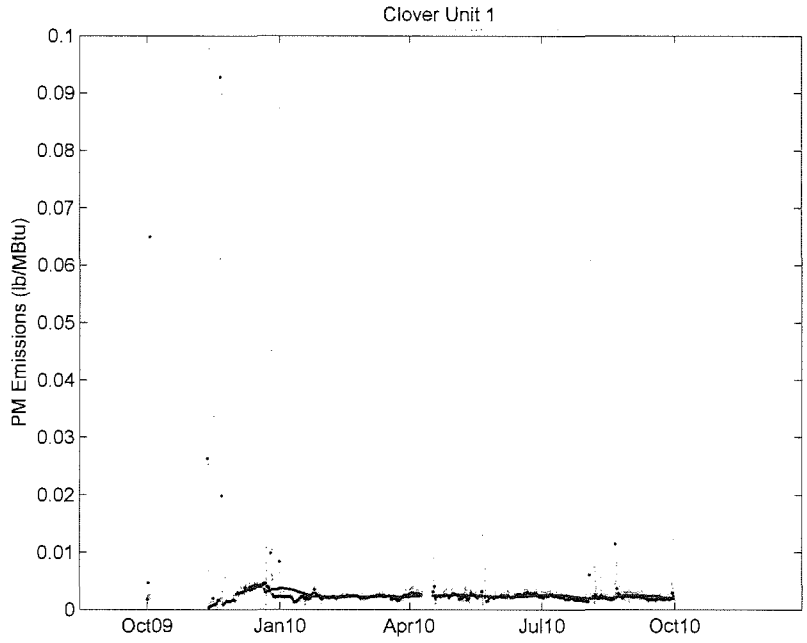


Figure 13: Clover Unit 1 PM emission averages.

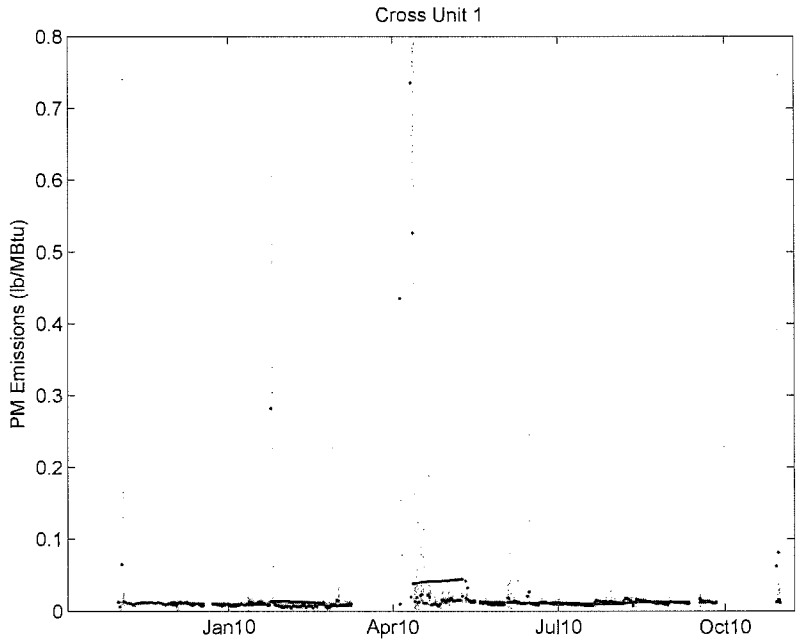


Figure 14: Cross Unit 1 PM Emission averages

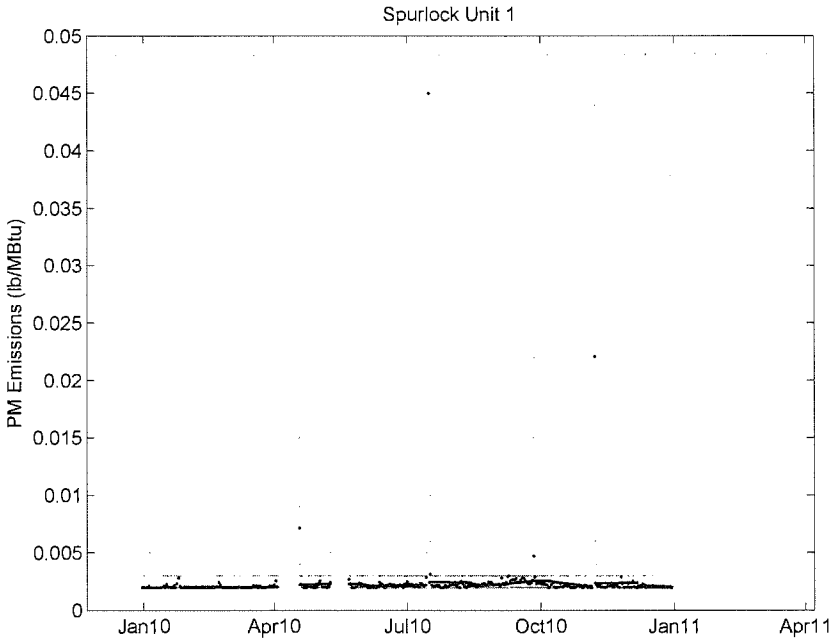


Figure 15: Spurlock Unit 1 PM emission averages.

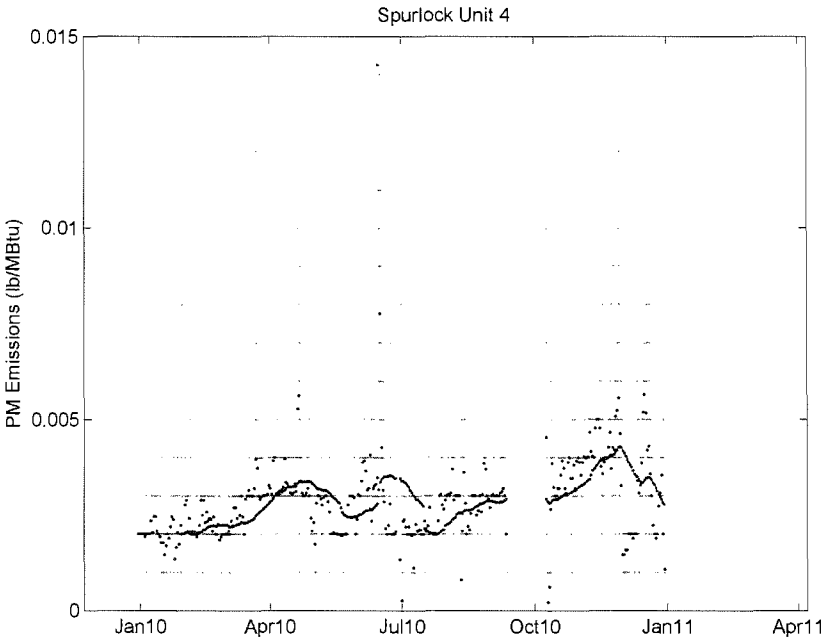


Figure 16: Spurlock Unit 4 PM emission averages.

Proposed long-term emission limits may be compared with historical 30-day emission averages (see Figures 17-20). The top panel in these figures shows the historical 30-day emission averages with a red line marking the 99th percentile. The distributions of 30-day PM emission averages for top performing units generally include right-hand tails attributable to PM emissions correlated in time and affected by startup events.

In addition to historical 30-day emission averages we also calculated the 30-day averages using randomly selected days. A large number (10^6) of averages were calculated for each unit (see Figures 17-20). The middle panel in these figures shows the random 30-day emission averages with a red line marking the 99th percentile. These random averages remove the effect of multi-day correlations of emissions including multi-day startup events on the long-term averages.

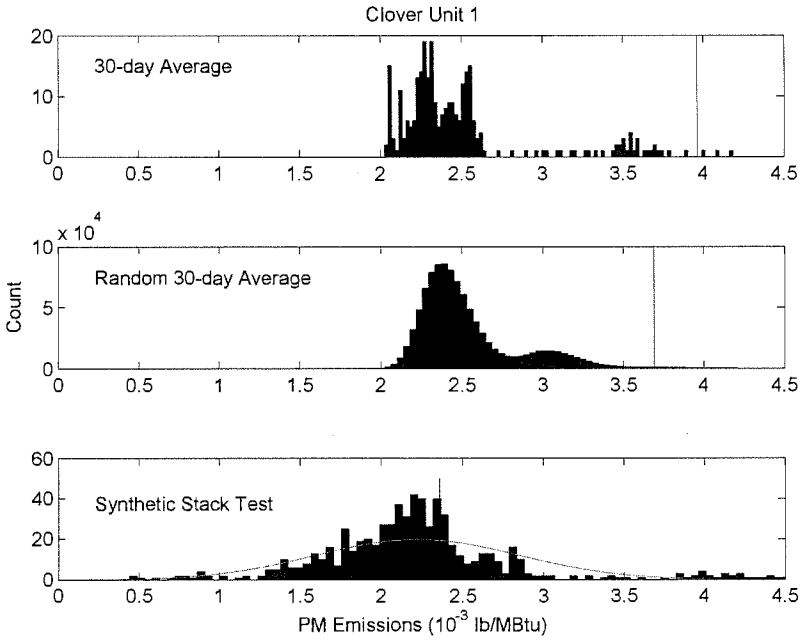


Figure 17: Clover Unit 1 distribution of PM emissions.

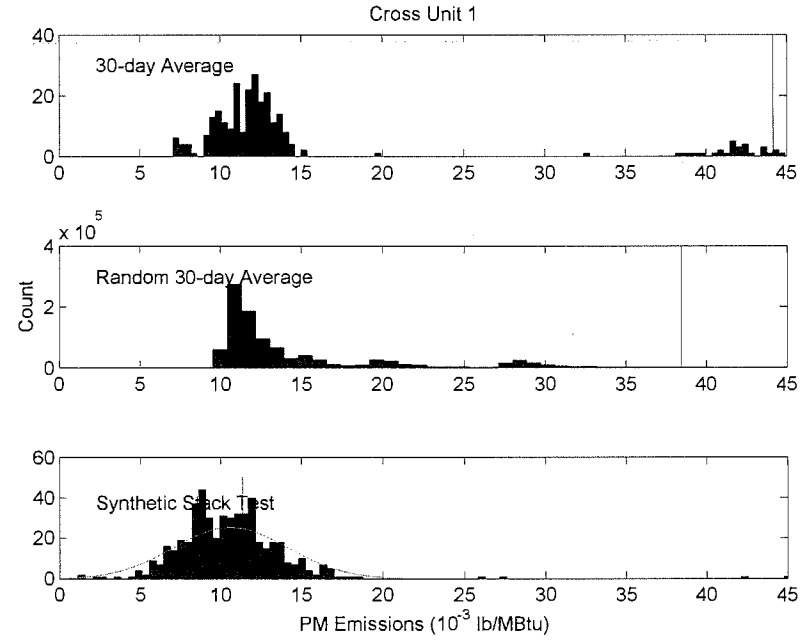


Figure 18: Cross Unit 1 distribution of PM emissions.

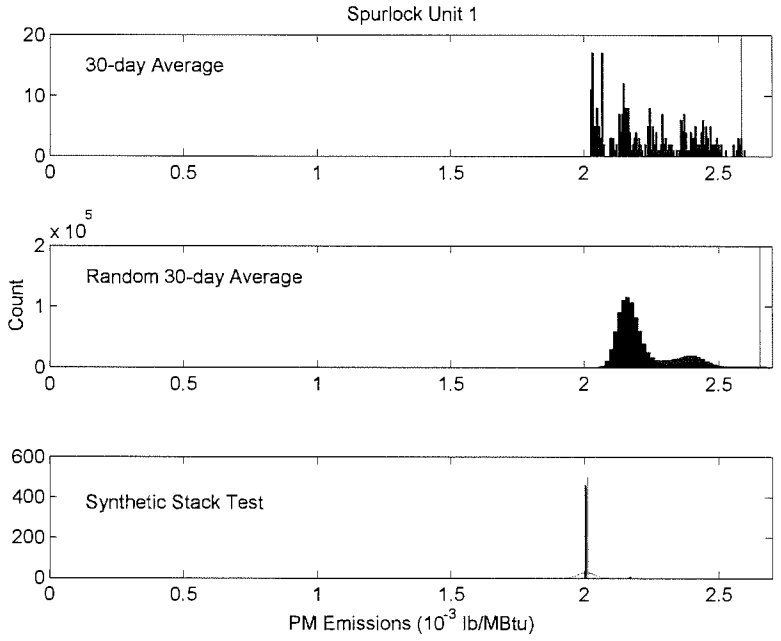


Figure 19: Spurlock Unit 1 distribution of PM emissions.

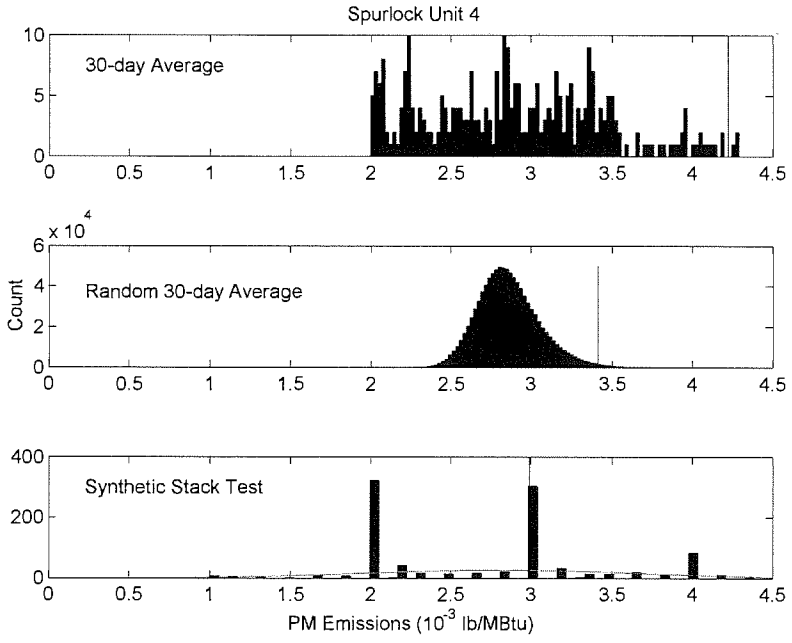


Figure 20: Spurlock Unit 4 distribution of PM emissions.

Synthetic Stack Test Emissions

In order to relate long-term emission averages to emission limits calculated from stack tests, we calculated "synthetic stack test" (SST) results from PM CEMS data. SST were calculated as the average of 6 consecutive hours of PM emissions when the unit was at or above a full load threshold set to approximately 90% of capacity (see Table 7). This procedure was designed to obtain data like that from a stack test using CEMS data, and is similar to procedures reported in Tyree and Allen (2010) and Allen, Looney, and Tyree (2011).

The SST data for each unit were then used to calculate t-statistics UPL values using Equation 4. The distributions of SST values are shown as blue bars in the bottom panels of Figures 17-20; fitted normal distributions are shown as green curves and the t-statistics UPL values as red lines. The SST t-statistics UPL values are emissions limits that would have been calculated had hundreds of consecutive stack tests been collected over 12 months at specific top performing units.

Table 7: Synthetic Stack Tests Calculated from PM CEMS Data.

Plant Name	Top Perform-ing Unit	Full Load Threshold (MW)	SST Statistics (lb/MBtu)			
			N	Mean	Std. Dev.	Range
Clover Unit 1	X	425	667	0.00222	0.00063	0.00044-0.00513
Clover Unit 2		425	751	0.00152	0.00029	0.00085-0.00237
Cross Unit 1	X	580	511	0.01060	0.00356	0.00115-0.04516
OCPP Unit 7		275	397	0.00395	0.00126	0.00114-0.01451
OCPP Unit 8		275	839	0.00849	0.00326	0.00293-0.02509
Spurlock Unit 1	X	300	468	0.00200	0.00003	0.00200-0.00250
Spurlock Unit 4	X	275	1039	0.00280	0.00088	0.00100-0.00683

Corrections to PM t-Statistics UPL

As mentioned above emission limits based on stack tests include no information about correlated emissions or emissions during infrequent events. Here we compare actual emissions averages with t-statistics UPLs calculated using SST for top performing units in order to assess the importance of these omissions and to propose empirical corrections to emission limits based on stack test results.

If PM emissions were independent and normally distributed, the SST-based UPL values should be close to the 99th percentile 30-day emission averages. In fact the ratios of 99th percentile historical 30-day emission averages to SST-based t-statistics UPLs are in the range 1.29-3.89 for the four top performing units studied here (see Table 8). We designate this ratio R. The extreme case is Cross Unit 1 which had high 30-day average emissions after a startup in April 2010. For other the three top performing units, R is in the range 1.29-1.69. Note that the SST values did not include startup periods because load was less than the full load threshold, and that startup emission are included in 30-day PM emission averages in the proposed rule.

High emissions from Cross Unit 1 during 11-12 April 2010 occurred during a startup period. Hourly heat inputs were positive while generated megawatts (GMWatt) were zero during this period; indicating that the coal was being burned in the unit, but no electricity generated. A

second set of UPLs were calculate for Cross Unit 1 including only those hours for which GMWatt was positive (see Table 8); for this data set R = 1.27. Note that Cross Unit 1 is the only unit for which we have heat input and PM CEMS data. Based on this limited data set, the likelihood that a top performing unit will have a high-PM startup event like the Cross Unit 1 April 2010 event is approximately once per year.

Further evidence of the importance of correlated emissions can be seen by comparing averages of 30 random days with historical 30-day averages. The random averages remove the effect of multi-day emission correlations on the long-term averages. The ratios of the 99th percentile of the historical to random 30-day emission averages were in the range 0.98-1.24 for the four top performing units studied here.

In order for emission limits to match the emissions achievable by a top performing unit, emission limits calculated as t-statistics UPLs should be multiplied by R. This correction is to account for correlated emissions and emissions during infrequent events. This approach is based on the assumption that CEMS PM measurements will scale to the PM emission metric. This correction is based on all of the top performing PM CEMS data available to us, which represents 4 of 131 the top performing units for PM emissions and 38 times more hours of emission data than were used in EPA's MACT floor calculations.

Table 8: Comparison of historical 30-day average PM emissions with t-statistic SST UPLs.

Plant Name	Top Perform-ing Unit	Historical 30-day Emission Avg.	Random 30-day Emission Avg.	t-Statistics UPL Based on SST	R
Clover Unit 1	X	0.00396	0.00369	0.00235	1.68
Clover Unit 2		0.00215	0.00214	0.00158	1.36
Cross Unit 1	X	0.0441	0.0384	0.0113	3.89
Cross Unit 1 (GMWatt > 0)	X	0.0144	0.0126	0.0113	1.27
OCPP Unit 7		0.00682	0.00494	0.00422	1.61
OCPP Unit 8		0.0104	0.00748	0.00918	1.13
Spurlock Unit 1	X	0.00259	0.00265	0.00201	1.29
Spurlock Unit 4	X	0.00422	0.00341	0.00299	1.41

Hg CEMS Data

As discussed above stack tests are limited to short term measurements of HAP emissions at optimal conditions on a single day. Long term Hg CEMS data were available for a few top performing units listed on EPA's MACT floor calculation spreadsheets. Hg CEMS instruments determine the concentration of Hg in the flue gas and so are directly comparable to the MACT floor Hg stack tests. Hg CEMS data were used to determine historical long-term emission averages while making no assumptions regarding the distribution of emissions data. This approach used long-term data to evaluate long-term emission standards using measurements like those which would likely be used to determine compliance. These historical long-term emission averages may in turn be compared to those predicted from stack tests in order to assess the accuracy of calculated UPLs.

Data Source and Quality Assurance

Hg CEMS data were extracted from EPA Access files which contained ICR Part ii and Part iii data (see Table 9). The Access files were retrieved from the docket website; these files were eu_icr_parti_partii.mdb and eu_partiii.mdb for Part ii and iii data, respectively. ICR Part ii and Part iii data included both hourly and daily averaged emissions. We designate these data sources P2H, P2D, P3H and P3D for Part ii hourly, Part ii daily, Part iii hourly, and Part iii daily data, respectively. Each data source included different fields, and so each data source was processed separately.

Table 9: Inventory of ICR Hg CEMS Data

Name	Source	Facilities	Data Points	Plant-Months
Part ii hourly Hg CEMS (P2H)	File: eu_icr_parti_partii.mdb Table: Hg_cem_hourly	112	747870	1039
Part ii daily Hg CEMS (P2D)	File: eu_icr_parti_partii.mdb Table: Hg_cem_daily	70	16265	542
Part iii hourly Hg CEMS (P3H)	File: eu_partiii.mdb Table: CEMS_Data_Hourly CEMS_Type = "Hg"	12	8640	12
Part iii daily Hg CEMS (P3D)	File: eu_partiii.mdb Table: CEMS_Data_Daily CEMS_Type = "Hg"	19	574	19

We have focused our analyses on the P2H data for top performing units (see Table 10). P2D data were not used because these consist mainly of a subset of the units and sampling periods in the P2H data set. Further, 30-day emission averages were calculated as averages of hourly data over 30 days, not averages of 30 daily averages. P3H and P3D data include only approximately 30 days of data for each unit, and so are not useful in assessing distribution of 30-day averaged emissions. Note that the Hammond Hg CEMS data relate to four top performing units listed on EPA's MACT floor spreadsheet.

The Hg CEMS and facility information were exported from the Hg_cem_hourly and facility_information tables in the Access file eu_icr_parti_partii.mdb. These were exported to comma separated value (CSV) text files which were imported into Matlab. The Hg_cem_hourly

table included total Hg emission factors (lb/MBtu) along with heat rate, load and "Operational Status". The operational status field included quality assurance information for some facilities.

Table 10: Inventory of ICR Part ii Hourly Hg CEMS Data for Top Performing Units

Facility - Stack Name	Data Period		Number of Avg. Periods		
	First Day	Last Day	Hours	Days	30-Day
Colbert - SK001	1 Nov 2009	6 Feb 2010	2002	94	65
Cross - C1	9 Aug 2006	31 Dec 2009	29627	1241	1128
Hammond - Scrubber Stack 1, 2, 3 & 4	1 Jan 2009	31 Dec 2009	8520	365	330
San Juan - Unit 4 Stack	1 May 2008	30 Apr 2009	4507	202	173
TS Power Plant - STK1	14 Apr 2009	31 Dec 2009	13251	627	444

Quality assurance information for the Hg CEMS data were supplied by the facilities and so differ among facilities. Invalid Hg emission data were determined separately for each facility. Invalid data were replaced with NaN (not-a-number) values and excluded from subsequent analyses. Colbert and Hammond data were used as retrieved. The Cross Hg emission data were excluded if the operational status field included "Emission Factor Invalid" or "Unit Offline". Cross Hg emission data from 22:00 and 23:00 on 14 March 2008 were greater than 40 lb/TBtu; these were immediately before a shut down and were deemed invalid. Hg emission data for San Juan Units 1-4 were available in the ICR Part ii data. San Juan Hg data were included only when Hg monitors were known to have had liquid nitrogen available (Robeson, 2011); therefore the San Juan data was limited to Unit 4 from 1 May 2008 through 30 April 2009. In addition, San Juan Hg emission data were excluded if the heat rate was less than 10 MBtu/h. The TS Power Plant Hg emission data were excluded if the operational status field included "CEMS Failed Calibration", "CEMS Maintenance", "CEMS Malfunction", or "DAHS Malfunction".

The Hg CEMS data comprise 57907 hours of measurements from top performing units. For comparison the EPA Hg MACT floor calculation is based on 40 6-h measurements (240 hours of data). EPA did include Hg CEMS data for some units, but only as a single average value which was weighted equally with single stack test results.

Hg Emission Averages

Daily emission averages were calculated as the mean of hourly Hg emissions including only operating hours. Thirty-day average emissions were calculated as the mean of hourly Hg emissions during operating hours during 30 days which included only days with at least one operating hour. Operating hours were those for which heat rate was positive.

Hg emission averages are shown for the five data series (see Figures 21-25). Hourly emission averages are shown as small green dots. Daily emission averages are shown as blue dots. Thirty-day emission averages are shown as red dots plotted at the end of each period. The date labels mark the start of each period; e.g. 'Jan09' marks the start of 1 January 2009. The y-axis scales have been selected to include the maximum daily emission average; hourly emissions greater than this value may not be displayed.

The Hg emission averages demonstrate that top performing unit's Hg emissions are correlated in time. Periods of correlated relatively high emissions are apparent for Cross Unit 1 in May 2009. This period of relatively high emissions is consistent with reemissions of Hg from the liquor of a flue gas desulfurization unit (Tyree and Allen, 2010).

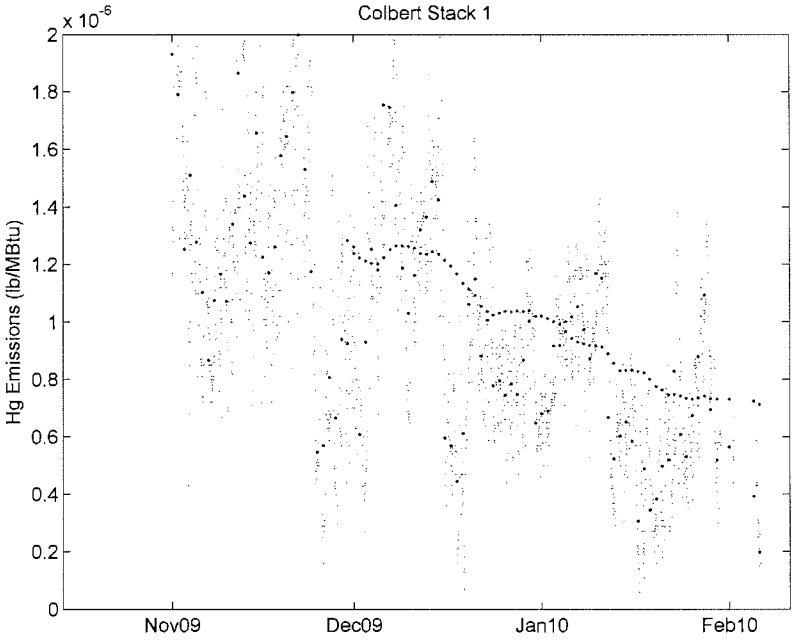


Figure 21: Colbert Stack 1 Hg emission averages.

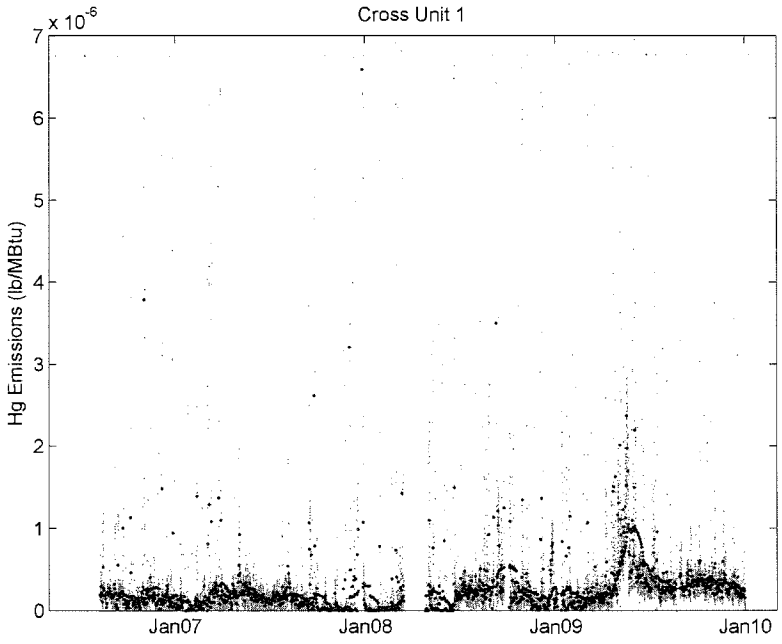


Figure 22: Cross Unit 1 Hg emission averages.

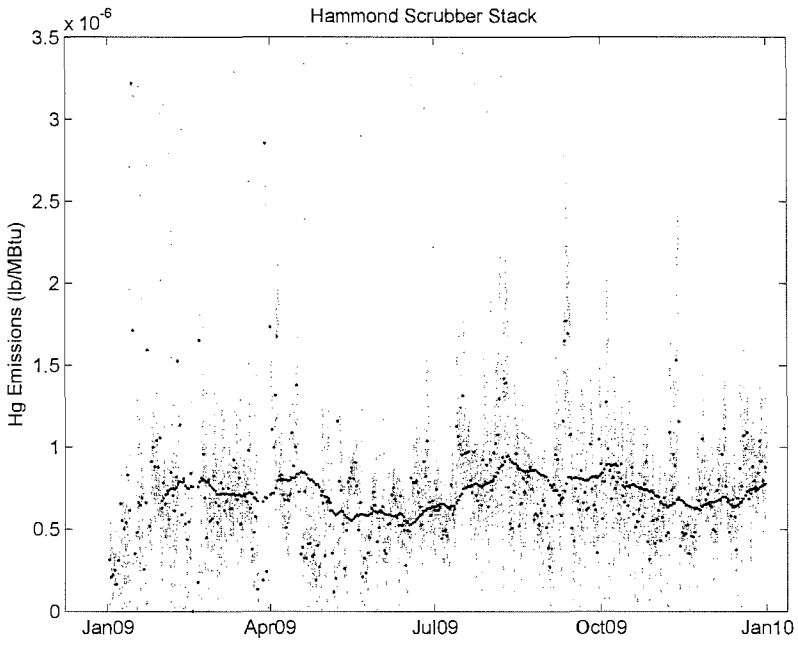


Figure 23: Hammond Scrubber Stack Hg emission averages.

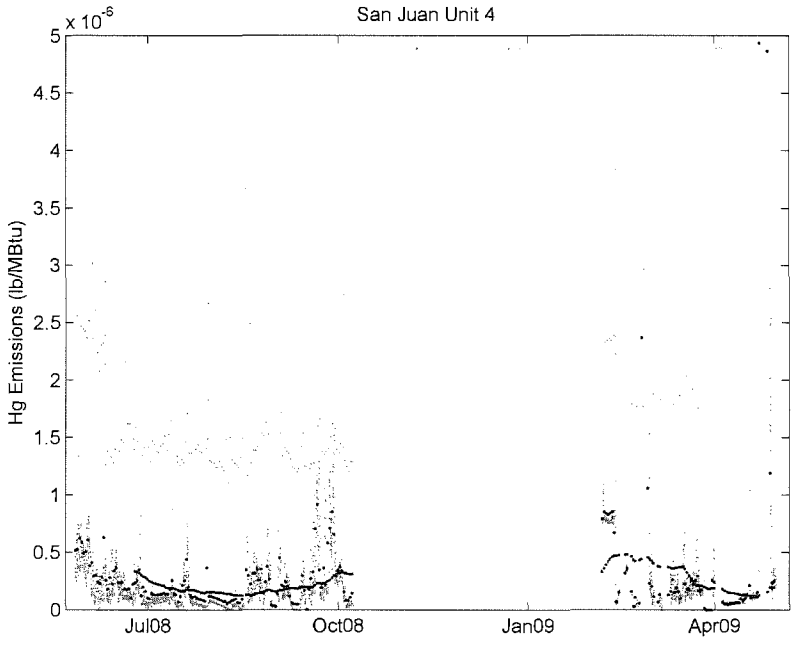


Figure 24: San Juan Unit 4 Hg emission averages.

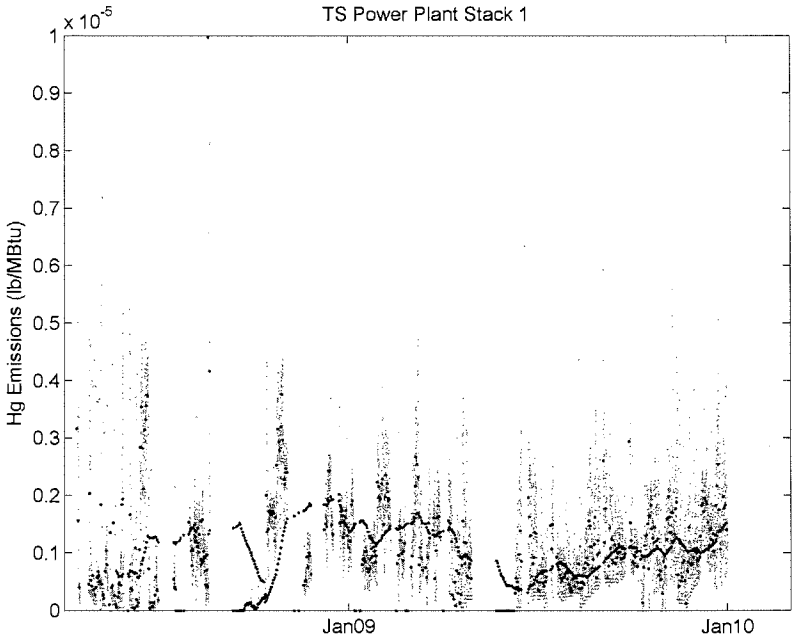


Figure 25: TS Power Plant Hg emission averages.

Proposed long-term emission limits may be compared with historical 30-day emission averages (see Figures 26-30). The top panel in these figures shows the historical 30-day emission averages with a red line marking the 99th percentile. The distributions of 30-day Hg emission averages for top performing units generally include right-hand tails attributable to correlated Hg emissions.

One can calculate a Hg emission limit directly as the maximum 99th percentile historical 30-day emission average among top performing units. Among the population of 127 top performing Hg units as compiled by RMB Consulting, valid CEMS data were available for five stacks (eight units). From these data, the highest 99th percentile historical 30-day emission average was 1.90 lb/TBtu. Note this value is above the UPL values calculated using stack test data (see Table 4). These emission limits are based on the Hg CEMS data available to us, which represent only a few of the top performing units; it is likely that higher 30-day emission averages would have been observed if more top performing unit CEMS data had been available.

In addition to historical 30-day emission averages we also calculated the 30-day averages using randomly selected days. A large number (10^6) of averages were calculated for each unit (see Figures 26-30). The middle panel in these figures shows the random 30-day emission averages with a red line marking the 99th percentile. These random averages remove the effect of multi-day correlations of emissions including multi-day startup events on the long-term averages.

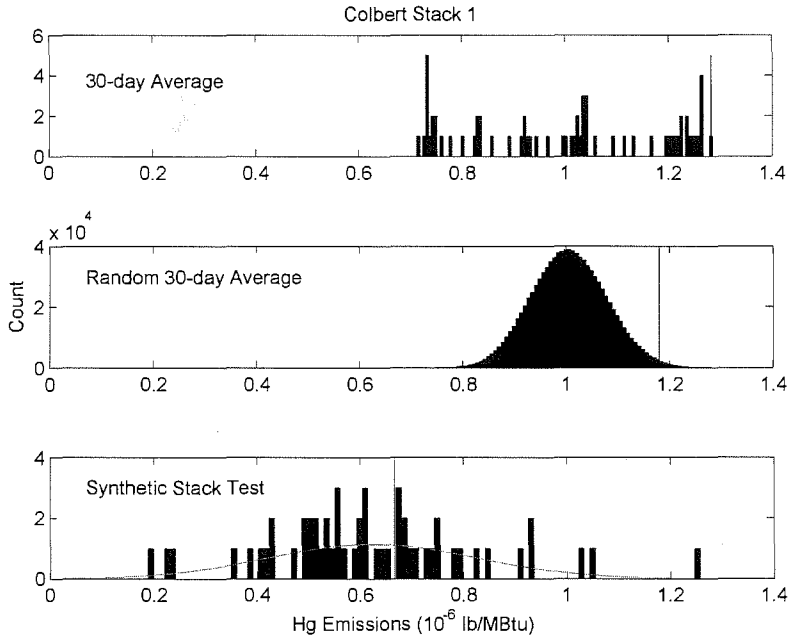


Figure 26: Colbert Stack 1 distribution of Hg emissions.

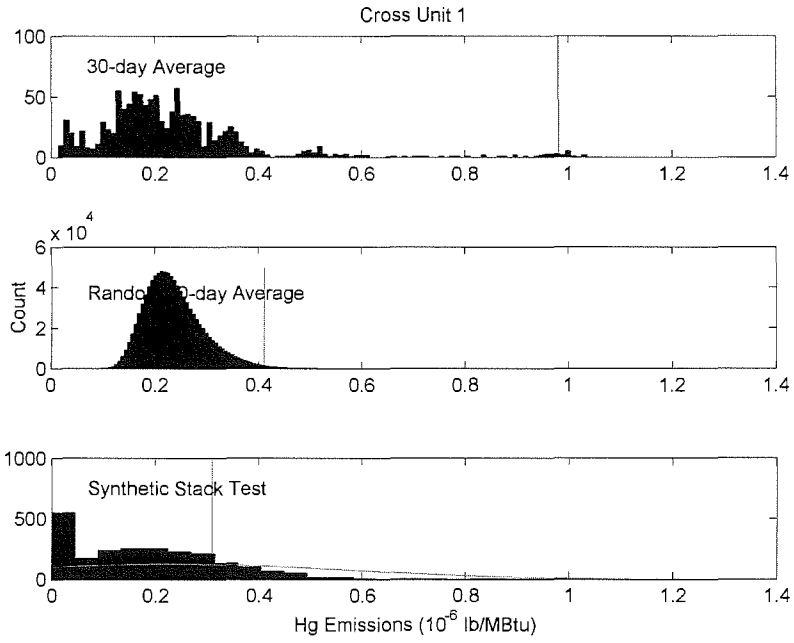


Figure 27: Cross Unit 1 distribution of Hg emissions.

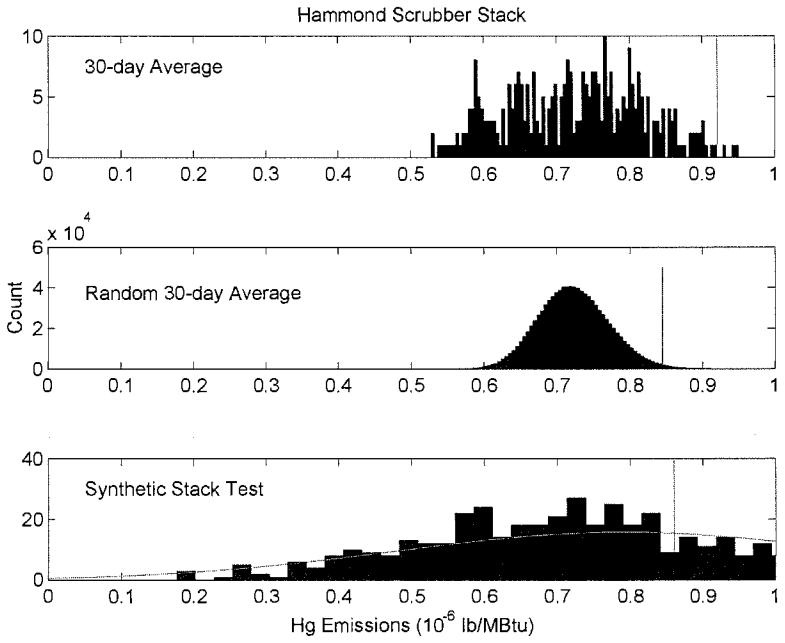


Figure 28: Hammond Scrubber Stack distribution of Hg emissions.

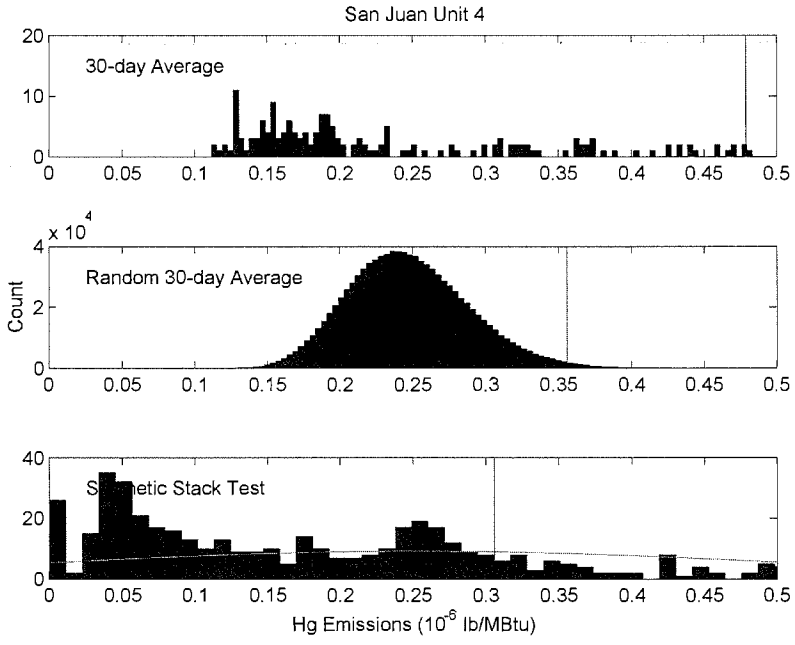


Figure 29: San Juan Unit 4 distribution of Hg emissions.

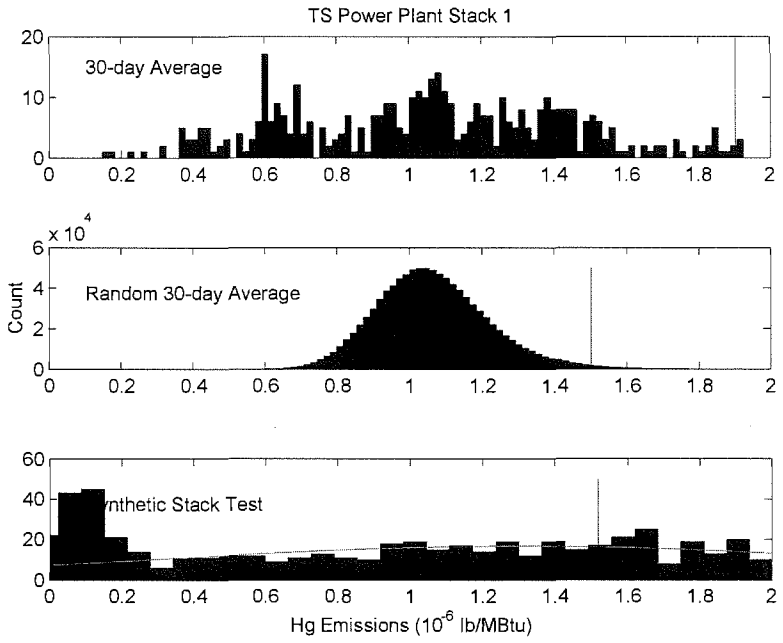


Figure 30: TS Power Plant distribution of Hg emissions.

Synthetic Stack Test Emissions

In order to relate long-term emission averages to emission limits calculated from stack tests, we calculated "synthetic stack test" (SST) results from Hg CEMS data. SST were calculated as the average of 6 consecutive hours of Hg emissions when the unit was at or above a full load threshold set to approximately 90% of capacity (see Table 11). This procedure is designed to obtain data like that from a stack test using CEMS data, and is similar to procedures reported in Tyree and Allen (2010) and Allen, Looney, and Tyree (2011).

The SST data for each unit were then used to calculate t-statistics UPL values using Equation 4. The distribution of SST values are shown as blue bars in the bottom panels of Figures 26-30; fitted normal distributions are shown as green curves and the t-statistics UPL values as red lines. The SST t-statistics UPL values are emissions limits that would have been calculated had many consecutive stack tests been collected over months at specific top performing units.

Table 11: Synthetic Stack Tests Calculated from Hg CEMS Data.

Plant Name	Full Load Threshold	SST Statistics			
	(MW)	N	Mean (lb/TBtu)	Std. Dev. (lb/TBtu)	Range (lb/TBtu)
Colbert Stack 1	650	54	0.622	0.203	0.188 - 1.26
Cross Unit 1	575	2464	0.236	0.345	0 - 4.50
Hammond Scrubber	450	485	0.793	0.315	0.177 - 2.73

Stack, Units 1-4					
San Juan Unit 4	500	550	0.253	0.247	0 - 1.13
TS Power Plant Stack 1	215	667	1.30	1.01	-0.104 - 6.28

Corrections to t-Statistics UPL

As mentioned above emission limits based on stack tests include no information about correlated emissions or emissions during infrequent events. Here we compare actual emissions averages with t-statistics UPLs calculated using SST for top performing units in order to assess the importance of these omissions and to propose corrections to emission limits based on stack test results.

If Hg emissions were independent and normally distributed, the SST-based UPL values would be close to the 99th percentile 30-day emission averages. In fact the ratios of 99th percentile historical 30-day emission averages to SST-based t-statistics UPLs are in the range 1.07-3.17 for the top performing units studied here (see Table 12). We designate this ratio R. Further evidence of the importance of correlated emissions can be seen by comparing averages of 30 random days with historical 30-day averages. The random averages remove the effect of multi-day emission correlations on the long-term averages. The ratios of the 99th percentile of the historical to random 30-day emission averages were in the range 1.08-2.38.

In order for emission limits to match the emissions achievable on any day by a top performing unit, emission limits calculated as t-statistics UPLs should be multiplied by R. This correction is to account for correlated emissions and emissions during infrequent events. The correction is based on all of the top performing Hg CEMS data available to us, which represents 8 of 127 the top performing units for Hg emissions and many more hours of emission data than were used in EPA's MACT floor calculations.

Table 12: Comparison of Historical 30-Day Average Hg Emissions with t-Statistic SST UPLs.

Plant Name	Historical 30-day Emission Avg.	Random 30-day Emission Avg.	t-Statistics UPL Based on SST	R
Colbert Stack 1	1.28	1.18	0.666	1.92
Cross Unit 1	0.980	0.411	0.309	3.17
Hammond Scrubber Stack, Units 1-4	0.921	0.844	0.860	1.07
San Juan Unit 4	0.478	0.356	0.306	1.57
TS Power Plant Stack 1	1.90	1.50	1.52	1.25

References

J. O. Allen, M. B. Looney, and C. A. Tyree, "Statistical estimates of long-term mercury emission limits for coal-fired power plants", 2011 AWMA Meeting, Paper # 2011-A-636-AWMA.

C. M. Jarque, and A. K. Bera, "A test for normality of observations and regression residuals." International Statistical Review. 55: 163-172, 1987.

R. Robeson, RMB Consulting, Personal Communication, 2011.

C. A. Tyree and J. O. Allen, "Determining AQCS mercury removal co-benefits", Power, 154(7):26-32, 2010.

ATTACHMENT 3

RMB Consulting & Research, Inc.

5104 Bur Oak Circle
Raleigh, North Carolina 27612

Phone (919) 510-5102
FAX (919) 510-5104

MEMORANDUM

TO: Utility Air Regulatory Group

FROM: Ralph L. Roberson, P.E.

DATE: August 3, 2011

SUBJECT: Technical Comments on EPA's Proposed Electric Generating Unit Rule

I. INTRODUCTION

On May 3, 2011 EPA proposed its National Emission Standards for Hazardous Air Pollutants from Coal- and Oil-Fired Electric Utility Steam Generating Units.[1] Because the emission standards set forth in this NESHAPs are based on emission reductions assuming application of maximum achievable control technology (MACT), such rules are often referred to as "MACT rules" or "MACT standards." The Utility Air Regulatory Group (UARG) asked me, Senior Consultant with RMB Consulting & Research, Inc. (RMB) to review and to provide technical comments on EPA's proposed EGU MACT Rule. UARG asked me to focus its review on the key components of any MACT rulemaking such as choice of surrogates; appropriate treatment of emission variability; calculation of MACT floors and compliance determinate requirements.

II. EPA'S APROACH TO PM SHOULD BE REVISED

Total PM Is Not Appropriate Surrogate

EPA proposes to regulate total PM, which is defined as the sum of filterable PM and condensable PM, solely on the basis of the behavior of selenium (Se). I disagree with EPA's decision for several reasons. First, there is overwhelming data (both historical and the 2010 EGU ICR) that support using filterable PM as the surrogate for antimony, arsenic, beryllium, cadmium, chromium, cobalt, lead, manganese, and nickel. While there is variability in the Se results, EPA's own data show exceptionally high removal percentages for all of the metals for all coals and all control technology configurations.[2] EPA states that the results for Se removal were less consistent. However, when we examine EPA's results closely, it appears that EPA is trying to distinguish Se where there is very little real difference. For example, EPA states that the results for Se control were consistently very good when subbituminous coal was fired. EPA also states that when a fabric filter was the primary control device, Se control was consistently good. Thus, the only questionable configuration for Se control appears to be when bituminous coal is fired and an electrostatic precipitator (ESP) is the only control technology.

[1] 76 Fed. Reg., 24,976 (May 3, 2011).
[2] 76 Fed. Reg. 25,038, col. 3 (May 3, 2011).

I believe EPA has unnecessarily complicated the control and regulation of non-Hg HAP metals based on shaky technical grounds. My analysis of the ICR data leads us to conclude that a unit cannot comply with the emission limits in the proposed rule while burning bituminous coal and only having ESP control technology. EPA's own analysis projects the installation of fabric filters for 166 GW of capacity.[3] Likewise, an owner/operator cannot burn bituminous coal and expect to comply with the proposed acid gas limitation without installing either wet or dry scrubbing. A simple calculation demonstrates that bituminous coal with a nominal chlorine content equal to 750 ppm will require approximately 97 percent removal to comply with the proposed 0.002 lb/10^6 Btu limit. This is a significant scrubbing requirement and will almost certainly require wet scrubbing.

$$HCl\ inlet = \frac{750\ ppm}{12,000\ Btu/lb} \times \frac{36.5\ lb\ HCl}{35.5\ lb\ Cl} = 0.064\ {}^{lb}/_{10^6\ Btu}$$

$$Removal = \frac{In - Out}{In} = \frac{0.064 - 0.002}{0.064} \times 100\% = 97\%$$

I believe all stakeholders would be better served with a filterable PM limit as the single non-Hg HAP metals surrogate. EPA should be aware that to the extent Se is neither controlled nor adequately characterized by filterable PM, the Agency certainly has not demonstrated that Se is collected in the condensible PM fraction. EPA's concluded that Se removal percentages were not consistent when burning bituminous coal with only ESP control.[4] But this conclusion is based on the observation that some Se was collected in the Method 29 impingers and not on the Method 29 filters. However, EPA Method 29 impingers are not the same solutions as Method 202 impingers, which is the EPA reference method for condensable PM. In other words, EPA has failed to provide any data that demonstrate Se is present in condensable PM.

Condensable PM Measurement Issues

As noted above, total PM consists of two components, filterable PM and condensable PM. Since no single EPA method measures both filterable and condensable PM, a minimum of two different EPA sampling methods must be utilized to determine total PM emissions. For the ICR, EPA specified OTM-28 for condensable PM measurement. Since the section 114 ICR letters were mailed by EPA to EGUs (December 2009), the requirements of OTM-28 have been incorporated into EPA Method 202, which is one of the proposed compliance methods. Method 202 has been flawed since it was issued by the Agency 20 years ago. Despite recent cosmetic changes to Method 202 by the Agency, the method remains flawed and yielded very inconsistent ICR test results. As EPA is aware, the Electric Power Research Institute (EPRI) has conducted numerous analyses on the EGU ICR data, and EPRI will be submitting detailed comments under its own cover. Among the EPRI results I am privy to are a series of regression analyses of the individual metals versus the various PM fractions (i.e., filterable, condensable and total). The PM component with clearly the least explanatory power was condensable PM. The reason for

[3] *Regulatory Impact Analysis of the Proposed Toxics Rule*, U.S. Environmental Protection Agency, p. 8-14, March 2011.

[4] Note that EPA's did not evaluate Se control with an ESP followed by a wet scrubbing system, which is quickly becoming the most prevalent configuration for burning bituminous coal.

lack of correlation is likely not only due to the fact that non-Hg metals are predominately solids (filterable PM) at stack temperatures, but also because of the poor quality of condensable PM data collected with EPA Method 202.

Suggested Approaches for Establishing Filterable PM Emission Limit

One alternative would be to follow the approach contained in EPA's PM Floor Spreadsheet,[5] but focus on filterable PM instead of total PM. Starting with the above-referenced EPA spreadsheet file, I identified all of the filterable PM data and sorted the units based on each unit's minimum reported filterable PM test average.[6] This is consistent with the procedure EPA has used for Hg and HCl for all of the proposed subcategories. I constructed a new filterable MACT pool using the 131units with the lowest reported filterable PM test averages. I computed $\bar{x}$ based on the average filterable PM emission value for each of the 131 units in the MACT pool, and I computed the variance of all the test values (i.e., 221 averages) in the pool. This approach yielded a upper prediction limit (UPL) = 0.013 lb/10^6 Btu.[7] As documented in Section III of this memorandum and also discussed in Allen Analytics Report to UARG, EPA's UPL approach fails to properly account for variability. Using an "R" multiplier value equal to 1.69, as recommended in the Allen Analytics Report,[8] yields a recommended filterable PM limit equal to 0.022 lb/10^6 Btu.[9]

As a second approach to developing a filterable PM limit, I wanted to more closely follow the approach EPA used for its Hg floor analysis. The key to this approach is to consider all of the Part II and Part III filterable data in the variance calculation, rather than just the minimum values that were used above and are included in the EPA PM floor calculations. By including all of the Part II filterable PM data along with the Part III filterable PM data, we may characterize enough hours of plant operation to eliminate the need for variability adjustment factor described in the Allen Analytics Report. Starting with the 131 best performing units in the MACT pool (determined by minimum test values), I went into EPA Microsoft® Access Database[10] to retrieve all of the additional filterable PM data for the MACT pool units. I added all of available Part II filterable PM data to the 131 MACT pool units, which were defined by the lowest reported filterable PM test averages. First, I computed $\bar{x}$ based on the average filterable PM emission value for each of the 131 units in the MACT pool, and I computed the variance of all the test

[5] See the EPA file, *floor_analysis_coal_pm_031611.xlsx.*
[6] It is difficult to divine exactly the procedure used by EPA in computing total PM from the filterable and condensable PM measurements provided in the above-referenced spreadsheet file. EPA may have inadvertently omitted the filterable PM data contained in Column "CS" of Tab "PM_coal_MMBtu." Regardless, I used all available filterable PM data from EPA's spreadsheet file for my analysis.
[7] Details of EPA's UPL approach are discussed in Section III of the memorandum.
[8] *Evaluation of Proposed EGU MACT Floor Emission Limits for Hg, PM, and HCl*, prepared by Jonathan Allen, Sc.D., P.E., Allen Analytics LLC, Tucson, AZ, July 30, 2011.
[9] The Allen report also contains a maximum recommended "R" values of 3.89, which appears to be driven by a PM CEMS dataset that contains an inordinately long start-up event. Given that most of the other R values are in the 1.3 to 1.7 range plus the fact that EPA's proposed rule provides for a diluent cap during start-up and shutdown events, I believe the "R" = 1.69 recommendation is reasonable, albeit it perhaps conservative.
[10] See the EPA file, *eu_partiii.mdb.*

values (i.e., 460 averages) in the pool. This approach yielded a UPL = 0.033 lb/10^6 Btu.[11] Second, I computed $\bar{x}$ based on the minimum filterable PM emission value for each of the 131 units in the MACT pool, and I computed the variance of all the test values (i.e., 460 averages) in the pool. This approach yielded a UPL = 0.029 lb/10^6 Btu. These two UPL results are driven by the variance term, which may be unrealistically large. I did not screen any of the Part II filterable results based on either outlier or data quality concerns. Visual inspection of the filterable PM results for some of the units revealed more variability than one would expect for "normal" operating conditions. Some of the PM test results may have been for specialized situations such as fuel tests or testing to develop a compliance assurance monitoring (CAM) plan. I simply did not have time to review individual test reports in effort to ferret out unrepresentative test results, which would serve to reduce the variance term as well as the calculated UPL.

I believe both of the two different computational approaches described and implemented above illustrate that EPA's UPL approach based on limited datasets fails to capture long-term variability. Both approaches indicate that EPA's UPL approach does not yield emission limits that are truly achievable on a continuous basis. This conclusion is confirmed by the results obtained from using the variability adjustment factor suggested in the Allen Analytics Report as well as from using EPA's UPL approach, but only when used with adequate datasets to sufficiently characterize variability.

III. CRITIQUE OF EPA'S VARIABILITY ANALYSIS

Overview of EPA's Variability Analysis

EPA's attempt to address emission variability through the use of an upper prediction limit (UPL) is fundamentally flawed. The UPL approach does not accomplish what the Agency purports it to accomplish. Failing to address variability correctly means EPA's proposed rule is technically deficient and also at odds with several rulings by the D.C. Circuit Court of Appeals.[12]

It is not obvious where within our comments is the optimum location to introduce this issue, but I believe all of the EPA floor calculations are fatally flawed because of EPA's decision to use the lowest measured value for each unit, when multiple test values were available. This is particularly an egregious error in the total PM calculations because EPA often had simultaneous or near simultaneous measurements of filterable PM employing different EPA methods and the Agency ignore variability and simply combine the lowest filterable PM concentration with a condensable PM concentration to arrive at a total PM concentration. (I address the PM measurement issue of multiple methods in Section IV of this memorandum.)

---

[11] This UPL result is driven by the variance term, which may be unrealistically large. I did not screen any of the Part II filterable results based on either outlier or data quality concerns.

[12] See, for example, *National Lime Association v. EPA*, 627 F.2d 416 (DC Cir 1980) (holding that EPA failed to show how the standard proposed was achievable under the range of operating conditions that might affect the emission that was being regulated).

I decided to introduce this topic in the variability section to illustrate how the Agency's decision biases the floor results and completely fails to address within unit variability. Based on one of the EPA spreadsheet files, apparently eight complete Hg test series were submitted for the BL England facility.[13] The test set averages ranged from a low of 0.0296 lb/10^{12} Btu to a high of 2.43 lb/10^{12} Btu – almost two orders of magnitude. Yet, EPA used only the lowest value (0.0296) to compute the mean of the best performing units for mercury. Based on the eight sets of BL England Hg data, the best or central estimate of performance is given by the mean, which is equal to 0.578 lb/10^{12} Btu.

What UPL Does and Does Not Accomplish

It would be perhaps gratuitous to quibble with some of EPA's statements leading up to the Agency's UPL analysis. For example, EPA incorrectly applies the Central Limit Theorem and concludes that the Agency can assume normal distributions because the sample size (ICR datasets, in this case) is 15 or larger. In point of fact, when we compute the kurtosis and skewness statistics for any of the ICR datasets (PM, HCl and Hg), it becomes abundantly clear that the data distributions are not normal. Also, as mentioned above, EPA does not actually calculate the correct average ($\bar{x}$) performance of the MACT pool because the Agency elects to use the minimum value of multiple test results. But such quibbles are not at the heart of EPA's misdoings.

Issues With EPA's Variability Analysis

EPA's attempt to address emissions variability through the use of an upper prediction limit (UPL) is fundamentally flawed. The UPL approach does not accomplish what the Agency purports it to accomplish. Failing to address variability correctly means EPA's proposed rule is technically deficient. EPA used the following formula to estimate the UPL for the best performing unit:

$$UPL = \bar{x} + t(0.99, n-1) \times \sqrt{s^2\left(\frac{1}{n} + \frac{1}{m}\right)}$$

Where:

n = the number of test runs for best performing source
m = the number of test runs in the compliance average
$\bar{x}$ = mean of the data for top performing unit
t(0.99, n – 1) = 99th percentile of the T-Student distribution with n – 1 degrees of freedom
s^2 = variance of the data from the top performing source.

The problem with EPA's approach is that the Agency is applying the UPL formula to very incomplete data, especially for the new unit analysis. For each HAP, EPA typically used three sampling runs that were conducted under relatively constant operating conditions and performed very close in time (i.e., at a maximum, over 3 consecutive days) for the single, best performing

[13] See, *floor_analysis_coal_hg_051811.xlsx.*

unit. The variance (s^2) that EPA calculates using the formula above is only representative of a very limited set of operating conditions and probably little, if any, fuel variability. Thus, EPA is only predicting the 99th percentile of a very limited range of operation and not necessarily a level that can be complied with at all times and under all operating conditions.

For total PM and HCl, EPA determined that there should be 131 units in the MACT pool for the category of coal-fired units designed for coal > 8,300 Btu/lb. Because EPA preselected among the best performing units for ICR stack testing, I agree with EPA's decision regarding the size (12% of 1,091 = 131) of the MACT pool for PM and HCl. The problem is that EPA's approach, at best, estimates the 99th percentile of 131 individual data snapshots. EPA has no idea how representative any single ICR snapshot is or what frequency of time a given unit can achieve the level at which operated during the ICR test.

Our concern with EPA's incomplete treatment of emission variability includes both the new unit and existing unit analysis. It may be easier to comprehend our criticism in the context of EPA's new unit analysis because we are only working with the emissions from a single, best performing unit.

EPA's UPL approach is flawed because the Agency's analysis fails to address how representative any perceived best performing ICR stack testing results are relative to long term operation and performance. The ICR instructions specifically directed (1) the stack tests to be representative of the fuel that is routinely burned and (2) all pollution control equipment to be operated in accordance with manufacturers' specifications for proper operation during emissions testing.[14] What the ICR test results do not tell us is how representative are the fuel and operating characteristics for each unit in the database of longer term performance. Is the ICR data snapshot indicative of : (1) a really good day(s) of fuel and operating parameter characteristics; (2) an average day(s); or (3) performance that can be achieved a relatively high percentage of the time? The plain truth is that EPA idea has no idea how representative any give ICR data snapshot is of long term performance. Since EPA cannot determine where along a unit's performance "curve" any given ICR test result resides, the Agency's claim that the proposed emission limits are based upon a 99th percent UPL is simply speculation and not supported by an information or analysis in the rulemaking docket.

My analysis of CEMS data, which will be presented later in this memorandum, demonstrates just how inadequate EPA's variability approach actually is in the context of the proposed limits being continuously achievable. As we noted earlier, some of the issues are easier explained in the context of proposed emission limits for new units because only a single unit is involved. EPA's proposed HCl limit for new units is 0.30 lb/GWh. The proposed limit is supported by the Agency's UPL analysis of HCl emission data from the Logan Generating Plant.[15] However, the same EPA spreadsheet lists a Part II HCl test result for the Logan Generating Plant equal to 0.66

---

[14] *Supporting Statement for OMB Review of ICR No. 2362.01 (OMB Control Number 2060-0631): Information Collection Request for National Emission Standards (NESHAP) for Coal- and Oil-Fired Electric Utility Steam Generation Units*, Part B, p. 7,U.S. Environmental Protection Agency, Research Triangle Park, NC, December 24, 2009 (cited hereafter as "2010 ICR Supporting Statement").

[15] See EPA Spreadsheet, *floor_analysis_coal_hcl_031611.xlxs*, Tab = HCl_New_MW, Cell = C102.

lb/GWh, which is over twice the proposed limit. In other words, the unit upon which EPA based the proposed HCl emission cannot meet that limit based on a single additional stack test. One can only imagine how the Logan Generating Plant would fare under the proposed HCl limit using CEMS data, including periods of start-up and shutdown.

Analysis of PM CEMS Data

In anticipation of EPA's proposed EGU MACT rule, I solicited PM CEMS data from a number of utility companies. However, in requesting PM CEMS data, I limited my requests to those units selected by EPA for mercury, non-Hg metals and PM testing in Part III of the 2010 ICR. RMB's request was very basic; we asked for a minimum of one calendar year of all quality assured, hourly PM emission averages.

I conducted several different analyses of the PM CEMS data. Since we received and began analyzing the PM datasets before Administrator Jackson signed the proposed EGU MACT rule (March 16, 2011), I decided to examine several different averaging times, ranging from 3-hour rolling averages to 30-day rolling averages. This type of analysis illustrates how the length of averaging time impacts the ability of a unit to comply with a given numerical limit. My results are summarized in a series of tables presented in Appendix A to this memorandum.
The values presented in the cells of Tables A1 – A6 represent the percentage of time the unit's emissions were below the hypothetical numerical limit shown as the each header. The reader should be aware that the number of compliance periods in a calendar year is dependent on the averaging time. I enumerated the number compliance periods in a year, using the unrealistic operating scenario[16] that a unit operates 24 hours per day for 350 days, which equals 8,400 operating hours.

Averaging Time	Number of Compliance Periods
8-hour rolling	8,393
24-hour rolling	8,377
Daily (24-hour discrete)	350
Weekly (168-hour discrete)	50
30-day rolling	321

With reference to the tables, if a unit complies with a numerical limit 90 percent of the time, then it would have 838 (0.10 x 8,377) exceedances based on 24-hour rolling averages but only 32 (0.10 x 321) exceedances based on 30-day rolling averages. The results shown in the tables are somewhat counter-intuitive results because *conventional wisdom* is that as the averaging time increases, so does the compliance rate. I believe the results shown in the tables may be partially the result of the vagaries of the computations (i.e., the number of potential exceedances in a year is a function of the averaging period examined). However, I believe there is an "operational" explanation for the observed results. That is, when an excursion is high enough to cause an exceedance of a longer-term average (e.g., 30-day rolling) that event remains in future

[16] It is unrealistic to assume a unit never experiences a forced outage during the year, and thus operates a full 24-hour cycle each day. Nonetheless, this example illustrates how the number of compliance period in a year vary as a function of assumed averaging time

calculation and results in additional exceedances. Thus, notwithstanding conventional wisdom regarding the lengthening of averaging time, my analysis shows that the percent compliance does not increase significantly as the averaging time increases. Thus, while the number of exceedances decrease with increasing averaging time, the percentage of time does not decrease very significantly.

PM CEMS Data Illustrates Emission Variability is Not Accounted for Properly

One of the compliance options for non-Hg metallic HAPs in the proposed rule is to achieve a total (filterable plus condensable) PM emission limit of 0.03 lb/10^6 Btu based on periodic stack tests. The owner/operator may then choose to demonstrate continuous compliance by installing and operating a PM CEMS in accordance with proposed §63.10010(g).[17] [In another section of comments, I have explained that proposed approach to establishing the filterable PM limit is unworkable and therefore must be corrected.] Since EPA's proposed approach is based on using 30-day rolling averages, I used the hourly PM CEMS data to calculate a series of 30-day rolling averages. Next, I prepared a cumulative distribution frequency (CDF) plots of each PM CEMS dataset. The CDF plots are presented as Figures B-1 through B-6 in Appendix B. The x-axis of the CDF plots represents filterable PM emissions in units of the standards, lb/10^6 Btu. The y-axis shows the cumulative or combined percentage of time that the 30-day rolling averages are less than the corresponding PM emission limits, shown on the x-axis.
Figures B-1 through B-6 show that an ICR stack test can be conducted at any point along a unit's emission distribution. EPA's UPL calculation fails to include or address this fact. For example, Figure B-1 indicates that the ICR test was conducted at an emission rate that Unit 7 operates at infrequently. Conversely, Figure B-2 shows that the ICR test was conducted at an emission rate that Unit 8 can achieve almost all the time. Interestingly enough, Units 7 and 8 are "sister" units at the Oak Creek Power Plant (OCPP), receive coal from the same coal pile and have similar PM control technologies. Again, EPA's UPL analysis is completely indifferent to this situation (i.e, one unit can achieve its ICR emission rate infrequently while a sister unit achieves its ICR rate almost all the time). As we stated earlier, EPA's UPL analysis focuses on identifying the 99th percentile of snapshot tests, but has no comprehension of how representative any single ICR snapshot test is of long-term performance.

The CDF plots in Appendix B clearly show, EPA's UPL analysis fails to account for how representative each ICF data snapshot is of longer term emissions. Moreover, an acceptable response is not, "we (EPA) proposed 30-day rolling averages and that solves the short-term data issue/problem." The CDF plots prove that EPA's UPL analysis has not demonstrated that the proposed emission limits are achievable, as required by the Clean Air Act and subsequent case law.

IV. EPA's PROPOSED TOTAL PM LIMIT MAY BE BASED ON BIASED DATA

Although not elaborated on in any of the technical support document, EPA's approach for determining total PM emission appears to bias the results low. Total PM consists of two components, filterable PM and condensable PM. Since no single EPA method measures both

[17] 76 Fed. Reg. 25.112, col. 2 (May 3,2011).

filterable and condensable PM, a minimum of two different EPA sampling methods must be utilized to determine total PM emissions. For the ICR, EPA specified OTM-28 for condensable PM measurement.[18] However, because all filterable methods may not be applicable to all types of stacks, the ICR included several filterable PM measurement methods (i.e., Method 5, OTM-27 and Method 29).

For the Part III ICR data, I concluded EPA selected the lowest filterable PM value among up to three options and combined this result with the condensable PM result from OTM-28. EPA implemented the same approach in processing any Part II PM data to arrive at total PM. Next, EPA selected the lowest total PM value between the Part III and Part II ICR results, and placed this value in a spreadsheet column titled, *Total_Particulate_Calc-min.*[19] EPA then sorted the total PM data, identified the best performing (lowest) 131 units, conducted its UPL calculation on these 131 values and arrived at 0.026 lb/10^6 Btu. At best, EPA has determined the 99th UPL of a group of minimum values.

Comparing ICR PM Test Results to Historical Test Results

As I began to review EPA's initial releases of the ICR Part III data (October 2010), I noted some of the lowest filterable PM results I have ever observed. The following antidotal experience caused me to research this issue in greater detail. There are two units in the database for which my company (RMB) has been working on a Pollution Control Upgrade Analysis for 2 to 3 years. RMB staff reviewed numerous stack tests for these units and conducted a number of computer simulation runs – all which resulted in us concluding that EPA's consent decree limit of 0.03 lb/10^6 Btu was not achievable other than by replacing the existing electrostatic precipitators (ESPs). I was surprised to find in EPA's ICR database a filterable PM result for one of the units equal to 0.013 lb/10^6 Btu. As a result of this experience, I began to contact utility companies and EPRI to inquire if others were observing unusually low PM concentrations.

The response to our calls was prompt and convincing. Following are two graphs, Figures 1 and 2, prepared by Southern Company Services. Figure 1 is for Plant Scholz, a facility with two coal-fired EGUs located in Florida. As Figure 1 shows, the two units have been individually tested a number of times over the last 5 years. The mean PM emission rate for Units 1 and 2, based on 16 tests (48 runs) is 0.018 lb/10^6 Btu. The probability of obtaining the Method 29 result of 0.004 lb/10^6 Btu is about 2.5 percent; the probability of obtaining the OTM-27 result of 0.002 lb/10^6 Btu is less than 1.5 percent. In other words, if the Scholz stack were tested 100 times, you would expect to obtain the ICR-reported results 1 to 3 times. The statistics for Plant Miller (Figure 2) are not as impressive as for Plant Scholz because the number of tests is much smaller. Nonetheless, Figure 2 shows that the ICR results are distinctly lower than previous PM test results.

---

[18] Since the section 114 ICR letters were mailed by EPA to EGUs (December 2009), the requirements of OTM-28 have been incorporated into EPA Method 202. Similarly, the requirements of OTM-27 have been incorporated into EPA Method 201A. See, 75 Fed. Reg. 80,118 (December 21, 2011).

[19] See EPA Spreadsheet, *floor_analysis_coal_pm_031611.xlxs*, Tab = PM_Coal_MMBtu.

Figure 1. PM Emissions Test Results for Plant Scholtz

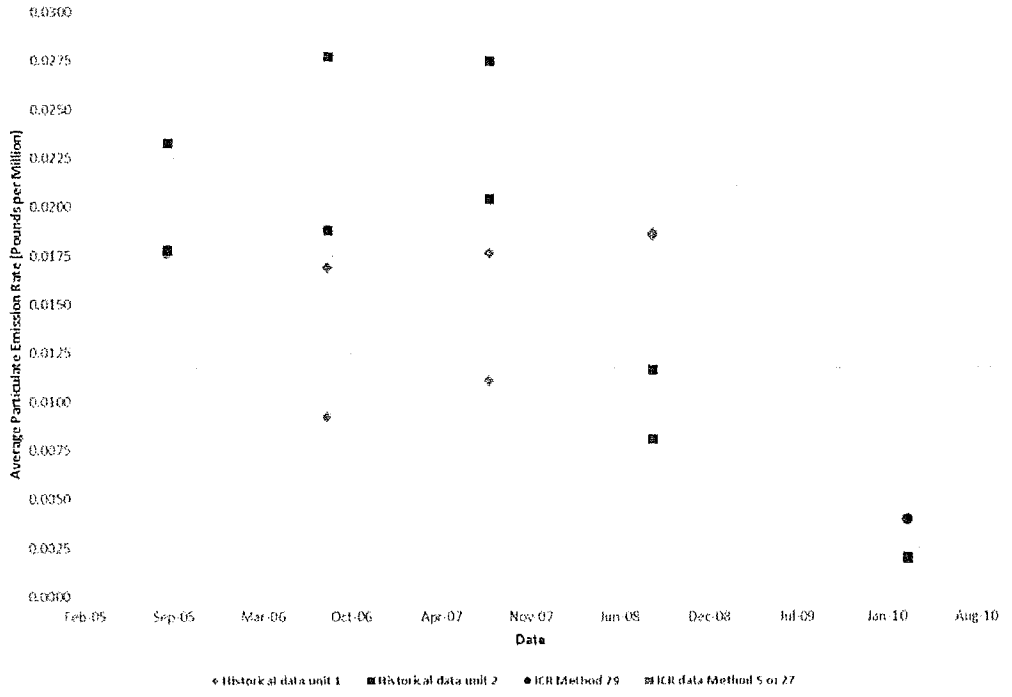
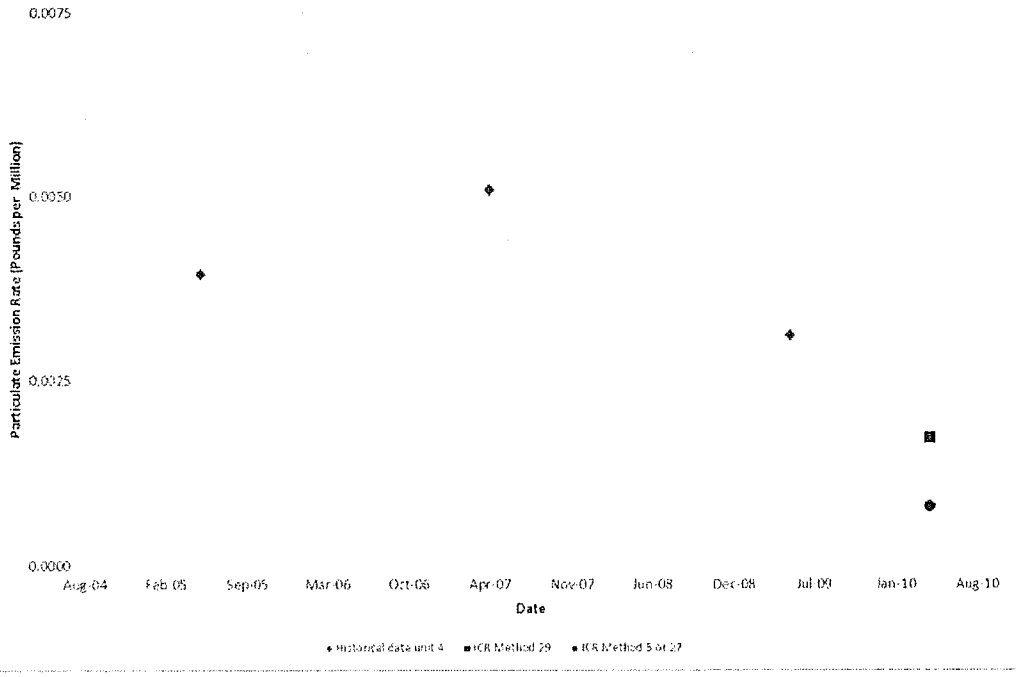


Figure 2. PM Emissions Test Resutls for Plant Miller



Developing and Testing Hypothesis

My thought process was to identify the differences between the required ICR testing and historical PM stack testing on EGUs. There were two obvious differences: (1) the ICR sampling runs were of longer duration (i.e., 4 hours versus 1- or 2-hour runs) and (2) the ICR specified Method 29 as the primary filterable PM method whereas historical tests relied on Method 5.

I know EPRI and the American Electric Power (AEP) undertook a field project to test the above-posited explanations. That work is reported in a standalone report, which will be submitted to EPA under separate cover. I understand that the AEP project was able to replicate the situation where Method 29 produced significantly lower PM measurements than did concurrent Method 5 results. Of course, the primary different between the two methods with respect to PM determinations is that Method 29 recommends a high purity quartz filter without organic binder whereas Method 5 specifies a glass-fiber filter without organic binder.

Additional Tests Designed to Specifically Evaluate Potential PM Bias

Earlier in 2011, I reviewed some of the above-discussed observations with representatives from Sunflower Electric Cooperative (Sunflower). Sunflower has a second unit permitted for its Holcomb station, but did not commence construction prior to EPA's May 3, 2011 EGU MACT proposal date. Presumably, Holcomb 2 would eventually be subject to whatever final MACT limits EPA may issue for new coal-fired units. Therefore, Sunflower is interested in any potential low biases in the ICR data that could unfairly lower the new source PM limits.

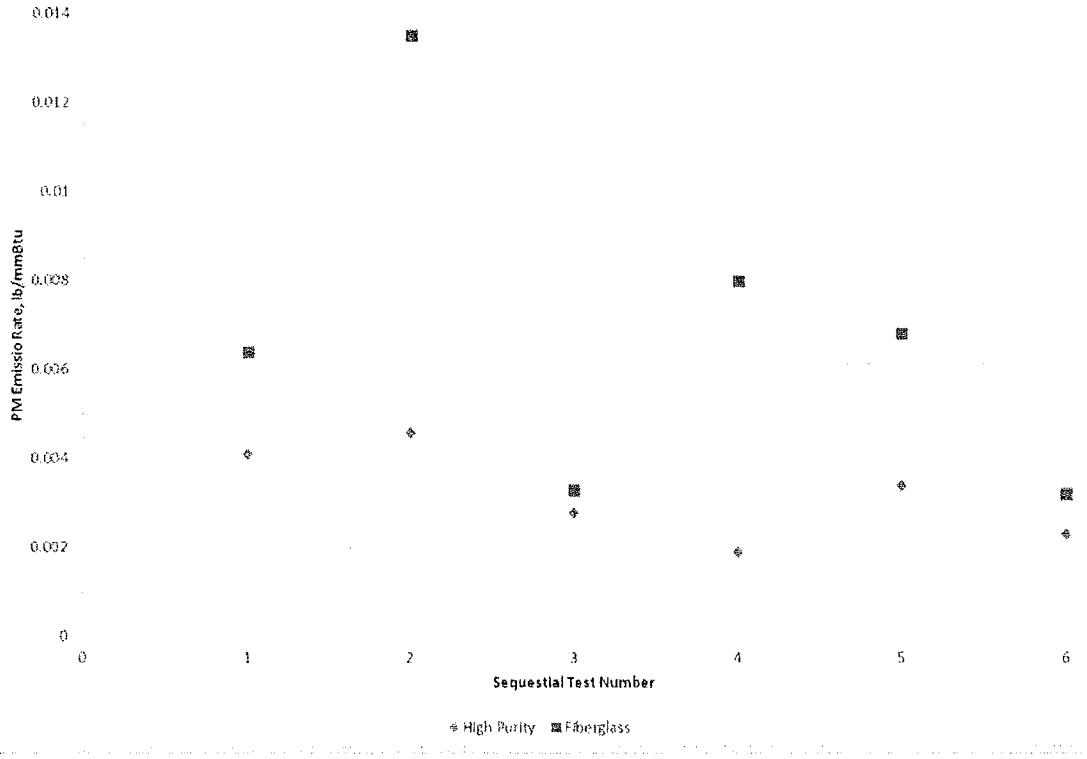
The following test matrix was developed and implemented at Holcomb 1 in April 2011. Although Holcomb 1 has been in commercial service 28 years, it is a well-controlled unit being one of the first units built subject to EPA's Subpart Da new source performance standards. The test matrix was implemented in duplicate on consecutive days. The Method 29 (M29) runs used high purity quartz filter without organic binder, and the Method 5 (M5) runs used a glass-fiber filter without organic binder. The results of these tests along with some interesting Method 29 findings (method detection limits versus some of EPA's proposed emission limits for new sources) are documented in a final test report.[20]

Run	Duration	Method	Temperature	Duration	Method	Temperature
1	4-Hr	M29	250 °F	4-Hr	M5	250 °F
2	2-Hr	M29	250 °F	2-Hr	M5	250 °F
3	2-Hr	M29	320 °F	2-Hr	M29	250 °F

I extracted filterable PM results from the above-referenced report and prepared Figure 3. Regardless of test duration (i.e., 2-hr versus 4-hr), the Holcomb results indicate that Method 29 results with a quartz filter consistently yields lower PM concentrations. While EPA may not be able to resolve this issue during the comment period, at a minimum EPA should revise Table 5 of the proposed rule to allow either Method 5 or Method 29 for PM testing.

[20] *Filterable Particulate and Trace Metals Emission Study Test Report,* prepared by Platt Environmental Services, Inc., Oak Brook, IL, prepared for Sunflower Electric Power, July 2011.

Figure 3. Comparative PM Measurements for Holcomb Unit 1



V. EPA's MERCURY MACT POOL IS NOT JUSTIFIED

EPA explains that the Agency targeted the best performing 15 percent of the coal-fired EGUs for ICR testing in the three HAP groups. Therefore, EPA concluded it had emissions data or information on the best performing 12 percent of all existing coal-fired EGUs. In the proposed rule, EPA used data from 12 percent of all units in the category for the two HAP groups. For Hg, EPA decided to only use "*the top 12 percent of the data obtained because, even though we required Hg testing for the units testing for the non-Hg metallic HAP, we did not believe those units represented the top performing 12 percent of sources for Hg in the category at the time we issued the ICR.*"[21]

This statement is inconsistent with the previous EPA statements and suggests some revisionist thinking. First, in responding to a comment received during the ICR notice and comment process regarding the selection of coal- and oil-fired units for testing. EPA states:

> *For the Hg and other non-mercury metallic HAP group, EPA believes that units with the newest PM controls installed represent those units meeting the lowest PM emission limits and, thus, <u>are believed to be among the top performers with respect to Hg</u> and other non-*

[21] 76 Fed. Reg. 25,023, col. 1 (May 3, 2011).

> *mercury metallic HAP emissions. Therefore, EPA has selected 175 units with the newest PM controls installed; of these 175, the newest 170 operating units will be required to conduct Hg and other non-mercury metallic HAP testing. This number includes units with ACI installed and other control configurations are also included (e.g., for SO_2 and NO_X control).*[22]

Then, again in the Agency's final Supporting Statement to the Office of Management and Budget (OMB), EPA discusses selecting units with activated carbon inject (ACI) and being identified as top performing units. *"The units selected also include a number with ACI installed. As units have been identified as meeting the criterion of being a "top performer" unit, substitution of units will not be permitted."*[23] Why would EPA have included "a number of units with ACI installed," if not trying to quantify the best performing units for mercury control? Referring to Attachment 11 of the 2010 ICR Supporting Statement, of the 170 units EPA selected for Hg and non-Hg metallic HAP testing, approximately 120 were equipped with fabric filters. As EPA undoubtedly knows from past experience, when a flue gas contains any measureable amount of unburned carbon, then a fabric filter becomes a highly efficient Hg control technology. The ICR data confirms the above statement. When I examine the best performing Hg units, as tabulated in one of EPA's spreadsheets,[24] I counted a grand total of <u>one</u> unit that does <u>not</u> have a fabric filter among the <u>top 60</u> units.

It is possible that when EPA initially examined a summary of the mercury ICR test results, the Agency may have questioned its success in truly identifying the best performing units. In fact, when I examined EPA's initially-posted spreadsheet, I was very surprised to observe how many of the apparent top performing units were not among the Part III ICR units. Quickly, I determined that EPA's initially-posted spreadsheet[25] contained a data conversion error that resulted in about 20 units having their results underreported by a factor equal to 1,000. This was communicated to EPA, and EPA subsequently revised its analysis.[26,27]

My review of EPA's revised Hg spreadsheet indicates that once EPA fixed its conversion error, the Hg data appear very similar to the other ICR data. That is, EPA did select the best performing units for Hg control so the Agency should have used 12 percent (0.12 x 1,061 = 127) of the coal-fired fleet in the Hg MACT pool.[28]

UARG asked me to recreate EPA's Hg analysis except to base the analysis on 12 percent of the entire subcategory of units designed to burn coal >8,300 Btu/lb, which we believe is 127 units.

---

[22] *Response to Comments Received on Proposed Information Collection Request (Published on July 2; 74 FR 31725), p.27,* U.S. Environmental Protection Agency, Research Triangle Park, NC, November 5, 2009 (emphasis added).
[23] See, 2010 ICR Supporting Statement.
[24] EPA's revised spreadsheet titled, *floor_analysis_coal_hg_051811.xlxs.*
[25] EPA's spreadsheet titled, *floor_analysis_coal_hg_031611.xlxs.*
[26] Letter to EPA Administrator Lisa P. Jackson from Lee B. Zeugin, Counsel to the Utility Air Regulatory Group, dated May 6, 2011.
[27] Letter to Lee B. Zeugin, Counsel to the Utility Air Regulatory Group from Gina McCarthy, EPA Assistant Administrator for Air, dated May 18, 2011.
[28] Since EPA proposed Subcategory 2 for Hg for 30 units designed for coal < 8,300 Btu/lb, the number of units in Subcategory 1 for Hg is equal to 1,091 – 30 = 1,061.

I began my analysis with EPA's latest (revised) spreadsheet, *floor_analysis_coal_051811.xlxs*. Using another spreadsheet file recently made available by EPA, *PartII_III_Hg.xlxs*, I was able to add units that EPA ranked 41st through 127th to the above-reference EPA spreadsheet. In terms of "numbers," I expanded the Hg MACT pool from 40 to 127 units[29] and the number of test averages from 80 to 263. Next, I replicated EPA's UPL approach exactly (i.e., computed $\bar{x}$ as the average of each unit's minimum test average and computed the variance of 263 test averages). I obtained a UPL = 2.10 lb/10^{12} Btu.

There are probably many variants to EPA's approach. A rather obvious approach would be to compute $\bar{x}$ based on the average emission for each of the 127 units in the MACT pool and to compute the variance of all the test values in the pool. This approach yields a higher UPL = 2.24 lb/10^{12} Btu. I have never been entirely comfortable with EPA's ranking and selecting the MACT pool units based on minimum test values. I believe most statisticians would say that extreme values (i.e., minimums or maximums) are the most uncertain measurements in a dataset. Accordingly, I computed the mean of all available Hg test averages, resorted the units based on mean Hg emissions and selected the best performing 127 units. Next, I computed $\bar{x}$ as the average of each unit's mean test average and computed the variance of 266 test averages). I obtained a UPL = 1.42 lb/10^{12} Btu. I believe this is the single, most statistically defensible results that can be obtained from the Hg ICR data.

It is interesting to observe that the last UPL result presented (1.42) does not have the minimum value for $\bar{x}$, but it does contain the minimum variance component. By taking the average before the units are ranked, I believe one is able to minimize the impact of some unusually low Hg values, which may not be able to be replicated. A unit with an unusual and perhaps randomly low Hg measurement gets placed in the pool, but subsequent tests are higher and more variable. I believe this is the fundamental reason why a pool based on minimum measured values does not guarantee or necessarily produce the minimum UPL result.

Lastly, UARG asked if I could provide any alternative results for EPA's Hg MACT pool, which only contains the lowest 40 emitting units. One obvious alternative would be to compute $\bar{x}$ from the mean of each unit's test averages (as opposed to the minimum test average) and to compute the variance from the same 80 test averages that EPA used. Using this approach, I calculate a UPL = 1.30 lb/10^{12} Btu.

VI. EPA'S APROACH TO NEW UNITS SHOULD BE OVERHAULED

For the reasons provided in the memorandum, I believe that the new unit emission limits in EPA's proposed MACT rule will mean that no new coal-fired EGUs can be constructed in the United States. There are several reasons that lead to this conclusion. First, as discussed in detail below, EPA's reliance on a "Franken-Plant" approach has produced a suite of emission limits that no existing unit either meets or can meet. Second, EPA's reliance on its UPL approach for new units fails miserably to account for variability. For new units, EPA has selected the unit with the lowest test average for each HAP and then used only three test runs to account for

[29] The pool size equal 127 is determined by taking 12 percent of the coal-fired fleet of 1,091 after subtracting the 30 units designed for burning coal with a heating value < 8,300 Btu/lb; that is, 0 .12(1,091 – 30) = 127.

variability. Three test runs, conducted at most over three consecutive days under basically identical operating conditions and probably lasting no more than 4 hours per test run. Does EPA seriously believe a facility that will operate for at least 40 years over a myriad of operating variables (e.g., seasonal challenges, varying loads, upset conditions, etc.) and probably receive coal from a numerous suppliers can adequately be characterized with three 4-hour snapshot tests taken over at most a 72-hour period?

Franken-Plant Approach

Just as Dr. Frankenstein's fictitious monster bore no resemblance to an actual human being, EPA's "Franken-Plant" bears no resemblance to the actual, best performing EGU. EPA's approach for setting emission limits for existing sources is equally flawed to the Agency's approach for setting emission limits for new sources. However, it is easier to demonstrate and comprehend the Agency's error for new sources. The relevant statutory provision is, *the maximum degree of reduction in emissions that is deemed achievable for new sources in a category or subcategory shall not be less stringent than the emission control that is achieved in practice by the best controlled similar source.*[30] Note that the statue refers to a single source – not multiple sources. If Congress had intended for EPA to set emission limits based on a "Franken-Plant" approach, the statue would read -- *best controlled similar sources.*
Using three EPA spreadsheets that the Agency posted on one of its web sites,[31] it is fairly straightforward to determine which individual unit EPA used to set the MACT floor for new units. Those units are listed in Table 1. From Table 1, it should be obvious that no existing unit meets all of the proposed emission limits for a new EGU.

Table 1. EPA's Franken Plant Approach For New Units.

Pollutant	Facility	99% UPL (lb/MWh)	Total Metal Ranking
Total PM	AES Hawaii	0.049	11th
Total Metals	Cedar Bay Unit A	3.3×10^{-5}	--1st --
Antimony (Sb)	AES Hawaii Unit 2	7.6×10^{-8}	11th
Arsenic (As)	Oak Grove Unit 1	1.6×10^{-7}	104th
Beryllium (Be)	Chamber Cogen Unit 2	2.2×10^{-8}	7th
Cadmium (Cd)	Walter Scott Unit 4	3.7×10^{-7}	3rd
Chromium (Cr)	PSEG Mercer Unit 1	1.7×10^{-5}	56th
Cobalt (Co)	Cholla Unit 3	7.2×10^{-7}	62nd
Lead (Pb)	Oak Grove Unit 1	8.8×10^{-7}	104th
Manganese (Mn)	Weston Unit 4	3.1×10^{-6}	3rd
Nickel (Ni)	Weston Unit 4	3.2×10^{-6}	3rd
Selenium (Se)	PSEG Mercer Unit 1	2.5×10^{-5}	56th
Hydrogen Chloride (HCl)	Logan Unit 1	2.6×10^{-4}	n/a

[30] 42 U.S.C. §7412(d)(3), emphasis added.
[31] See, *floor_analysis_coal_pm_031611.xlsx, floor_analysis_coal_hcl_031611.xlsx, and floor_analysis_coal_hg_051811.xlsx.*

Mercury (Hg)	Nucla Unit 1	1.7×10^{-5}	n/a

In the above table, we included dotted horizontal lines to differentiate among the optional emission limits for non-Hg metallic HAPs. The mercury and hydrogen chloride limits must be complied with individually, independent of the option chosen for non-mercury metallic HAPs. Table 1 clearly demonstrates the fallacy in the EPA's "Franken-Plant" approach. Consider the following EPA statements:

> *For the non-Hg metallic HAP, we chose to use PM as a surrogate. Most, if not all, non-Hg metallic HAP emitted from combustion sources will appear on the flue gas fly-ash. Therefore, the same control techniques that would be used to control the fly-ash PM will control non-Hg metallic HAP.*[32]

Oak Grove Unit 1 is EPA's basis for two individual metallic HAP emission limits, arsenic and lead. The key question is how or what control technology could the Oak Grove owners add to meet the other metallic HAP limits given (1) EPA's statement that the same control techniques that work for fly-ash PM also work for non-Hg metallic HAPs and (2) Oak Grove is already the best performing unit for not one but two non-Hg metallic HAPs. EPA's rejoinder to my comment may very well be that compliance with the individual non-Hg metallic HAP limits is an option and not a requirement. This is an inadequate response and misses the point. EPA should not be permitted to base a portion of a suite of emission limits upon the performance of a single unit when that same unit cannot comply with the other enforceable components of that same suite of emission limits. Moreover, the unit that formed the basis for one of the regulated HAPs (e.g., total PM) may not meet one of the other mandatory limits (e.g., HCl). AES Hawaii Unit 1 is the basis of the new unit total PM limit, but the lowest HCl emission results reported in the ICR data are approximately 50 times higher than the proposed new unit HCl limit.[33] Nucla Unit 1 is the basis of the new unit total Hg limit, but the total PM results reported in the ICR data are almost an order of magnitude higher than the proposed new unit total PM limit.

Speaking of AES Hawaii, it is the only coal-fired plant in Hawaii, and the plant has a generating capacity of 180 MW. Unit 1 is only capable of supplying one-half of the steam required by the 180 MW turbine/generator, so Unit 1 in effect has a capacity of 90 MW. The AES Hawaii unit burns coal, which is imported from Indonesia. To supplement the imported coal, the unit also burns old tires, used motor oil, and carbon from Board of Water Supply filters. While EPA is mandated to set limits for new sources based on the *maximum degree of reduction in emissions that is deemed achievable*, such degree of emission control must be *achieved in practice by the best performing similar source*. It is quite clear there is not or most probably will not be another similar source to AES Hawaii in the continental United States. Beyond the *similar source* issue, I believe EPA made a computational error in converting the AES Unit 1 total PM results from input units (lb/10^6 Btu) to output-based units (lb/MWh). EPA mistakenly assumed that both AES units have a capacity of 180 MW; in point of fact, the capacity of the two-unit plant is 180 MW. This error is easily verified in EPA's spreadsheet, because it shows Unit 1 to have a heat

[32] 76 Fed. Reg. 25,039, col. 3 (May 3, 2011).
[33] EPA's proposed HCl limit for new units is 0.0003 lb/MWh; the minimum ICR test result for AES Hawaii Unit 1 is equal to 0.0146 lb/MWh.

rate of 5.03 million Btu per MWh, when the correct value is exactly twice that or 10.06 million Btu per MWh. When I correct the heat rate or conversion error in three individual total PM runs and simply repeat EPA's UPL calculation, I obtain 0.10 lb/MWh. Even as an unrepresentative unit that AES Hawaii may be, it does not support a total PM emission limit of 0.05 lb/MWh.

Alternative Approach for New Units

Over the last several years, I have had the opportunity to consult for several potential EGU owner/operators that now find themselves in a quandary. I am referring to projects that have worked through the regulatory maze (e.g., obtained the necessary permits to construct such as prevention of significant deterioration (PSD) permits) and often withstood legal challenges (e.g., administrative law judge hearings) but now are faced with proposed MACT emission limits that cannot be achieved. Two projects with which I am familiar (i.e., Sunflower's Holcomb 2 and LP Power's Longleaf Plant) are both permitted as non-major (area) sources of HAPs. Since these two projects represent significant generating capacity (895 and 1,200 MW, respectively), we know the emission limits contained in the respective permits must be very low.

Before finalizing the EGU MACT limits for new units, I believe EPA should review the emission limits contained in these various state-of-the-art power projects. I understand that legal opinions exist that permitted emission limits do not equate to "achievable" emission limits. Likewise, I do not believe emission limits like those contained in EPA's proposed EGU MACT rule for new units, which control equipment suppliers cannot and will not guarantee, satisfy the definition "achievable" emission limits either. The simple fact is that after years of negotiation and/or litigation, the emission limits contained in these permits reflect the best level of performance that existing control technology can achieve. In other words, if lower emission limit guarantees were available from the equipment suppliers, the permit limits would be lower.

VII. WORK PRACTICE STANDARDS

During the approximate 6-month ICR Notice and Comment period, I was frequently critical of EPA's approach for dioxin/furan (D/F) testing and also for the non-D/F organic testing. Based on previous experience with the EPRI Power Plant Integrated Systems Chemical Emission Studies (PISCES), I believed that organic emissions from EGUs are predominantly below the method detection limits. Our experience also indicates that organic compounds that are measured at or above detection limits are often the result of contamination. Most of the EPA's stack sampling methods use copious quantities of organic solvents (e.g., acetone, methylene chloride) for recovering samples and/or sample train clean-up; therefore, it can be very difficult to avoid sample contamination when non-organic tests are being conducted concurrent with organic sampling. The ICR Data Review report prepared by EPRI and submitted to EPA concluded that many of the organics measured above the detection limits were likely the result of sample contamination.[34]

---

[34] *Data Quality Evaluation of Hazardous Air Pollutants Measurements for the US Environmental Protection Agency's Electric Utility Steam Generating Units Information Collection Request.* EPRI, Palo Alto, CA: 2010. 1021216.

All of the above said, I commend EPA on what I believe is the only appropriate means to deal with all of the non-detected organic results in the ICR database. EPA correctly concludes that both the D/F compounds and the non-D/F organic HAPs are routinely below the detection limits of the EPA test methods, even when extended sampling times and volumes were implemented. Work practice standards proposed pursuant to section 112(h) is the most effective and efficient approach for addressing organic emissions from EGUs. I should note that I am not providing comments on the specifics requirements of the work practice standards proposed in this rule as that would be outside my area of expertise; however the work practice concept is clearly superior to specific numerical limits.

VIII. CONFUSING COMPLIANCE REQUIREMENTS

The proposed compliance and testing obligations in EPA's proposed EGU MACT rule are among the most confusing I have ever attempted to review. I believe the most likely coal-fired configurations to survive EPA's proposed rule will (1) opt for the total PM limit and utilize PM CEMS for continuous filterable PM measurements; (2) employ wet or dry scrubbing for acid gas control and elect to use already-installed SO_2 CEMS; and (3) directly monitor mercury using either Hg CEMS or sorbent trap monitoring system. For the above-described scenario in which either all of the regulated HAPs or their surrogates are continuously monitored, **no** additional burdens (e.g., coal sampling, operating parameter limits, etc.) should be imposed. I would like to believe EPA understands this logic, but the manner in which the proposed rule is drafted is very unclear. For example, proposed §63.10007 (c) states:

> (c) *You must conduct each performance test under the specific conditions listed in Tables 5 and 7 to this subpart. You must conduct performance tests at the maximum normal operating load while burning the type of fuel or mixture of fuels that has the highest content of chlorine, fluorine, non-Hg HAP metals, and Hg, and you must demonstrate initial compliance and establish your operating limits based on these tests.*[35]

While proposed §63.10007 (c) continues beyond the above-quoted passage, there is no indication but what every affected facility is subject to the requirements. How any owner/operator is supposed to able to divine that "fuel with the highest content of chlorine, fluorine, non-Hg HAP metals, and Hg" is being burned during the performance test is not explained. However, a more important point is that such clairvoyance is not needed when either all of the regulated HAPs or their surrogates are continuously monitored.

Proposed §63.10008 lays out, in gory detail, a fuel sampling and analysis regime that requires two full columns in the Federal Register. Granted, proposed §63.10008(a) contains an *as applicable clause*, but its meaning is never explained. However, using the same logic as in my previous paragraph, the fuel information is neither meaningful nor relevant when either all of the regulated HAPs or their surrogates are continuously monitored.

---

[35] 76 Fed. Reg. 25,107, col. 3 (May 3, 2011).

IX. PM CEMS DILEMMA

EPA's proposed rule would require the use of PM CEMS for continuous compliance if the owner/operator elects the total PM option instead of complying with either the total or individual non-Hg HAP metal limits. I term this a dilemma because I believe the PM CEMS approach is much less burdensome than frequent stack testing coupled with rigid operating parameter limits that are mandated for either of the other two options (i.e., total or non-Hg individual HAP metals). The dilemma results from the status of PM CEMS technology. That is, commercially available continuous PM monitors, especially those that are based on the principal of light scattering, do not provide a direct measure of PM emissions. By direct measure, I mean that the instrument should measure the mass of PM and the volume of flue gas from which that mass of PM was sampled. Instead, EPA Performance Specification 11 (PS-11) requires that PM CEMS be correlated to a series of manual stack testing results, and the permissible "tolerance" between PM CEMS readings and corresponding stack test values allowed by PS-11 is significant. I have experience with some 20 or so PM CEMS installations on coal-fired EGUs. In general, complying with the initial PM CEMS certification (PS-11) requirements has not proven overly difficult. However, unlike SO_2 and NO_X monitoring systems, the utility industry has little, if any, experience with PM CEMS as continuous compliance devices. Because of this inexperience, it is difficult to forecast the level training and quality assurance support that any successful PM CEMS compliance monitoring program may require.

Measurement Issues During Start-up and Shutdown

Of course, this PM CEMS technology limitation poses another major problem with respect to start-up and shutdown emissions. That is, PM CEMS are not correlated during such times because, among other things, EPA typically requires normal, full-load operation for CEMS certification testing. Moreover, EPA manual test methods like the ones used for PM CEMS correlation testing (e.g., Method 5) were never designed to be used during transient plant operations. Such methods were designed and, if validated, were done so only during steady-state plant operation.

EPA did take a positive step by proposing the use of a 10 percent O_2 diluent cap during start-up and shutdown periods.[36] The proposed diluent cap will prevent some hourly PM values, when calculated in the unit of standard ($lb/10^6$ Btu), from being recorded at extremely high and unrepresentative levels. Unfortunately, EPA's proposed diluent cap does not address the underlying uncertainty and potential inaccuracy. Another constructive step EPA could consider in dealing with start-up and shutdown events (especially since the rulemaking record is silent with respect to any Agency data analysis pertaining to impact of such events) is to weight the hourly $lb/10^6$ Btu PM values by heat input. In other words, each hourly PM value in the 30-day rolling average would be multiplied times the heat input for that hour and the hourly summation

[36] 76 Fed. Reg. 25,106, col. 2 (May 3, 2011). Note that the overwhelming majority of gaseous CEMS currently installed and operating on EGUs measure CO_2 instead of O_2. Most, if not all, EGUs utilize a dilution/extractive technology, which has proven to be reliable and accurate under EPA's Part 75 Acid Rain Monitoring Program. However, it is not feasible to monitor O_2 with a dilution/extractive system because the diluent air contains 20.9 percent O_2 and would overwhelm the amount of O_2 in the sample extracted from the stack.

divided by the total heat input during the 30-day rolling average period. Weighting by heat input would help to normalize the impact of any potentially elevated PM concentrations during start-up events because heat inputs are necessarily lower during start-up events, and the actual mass of PM emitted is also lower because of a reduced firing rate. Emission weighting as suggested above is consistent with the rationale EPA provided in the preamble to proposed revisions to Subpart Da new source performance standards to address that very same start-up and shutdown issues.[37]

Continuous Compliance Issues

To my knowledge, no organization (e.g., EPA, EPRI, ASTM, etc.) have completed any studies that would inform a decision relative to using PM CEMS for determining compliance. However, EPRI has initiated a Tailored Collaboration (TC) project titled, "Comprehensive Continuous PM Monitoring Field Study." In November 2010, the EPRI project team installed five individual PM CEMS on a coal-fired power plant stack in the upper Midwest. The project scope of work includes both initial and final PM correlation testing, perhaps 6 to 8 months apart. The initial correlation test was conducted the first week in February 2011. In between the two correlation test periods, the EPRI team will monitor and document the operability and reliability of each PM CEMS and also evaluate potential alternative calibration techniques. The output of this project will certainly update and inform our knowledge of PM CEMS. Unfortunately, I do not expect those results will be available until the end of 2011.

Conflicting Requirements

Of course, for PM CEMS to be a viable option, EPA must straighten out the disarray of the PM CEMS approach contained in the proposed rule. We begin this discussion with contradictory language, which is probably a drafting error resulting from a hurried rulemaking schedule. Proposed §63.10011(d) states, *the operating limit will be the average of the PM filterable results of the three Method 5 performance test runs.* Meanwhile, Table 3, Line 8 states, *maintain the PM concentration (mg/dscm) at or below the highest 1-hour average measured during the most recent performance test demonstrating compliance with the total PM emission limitation.* This inconsistency can be resolved quickly; we assume EPA meant for the operating limit to be based on the highest 1-hour value recorded by the PM CEMS during the most recent performance test. However, resolving the inconsistency only solves a minor part of the PM CEMS problem. As noted above, PS-11 requires that the PM CEMS output be correlated to a series of manual stack testing runs (minimum of 15 is required). Two of the principal "performance" criteria are the widths of the confidence interval and the tolerance intervals about the correlation line that is drawn. The widths of both acceptable intervals are tied to the PM emission limit; the confidence interval (95%) must be within 10% of the PM emission limit and the tolerance interval must have 95% confidence that 75% of all possible values are within 25% of the PM emission limit. It is not clear how one would demonstrate compliance with PS-11 in the absence of a defined filterable PM emission limit. Thus, the PM CEMS compliance approach as proposed by the Agency is simply unworkable and internally incongruent with existing EPA requirements (i.e., PS-11).

[37] 76 Fed. Reg. 25,061, col. 3 (May 3, 2011).

X. CONCLUSIONS

Given all of the data collected and submitted to EPA by the utility industry, it is unfortunate that EPA management committed to a politically expedient rulemaking schedule and failed to allow EPA staff sufficient time to develop a quality proposed EGU MACT rule. The hurried and committee-like approach required to get a proposed rule signed by 3/16/2011 produced a document with numerous errors and inconsistencies – many of which have been discussed in this memorandum.

EPA's proposed approach to regulate total PM and then establish "custom" filterable PM limits based on performance testing is nonsensical and must be fixed. While I remain cautiously optimistic that EPA will significantly revise its PM approach between the proposed and the final rule, it is difficult to guess what the Agency may ultimately decide. Likewise, EPA must either significantly revise its approach to setting limits for new units or accept that this Administration has brought an end to constructing new, coal-fired EGU capacity in the United States. It has been frustrating trying to preparing meaningful comments on this proposed rule, because numerous revisions and/or corrections are warranted and there is perhaps limited opportunity to respond to those changes.

Table A-1. PM CEMS Emission Summary for Oak Creek Unit 7
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.004	<0.006	<0.008	<0.010	<0.012	<0.014
3-Hr Rolling	60.2%	90.2%	96.9%	97.8%	98.1%	98.6%
8-Hr Rolling	59.6%	89.9%	96.7%	97.6%	98.3%	98.8%
24-Hr Rolling	57.1%	89.8%	96.4%	97.4%	97.9%	99.2%
Daily	56.8%	89.5%	96.5%	97.5%	98.1%	99.4%
Weekly	52.9%	88.2%	94.1%	98.0%	98.4%	98.0%
30-Day Rolling	53.3%	85.1%	90.5%	96.4%	99.8%	99.8%
Long-Term Hourly Mean = 0.004 lb/10^6 Btu						

Table A-2. PM CEMS Emission Summary for Oak Creek Unit 8
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.007	<0.009	<0.011	<0.013
3-Hr Rolling	66.0%	82.7%	91.2%	95.5%
8-Hr Rolling	66.0%	83.9%	92.4%	96.2%
24-Hr Rolling	66.6%	86.0%	93.2%	97.2%
Daily	66.0%	85.3%	92.8%	96.8%
Weekly	66.7%	86.1%	95.8%	100%
30-Day Rolling	65.9%	92.7%	100%	100%
Long-Term Hourly Mean = 0.007 lb/10^6 Btu				

Table A-3. PM CEMS Emission Summary for Cross Unit 1
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.012	<0.015	<0.018	<0.020	<0.022	<0.025
3-Hr Rolling	68.5%	91.1%	96.7%	97.7%	98.1%	98.5%
8-Hr Rolling	68.1%	90.8%	96.3%	97.5%	97.9%	98.2%
24-Hr Rolling	66.6%	89.7%	95.7%	96.4%	96.8%	97.6%
Daily	66.8%	90.9%	96.2%	96.5%	96.9%	97.9%
Weekly	65.2%	87.0%	93.5%	93.4%	93.5%	93.5%
30-Day Rolling	58.5%	82.0%	82.0%	90.0%	90.3%	93.4%
Long-Term Hourly Mean = 0.012 lb/10^6 Btu						

Table A-4. PM CEMS Emission Summary for Clover Unit 1
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.002	<0.003	<0.004	<0.005	<0.006
3-Hr Rolling	26.8%	89.3%	96.0%	99.3%	99.6%
8-Hr Rolling	25.7%	89.8%	95.8%	99.1%	99.4%
24-Hr Rolling	24.0%	90.7%	95.4%	99.1%	99.2%
Daily	24.3%	90.8%	95.7%	99.3%	99.7%
Weekly	21.7%	89.1%	95.7%	100%	100%
30-Day Rolling	16.2%	94.0%	100%	100%	100%
Long-Term Hourly Mean = 0.002 lb/10^6 Btu					

Table A-5. PM CEMS Emission Summary for Clover Unit 2
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.002	<0.003	<0.004	<0.005
3-Hr Rolling	89.1%	99.6%	99.9%	99.9%
8-Hr Rolling	89.7%	99.7%	99.8%	99.8%
24-Hr Rolling	90.0%	99.6%	99.8%	99.8%
Daily	89.2%	99.4%	99.7%	100%
Weekly	94.3%	100%	100%	100%
30-Day Rolling	98.9%	100%	100%	100%
Long-Term Mean = 0.002 lb/10^6 Btu				

Table A-6. PM CEMS Emission Summary for Spurlock Unit 1
(Percent of Averages Less Than Each Numerical Limit)

Averaging Time	<0.003	<0.004	<0.005	<0.006
3-Hr Rolling	90.3%	99.6%	99.6%	99.7%
8-Hr Rolling	96.4%	99.3%	99.3%	99.4%
24-Hr Rolling	97.7%	98.7%	98.8%	99.0%
Daily	97.6%	98.8%	99.1%	99.1%
Weekly	98.0%	100%	100%	100%
30-Day Rolling	100%	100%	100%	100%
Long-Term Mean = 0.002 lb/10^6 Btu				

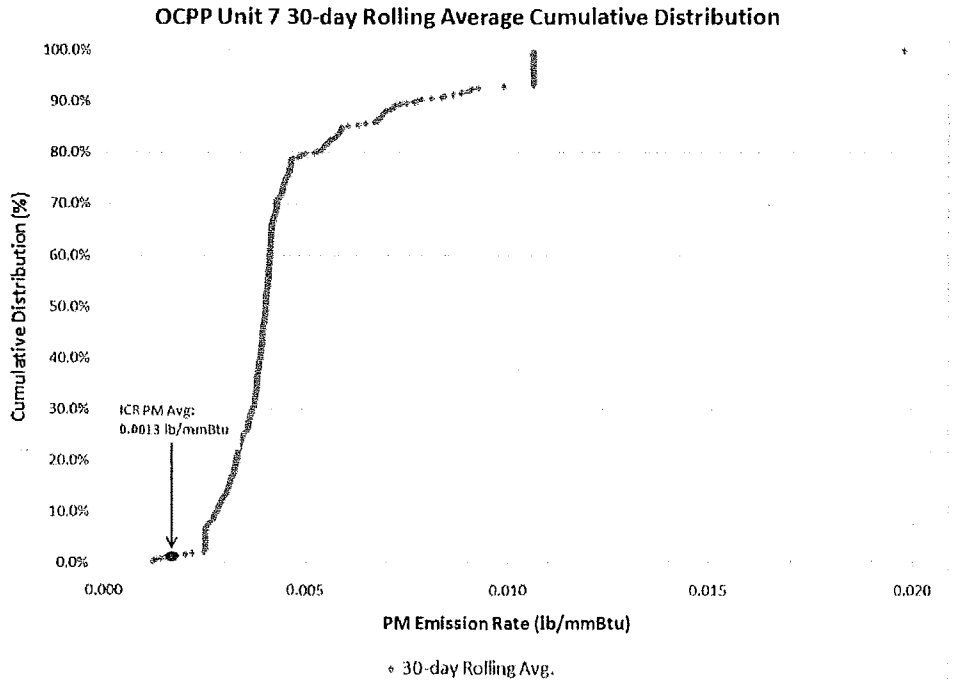


Figure B-1: Oak Creek Unit 7 30-Day Rolling Average Cumulative Distribution Plot

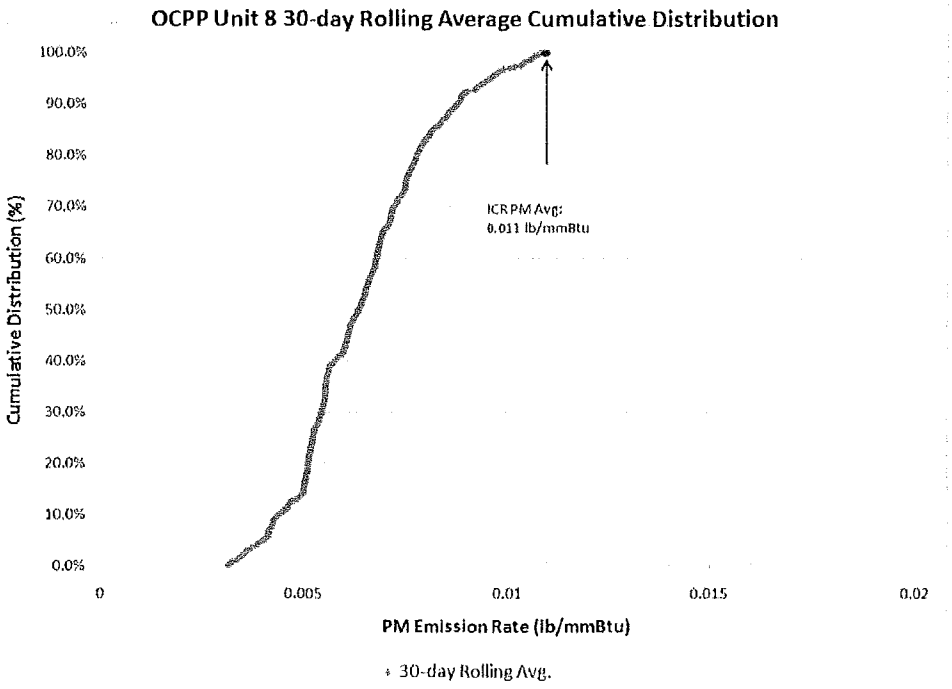


Figure B-2: Oak Creek Unit 8 30-Day Rolling Average Cumulative Distribution Plot

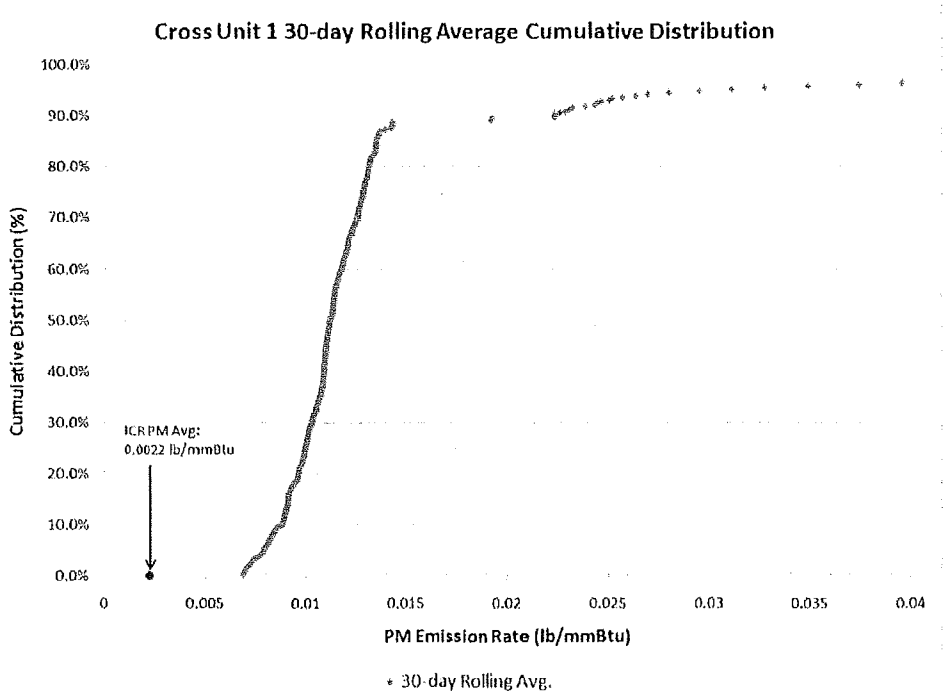


Figure B-3: Cross Unit 1 30-Day Rolling Average Cumulative Distribution Plot

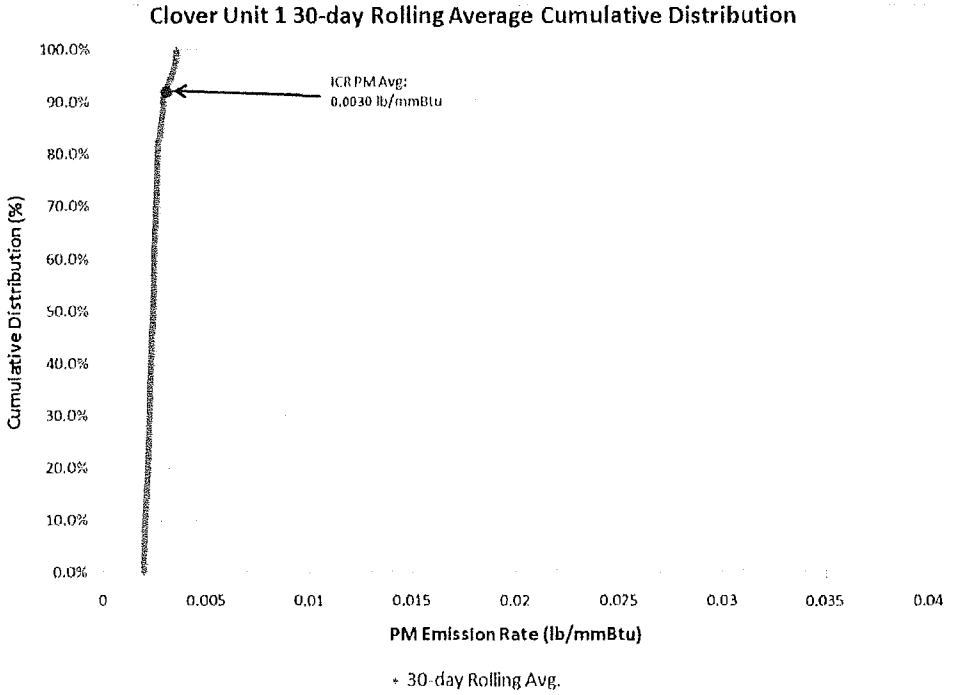


Figure B-4: Clover Unit 1 30-Day Rolling Average Cumulative Distribution Plot

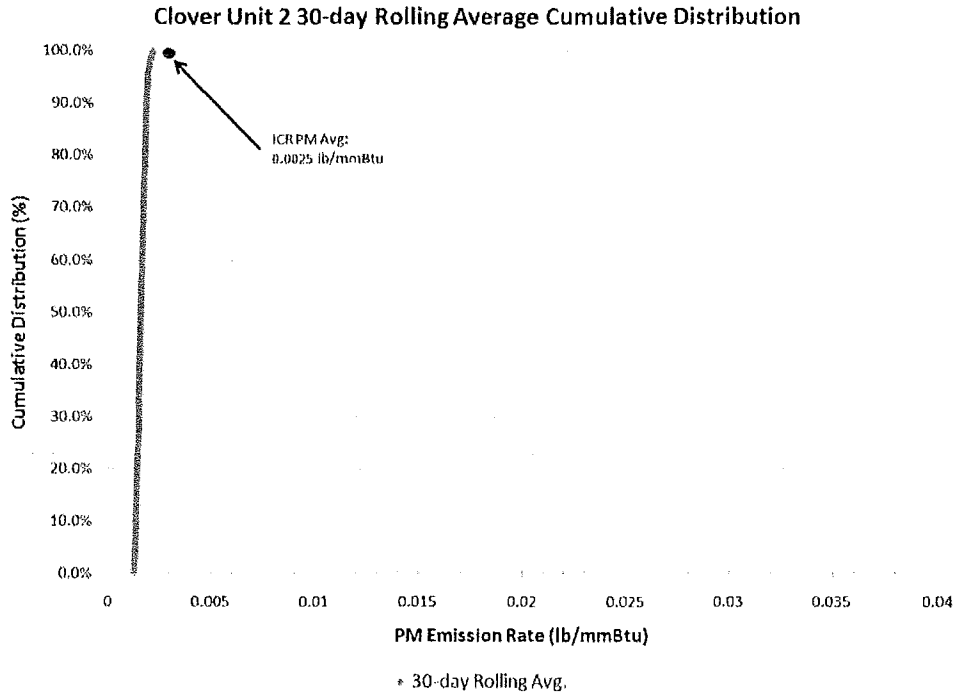


Figure B-5: Clover Unit 2 30-Day Rolling Average Cumulative Distribution Plot

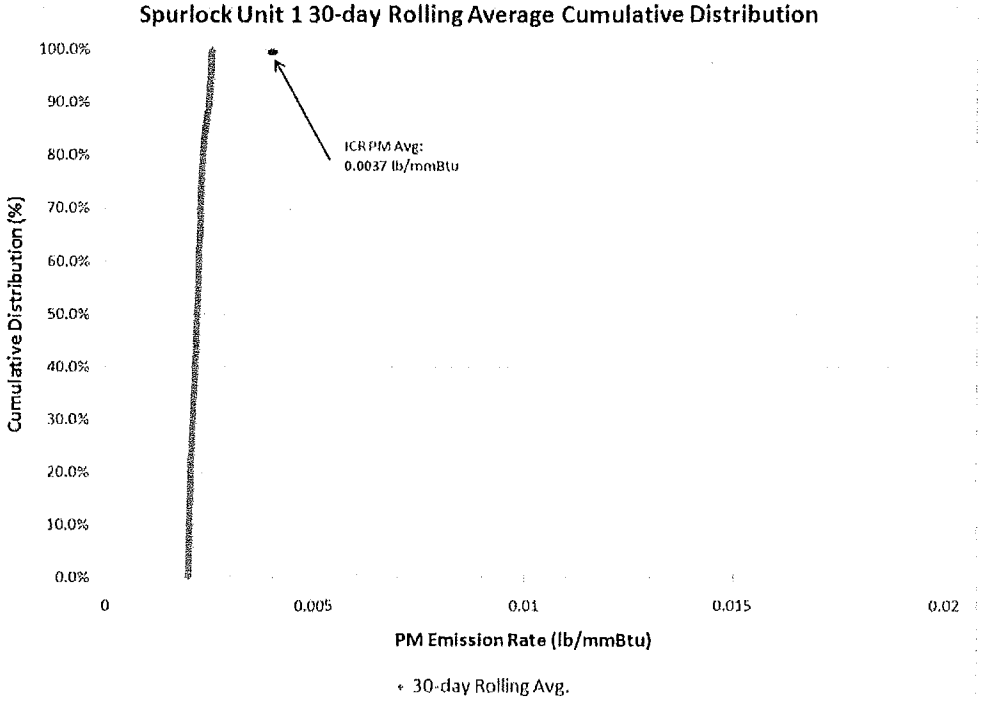


Figure B-6: Spurlock Unit 1 30-Day Rolling Average Cumulative Distribution Plot