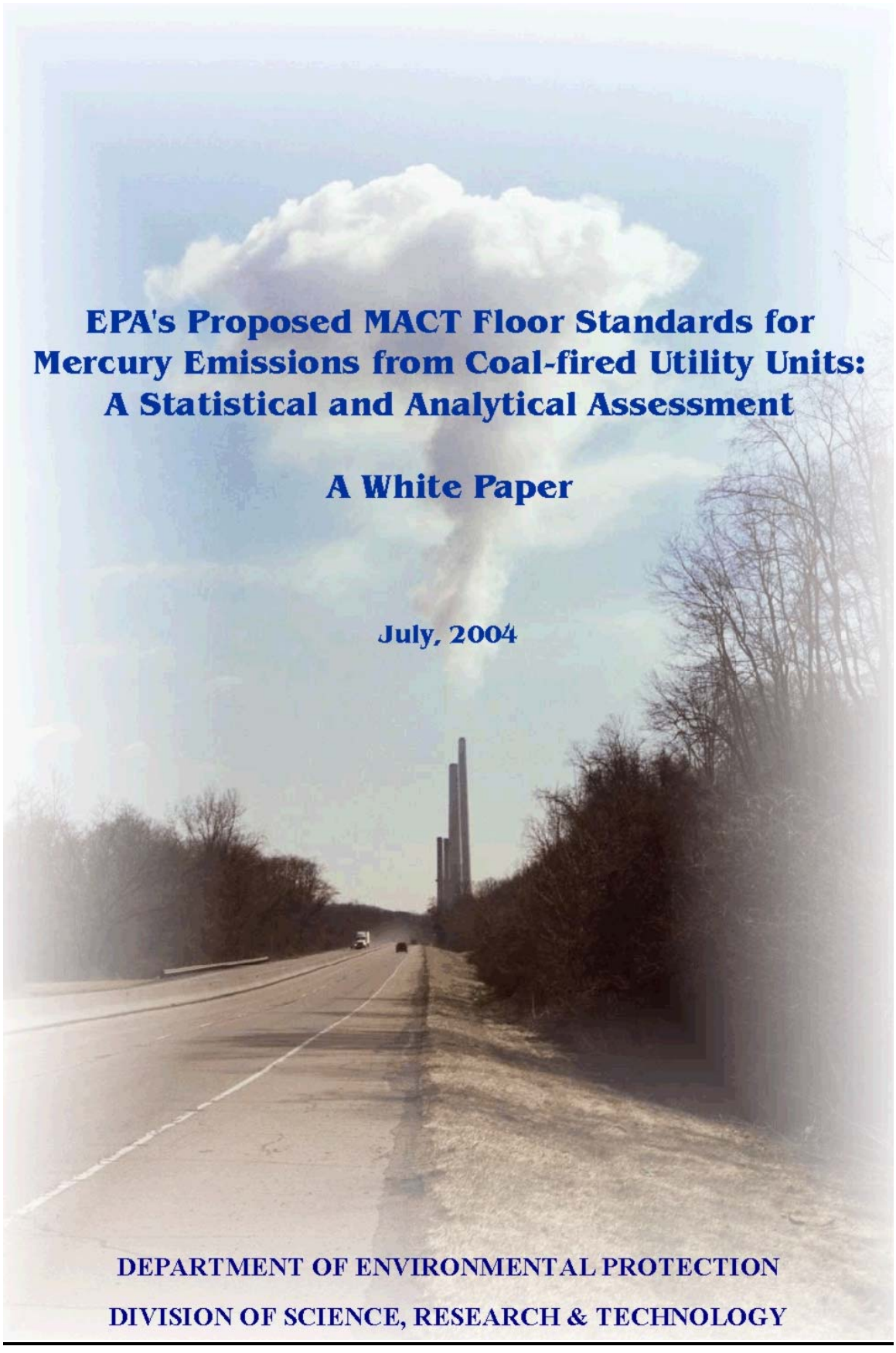


EPA's Proposed MACT Floor Standards for Mercury Emissions from Coal-fired Utility Units: A Statistical and Analytical Assessment

A White Paper

July, 2004

DEPARTMENT OF ENVIRONMENTAL PROTECTION

DIVISION OF SCIENCE, RESEARCH & TECHNOLOGY

EPA's Proposed MACT Floor Standards for Mercury Emissions from Coal-fired Utility Units
A Statistical and Analytical Assessment

A White Paper

July 2004

Prepared By:

Michael Aucott and Leo Korn

DEPARTMENT OF ENVIRONMENTAL PROTECTION

DIVISION OF SCIENCE, RESEARCH & TECHNOLOGY
PO Box 409
Trenton, NJ 08625-0409

(609) 633-6070

Table of Contents

Table of Figures

EPA's Proposed MACT Floor Standards for Mercury Emissions from Coal-fired Utility Units: A Statistical and Analytical Assessment

Introduction

On January 30, 2004, the United States Environmental Protection Agency (EPA) proposed Maximum Achievable Control Technology (MACT) standards for mercury emissions from coal-fired utility units. This white paper reviews statistical and analytical approaches that EPA applied as it developed the proposed MACT standards.

The Clean Air Act requires a MACT standard for existing sources of air pollution must be no less stringent than "the average emission limitation achieved by the best performing 12% of the existing sources" for which EPA has emissions information.[1] However, EPA did not take this straightforward approach. Instead, EPA stated that "it is essential that EPA be able to identify and quantify the level of variability" in emissions before it could determine what existing sources were actually achieving,[2] and attempted to develop a proposed standard with additional accommodations for variability.

The importance of variability of emissions depends on whether and to what extent variability could be controlled. Past experience with municipal solid waste incinerators showed that those sources could reduce their mercury emissions by 90% or more, even though their fuel (solid waste) tends to vary dramatically.

The importance of variability of emissions also depends on whether a standard could be crafted so that transient peaks in emissions would not automatically lead to violations. If compliance depended upon the results of a single stack test, a way of addressing variability would indeed be essential. However, if compliance depended upon an average of test results over a longer period, the importance of an isolated test that happened to capture peak emissions would be lessened.

Proceeding from the assumption that it was essential to identify and quantify variability in emissions before setting a MACT standard, EPA employed a series of statistical and analytical approaches to determine variability. Applying those approaches, EPA determined that after accounting for potential variability the average mercury emissions from the best performing 12% of existing sources were much higher than what those sources actually reported.

Specifically, the average emission rate of the best 12% of the bituminous units that were stack tested under EPA's Information Collection Request was 0.12 pounds per trillion Btu (lb/TBtu). To arrive at its proposed standards, EPA went beyond the actual stack test data. First, it estimated what mercury emissions might have been if many more stack tests had been conducted. Second, EPA made an upward adjustment of those estimates of what the mercury emissions might have been, based on the results of a series of analyses of the coal burned in the best performing units. Third, EPA used those adjusted estimates to estimate maximum emission rate expected to be exceeded only very rarely

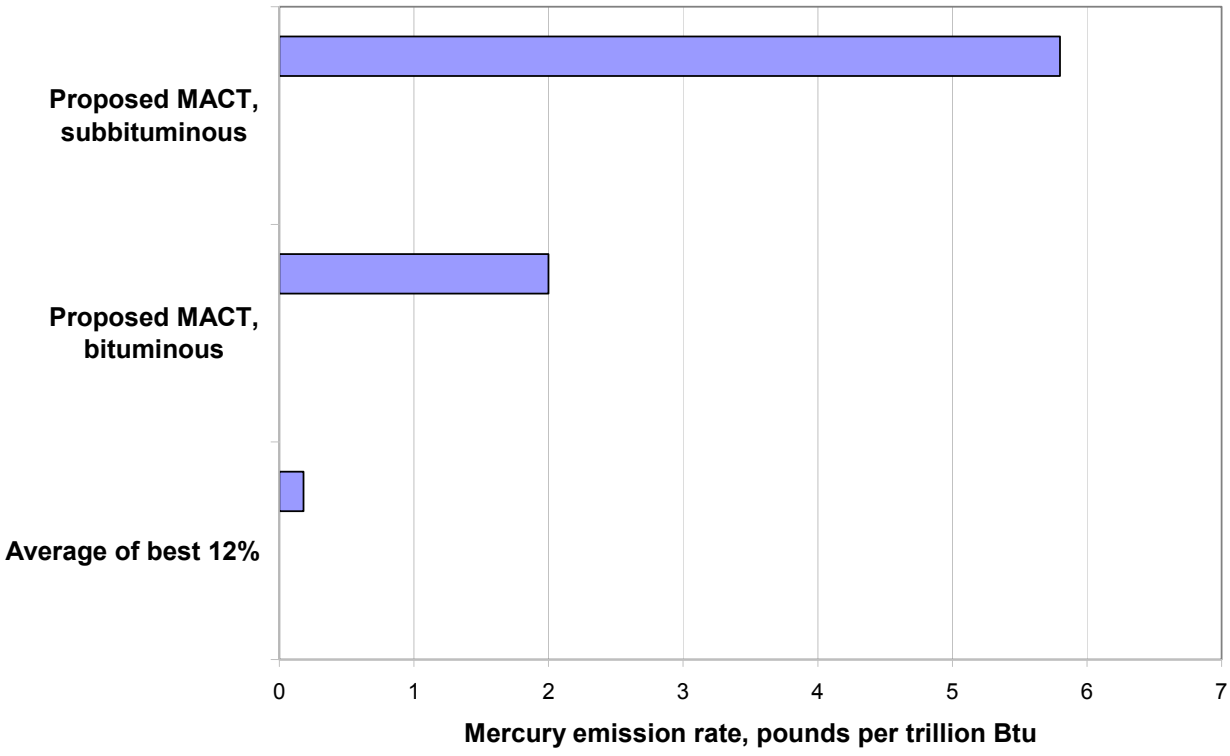
[1] Clean Air Act, Section 112(d)(3)(A).
[2] 69 F.R. 4670, January 30, 2004.

(no more than 2.5% of the time), and concluded that this estimate (the 97.5th percentile value) was representative of how the unit would perform under conditions reasonably expected to occur. Finally, EPA raised the number higher by calculating a 97.5% confidence level upper limit for the average of the 97.5th percentile values.

As a result, the proposed MACT standards would allow substantially greater emissions of mercury than what the best performing 12% are achieving today. EPA proposed a MACT standard more than 30 times greater than 0.18 lb/TBtu average emission rate that the best performing 12% achieved in practice: 5.8 lb/TBtu for units currently firing subbituminous coal, or for units that switch to subbituminous coal from bituminous coal. For units that continue to fire bituminous coal, EPA proposed a standard of 2.0 lb/TBtu, more than 11 times greater.

The following chart illustrates the difference between what the best performing 12% are achieving today and the standards that EPA proposed.

Figure 1. Resizing MACT; bituminous and subbituminous coal-fired utility units



In the pages that follow, the numerous statistical and analytical difficulties with the approaches that EPA used are discussed in more detail, and related actions taken by EPA are noted. With each approach or action, associated difficulties, including questionable assumptions, inappropriate statistical operations, or failure to consider important aspects, are described.

I. In concluding that it was essential to identify and quantify variability, EPA failed to consider data on management of mercury variability from operating facilities.

In its variability, coal is not unlike other mercury-containing fuels such as petroleum and municipal solid waste (MSW). MSW incinerators in New Jersey experience variations of inlet mercury concentrations spanning a range greater than a factor of 100. Nonetheless, New Jersey's five MSW incinerators have been controlling mercury emissions with carbon injection since 1994.

The experience of these MSW incinerators at least raises the possibility that mercury variability can be managed successfully. EPA did not seek actual emissions data to evaluate this possibility. Instead, the assumption that variability was key led EPA instead to embark on a path involving a variety of statistical and analytical manipulations.

II. Variability is best addressed by tailoring the form of the standard to account for variability.

EPA has previously developed effective standards to regulate emissions that derive from variable constituents. Such standards have the form of a concentration limit and percent reduction limit, where the facility may choose the less restrictive. The New Source Performance Standard (NSPS) for sulfur dioxide from electricity generating units (EGUs) has the form of "x lbs. per million Btu or y percent reduction." The mercury limit for MSW incineration has the form of "w micrograms per dry standard cubic meter or z percent reduction." This form allows the concentration limit to be based on the average level of the constituent, because the percent reduction limit can be used for situations where the constituent is much higher than the average.

III. EPA attempted to quantify variability in <u>short-term peak</u> emissions, even though the essential question for compliance with EPA's proposed standards is variability in <u>long-term average</u> emissions.

If EPA had proposed a standard that based compliance on a unit's short-term mercury emissions, it would be relevant to identify and quantify the worst case short-term emission rate for all units affected by a MACT standard, especially if there were no reason to believe that variability could be managed successfully and if it were impossible to use a form of a standard that accounted for variability.

However, EPA proposed MACT standards based on a 12-month rolling average. If it were indeed essential to identify and quantify variability in emissions, the relevant measure of variability would be what was experienced over 12 months, not what was experienced in any single stack testing event. A transient peak in emissions during the 12-month period will not cause noncompliance with the standard, because lower emissions during the rest of the period will average out the effect of the peak. For that reason, using the short-term worst-case scenario to set an annual standard immediately makes the standard unnecessarily lax. If compliance is to be based on a long-term

average, EPA should be estimating long-term averages, not worst case emissions, to determine the MACT floor.

IV. Although EPA relied heavily on the chlorine content of coal in estimating mercury emissions, there is no evidence of a significant correlation between the chlorine content of a sample of coal and the ability to control mercury emissions when that sample is burned.

EPA considered several studies of variability and how best to accommodate it. These include reports by Research Triangle Institute (RTI)[3], Science Applications International Corporation (SAIC)[4] and a study prepared for WEST Associates (WEST) by ENSR Corporation.[5] This latter report appears to have been used, in some cases virtually word for word, as the basis for the approach presented in EPA proposal.

EPA, apparently basing its conclusions on the WEST Associates report, states that a relationship exists between chlorine and the ability to control mercury. More specifically, the agency states that the chlorine content of coal can be used to predict mercury emissions, and that "the higher the Cl:Hg ratio, the more likely the formation of mercuric chloride (ionic or oxidized Hg) that is more readily captured by existing controls," and "This Cl:Hg ratio is independent of the coal rank as an indicator of Hg controllability."[6]

Analysis of ICR III[7] data by the NJDEP suggests that a relationship between mercury control and the portion of inlet mercury that is either oxidized or particle-bound is plausible. See Figure 2. However, EPA offers no evidence and cites no sources to demonstrate a significant correlation between the chlorine:mercury ratio and the ability to control mercury emissions. As Figure 3 shows, there is no apparent correlation between the chlorine:mercury ratio of all coals and the portion of the inlet concentration that is oxidized.

---

[3] Cole, Jeffrey, 2002, Memorandum from Jeffrey Cole, Research Triangle Institute (RTI), to William Maxwell, EPA, Statistical analysis of mercury test data variability in support of a determination of the MACT floor for the regulation of mercury air emissions from coal-fired electric utility plants, August 28, 2002.

[4] SAIC, 2003, Calculation of possible mercury MACT floor values for coal-fired utilities: Influence of variability and approach, prepared by Science Applications International Corporation (SAIC), 11251 Roger Bacon Drive, Reston, VA, for U.S. Dept. of Energy, National Energy Technology Laboratory, Pittsburgh, PA, December, 2003.

[5] WEST Associates, 2003, Multivariable Method to Estimate the Mercury Emissions of the Best-performing Coal-fired Utility Units Under the Most Adverse Circumstances which Can Reasonably Be Expected to Recur, ENSR Corporation, West Associates, Tucson, AZ, March 4, 2003.

[6] Maxwell, William, 2003, Letter to Utility MACT Project Files from William Maxwell, USEPA, CG/ESD (C439-01), November 26, 2003.

[7] "ICR-III" is the designation for stack-test data for 81 coal-fired power plants obtained pursuant to EPA's Information Collection Request (ICR) concerning mercury emissions from those plants.

Figure 2. Portion of Hg inlet conc. that is particle-bound or oxidized vs. control efficiency all tests for all units

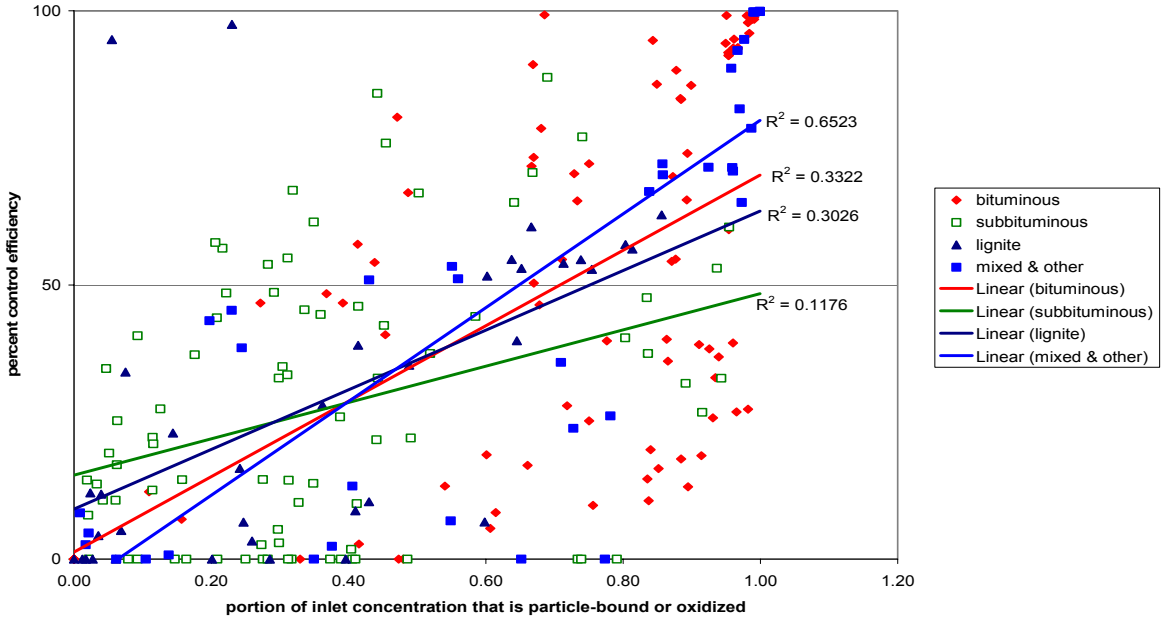
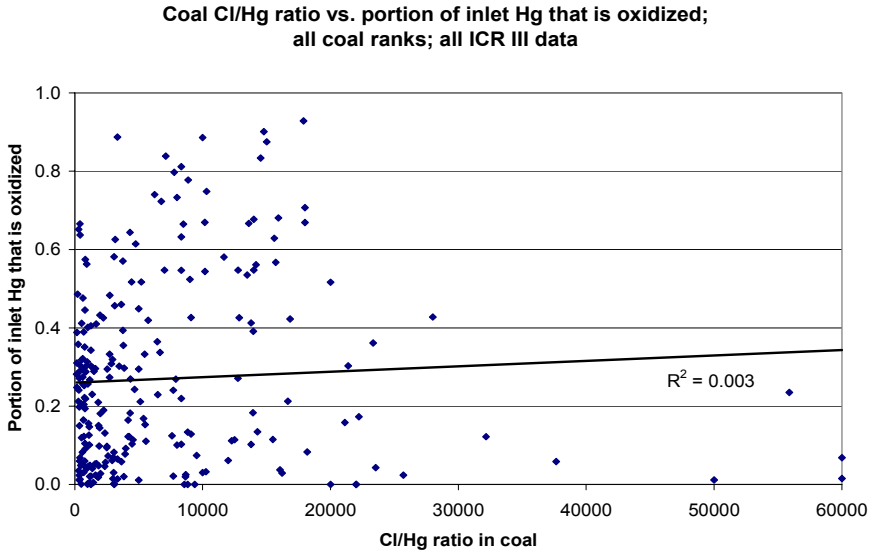


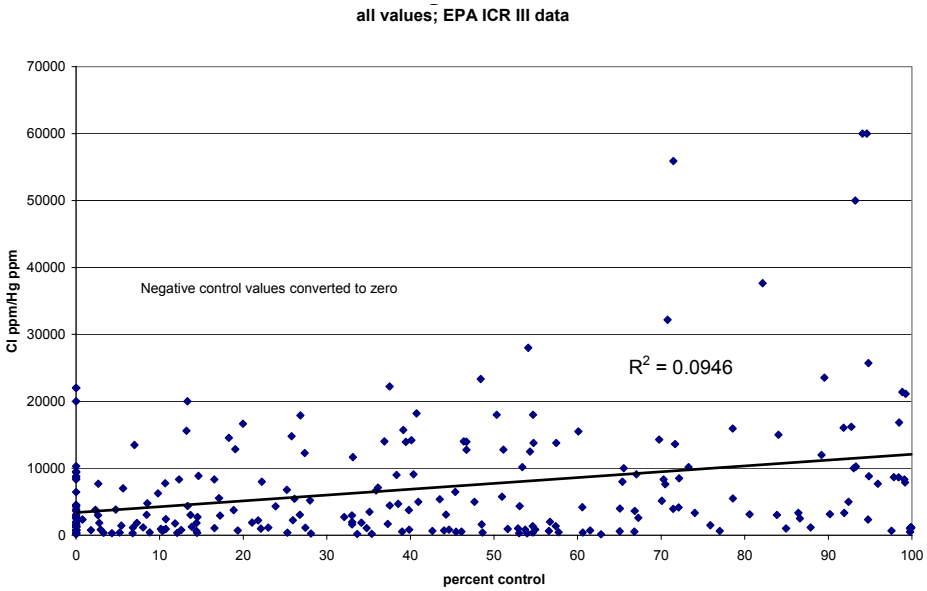
Figure 3. Coal Cl/Hg ratio vs. portion of inlet Hg that is oxidized; all coal ranks; all ICR III data



As Figure 4 shows, the highest chlorine:mercury ratios are associated with relatively good control. However, as is also clear from Figure 4, there are just as many, if not more, units with good control that are burning coals with low Cl/Hg ratios. Clearly other

factors are involved in the achievement of good control than coal's chlorine content. For these reasons, it is unlikely that there is anything more than a weak correlation between the chlorine/mercury ratio of the coal to the percent of mercury emissions controlled.

Figure 4. Ratio of Cl to Hg vs. control efficiency



V. EPA's estimation of mercury emissions using equations based on coal chlorine content is invalid.

EPA offered no evidence and cites no sources to demonstrate a significant correlation between the chlorine:mercury ratio and the ability to control mercury emissions Nonetheless, without demonstrating that a significant correlation between coal chlorine content and degree of control actually exists, EPA used equations developed for WEST Associates[8] to quantitatively relate Hg emissions to the amount of Cl in coal.

The first step in WEST's approach to developing these equations relating coal chlorine content and mercury emissions appears to have been a grouping of units in the ICR III database by type of control. In one such case, WEST identified 10 units that use a combination of a fabric filter (baghouse) and a spray dryer absorber (FF/SDA units). Four of these units burn bituminous coal, including two that are among the four bituminous burners with the lowest mercury emissions (WEST actually looked at the five best-performing units). The other six units in the FF/SDA grouping burn either subbituminous or lignite.[9]

[8] WEST Associates, 2003; see especially, chapter "Development of Mercury Removal Correlation Equations for Each Particulate/SO2 Control Combination"

[9] NOTE: Grouping coals of different rank on the basis of control technology is at odds with EPA's assertion, stated in the proposal, that coal rank is of overriding importance and that limits would be established for Hg depending on the rank of coal. It is unclear why WEST in this part of its analysis considered that control technology configuration superseded coal rank as a valid consideration, and it is unclear why EPA accepted this approach given its decision to base regulation of coal units on coal rank.

An equation form of $y = 1 - \beta e^{-\alpha x}$ was selected. WEST states that in the equation form selection process "care was taken that the mathematical expression accurately reflected the physical and chemical process by which chlorine contributes to the controllability of stack mercury emissions."[10] However, WEST provided no evidence of work efforts, research reviewed, etc. that might demonstrate that such care was in fact taken. Nor is there any discussion of whether other equation forms were considered, or if so, why they were rejected.

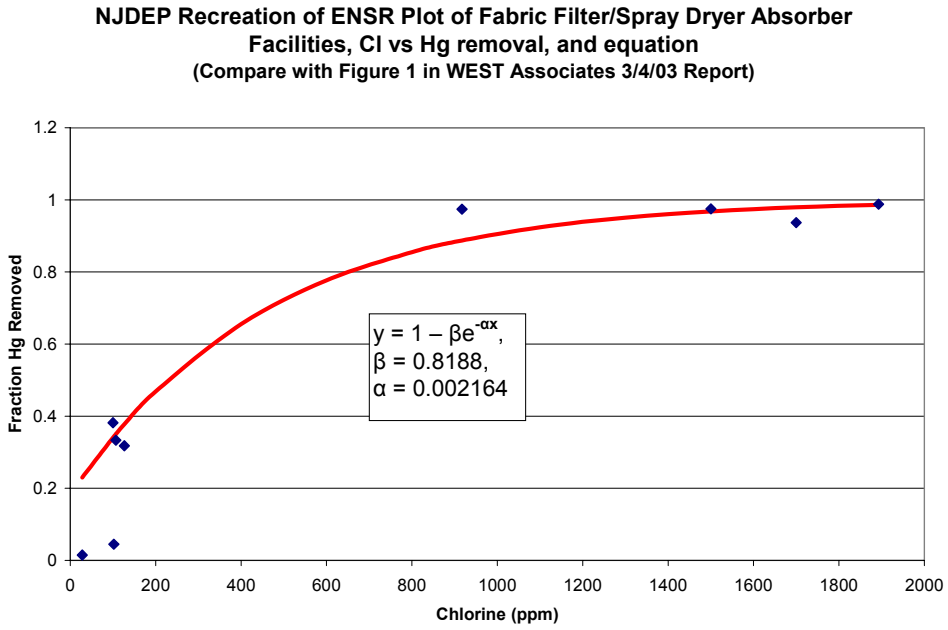
In some cases, WEST found that an equation of the form chosen could not be found that fit the actual data sufficiently well to justify its use. In those cases, WEST elected to use the average removal efficiency as reported in the ICR III database.

In other cases, WEST used equations that it considered to have sufficient predictive value based on the R^2 value. One of these equations, the one chosen to represent the relationship between chlorine content and mercury emissions for the fiber filter/spray dryer absorber (FF/SDA) units, was used by EPA to estimate mercury emissions of two of the four best-performing bituminous units, and two of the five best-performing lignite units.

WEST claims that the best-fitting equation, $F_{rem} = 1 - 0.8188 * e^{(-0.002164 * Clppm)}$ is associated with an R^2 value of 0.935. The NJDEP used the data listed by WEST[11] to develop its own plot of the data, and concluded that WEST overstated the R^2 value and therefore also overstated its value in predicting mercury emissions. A recreation of the WEST analysis by the NJDEP (Figure 5) found that the R^2 value is only about 0.85, not 0.935 as claimed. Furthermore, although WEST claimed to use the data from 9 of the 10 units,[12] it appears that WEST actually plotted only 8 of the 10 units, further detracting from claims of predictive value with respect to mercury emissions.

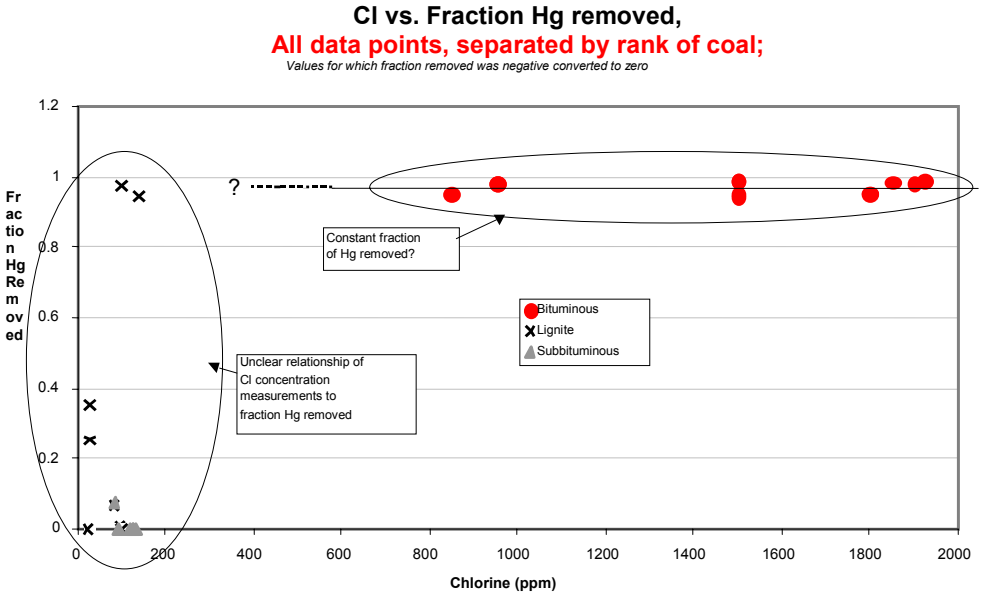
[10] WEST Associates, 2003, Chapter "Statistical Approach", p. 7.
[11] WEST Associates, 2003, Chapter "Statistical Approach", pp. 10 & 11.
[12] WEST stated that it did not use the data from one unit in graphing the data, because the coal chlorine measurements for that unit were so low as to be of questionable accuracy.

Figure 5. NJDEP Recreation of ENSR Plot of Fabric Filter/ Spray Dryer Absorber



Closer analysis reveals the model fit is even poorer. See Figure 6. This figure shows all the data for the FF/SDA units instead of the averages of the three values for each unit as WEST did. The figure also separates the coals by rank instead of lumping them together as WEST did. Inspection of the figure suggests that there could be other models that might fit the data better than the equation form used. It could be that there are in fact two subgroups of units in this group, bituminous coal burners and those burning other ranks of coal. The best approach might be to fit a separate equation to each group. One such equation, for bituminous burners, might be virtually a horizontal straight line, with control efficiency above 90% regardless of chlorine concentration.

Figure 6. NJDEP Version of Plot of Fabric Filer/Spray Dryer Absorber Facilities



Particularly problematic is the fact that, as shown in Figure 6, the range of values corresponding to the fraction of mercury removed for coals with chlorine concentrations between approximately 150 ppm and 850 ppm is totally undefined by the data. Nevertheless, the equation used by EPA, corresponding to the values depicted in Figure 5, was used to predict mercury emissions from some units burning coal with chlorine content corresponding to these undefined values.

Problems evident in the development of the above equation indicates that it is invalid as a basis for estimation of mercury emissions. There is another equation that was also used by EPA to estimate mercury emissions based on coal chlorine content for one of the bituminous units.[13] Because it was apparently developed through similar processes it may also be invalid.

VI. Because the equations based on coal chlorine did not have strong value in predicting mercury emissions, the EPA's distributions of emissions estimates based in part on those equations are questionable.

EPA next proceeded to use the ICR II data[14] to develop estimates of what mercury emissions might have been at certain times, even though no stack tests had been conducted at those times. EPA took the mercury content of the coals purchased by the best-performing 12% of units in the chosen subcategories and multiplied each of these values by a value representing the fraction of that mercury that EPA estimated would be removed by the unit's control systems prior to emission. This value, the fraction removed (Fr), was derived by EPA in one of two ways. It was either:

1) the average fraction removed by the unit in the three tests reported in the ICR III database, or,
2) the Fr value calculated by a WEST-derived equation, as discussed above, for those units.

Approach 2) was used for three of four bituminous units and two of five lignite units where EPA assumed that the equation sufficiently represented the relationship of mercury in the coal to mercury emissions.[15] As discussed above, these equations do not sufficiently represent the relationship of mercury in the coal to mercury emissions. Claims of the predictive value of the equations turned out to be overstated. Accordingly, Approach 2) should not have been used. Approach 1) arguably has a greater potential to be valid, but that approach is based on an assumption that emissions of mercury are linearly related to the mercury content of the coal burned. This assumption is unlikely to be valid in all cases.

The result of this multiplication process was a series of "distributions" for each unit. These distributions are estimated mercury emissions for that unit assuming that the coal it

[13] This is of the same $y = 1 - \beta e^{-\alpha x}$ form but has a value for α (alpha) of 0.0069 and a value for β (beta) of 0.3186.
[14] "ICR-II" is the designation for coal quality data for coal delivered to essentially all large U.S. coal-fired power plants in 1999. Pursuant to the EPA's Information Collection Request, in general every sixth delivery was analyzed for mercury content, chlorine content, heat content, and certain other parameters.
[15] The WEST-derived equations use the Cl content of coal in the ICR III data to estimate a Fr. Then this Fr is multiplied by the Hg content of the coals in the ICR II data to estimate Hg emissions from the unit.

would burn over the course of a year would be the same as the coals reported in the ICR II database for that unit.

EPA assumed that these distributions represent the mercury emissions associated with each separate coal shipment likely to be used by each unit over the course of a year. EPA also assumed that the same group of units would still represent the best-performing units after these calculations.

VII. EPA's selection of 97.5th percentile values could potentially be helpful in projecting peak emissions, but substantially overstates expected the average annual emissions that are relevant to the standard EPA actually proposed.

After developing the distributions of mercury emissions estimates, EPA then determined the 97.5th percentile of these distributions for each unit considered. This represents a value that, based on these data, would likely be exceeded only 2.5% of the time. This approach captures likely peak emissions, and therefore would be relevant only if compliance were to be based on a short-term test.[16] Under EPA's proposal to base compliance on an annual average, a lower percentile, such as the 50th percentile, would be far more relevant.

It is interesting that, while basing much of its procedure on the procedure described in the WEST report to relate Cl to Hg emissions, WEST recommended using a 95th percentile value, stating that this represented "the operation of the unit under 'worst conditions'."[17] EPA, however, used 97.5th percentile value of the cumulative distributions, and stated that the resulting number represents "the operation of the unit under conditions reasonably expected to occur at the unit."[18]

Given that the proposed rules indicate that a rolling average of a year's worth of data will be used to determine compliance, it is unclear why EPA considered selecting a 97.5th percentile value. A 95th or 97.5th percentile value might be appropriate if compliance were to be determined with a once-a-year stack test, but is inappropriate for a year-long averaging process.

A variety of other approaches that were not based on questionable assumptions could have been taken.[19] Nevertheless, EPA selected the 97.5th percentile values. These,

---

[16] For example, three test runs of one hour each in one day.

[17] WEST Associates, 2003, chapter, "STATISTICAL APPROACH", p.14. It is possible that WEST assumed that compliance would be determined with short-term tests and hence recommended a 95th percentile value.

[18] USEPA, 2004; Federal Register, Vol. 69, January 30, 2004, page 4673.

[19] One approach would have multiplied the Hg content of the coal in the ICR II database by the average removal efficiency in the ICR III database for all units considered. In the case of the four best-performing bituminous units, this multiplication process generates a mean emission rate of 0.28 lbs Hg/TBtu. A basic statistical approach might suggest that the 95 percent confidence limit of this mean be determined with use of the student's t statistic method. However, a more appropriate approach would account for the fact that each of the four values is in fact a mean derived from numerous values (the distributions). The within-unit variance of these means can be estimated, and then the variance of the across-unit mean can be estimated from these variances, and used to calculate the upper limit of the 95% confidence interval. The DEP's use of this approach results in a 95% confidence interval with upper bound of 0.31 lbs/TBtu. This value could

coupled with the use of distributions discussed above that, in some cases, predict much higher Hg emission rates than a unit's (ICR III) stack test data show, result in estimated emission rates in many cases much higher than the stack test data for the best-performing units.

VIII. By treating EPA's 97.5th percentile value estimates as "data," and then determining means and upper confidence interval limits of those "data," EPA further overstated mercury emissions from the best performing 12%.

After developing the 97.5th percentile values, EPA then treated these values as if they were data points. EPA then determined the average 97.5th percentile value and 97.5% upper confidence limit of this average for each set of units. Because there are only a small number of units under consideration (e.g., in the case of bituminous and subbituminous, only four units each), the use of statistical procedures that are intended for normally distributed data inappropriately results in a still higher value.

Since these percentile values are statistics derived from data, the variance estimate for each of these statistics should have been used when constructing confidence intervals. This would remove the variability due to unit, which should not have been included in the error variability. It should also be noted that such a high percentile value is difficult to estimate with data sets of moderate sample size, such as those used here. Considering that the data come from imprecise modeling and non-random sampling, these percentiles are likely highly inaccurate.

This last data manipulation completes a tortuous process and results in proposed limits that are, in the case of bituminous coal burners, seventeen times higher than the average of the best-performing 12% of tested units. This is an inappropriate outcome, especially considering the flaws and questionable assumptions in EPA's analysis.

IX. Even though a bituminous-fired unit could become a subbituminous-fired unit with little or no modification, EPA set a substantially higher emission limit for subbituminous-fired units, encouraging fuel switching that would bring substantially higher mercury emissions.

EPA proposed separate standards for units burning bituminous coal and units burning subbituminous coal.[20] However, units burning bituminous coal and units burning subbituminous coal do not appear by their nature to be qualitatively different. Many units that currently burn bituminous coal exclusively could easily burn subbituminous coals or blends of subbituminous and bituminous coals. Fuel switches like these have already been implemented successfully at a number of units.

For that reason, the rationale for establishing separate limits for bituminous-fired units and subbituminous-fired units appears questionable. Since the proposed limits for

---

be, in the DEP's view, an appropriately derived standard for bituminous coal combustion. Use of 99% confidence interval results in a value of 0.32 lbs/TBtu.

[20] EPA also proposed separate standards for units burning lignite and for units burning waste coal, as well as for units using integrated gasification combined-cycle technology.

subbituminous-fired units are nearly three times what was proposed for bituminous-fired units, a fuel switch may enable a unit to comply without installing mercury control technology, while emitting much more mercury.

Michael Aucott, Ph.D. and Leo Korn, Ph.D., Research Scientists
Division of Science, Research, and Technology
New Jersey Department of Environmental Protection
Trenton, New Jersey 08625-0409

July, 2004