REVIEW AND CRITIQUE OF DATA AND METHODOLOGIES USED IN EPA PROPOSED UTILITY MERCURY MACT RULEMAKING:

40 CFR Parts 60 and 63 Proposed National Emission Standards for Hazardous Air Pollutants; and, in the Alternative, Proposed Standards of Performance for New and Existing Stationary Sources: Electric Utility Steam Generating Units; Proposed Rule

Prepared for:

National Mining Association
101 Constitution Ave., N.W. Suite 500 East
Washington, D.C. 20001

Prepared by:

AEMS, LLC
2020 North 14th Street Suite 210
Arlington, VA 22201

RWCrawford Energy
2853 S. Quail Trail
Tucson, AZ 85730

April 2004

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY v
2. EVALUATION OF DATA ON MERCURY EMISSIONS FROM COAL-FIRED POWER PLANTS 1
  2.1 Introduction 1
  2.2 Adequacy of Data for EPA Rulemaking 2
    2.2.1 Non-Proportional Representation of US Population 2
    2.2.2 Bias in Testing Conditions 4
    2.2.3 High Variability Present in the Data 5
    2.2.4 Mismatch of Test Data to Form of the Proposed Standard 8
    2.2.5 Mismatch of Test Method to Compliance Monitoring Method 10
  2.3 Revisions to the Structure of the EPA Database 11
    2.3.1 Classification of Units by Boiler Technology 11
    2.3.2 Classification of Units by Coal Rank 12
  2.4 Statistical Review of Data Quality 16
    2.4.1 Review of Missing Data 17
    2.4.2 Precision of Stack Gas Mercury Measurements 18
    2.4.3 Validity of Tests Indicating Negative Mercury Removal 21
3. CRITIQUE OF EPA METHODOLOGY FOR DETERMINATION OF MERCURY MACT FLOORS 25
  3.1 Introduction 25
  3.2 Counting Top Performing Units 26
  3.3 Ranking Units Based on Emissions 27
  3.4 Exclude Incomplete or Invalid Data Before Ranking Units 29
  3.5 Rank Units Based on Annual Emissions for Boiler Technology Subcategories 29
  3.6 Define the Adverse Conditions to be Considered in MACT Floors 30
  3.7 Estimate MACT Floors, Considering Unit Performance Variability 31
  3.8 Conversion of EPA's MACT Floors to Match the Form of the Standard 33
4. STATISTICAL ASSESSMENT OF MERCURY CONTROL PERFORMANCE OF DATA AND METHODOLOGIES USED IN VARIOUS CONTROL SYSTEMS BASED ON ICR PART III DATA 34
  4.1 Introduction 34
  4.2 Statistical Methodology 36
  4.3 Results of Correlation Analysis 39
    4.3.1 Fabric Filter Technology Group 40
    4.3.2 Wet FGD with ESP 44
    4.3.3 Other Mercury Control Technologies 47
  4.4 Assessment of Unit and Test Variability 48
  4.5 Adequacy of Correlations to Predict Performance of Lignite Units 50
5. DEVELOPMENT OF ALTERNATIVE MACT FLOORS FOR EXISTING UNITS 52
  5.1 Introduction 52
  5.2 Counting Top Performing Groups based on Population Statistics 53
  5.3 Ranking Units Based on Emissions when Firing an Average Coal 54
  5.4 Accounting for Uncertainty and Variability 58
6. DEVELOPMENT OF MACT FLOORS FOR REGIONAL LIGNITE COALS 64
  6.1 Introduction 64

6.2 Generating Unit Populations 66
6.2.1 Fort Union Lignite Units 66
6.2.2 Gulf Coast Lignite Units 68
6.3 Emissions of Top Performing Units under Adverse Circumstances 68
6.4 MACT Floors for Regional Lignite Coals 70
7. ESTIMATING MACT LIMITS FOR NEW UNITS 73
7.1 Introduction 73
7.2 New Unit MACT 73
8. ABILITY OF VARIOUS COALS TO COMPLY WITH EPA's PROPOSED MACT STANDARDS 76
8.1 Existing Units 76
8.2 New Units 78
9. CRITIQUE OF THE USE OF ICR PART III DATA IN RECENT EPA PAPER ON CONTROL OF MERCURY EMISSIONS 81
9.1 Introduction 81
9.2 The Limitations of Small Samples 82
9.3 Short-term Nature of Testing and Operating Bias 86
9.4 High Variability Present in the ICR Data 87
9.5 Conclusion 88

LIST OF TABLES

Table 1.1 Emission Limits for Existing Coal-Fired Utility Steam Generating Units v
Table 2.1 Representational Bias Introduced by EPA Stratified Sampling Approach 3
Table 2.2 Classification of Coal Rank for Units Potentially Firing a Fuel Blend During ICR Part III Testing 14
Table 2.3 Exclusion of Tests due to Missing Primary Data 18
Table 2.4 Units Exhibiting High Variability in Repeat Tests of Total Stack Gas Mercury Concentrations 21
Table 2.5 Units Exhibiting Statistically Significant Negative mercury Removals 23
Table 2.6 Exclusion of Generating Units Based on Review of Data Quality 24
Table 2.7 Exclusion of Individual Test Results Based on Data Quality 24
Table 4.1 Correlation for All Fabric Filters With or Without Scrubbing 41
Table 4.2 Correlation for Wet FGD + ESP (Cold- or Hot-side) 45
Table 4.3 Summary of Other Technologies (Not Chlorine Sensitive) 48
Table 4.4 Components of Uncertainty in Emissions Estimates by Technology (Relative Standard Deviations, percent) 50
Table 5.1 Counting Top Performing Units based on Population Statistics 54
Table 5.2 Characteristics of Average Coal 55
Table 5.3 Identification of Top Performing Units and their Control Technologies for Conventional Coal Subcategories (Without Consideration of Uncertainty or Variability) . 56
Table 5.4 Characteristics of the Adverse Coal by Rank 57
Table 5.5 Expected Emissions under Adverse Conditions for Top Performing Units (Without Consideration of Uncertainty or Variability) 58
Table 5.6 Determination of MACT Floors for the Annual-Average Form of the Standard 62
Table 5.7 MACT Floors for Conventional Coal Units 63

Table 6.1 Average Lignite Coal Characteristics Based on ICR Part II Shipments Data 65
Table 6.2 Population of Fort Union Lignite Units 67
Table 6.3 Population of Gulf Coast Lignite Units 67
Table 6.4 Characteristics of Adverse Lignite Coals By Region 69
Table 6.5 Expected Emissions under Adverse Conditions for Regional Lignites Based on Correlations Curves for all Ranks (Without Consideration of Uncertainty or Variability) . 70
Table 6.6 Determination of MACT Floors for Regional Lignite Coal Annual-Average Form of the Standard 71
Table 6.7 MACT Floors for Regional Lignites 71
Table 7.1 Assessment of New Unit MACT 74
Table 8.1 Ability of Coal to Achieve Compliance with EPA Proposed MACT Limits with 97.5% Confidence When Fired in Units with Top Performing Control Technology (Percent of Btus) 77
Table 8.2 Coal <u>Not</u> Achieving Compliance with EPA Proposed MACT Limits with Stated Confidence When Fired in Units with Top Performing Control Technology (Percent of Btus) 78
Table 8.3 EPA's Proposed MACT Limits for New Units (Assuming Heat Rate of 9,500 Btu/kwh) 79
Table 8.4 Ability of Coal to Achieve Compliance with EPA Proposed MACT Limits with 97.5% Confidence When Fired in New Units with Top Performing Control Technology (Percent of Btus) 79
Table 8.5 Coal <u>Not</u> Achieving Compliance with EPA Proposed MACT Limits with Stated Confidence When Fired in New Units with Top Performing Control Technology (Percent of Btus) 80
Table 9.1 Sample Sizes for Technology Groups Tested in ICR Part III 83

LIST OF FIGURES

Figure 2.1 ICR Test Results for Antelope Valley Station B1 7
Figure 4.1 Mercury Removal Performance for Fabric Filter Technologies 41
Figure 4.2 95 Percent Confidence Intervals on the Performance of Fabric Filter Technologies . 44
Figure 4.3 Mercury Removal Performance for Wet FGD + ESP Technologies 46
Figure 4.4 95 Percent Confidence Intervals on the Performance of Wet FGD with ESP Technology 47
Figure 6.1 Distribution of Mercury and Chlorine Content for Regional Lignite Coal 65

1. EXECUTIVE SUMMARY

This report presents a review and critique of the data and methodologies used by the US Environmental Protection Agency (EPA) in support of its proposed rule to establish national standards under Section 112 of the Clean Air Act, known as Maximum Achievable Control Technology (MACT) limits, for mercury emissions from electric utility steam generating plants. The proposed rule would establish the emissions limitations based on the type of coal burned and certain technologies shown in Table 1.1.

Table 1.1 Emission Limits for Existing Coal-Fired Utility Steam Generating Units

Subcategory	Annual Emissions Limitation lbs/TBtu
Bituminous	2.0
Subbituminous	5.8
Lignite	9.2
IGCC	19.
Coal Refuse	0.38

The scope of the review and critique included:

- The quality and adequacy of the data used in support of the proposed standards.
- A review and critique of the analytical methods used by EPA to develop the "floor" for the existing source MACT standards, and the development of alternative approaches to address the identified concerns.
- A new analysis of the mercury removal performance of existing emission controls on coal-fired power plants, with a specific emphasis on determining the uncertainty in the estimates of the emissions performance of these control technologies.
- The derivation of alternative MACT floors based on a new assessment of available data and a fuller consideration of the effects of uncertainty and variability

The conclusions of this assessment are:

1. Based on our review, the ICR Part III database are inadequate to characterize the mercury emissions and control performance of existing units for the purpose of setting meaningful MACT limitations, for the following principal reasons:

    - The data over-represent units with more sophisticated controls or with combinations of controls, and under-represent units equipped with cold side ESPs and hot side ESPs; therefore, the sample group does not represent a cross section of US coal-fired generating units.

    - The data are affected by a bias in testing conditions, because the testing was done during high-load and steady-state operations. The data provide no evidence of the emissions during partial load, transient operations or during maintenance events, all of which are covered by the proposed standards.

    - The data have a very high degree of variability and uncertainty, even for properly conducted tests, as a result of the complexity of the test methods and the measurement procedures employed. The high degree of variability and uncertainty is not adequately attenuated by the small number of tests performed on each unit. As a result, we do not know whether the units that appear to be top performers, based on the data, would again demonstrate low emissions if retested at a later date.

    - The data consist of "snapshot" measurements of mercury emissions from short-term testing and do not provide information on the levels of emissions over a full year. As a result, relatively little is known about the long-term mercury emissions performance of the units that were tested, and the data cannot be used to identify specific units that would be top performers on the annual-average basis that EPA has adopted for its proposed standards.

    - The data were gathered using a test method that is very different from what is proposed for compliance demonstration under the rule and no effort has been made to translate the proposed standards that were developed from the data to the basis of the test methods proposed for compliance demonstration.

    - In applying these data to the setting of standards for specific types of coals and boilers, EPA has not separated the FBC and stoker units, both of which are viewed as having very different combustion, emissions, and control characteristics, from the analysis of conventional coal boilers.

    As a result of these inadequacies, the ICR Part III data cannot be used to identify and evaluate the emissions performance of specific units that would rank in the top 12 percent of their subcategory (or among the 5 best) on an annual basis.

2. A review of the reported rank of the coal burned during each test identified an error in the classification of test coals and many instances in which the test coal rank could not be

conclusively determined. Units firing blends of coal during testing, or firing a fuel blend of unknown proportions, should be separated from the database that is used to establish standards applicable to specific coal ranks.

3. On the basis of missing data and the results of statistical tests that call into question the validity of certain results, test data for 10 units should be excluded entirely, and data for 7 individual test runs at 6 units should be excluded.

4. The objective of EPA's MACT floor analysis is to determine the average emission limitation that the best performing units in a subcategory can achieve "under the most adverse conditions which can reasonably be expected to occur".[1] In addressing the MACT floors, EPA correctly recognized the importance of data variability and uncertainty in determining what level of emissions was achievable by the top performing units, and EPA made an effort to account for the variability in unit performance in deriving floors that could be met by the best performing units. Nevertheless, the EPA methodology has these key weaknesses and errors:

    - Important sources of uncertainty and variability that affect the derivation of the MACT floors were not addressed fully. No consideration was given to the confidence bounds (uncertainties) of the correlations and average removal rates used to estimate the emissions performance of units under adverse circumstances.

    - Further, the method used to account for unit performance variability is based on a narrow and incorrect consideration of the performance of top performing units (only) that is subject to a "self-selection" bias and that undercounts the actual variability of the emissions of top performing units.

    - The definition of "most adverse conditions" for compliance of the top performing units is limited to variations in coal characteristics and did not address the effect of variations in unit operating conditions.

    - EPA has not adjusted its MACT floors for the shift from Ontario Hydro stack tests to CEMs for demonstrating compliance.

5. Based on a new analysis of the performance of existing emission controls, we conclude:

---

[1] National Lime Association v. EPA, US Court of Appeals for the District of Columbia Circuit, 627 F.2d 416, Decided May 19, 1980.

- The best performing technologies for mercury removal are fabric filters (with or without scrubbing) and wet scrubbers with (cold- or hot-side) electrostatic precipitators (ESPs). The mercury removal capability of these technologies is found to be correlated -with coal chlorine content.

- No statistically significant differences can be detected in the mercury removal performance among the three configurations of fabric filter controls alone or combined with wet or dry scrubbers, at least within the modest power of the ICR Part III test data.

- Similarly, no statistically significant differences can be detected in the mercury removal performance among cold- and hot-side ESPs combined with wet scrubbers.

- The performance of other emission control technologies does not appear to be sensitive to chlorine content.

6. The performance of the best technologies is substantially reduced and highly variable when firing coals with low chlorine content. Thus, we cannot have a high level of confidence that the best performing technologies will reduce mercury emissions to a significant degree when units fire coals of relatively low chlorine content.

7. A revised methodology has been applied to the derivation of MACT floors for conventional units burning bituminous, subbituminous and lignite coals. The methodology gives consideration to the statistical uncertainties in the derivation of MACT floors and, specifically, the confidence limits of the resulting predictions. The alternative floors reflect the annual emissions performance of the top-performing units and technologies that would be achieved 97.5 percent of the time when firing an adverse annual coal supply that is at the 90th percentile of emissions. The floors also reflect a full and appropriate consideration of the effect of uncertainties and of variability on the emission levels that can be achieved on an annual basis.

8. The MACT floors developed in this study for existing units for an annual standard are:
    - 6.9 lbs/TBtu for bituminous units
    - 7.9 lbs/TBtu for subbituminous units
    - 34 lbs/TBtu for all lignite units
    - 9.1 lbs/TBtu for Fort Union lignite units

- 34 lbs/TBtu for Gulf Coast lignite units

While these levels are higher than those derived by EPA, they do not reflect adjustments that should have been made, but could not be made, to address the uncertainty regarding: (1) emissions during load swings, low load and maintenance activities; and (2) the transformation from the short-term Ontario Hydro test method to a continuous emissions monitoring method.

9. The MACT limits developed in this study for new units for an annual standard are:
    - 5.1 lbs/TBtu for bituminous units
    - 7.4 lbs/TBtu for subbituminous units
    - 32 lbs/TBtu for all lignite units
    - 8.5 lbs/TBtu for Fort Union lignite units
    - 32 lbs/TBtu for Gulf Coast lignite units

10. The high levels of the MACT floors and limits derived here reflect the great degree of uncertainty about emissions performance that is present in the test data and the inadequacy of the data base for use in deriving a floor or a standard in which one can have an appropriate level of confidence of compliance.

11. The uncertainty in the analysis of mercury removal is sufficiently large that, at the present time and based on the ICR Part III data, one cannot say with acceptable statistical confidence that the best performing existing units will consistently meet a floor-based standard that is below the mercury present in the coal, when the chlorine content of the coal is low.

12. Based on the analysis of emissions performance and variability prepared in this study, an assessment was made of the ability of US coals to be used in compliance with the MACT standards proposed by EPA, when fired in the best performing units. A large portion of the US coal supply – 49 percent of bituminous coals, 41 percent of subbituminous coals, and 62

percent of lignite coals on a Btu basis (71 percent for Gulf Coast lignite coals and 37 percent for Fort Union lignite coals) – will be unable to comply with the proposed standards with high statistical confidence (97.5 percent). Very large portions of the US coal supply will be unable to comply with high confidence with EPA's proposed standards for new units, including more than 80 percent of bituminous coals, more than 90 percent of subbituminous coals, and more than 75 percent of lignite coals. As a result, EPA's proposed mercury standards are likely to have substantial impacts on US coal supply and the coal-based electric industry.

13. The EPA paper "Control of Mercury Emissions from Coal-Fired Electric Utility Boilers" presents a narrow and misleading view of the mercury capture performance of conventional SO2 and particulate control technologies. If the purpose of the paper was to communicate what is and is not known about mercury control, the paper should have discussed the limitations of the data from which conclusions were drawn, the variability and uncertainty of the results in that data, the performance that can be expected over a range of coal types, the confidence intervals for those estimates and what EPA is doing to improve the state of knowledge on the effectiveness of conventional as well as advanced control systems.

Chapter 2 addresses the adequacy of the database used by EPA and presents a statistical review of data quality and validity. Chapter 3 reviews the EPA methodology for determining MACT floors. Chapter 4 presents a statistical assessment of mercury removal performance of various control technologies based on the ICR Part III data. Chapter 5 develops alternative MACT floors for existing units using an improved methodology that adjusts for the concerns with the EPA methodology. Chapter 6 develops alternative MACT floors for two regional lignite subcategories – the Fort Union and Gulf Coast lignite coals. Chapter 7 develops MACT limits for new units using the methodology employed for existing units. Chapter 8 addresses the ability of US coals to comply with EPA's proposed standards. Chapter 9 presents a review and critique of EPA's assessment of existing controls in the paper entitled "Control of Mercury Emissions from Coal-Fired Electric Utility Boilers."

2. EVALUATION OF DATA ON MERCURY EMISSIONS FROM COAL-FIRED POWER PLANTS

2.1 Introduction

To support the assessment of mercury emissions from coal-fired power plants, EPA required the reporting of data under a three-part Information Collection Request (ICR):

- Part I: General Facility Information, identifying the characteristics of every coal-fired power plant in the US, including the fuels fired and existing emissions controls.
- Part II: Coal Analyses, measuring the characteristics of the coals fired during 1999, including mercury and chlorine contents, for approximately every sixth coal shipment received at all power plants.
- Part III: Mercury Emissions Testing, based on the Ontario Hydro method at the inlet and outlet of the last stack control device for a selected sample of coal-fired units. Three tests were conducted on a one-time basis at each selected unit. The database in the docket contains completed emissions testing reports for 80 generating units.

These data were used as the primary basis for EPA's assessment of mercury emissions, MACT floor determination, and the mercury emissions standards proposed in its rulemaking. Because the Part III data on mercury emissions are central to EPA's analysis and proposed standards, the data have been subjected to an in-depth review to determine their adequacy for use in this rulemaking. The review has identified inadequacies in the data and the need for adjustment or exclusion of portions of the data that reflect incomplete or invalid testing.

When a regulatory finding has been made that requires regulation of a pollutant under the MACT provisions, EPA must establish an emissions limitation for existing units that is not less stringent than the MACT floor. The floor is the emissions limitation that could be met by the top-performing units in the population under the most adverse circumstances that can reasonably be expected to occur. Top performing units are the top 12 percent in populations of 30 or more units, and the top 5 units in smaller populations. Thus, the floor depends on several determinations: how large is the population; which are the top-performing units in the population; what are the adverse circumstances that can reasonably be expected to occur; and what is the achievable emissions limitation of the top-performing units when operating under the adverse circumstances. The review and critique presented in this report have the purpose of assessing whether the ICR Part III data, and EPA's analysis of them, provide an adequate and reliable basis for these determinations.

2.2 Adequacy of Data for EPA Rulemaking

Five concerns with the use of this database for setting of MACT standards have been identified:

- Non-proportional representation of the US population of coal-fired generating units
- Bias in testing conditions
- High variability in the data
- Mismatch of test data to the form of the proposed standard
- Mismatch of test method to the proposed compliance monitoring method.

2.2.1 Non-Proportional Representation of US Population

The units tested in ICR Part III reflect a stratified survey in which EPA selected generating units to assure representation and coverage of specific combinations of the emission control technologies that were thought at the time of the survey design to be relevant to mercury emissions control. In designing the survey, EPA first identified 36 sample cells based on the combinations of three factors:

- SO_2 control technology (not scrubbed, wet scrubber, dry scrubber)
- PM control technology (cold-side ESP, hot-side ESP, other PM control)
- Rank of coal fired (bituminous, subbituminous, lignite, blends)

In the survey design, fabric filters were combined with mechanical PM controls in the "other PM control" group, although fabric filters were subsequently shown to be relatively effective in removing mercury from stack gases when used in certain conditions.

Up to 3 units in each cell were selected for testing, and a total of 80 units completed testing and filed emissions reports that EPA has used. Some of the units were selected for testing by random selection procedures, while other units were selected because they had previously been tested in work sponsored by the Electric Power Research Institute (EPRI) and The US Department of Energy (DOE). The EPA survey takes the form of a stratified random sample, but it includes data selected by other means.

Stratified survey design is a widely used research technique, but it means that the Part III dataset does not itself represent coal-fired generating units in proportion to their numbers in the US population. Table 2.1 demonstrates this by comparing the representation of selected technology combinations in the US population and in the ICR Part III sample. The population estimates are based on data used by EPA in survey design that give the estimated number of generating units in the US represented by each survey cell. Nearly 59 percent of the US generating unit population is not scrubbed for SO_2 control, has ESPs in place for PM control, and fires bituminous coal, subbituminous coal, or a blend of these ranks. As a result of the stratified survey design, however, such units constitute only 3.5 percent of the ICR Part III dataset. On the other hand, units having fabric filter controls are approximately 9 percent of the US generating unit population, but constitute 29 percent of the ICR sample.

Table 2.1 Representational Bias Introduced by EPA Stratified Sampling Approach

		Percent of US Population	Percent of Sample
3 Most Populous Categories in US Population			
	Bituminous / No Scrubber / cold-side ESP	39%	3.5%
	Subbituminous / No Scrubber / cold-side ESP	11%	3.5%
	Bit-Sub Blend / No Scrubber / cold-side ESP	9%	3.5%
	Total	59%	11%
All Fabric Filter Technologies		9%	29%

In general, the ICR Part III data over-represent units with more sophisticated controls and with combinations of controls that have the potential for increased mercury removal, but which are infrequently encountered in the US population. Other the other hand, the ICR Part III data under-represent units with less sophisticated or fewer emission controls in comparison to the population. Any statistic calculated from the ICR Part III sample without controlling for its non-proportional representation of the population will be biased by the over-representation of units having a generally higher level of controls than exists in the US population.

EPA relied on the test database in its analysis of the MACT floor for each subcategory and made no adjustment to account for this representational bias in its identification of the top-performing units in each subcategory. Instead, EPA counts units in terms of the sample proportions – e.g., the top-performing bituminous units are the four units that constitute the top 12 percent of the bituminous units tested in ICR Part III. Section 112 of the Clean Air Act states that the assessment of top performing units is to be based upon units "for which the Administrator has emissions information." The Administrator has emissions information on the full population of coal fired units in the US. The ICR Part III dataset covers all US units, and it was specifically collected to have information on the emissions of the entire population of US units. EPA has the information needed to expand the ICR Part III data to represent emissions of the full US population on a proportional basis, and it can identify the top-performing portions of the entire US population using the knowledge that can be derived from the test results. By failing to correct for the non-proportional representation, EPA identifies top-performing units that represent smaller proportions of the population and introduces a bias toward a top-performing population that is less than 12 percent of the units.

2.2.2 Bias in Testing Conditions

EPA notes in the preamble to the proposed rule that the ICR Part III emissions testing was conducted under test conditions of high load and near steady-state operation. As a result, these data do not provide any information on mercury emissions that occur under part load, transient loads, and start up and shutdown of the unit or when maintenance activities are underway. Yet, the MACT floors and the proposed standards developed from these data are to apply as an annual average emissions limitation covering all of these operating conditions. EPA acknowledges this limitation of the data, noting that "… its results are likely to underestimate the reasonable worst-case emissions of the best performing facilities."[2]

Because emissions under part-load or transient operating conditions are not reflected in the ICR Part III data, it is not possible to say with analytical confidence what effect the omission has on EPA's assessment of mercury emissions and their control. That the effect is likely to be significant can be seen from the test data on the one unit (Gibson Generating Station 3) that was tested on two separate occasions. At its first test, the unit was conducting soot blowing (a common periodic operating practice) during testing and reported average emissions of 31 lbs of

---

[2] OAR-2002-0056-0006. Analysis of variability in determining Maximum Achievable Control Technology (MACT) floor for coal-fired electric utility steam generating units." Memorandum from William H. Maxwell to Utility MACT Project Files.

mercury per Trillion Btus (TBtus), or more than three times the mercury content of the coal fired during the test. When retested several months later, the unit's stack emissions were both much lower and were consistent with the mercury content of its coal. By failing to sample emissions during part-load or transient operating conditions and when maintenance activities are underway, it is likely that the ICR Part III data understate the emission levels that will occur over an entire year of operation.

There is no analytical basis to correct for this limitation when determining the MACT floors based on the ICR Part III data. The only solution would be to collect data reflecting the effects of varying operating conditions on the mercury removal performance of emission control devices on a representative population of coal boilers.

2.2.3 High Variability Present in the Data

The ICR Part III data are characterized by a high degree of variability in measured mercury emissions and control efficiencies, even for properly conducted tests, as a result of the complexity of the test methods and the measurement procedures employed, as well as the high degree of variation identified in the performance of units with the same control configurations. For low emitting units, among which are those identified as top performers when setting the floor, the Ontario Hydro mercury emissions test method has a reported relative standard deviation of 34 percent in individual tests for the measurement of the total stack gas mercury concentrations[3]. Additional measurements of stack gas flow rate, coal feed rate, and heating value must be made to compute a unit's stack emissions in units of pounds of mercury per Trillion Btus of coal input (lbs/TBtu). Each of those measurements contributes added variability.

In the statistical review of data quality reported later in this chapter, an expected variability was determined for the emission rate measured in each ICR Part III test, based on the ASTM repeatability standards for the test methods employed. From that review, one finds that, when all sources of measurement variability are accounted for, the determinations of stack emission rates in lbs per TBtu for low-emitting units are subject to uncertainties of 25 percent to 35 percent.

---

[3] ASTM D6784-02. Standard Test Method for Elemental, Oxidized, Particle-Bound and Total Mercury in Flue Gas Generated from Coal-Fired Stationary Sources (Ontario Hydro Method)

The determinations of coal mercury content are subject to uncertainties of 15 percent to 25 percent for average coals and to uncertainties of approximately 100 percent for the coals with the lowest mercury levels. These uncertainties are based on the repeatability standards for testing that is conducted in accordance with ASTM procedures and, therefore, reflects the best outcomes that can be expected given the test procedures involved.

The high variability present in the data is illustrated by the fact that "negative" mercury removal rates were measured in 24 percent of the tests (59 of 244 stack measurements), meaning that the stack emission rate was found to exceed the mercury content of the coal. EPA has attributed this result to the independent variation of the measurements for coal and stack gas mercury contents in units with low or zero removal rates. Based on the EPA analysis, fabric filters are a technology that achieves high mercury removal, yet for the 17 conventional boilers with fabric filter controls that were tested, ten of the 51 tests (20 percent) reported "negative" mercury removals, and 4 of the 17 units reported negative removals for the average of their three tests. The reported average removals for fabric filter units ranged from –79 percent (Valley 2) to 98.5 percent (Mecklenburg Cogeneration 1).

The high variability present in the data is not adequately attenuated by the small number of tests that were performed on each unit. For a sample of three tests, the 95 percent confidence interval of the mean is essentially as wide as the range of the individual data points. Figure 2.1 shows the individual test results for Antelope Valley Station B, which fires lignite coal and is equipped with a fabric filter and dry scrubber (spray drier absorber) for emission control. In the first test, the unit emitted 0.3 lbs of mercury per TBtu and achieved 95 percent removal of the mercury present in the coal. In repeat tests 2 and 3, emission rates exceeded 5 lbs/TBtu with less than 10 percent mercury removal. Thus, we see that the performance of existing controls in removing mercury, and the resulting emission rates of units, can be highly variable from one test to the next and will not be adequately characterized by the average of only three tests. The later review of data quality identifies this test sequence as showing excess variability in comparison to the ASTM standards, when interpreted as repeat tests conducted under consistent test conditions, and recommends the exclusion of the data. It is possible, however, that the sequence reflects valid testing in circumstances where an aspect of unit operations or test conditions that is important to mercury emissions changed between tests 1 and 2.

As will be shown in a re-analysis of the ICR Part III data later in this report, the net effect of high variability is to produce wide confidence bounds for mercury emissions and removal rates for the best performing units associated with the MACT floor. As a result, these data provide relatively little specificity as to the emissions characteristics of the units tested because of the high variability in the data, the small number of tests conducted, and the artificially constrained conditions under which testing was performed.

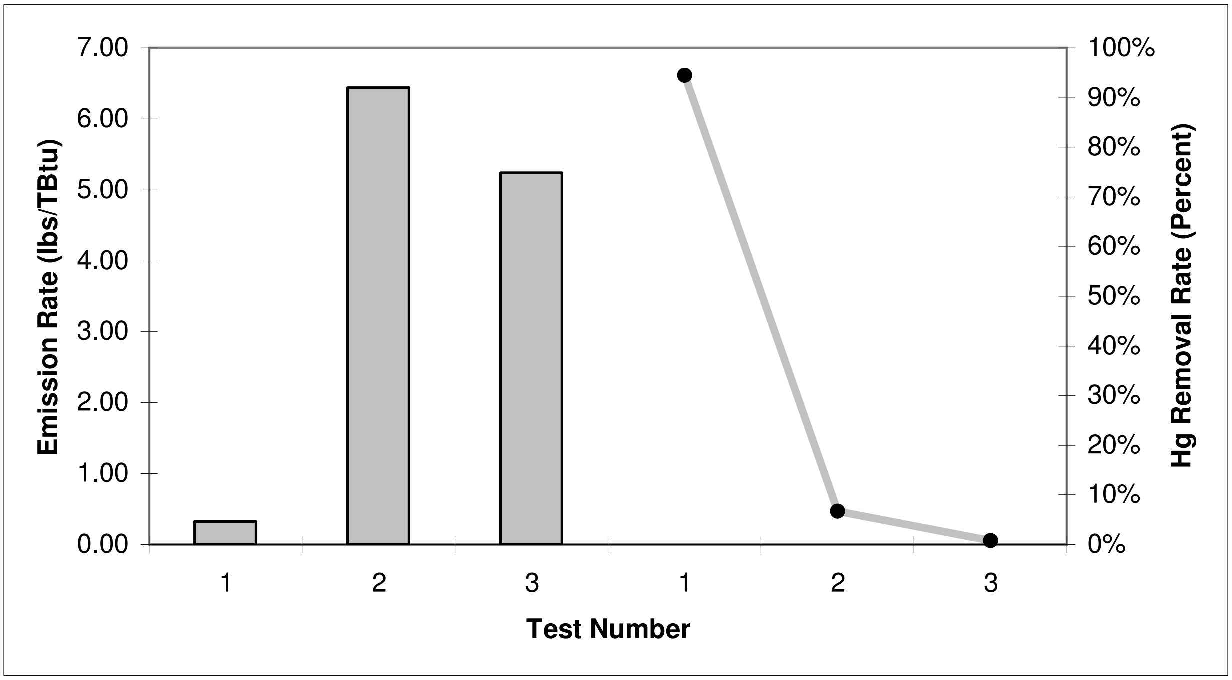

Figure 2.1 ICR Test Results for Antelope Valley Station B1

2.2.4 Mismatch of Test Data to Form of the Proposed Standard

EPA has proposed annual-average emissions limitations for units as an outcome of the use of these test data to derive MACT floors. For purposes of setting an annual emissions limitation and the associated MACT floors, the population of relevance is the units whose *annual* emissions performance would place them in the top 12 percent of all units in their subcategory. However, no data on the emissions of units over an extended period have been employed in the development of the MACT floors and the resulting proposed standards. The ICR Part III database used for this purpose is composed entirely of data acquired in short-term testing of units on a one-time basis. It does not provide information on the annual emissions of units and cannot be used directly to identify specific units that would be top-performing units on an annual basis.

Operational, environmental, and other (presently unidentified) factors cause generating units to exhibit substantially varying mercury emissions performance over time. If annual emissions monitoring data were available, the top-performing annual units could be straightforwardly identified. What the ICR Part III data provide, however, is a "snapshot" on the performance of individual units tested on specific dates and in specific circumstances. The selection of units for the ICR Part III testing, and the specific dates and circumstances chosen to conduct the tests, result in a single "roll of the dice" in this database that will place some truly top-performing units in circumstances to exhibit high emissions during the short-term testing, and thereby fail to be classified as top performers, while it places other units that are not truly top annual performers in circumstances to exhibit low emissions during testing, and thereby be wrongly classified as top performers according to the short-term data.

W. H.Sammis Unit #1 provides a clear example of this problem. During testing for ICR Part III, the unit achieved an average 87 percent reduction in mercury from coal to stack. This removal is comparable to the removal achieved by one of the four units that EPA identified as best performers in the bituminous subcategory. Yet, when this Sammis unit was retested during June 2002, its mercury removal varied from 49 percent to 75 percent in a sequence of tests over a three week period[4]. Thus, the single emissions snapshot captured in the ICR Part III data for this

---

[4] U.S. Energy Policy: Mercury MACT Discussion. National Mining Association Briefing with the Environmental Protection Agency. Research Triangle Park, NC. April 10, 2003.

unit is clearly inadequate to characterize its emissions performance over a longer term period. The same will be true for the units that EPA has identified as top performers and for other units that might be nominated as top performers in other analyses.

As a result, the limited test results are not a good indicator of a unit's true performance characteristics and the ranking of units based on the short-term testing in EPA's database cannot actually identify the units that will exhibit the best emissions performance on annual basis. Attempts to do so run the very real risk of identifying, as top performers, units whose performance was biased towards low emissions on the day of the ICR Part III testing and which would not exhibit low emissions on an annual basis. It is not possible to use the ICR Part III data to identify specific units as top-performing units and, in fact, the ICR Part III data are not adequate to differentiate high and low annual emissions performance among units that are otherwise similar in terms of emissions control configuration and coal characteristics.

If EPA chooses to limit its identification of *specific* units as top-performing units based solely on these test data, then it must limit its consideration of MACT standards to ones that are based on the same conditions, namely the average of three Ontario Hydro stack tests conducted back-to-back. Because of the high variance present in "snapshot" emissions data, units tested at randomly chosen dates and operating circumstances to demonstrate compliance can show widely varying emissions rates. Chapter 5 of this report develops alternative MACT floors for a standard based on the average of three tests and shows that short-term standards will necessarily reflect high emission rates and must even admit the possibility of measuring mercury emissions far in excess of the coal mercury content.

On the other hand, if EPA wishes to proceed with the development of an annual-average mercury limitation based on the ICR Part III data, it must take an approach that focuses on annual emissions performance. As shown later in this document, this can be done through an analysis of emissions data for groups of units sharing a common control technology, thereby identifying units through their technologies and setting an annual floor for the standard based on the performance of the best performing units in the top 12 percent under reasonably foreseeable adverse conditions. Chapter 4 of this report presents an analysis of technology performance to support this technology-based approach, and Chapter 5 applies it to develop alternative MACT floors.

2.2.5 Mismatch of Test Method to Compliance Monitoring Method

EPA proposes to determine compliance with the MACT standards using (unspecified) Continuous Emissions Monitors or CEMs. Because the MACT floors and resulting MACT standards are based on the existing stack test data, EPA logically must address the differences between the Ontario Hydro test method and the CEMs in two regards: differences in the average values measured by the tests (bias) and differences in the test variability (precision). EPA has not addressed these differences in the proposed rule. The situation is compounded by the fact that mercury CEMs are not yet in commercial operation on coal-fired power plants in the US and there is no approved device for this purpose. Therefore, there is no direct basis for assessing the performance of CEMs in comparison to the Ontario Hydro test procedure.

This study does not attempt to assess the differences between Ontario Hydro and the several CEMs technologies that are in development or the potential effect such differences have on determining the proper MACT floor for a standard dependent on the use of CEMs. Based on a review of the two studies[5] that compare CEMS to Ontario Hydro, it can be said that one CEM method appears to be biased high relative to Ontario Hydro and both leading CEMs methods add variability to that of the Ontario Hydro test.

One report[6] in the docket for this rulemaking addresses this issue of the bias and added variability of the two leading CEMs methods in comparison to Ontario Hydro. In the analysis, units were assumed to retrofit controls to reduce their emissions as needed to reach the EPA-proposed limitations for their subcategory. The relative bias and variability of CEMs methods were then simulated over various averaging periods using Monte Carlo techniques. The study concluded that generating units would face high risks of non-compliance – up to a 50 percent

---

[5] S. Sjostrom, "Simplified Approach for Sampling Mercury Emissions in Flue Gas", CEM Users Group Meeting, San Diego, CA (EPRI, May 15, 2003). S. Klam and J. Hosenfeld, "Mercury CEMS Field Observations: September-December 2002", Midwest Research Institute. (2003)

[6] Docket ID OAR-2002-0056-0007. "Averaging Time Study. Addendum to: Multivariate 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." (2003)

chance of violation in many cases – if they do not build in a safety margin to the design of compliance strategies to account for the Ontario Hydro versus CEMs differences.

2.3 Revisions to the Structure of the EPA Database

2.3.1 Classification of Units by Boiler Technology

EPA has used the ICR Part III data to propose separate MACT standards for five subcategories of generating units:

- Bituminous coal
- Subbituminous coal
- Lignite coal
- Integrated coal gasification and combined-cycle units (IGCC)
- Fluidized bed combustors (FBCs) firing waste coals exclusively.

In the first three subcategories, coal-fired units may employ any boiler technology, including conventional pulverized coal (PC) boiler configurations, FBCs firing non-waste coals, and stoker boilers. The US population in these subcategories is dominated by conventional PC boilers.

Although EPA recognizes the impact of boiler technology in the case of IGCCs and FBC/waste coals, it fails to separate FBCs and stokers from conventional coal units in the first three subcategories. The one stoker unit tested (Dwayne Collier 2B) and an FBC unit (Stockton 1) are among EPA's top-performing units in the bituminous category. One FBC unit is also counted among EPA's top performers in each of the subbituminous and lignite subcategories.

Because stokers and FBCs process coal in fundamentally different ways from conventional boilers, they would logically be removed from the database used to set standards for conventional coal units. In the analysis presented later in this report, FBC and stoker units have been separated from the data on conventional coal units and excluded from the analysis. Whether and how EPA should set MACT standards for these two other boiler technologies are separate questions not addressed in this analysis.

2.3.2 Classification of Units by Coal Rank

In the review of data quality, it was determined that a number of units fired coal blends during the ICR Part III testing. Coal blending is common in the industry, and EPA has proposed methods for determining a weighted-average limitation for units firing blends, based on the limitations applicable to the coal ranks individually. However, it appears that EPA has made its coal rank determination for units based on the unit's predominant coal supply (largest portion of annual Btus), rather than specific information on the mix of fuels fired during ICR Part III testing. Because EPA decided to have subcategories based on coal rank, the structure of the EPA database should be revised to accurately reflect the coal rank(s) fired during ICR Part III testing and to separate units firing blends of coals into a separate category.

A review of the reports filed with EPA by the utilities was conducted for the purpose of determining the coal rank or blend fired by units during the ICR Part III testing. From the review, we find that:

- There is one instance of apparent EPA error in transcribing the utility reports into the database

- About one-third of utility reports are silent or ambiguous on the rank of coal fired during testing.

- Most of the utility reports appear to describe the coal(s) *generally fired* at the unit, rather than *specifically fired during the testing*.

In 14 cases, the reports do not explicit state the type of coal fired by the unit, while in nine more cases the reports cite blends of coals and other fuels without indicating the proportion of each fuel that is fired. Further, while characteristics of the test coals are reported (e.g., heating value, mercury and chlorine content), the coals were not classified by rank based on the assay. As a result, there exists considerable uncertainty regarding the actual coal rank fired during testing, whenever a unit is known to receive or fire a mix of coals during 1999 based on the information on coal receipts from the ICR Part II or other data.

While the development of rank-specific MACT floors and standards for conventional coal units should ideally be based only on units firing a single rank of coal during testing, the blending of small quantities of other fuels is common in the industry, and it may be appropriate to use data from units with a *de minimis* level of blending. In this study, a unit known to have at least 90 percent of its annual coal supply of a single rank was accepted as firing coal of that rank during testing. Units blending a larger proportion of another coal or fuel, or for which the blending proportion is not known, have been separated into a subcategory of coal blends.

An effort was made in this study to clarify the coal rank fired during testing for units that potentially fired coal blends, as summarized in Table 2.2. Three types of ambiguity were investigated: cases in which EPA may have erroneously reported the coal rank; cases where units fired blends during the Part III testing, but can be treated as firing a single rank because the extent of blending was minor (less than 10 percent of annual Btus); and cases of units firing coal blends that must be classified as such because the blending exceeded threshold level of 10 percent of annual fuel supply. The table shows the coal rank classification adopted for this study in comparison to the coal rank identified by EPA and, for units firing blends, the coal or other fuel(s) fired at the unit. The last column in the table gives the source of information on which our coal rank classification was based or the data on the mix of annual fuel use that was used to

Table 2.2 Classification of Coal Rank for Units Potentially Firing a Fuel Blend During ICR Part III Testing

Classification Group	Coal Rank Classification For this Study	Coal Rank Identified by EPA	Fuel(s) Actually Fired by Unit	Determinative Information Source or Annual Coal Mix[7]
EPA Error				
Navajo 3	Sub	Bit	Utility reported subbituminous coal	
Units Classified as Firing a Single Rank				
Bailly 7 and 8	Bit	Bit	Bit/Pet. Coke (PC)	Bit = 99%
Cholla 2 and 3	Sub	Sub	Sub/Bit	Sub = 95% (industry database)
Craig C1 and C3	Sub	Sub	Sub/Bit	Sub = 97%
GRDA 2	Sub	Sub	Sub/Bit	Sub = 96%
Jim Bridger BW74	Sub	Sub	Sub/Bit	Sub > 99%
La Cygne 1	Sub	Sub	Sub/Bit	Sub =90%
Newton 2	Sub	Sub	Sub/Bit	Sub = 96%
Wabash River 1+1A	IGCC/Bit	IGCC	Bit/PC	Bit = 96%
Units Classified as Firing Coal Blend				
Bay Front 5	Blend	Bit	Bit/Sub	Sub = 85% (industry database)
Clifty Creek 6	Blend	Sub	Sub/Bit	Sub = 62%
Lawrence 4	Blend	Sub	Sub/Bit	Sub = 65%
Meramec 4	Blend	Sub	Sub/Bit	Sub = 55%
Monticello 3	Blend	Lig	Lig/Sub	n/a
Presque Isle 5 and 6	Blend	Bit	Bit/PC	PC = unknown
Shawnee Fossil 6	Blend	Bit	Bit/Sub	Bit = 89%
St Clair Power 4	Blend	Sub	Sub/Bit	Sub = 64%
Stockton Cogen GEN1	Blend	Bit	Bit/PC	PC = 89%
Valley 2	Blend	Bit	Bit/PC	PC = unknown
Valmont 5	Blend	Bit	Bit/Sub	Bit = 79% (industry database)
Polk Power	IGCC/Blend	IGCC	Bit/Sub	Bit = 68%

---

[7] Information on annual coal mix are from the ICR Part II data unless otherwise noted.

assess the probable extent of blending. The reports filed by the utilities with EPA on the ICR Part III testing were given greatest weight in this determination, but information from the ICR Part II (and other sources) on the 1999 annual coal receipts was used in conjunction with an approximate classification[8] of Apparent Rank based on ASTM D388-99 and selected contacts with coal industry personnel.

For one unit (Navajo 3), we find that the utility report filed with EPA identified a different coal rank than has been coded in the EPA database, and this unit has been reclassified to the rank reported by the utility. Ten units (at seven plant sites) are known to fire blends of bituminous and subbituminous coal, or bituminous coal mixed with petroleum coke. These units were classified as firing a single coal rank because the primary fuel constituted 90 percent or more of the annual fuel mix. Although the ICR Part II data indicates that the annual coal supply for Cholla 2 and 3 fails to reach the 90 percent threshold, an industry database that provides an accounting of coal supply based on the Form 423 data collected by DOE, reported that 95 percent of the 1999 annual coal supply is properly classified as subbituminous coal[9]. Therefore, the EPA classification of the Cholla units as subbituminous has been retained.

Fourteen units have been classified as firing coal blends during the ICR Part III testing. These units are ones for which the primary coal rank fails to reach the 90 percent threshold or for which the percentage mix of fuels could not be determined. For units firing blends of bituminous coal and petroleum coke, coke can constitute more than 10 percent of annual Btus. Therefore, units firing coal/petroleum coke blends in unknown percentages have been classified as firing blends. For Bay Front 5, the utility reports that the unit fires bituminous coal from an Arch Coal mine. In consultation with a representative of Arch Coal[10], it was determined that the coal from Arch of Wyoming's Seminoe 2 Mine is properly classified as subbituminous A. On this basis, and after review of annual coal supply data from an industry database, Bay Front 5 has been reclassified as firing a coal blend. The Valmont 5 unit has been classified as firing a coal blend based on information indicating that portions of its coal supply from the Kennecott Colowyo mine is

---

[8] The classification is approximate because ash content was reported on a total basis, and not adjusted for SO_3, and because no information is available on agglomerating character for differentiating some bituminous and subbituminous coals.

[9] Personal communication from Greg Schaefer, Vice President of External Affairs, Arch Coal.

conventionally, but erroneously, classified as bituminous, and should properly be classified as subbituminous.[11] Monticello 3 is classified as a lignite unit by EPA, but it fires a mix of minemouth lignite coal and subbituminous coal received at the Monticello plant site; Monticello units 1 and 2 fire only lignite coal.[12]

2.4 Statistical Review of Data Quality

A statistical review of the ICR Part III data was conducted to identify data that should be excluded from the analysis. The review covered the primary and secondary data used to compute average stack emissions in units of lbs/TBtu by coal rank, as adjusted for coal characteristics. The primary data are:

- Mercury concentration of stack gases (µg/dscm)
- Parameters used to convert concentrations to units of lbs/TBtu, including outlet gas flow rates, coal feed rates, and coal heating value.

The secondary data are the mercury and chlorine contents of the test coals, which are used for purposes of correlation studies in EPA's analysis.

Because EPA has used only the total mercury emissions, some data quality issues are not relevant to the proposed final rule and have not been addressed in this study:

- Precision of the individual speciated mercury concentrations, except as they determine the precision of the total stack gas mercury concentration
- Material balance between the furnace outlet and control device inlet, since EPA has conceded that the inlet gas flow rates are unreliable.

In addition, the ICR Part II data on coal characteristics have not been subjected to a thorough review of data quality. The Part II data have been accepted as reported by EPA, except where other available data clearly indicate an omission or error.

Three statistical reviews were conducted:

- Review of missing values for primary and secondary data items

---

[10] Personal communication from Greg Schaefer, Vice President for External Affairs, Arch Coal.

[11] Personal communication from Greg Schaefer, Vice President for External Affairs, Arch Coal.

[12] Personal communication from Usha Turner, TXU Energy.

- A comparison on the variability of stack gas mercury concentrations against the expected precision of the Ontario Hydro test based on ASTM D6784-02

- An evaluation of whether individual ICR Part III tests that report negative mercury removal (stack emissions greater than the mercury content of the coal) were indicative of invalid testing.

The reviews were conducted on the data originally extracted by EPA[13] from the test reports filed by the utilities participating in ICR Part III. These are the data on individual tests that predate any imputations or adjustments that may have been made by EPA.

2.4.1 Review of Missing Data

An ICR Part III emissions test should be excluded when any of the primary data items are missing, because it is then not possible to compute the total stack gas mercury concentration. The most common deficiency in the primary data is the loss of the stack gas concentration value for the particle-bound mercury fraction due to problems in conducting the Ontario Hydro test. A total of 15 tests involving 7 units were affected in this way. In two cases, the mercury concentration value for the oxidized fraction was missing, and in one case was missing for the elemental fraction.

Among the parameters required to convert gas concentrations to emission rates in units of lbs/TBtu:

- The stack gas flow rate at the outlet of the last control device was reported in all cases
- The coal feed rate was missing in all three tests for the Jim Bridger BB74 unit
- The coal heating value was reported in all cases, as were the coal mercury and chlorine contents.

It is not clear what EPA has done in the instances of missing data, but the agency appears to have estimated or inferred at least some information to retain the test results.

In addition, we reviewed the treatment of selected tests for which information such as coal mercury and chlorine content was reported as “not detected”. In general, every measurement method has a sensitivity threshold value below which it is not possible to report a quantitative value. When this occurs, it is known that the true value is small even if it is not possible to quantify it. EPA appears to have imputed a value for non-detections at one-half the sensitivity threshold of the test. Even though it is undesirable to have values imputed for non-detections recorded in the ICR Part III data, the method of imputation appears to be reasonable, assuming that the determination was done using a method of adequate sensitivity.

Based on the review for missing data, five units tested in ICR Part III must be excluded entirely, while the number of tests is reduced to 2 for three units. Table 2.3 summarizes the exclusions made based on missing data.

2.4.2 Precision of Stack Gas Mercury Measurements

The quality of the ICR Part III testing was evaluated by comparing the observed variability of emissions across the three repeat tests for each unit with the expected precision of the Ontario Hydro test. When repeat tests are conducted at a unit, the stack gas mercury concentration measurements will vary based on differences in test conditions and random variation in the outcome of each test. The testing for a unit is invalid if the observed variation from test-to-test is significantly greater than can be accounted for by known differences in test conditions and random variation.

Table 2.3 Exclusion of Tests due to Missing Primary Data

Generating Unit	Missing Stack Gas Mercury Concentrations			Missing Coal Flow Rate	Number of Tests Excluded
	Particle Bound	Oxidized	Elemental		
AESCayuga (NY) 2	1	-	-	-	1
Big Bend BB03	3	-	-	-	3
Brayton Point 3	1	-	-	-	1
Cholla 2	3	-	-	-	3
Cholla 3	3	1	-	-	3
Laramie River 1	3	-	-	-	3
Leland Olds 2	1	1	1	-	1
Jim Bridger BB74	-	-	-	Yes	3

---

[13] The data reviewed are those listed at http://www.epa.gov/ttn/atw/combust/utiltox/utoxpg.htm under the heading

To permit the comparison of the observed variability to the expected precision of the Ontario Hydro test, the variation in coal mercury content among tests was first removed from the data for each unit by adjusting mercury gas concentrations in each test to the average mercury content of the unit's test coals. This approach implicitly assumes a constant removal rate for each unit. Operating conditions were assumed to be consistent across the repeat tests because the ICR Part III testing was generally conducted under high load and steady-state conditions. If unit operations (or other test conditions) changed during the sequence of three tests, then it is possible that the data identified as showing excess repeat test variability could reflect valid testing.

The observed variability in repeated tests was determined by computing the total adjusted stack gas mercury concentration (sum of the adjusted speciated mercury concentrations) for each test, then calculating the mean mercury concentration and the variance and standard deviation of the mercury concentrations for the sample of three tests for each unit.

ASTM D6784 gives the precision of the Ontario Hydro test as follows in the form of the relative standard deviation (RSD):

- RSD <= 11 percent for concentrations > 3 μg/dscm
- RSD <= 34 percent for concentrations <= 3 μg/dscm

The RSD is the standard deviation of repeat tests expressed as a percentage of the mean value across the tests. The expected variability for each unit's testing is determined by applying the ASTM standard to the individual (adjusted) speciated values to compute the expected variances. The expected variance of the total stack gas mercury concentration is then the sum of the expected variances for the individual species.

If the tests are conducted in accordance with the ASTM methods, the observed variances of the test results for individual units will cluster around the variance that is expected from ASTM. The null hypothesis that all tests were conducted in conformance with ASTM can be evaluated using a χ^2 (chi-squared) test to determine whether the observed variance differs from the expected variance by a sufficient margin that the difference is not likely to occur by chance.

When a sample of size n is drawn from a population of known mean μ and standard deviation σ, the sample variance S^2 follows a χ^2 distribution with lower and upper bounds L and U for a 100(1-α) percent confidence interval given by:

$L = (n-1)\ S^2 / \chi^2_{\alpha/2}$

$U = (n-1)\ S^2 / \chi^2_{1-\alpha/2}$

For example, if the population mercury concentration is 2 μg/dscm, its standard deviation is $\sigma = 0.34 * 2 = 0.68$ μg/dscm, and its variance is $\sigma^2 = 0.462$, then the sample variance for three tests will lie in the following interval 95 percent of the time (95 percent confidence interval):

$L = (3-1)\ 0.68^2 / 9.35 = 0.099$ with $\chi^2_{\alpha/2} = 9.35$

$U = (3-1)\ 0.68^2 / 0.22 = 4.204$ with $\chi^2_{1-\alpha/2} = 0.22$

For a sample of n=3, the χ^2 distribution is relatively wide and the sample variance can depart significantly from that of the population due to chance. In this case, the statistical test has relatively low power for discriminating between valid and invalid tests because of the small sample size.

Table 2.4 reports the results of the χ^2 test. Three units, all firing lignite coal, display an observed variability that exceeds the upper limit of the 95 percent confidence interval. No units displayed an observed variability that fell below the lower bound of the 95 percent interval. Because there is a 2.5 percent chance that a valid test would exceed the upper 95 percent confidence limit by chance alone, one should exclude only those clearly in excess of the limit. On this basis, two units (Antelope Valley B1 and Coyote 1) should be excluded based on excessive variability in the measurements of total stack gas mercury concentrations. Monticello 1 displays an observed variance that is only marginally in excess of the upper 95 percent confidence interval and can be retained as valid testing.

Table 2.4 Units Exhibiting High Variability in Repeat Tests of Total Stack Gas Mercury Concentrations

Generating Unit	Coal Rank	Confidence Interval		Observed Sample Variance	Conclusion
		Lower Limit	Upper Limit		
Antelope Valley B1	Lignite	0.21	9.20	13.0	Exclude as Invalid
Coyote 1	Lignite	0.53	22.92	50.8	Exclude as Invalid
Monticello 1	Lignite	0.40	17.33	18.3	Accept as Valid

2.4.3 Validity of Tests Indicating Negative Mercury Removal

Twenty four percent of the ICR Part III tests reported stack emissions of mercury that were in excess of the coal mercury content. EPA has argued that this result is statistically possible in generating units with low (or zero) removal rates due to the independent variation of the coal and stack gas mercury measurements. The validity of these tests was evaluated by computing the expected variation in the measured mercury removal from coal to stack based on the precision of the testing involved. Mercury removal is defined as the mercury content of the coal less the mercury content of the stack emissions, both measured in lbs/TBtu. In a unit with zero mercury removal, the measured test result can vary from zero by an amount that depends on the variance expected in the measurement.

The mercury content of the coal depends on the mercury content in ppm and the calorific value HHV in Btu/lb according to Eq (1):

$$\text{Coal mercury}_{\text{lbs/TBtu}} = 10^6\, \text{Hg}_{\text{ppm}} / \text{HHV} \qquad (2\text{-}1)$$

The ASTM standards for precision are:

ASTM D6722: $\sigma(\text{Hg}_{\text{ppm}}) = 0.008 + 0.06\, \text{Hg}_{\text{ppm}}$

ASTM D5865: $\sigma(\mathrm{HHV}) \sim 0.5\%$ for bituminous coal

$\sim 0.7\%$ for subbituminous and lignite

Coal Hg $_{\mathrm{lbs/TBtu}}$ has essentially the same RSD as $\mathrm{Hg}_{\mathrm{ppm}}$, because the HHV value can be taken as nearly error free.

The mercury content of the stack gas in lbs/TBtu depends on the total mercury concentration $\mathrm{Hg}_{\mathrm{con}}$ of the outlet gas, the coal flow rate (C_{FR}), the outlet gas flow rate (O_{FR}), and the calorific value (HHV):

$$\text{Stack Hg}_{\mathrm{lbs/TBtu}} = 10^3\,[\,\mathrm{Hg}_{\mathrm{con}}\,O_{FR}\,]\,/\,[\,C_{FR}\,\mathrm{HHV}\,]$$

The ASTM standards for precision are:

ASTM D6784: $\sigma(\mathrm{Hg}_{\mathrm{con}})$ was determined for each test in the prior statistical review

ASTM D3154: $\sigma(O_{FR}) \sim 3$ percent

A precision of ~3 percent was assumed for C_{FR}, while HHV was taken to be nearly error free.

Based on the foregoing, the expected precision of the estimated total mercury removal is then:

$$\sigma^2\,(\text{Hg Removal}_{\mathrm{lbs/TBtu}}) = \sigma^2\,(\text{Coal Hg}_{\mathrm{lbs/TBtu}}) + \sigma^2\,(\text{Stack Hg}_{\mathrm{lbs/TBtu}})$$

For tests indicating negative removal, the null hypothesis is that the true (population) removal rate is zero and the measured negative removals occur only due to chance. The lower confidence limit (LCL) for a one-sided 97.5 percent confidence interval is the value that test results would fall below only 2.5 percent of the time. This is given by:

$$\mathrm{LCL} = -1.96 * \sigma\,(\text{Hg Removal}_{\mathrm{lbs/TBtu}})$$

In a dataset of this size, one would expect five to six tests to fall below the LCL, even if all tests were valid. Therefore, only tests substantially below the LCL should be excluded. Table 2.5 summarizes the ICR Part III tests that show negative removals of sufficiently large size to be statistically significant. Five tests (for Coyote 1, Presque Isle 9, and Wyodak BW91) show negative removals that are only marginally below the LCL and can be retained as valid. Tests to be excluded include: all three tests from the initial test sequence at Gibson 3, one subsequent test at Gibson 3, one test at Monticello 1, and two tests at Platte 3.

The initial testing at Gibson 3 showed mercury emissions at a level three times that of the coal mercury content. The large deviation in mercury measurements and the results of this review suggest that the mercury mass balance (mercury in = mercury out) was being violated at the time of the initial testing.

Tables 2.6 and 2.7 summarize the units and individual tests that should be excluded from the database based on the reviews of data quality.

Table 2.5 Units Exhibiting Statistically Significant Negative mercury Removals

Generating Unit	Test Run	Observed Mercury Removal lbs/TBtu	97.5% LCL lbs/TBtu	Conclusion
Coyote 1	3	-3.80	-3.7	Accept as Valid
Gibson 3 (10/99 testing)	1, 2, 3	< -18.0	<-5.8	Exclude as Invalid
Gibson 3 (03/00 testing)	3	-6.1	-4.0	Exclude as Invalid
Monticello 1	2	-28.4	-14.2	Exclude as Invalid
Platte 3	2, 3	-8.0	-3.5	Exclude as Invalid
Presque Isle 9	1	-2.7	-2.5	Accept as Valid
Wyodak BW91	1, 2, 3	< -3.8	-2.5	Accept as Valid

Table 2.6 Exclusion of Generating Units Based on Review of Data Quality				
Generating Unit	**Coal Rank**	**Review #2: Missing Primary Data**	**Review #3: Precision**	**Review #4: Negative Removal**
Polk Power 1	Bit			
Lawrence 4	Sub			
Big Bend BB03	Bit	Speciated mercury		
Cholla 2	Bit/Sub	Speciated mercury		
Cholla 3	Bit/Sub	Speciated mercury		
Laramie River 1	Sub	Speciated mercury		
Jim Bridger BB74	Sub	Coal Flow Rate		
Antelope Valley B1	Lig		Exclude	
Coyote 1	Lig		Exclude	
Gibson 3 (10/99)	Bit			Exclude

Table 2.7 Exclusion of Individual Test Results Based on Data Quality				
Generating Unit	**Coal Rank**	**Test Run**	**Reason for Exclusion**	**Tests Remaining**
AES Cayuga (NY) 2	Bit	1	Missing Primary Data	2
Brayton Point 3	Bit	3	Missing Primary Data	2
Leland Olds 2	Lig	3	Missing Primary Data	2
Gibson 3 (03/00)	Bit	3	Precision	2
Monticello 1	Lig	2	Precision	2
Platte 3	Sub	2, 3	Precision	1

3. CRITIQUE OF EPA METHODOLOGY FOR DETERMINATION OF MERCURY MACT FLOORS

3.1 Introduction

In its analysis memorandum[14] EPA presents the methodology used to determine mercury MACT floors for each of the existing unit subcategories. The objective of the analysis is to determine the average emissions limitation that the best performing units in a subcategory can achieve "under the most adverse conditions which can reasonably be expected to recur".[15] EPA understood that there is a high degree of variability in the ICR Part III data and made an effort to account for unit performance variability in deriving floors that could be met by the best performing units. Nevertheless, the EPA methodology has certain analytical weaknesses and errors:

- EPA did not use the full set of data available on the US population to identify the top performers. EPA did not use information on the frequency of control configurations in the population to adjust the ICR Part III sample to be proportionately representative of all units. Further, EPA did not utilize stack emissions data generated in field tests of CEMs to characterize unit emissions over extended periods of time.

- EPA's definition of "most adverse conditions" is limited to variations in coal characteristics and does not address variations in unit operating conditions that are not sampled in the testing program.

- EPA develops emissions control performance correlations from the ICR Part III data and applies them to predict emissions of individual units, but it does not account in its emissions estimates for the uncertainty associated with those correlations.

- EPA underestimates the degree of variability in control performance of the top-performing units. The unit performance variability is determined solely based on the emission test results for the units selected as top performers, without considering the evidence of performance variations present in the results of testing for other units with the same technology. This approach is therefore subject to self-selection bias – i.e., some of the units selected as top performers are so treated only because the control performance was favorable at the time of the tests. The variability has been underestimated because units in one tail of a distribution are more like each other than other units drawn from the same population and found elsewhere in the distribution.

---

[14] OAR-2002-0056-0006. Analysis of variability in determining Maximum Achievable Control Technology (MACT) floor for coal-fired electric utility steam generating units." Memorandum from William H. Maxwell to Utility MACT Project Files.

[15] National Lime Association v. EPA, US Court of Appeals for the District of Columbia Circuit, 627 F.2d 416, Decided May 19, 1980.

- EPA has not adjusted the proposed MACT standards for the shift from Ontario Hydro stack tests to CEMs for demonstrating compliance.

The EPA MACT methodology is reviewed step-by-step in this chapter, and improved analytical methods are discussed to address limitations in the EPA methodology.

3.2 Counting Top Performing Units

EPA only considers units for which Part III test results exist when identifying the top performers in each category. EPA counts as top performers those units that constitute the top 12 percent *of the ICR Part III sample* when a subcategory has 30 or more units; and the top five units *in the sample* when a subcategory has fewer than 30 units. Hence, the top performers for the bituminous and subbituminous subcategories amount to four units each based on 12 percent of the sample counts, and the top five sample units are selected for the lignite subcategory. This view appears to derive from the statutory language regarding units "for which the Administrator has emissions information".

A sample-based counting procedure is inappropriate for data from a stratified survey that, by definition, does not represent the population proportionately. As was shown in Chapter 2, the control technology cells into which nearly 60 percent of the US generating unit population falls account for only 11 percent of the ICR Part III sample, while technologies representing only nine percent of the US population account for 29 percent of the sample. In such circumstances, the selection of the top 12 percent *of the sample* will necessarily represent a much smaller proportion of the actual US population of generating units.

When data are derived from a stratified sample of a population, accepted statistical practice does not permit treating the sample as directly representative of the population. Instead, whenever the data is used to draw conclusions about the population, expansion factors must be defined based on the sample stratification plan and used to adjust sample counts so that they represent proportions in the population. From an analytical viewpoint, the sample data give emissions information that characterize all units in the population, and we are consistent with the statutory language when counting units based on the population. Further, since the standard is an annual emissions limit, there is no reason to use the short-term test results as the only relevant emissions information. Indeed, those data provide no direct evidence of annual performance of those units.

The use of the ICR Part III test results to develop knowledge of emissions performance and then apply that knowledge to the available detailed information on the entire population is substantiated by the fact that EPA has created a dataset that estimates emissions in 1999 for all units in the US based on the application of mercury removal rates by technology (derived from the ICR Part III testing) to an inventory of all power plants in operation. The same concept can be applied to characterizing the top performers in the entire population.

An improved approach would be to use population counts as the basis to identify the best performing units. That is, one should count as top-performing units those units in the ICR Part III sample (however many) that represent the best performing 12 percent of the US generating unit population in the subcategory, whenever there are 30 or more such units in the population. When there are fewer than 30 units in the US population, one should choose those units in the sample that represent the same percentage that the five best units would represent in the population. This approach is consistent with the idea of identifying "best practices", and it eliminates the bias that is inherent in counting based on the non-representative proportions of the stratified test sample.

3.3 Ranking Units Based on Emissions

The method chosen for ranking units in a subcategory is an important step in identifying top performers. EPA ranks units based on the average emissions during the three tests conducted for ICR Part III, thereby ranking based on the average coal fired during the tests and the performance of the units on the day of the test.

EPA's choice is only one of many alternatives. A unit that happens on that day to fire a coal having favorable emissions characteristics has an increased chance of appearing among the top performers and, thereby, have its control technology identified as among the best performers. Further, a unit that happens to exhibit favorable emissions performance due to the operating and environmental circumstances prevailing on the day of the test has an increased chance of appearing among the top performers, even if repeat testing over an extend time period might reveal that the unit would exhibit poorer long-term emission performance than other units.

The ICR Part III results amount to a snapshot in time on the mercury emissions performance of the units tested. A substantial part of the differences that are observed among units in the testing is actually related to operational, environmental, and other (presently unidentified) factors that caused some units to display high emissions when tested and other units to display low emissions. Contributing factors include:

- Differences in coal characteristics among units that influence mercury emissions and are not controlled for in the analysis as done for chlorine content.
- Differences in unit-specific operating conditions, such as stack gas temperature or ash carbon content, or in environmental conditions, such as ambient temperature, pressure, and humidity.
- Detailed differences in control system design or operations among unit, such as the length of duct runs or the fabric cloth type or fabric cleaning methods or methods.

The two independent test sequences on the Gibson 3 unit demonstrate the extent to which unit performance can vary from one time period to another based on operational factors. When the objective is to establish an emissions limitation that will apply on an annual-average basis, the EPA method of ranking units based on the average of three short-term tests does not appear to be a defensible approach.

An improved approach would be to rank units based on their estimated emissions performance over an annual time period and select those units that display the best emissions performance on that basis. The collection of actual emissions test data on a variety of units and technologies operated over an annual period is clearly the preferred method for identifying the best emissions performance on an annual basis. In the absence of such test information, one may utilize statistical methods to estimate annualized unit performance based on the emissions-related characteristics of the coals that would be fired during a year and estimates of emission performance for technology-based groups of units. The statistical methods used for estimation may be simple – as in the use of average mercury removal factors for groups of units sharing a common emission control technology configuration – or they may involve the use of correlation equations that predict mercury emissions as a function of coal characteristics and, potentially, other relevant factors.

Given the high degree of variability in the performance of individual units on short-term tests, statistical techniques applied to data for groups of units are more likely to provide a reliable basis for assessing the emissions performance that can be achieved over an annual period, than any method for ranking based on emissions during the ICR Part III testing. In subsequent chapters of this report, MACT floors are determined based on ranking units according to the annual emissions that would be expected when firing representative coal of the given rank. This approach separates the performance of the unit from the coal that it happened to fire on a given day or a given year. By doing so, we evaluate all units on the same basis to determine which perform best in removing mercury from a particular coal.

3.4 Exclude Incomplete or Invalid Data Before Ranking Units

EPA's memorandum does not describe efforts to evaluate data quality and to exclude from the analysis any test data that are incomplete or invalid, and we are aware of no instances in which EPA has excluded ICR Part III test data from its analysis. Chapter 2 of this report described the efforts made to qualify the ICR Part III data based on a comprehensive review of data quality. The specific exclusions of test information from the analysis presented later in this report were summarized in Table 2.6 and 2.7.

3.5 Rank Units Based on Annual Emissions for Boiler Technology Subcategories

EPA's analysis develops MACT floors for five subcategories of units, specifically bituminous, subbituminous, lignite, IGCC, and FBC/Waste Coals. As discussed in Section 2.3, we believe that EPA should restructure its database and analysis to better account for coal-fired boiler technologies and the prevalence of coal blending in the power generating industry. Specifically, it is more appropriate to IGCC, FBC, and stoker boilers from conventional boilers when defining subcategories and determining MACT floors for the three coal ranks.

Conventional units include pulverized coal and cyclone boilers that constitute the large majority of generating units in the US population. Units firing a blend of coals, in which less than 90 percent is accounted for by a single rank, should be separated in the database and not used in the analysis conducted for these three subcategories. We do not present analyses of performance characteristics or MACT floors for IGCC, FBC, stoker boilers or for conventional units firing coal blends.

3.6 Define the Adverse Conditions to be Considered in MACT Floors

Having selected units and technologies that display the best annual emissions performance in the subcategories, one must then consider the most adverse conditions under which the units may reasonably be expected to operate. EPA defines the worst-case condition as operating on the coal that would give emissions at the 97.5th percentile, considering all shipments received by the unit in 1999. Thus, the most adverse conditions are defined as firing a near-worst case coal without regard to the occurrence of adverse operating conditions that were, in general, not adequately sampled in the ICR Part III data.

Selecting a single shipment from among the coals used within a year would be appropriate if compliance were determined by short-term testing – i.e., the average of three Ontario Hydro tests – because a unit might operate on such a coal during the compliance test. However, for purposes of an annual-average emissions limitation, EPA's choice means that the adverse condition is implicitly that a unit operates for an entire year on a coal with emissions characteristics that were encountered with only 2.5 percent of the coal supply for the year 1999, as represented by the ICR Part II data.

Given the preceding recommendations that units be ranked based on estimates of annual emissions for purposes of developing an annual-average standard, we believe that the adverse coal is more properly defined as an annual coal supply having adverse characteristics that lead to increased emissions on an annual basis. Because of the snapshot nature of the ICR Part III testing, and the resulting inability to identify specific units as being top performers on an annual basis directly from that data, it is more appropriate to define the adverse coal conditions in terms of actual annual coal supplies received by units during 1999. For the MACT analysis presented later in this report, the ICR Part II data were used to identify an annual coal supply in each rank that, if fired in a top-performing unit, would produce an annual emissions level at the 90th percentile of the emissions distribution. In this approach, one can be assured that the assumed adverse coal (and worse) has been encountered in practice, because it was actually received by a plant during 1999.

3.7 Estimate MACT Floors, Considering Unit Performance Variability

In the EPA analysis, the measured emissions of top-performing units are adjusted to reflect adverse coal conditions using a combination of correlation equations and average removal rates and the characteristics of the adverse coal. Correlations are used for two technology groups[16] where the statistical evidence indicated that coal characteristics (in practice, limited to chlorine content) influenced mercury removal rates in the control devices and the correlations were deemed by EPA to be of sufficient explanatory power. Otherwise, a fixed, average mercury removal rate determined from the ICR Part III data was used. To determine the MACT floor, EPA computes a mean value and the standard deviation of the estimated emission rates when firing the adverse coal for the group of top-performing units. An effort is then made to account for variation among these units by computing the upper 97.5th confidence limit of the mean – that is, the emissions level that would be exceeded only 2.5 percent of the time for the average of the top-performing units.

The use of statistical correlation equations to make predictions for a population that has been sampled is a common research technique. Whether the use of correlations is appropriate in any context depends largely on the strength of the statistical evidence for the correlation, as supported by an understanding of the mechanism(s) represented by the correlation, and the extent to which the correlation is used in circumstances that fall within the range of the original data.

The most important weaknesses of EPA's analysis in this area are:

- There is no accounting for the uncertainty bounds of its correlations (and average mercury removals) when estimating the performance of units under adverse conditions
- The reliance on a very narrow view of unit performance variability that substantially undercounts the variability actually present in the ICR Part III data and the sample of units tested.

---

[16] Correlations are developed by EPA for the technology groups of fabric filters with dry scrubbers and FBC units firing non-waste coals. The EPA analysis considers only technology groups represented in its top performing units and is not a comprehensive analysis of technology performance.

We will not comment specifically on the correlations developed by EPA, because a later chapter in this report presents a new, more comprehensive analysis of emissions performance by technology, but we will discuss the issue of unit performance variability.

EPA takes a narrow view of unit performance variability by asserting that only the variation observed among the top-performing units (adjusted to worst case conditions) can be considered. This is a statistically flawed view that can be shown to suffer from “self-selection” bias. That is, take a population of units that display inherent, random variation in a characteristic. Draw a sample from the population and select those units lying at one tail of the sample distribution for the characteristic. The computed standard deviation for the selected units will always be substantially less than the standard deviation that actually characterizes the population. This is because chance has caused some units to fall into the tail of the distribution, and those units are much more similar to each other than they are to units falling elsewhere in the full distribution.

Top performing units on an annual basis will display varying emissions levels in short-term testing and may be found throughout the ICR Part III database. The variability among top annual performers will be substantially greater than observed within the small group of units having the lowest emissions rates in the short-term testing.

As will be shown in the later analysis of technology-based emissions performance, EPA’s method for determining unit variability substantially understates the actual unit variability that exists in the data and that will be characteristic of the top annual performing units when subjected to short-term testing. Given the characteristics of the ICR Part II data, the unit performance variability in short-term testing will be best estimated by the variation of individual units around a correlation line (or average mercury removal) that represent the expected performance of their technology group. Chapter 5 of this report presents an alternative derivation of MACT floors for existing units that provides a full and proper consideration how the uncertainties in the analysis and unit performance variability affects the emissions limitations that can be met by the units that are top performers on an annual basis.

3.8 Conversion of EPA's MACT Floors to Match the Form of the Standard

EPA's MACT floor analysis is based on emissions testing conducted according to the Ontario Hydro test. Yet, EPA proposes to determine compliance on a rolling, annual-average basis using continuous emissions monitors (CEMs). Logically, the shift from a short-term measurement to the annual average of a continuous measurement using a different test procedure obligates EPA to adjust the MACT floors for:

- The bias (if any) of the CEM methods relative to the OH test. That is, the extent to which the CEM methods would give different values for mercury emissions on average.
- The additional variability introduced by CEMs measurements, relative to the Ontario Hydro test, as appropriately attenuated by the averaging of measurements taken over an annual period.

The goal should be that a unit's chances of demonstrating compliance should not be changed by the shift to CEMs.

In addition, EPA did not account for the attenuation of the repeat test variability that remains in its MACT floors (from the use of data based on the average of three tests) as the result of shifting to an annual average. Instead, EPA has equated its MACT floors, as developed from the analysis of the ICR Part III data, to an annual-average emissions limitation as measured by an (unstated) CEM method.

4. STATISTICAL ASSESSMENT OF MERCURY CONTROL PERFORMANCE OF DATA AND METHODOLOGIES USED IN VARIOUS CONTROL SYSTEMS BASED ON ICR PART III DATA

4.1 Introduction

The physical and chemical processes of mercury liberation, capture, and emission in the stack gases of coal-fired power plants are complex. The current understanding of these processes[17] is that mercury vapor is liberated during the combustion of coal and undergoes various chemical reactions that transform vapor into a mix of elemental, oxidized, or particle-bound mercury compounds (or species). Elemental mercury is thought to be the most difficult to remove from the stack gas stream, with its capture largely dependent on adsorption onto an active surface. Oxidized mercury is thought to be more easily removed from stack gases through existing control devices – typically SO_2 scrubbers or PM control devices – with an efficiency of capture that depends on the compounds formed by reaction with other stack gas constituents such as chlorine. Particle-bound mercury can be removed by devices that are designed for PM control. The speciation of mercury – the mix of elemental, oxidized, and particle-bound compounds – may be affected by the combustion characteristics of high and low rank coals and by emission control devices such as Selective Catalytic Reduction (SCR) that may catalyze changes in combustion gases.

A series of research studies by EPA and industry parties have used the ICR Part III data to examine the mercury emissions control performance of various technologies. The performance assessments have taken the forms of average mercury removal rates and correlation equations that relate mercury removal to other factors, most often the chlorine content of the coal. For some technologies, coal chlorine content appears to be a statistically significant predictor for mercury removal efficiency. This effect is conventionally explained as indicating the formation of oxidized mercury compounds in the stack gas, through the action of chlorine as a catalyst, that are removed by control devices. Correlation equations can contribute empirically to the understanding of mercury emission control processes and the assessment of emissions control technologies and are widely used as predictive tools.

In light of the potential importance to the understanding of mercury control, a new statistical analysis of the mercury control capability of existing control devices has been conducted. While this topic has been studied extensively, we find relatively little consideration in past studies of the uncertainties of the correlations and essentially no treatment of the confidence bounds of the resulting predictions. These uncertainties affect the ability of the analyses to estimate emissions performance for top-performing units or technologies under adverse conditions, and therefore, they are highly relevant to the development of MACT floors. The full and proper treatment of predictive uncertainty, based on the confidence bounds for the correlations and the residual (unexplained) variation in the data, is the primary motivation for the analysis presented here.

That "correlation does not imply causation" is an analytical truism that is honored more in theory than in practice. This truism is a shorthand expression of the reality that statistical studies may determine that a particular factor or variable exhibits a statistically significant association with a dependent variable of interest, but that fact alone does not establish that the factor or variable exerts a direct effect on the dependent variable. This is because the particular factor being studied may be only a surrogate for other, unidentified factors that actually exert the direct effect. In such cases, the factor being studied is correlated with the unidentified causal factors and, as the studied factor is varied in data, the actual causal factors also tend to vary. Therefore, statistical studies are subject to the potential misidentification of the true causal factors affecting the dependent variable of interest. The level of scientific certainty usually cannot be reached unless the results of statistical studies are explained by fundamental theory and are confirmed by controlled experiments, in which the identified factors are varied independently of all other factors to demonstrate a direct effect.

To our knowledge, it has not been demonstrated that the independent variation in coal chlorine content has a direct effect on mercury removal performance of existing emissions controls. In the context of mercury control, this means that the body of evidence indicating that high coal chlorine content is favorable for the removal of mercury by existing control devices does not

---

[17] "Control of Mercury Emissions from Coal-Fired Electric Utility Boilers". Environmental Protection Agency, Air Pollution Prevention and Control Division. (NC 2004).

reach the level of scientific certainty. Given the absence of scientific certainty in regard to the effect of chlorine (or other coal characteristics) on mercury removal and emissions, it is critical that statistical correlations be used only in conjunction with a full appreciation of their confidence bounds.

4.2 Statistical Methodology

The analysis of mercury removal performance as a function of coal chlorine content (and potentially other factors) is based on the application of conventional regression analysis to the ICR Part III data. Technology groups are defined from the combinations of SO_2 and PM controls on the tested units. SO_2 controls are categorized as: Unscrubbed; Wet FGD; or Dry FGD (spray drier absorber), while PM controls are categorized as: Fabric Filters (FF); cold-side ESP (ESPc); hot-side ESP (ESPh); and mechanical PM controls.

A mathematical model of the following form is used to represent the potential dependence of mercury removal on chlorine content:

$$\text{Hg Removal} = 1 - A * \exp(B * Cl_{lbs/TBtu}) \qquad (4\text{-}1)$$

This form is mathematically similar to that used by EPA, but we choose to measure the coal chlorine content in units of lbs/TBtu, rather than ppm, to normalize for the varying Btu content of coals. The behavior of the functional form is that it permits a non-zero mercury removal rate at zero chlorine content, which rises with increasing chlorine content to reach an asymptotic value at high chlorine contents. Mercury removals greater than 1.00 are not permitted by the functional form, but negative removals are allowed and the mercury removal rate at zero chlorine content may be either negative or positive.

The coefficients of the model are estimated by the application of ordinary least squares to data that have been transformed to the space of natural logarithms in the dependent variable. The natural log transformation linearizes the regression equation and has the effect of tending to normalize what is otherwise a right-skewed (log normal) distribution of emissions values. The regression equation is:

$$\ln (\text{Stack emissions} / \text{Coal Hg content}) = \ln(A) + B * Cl_{lbs/TBtu} \qquad (4\text{-}2)$$

in which the dependent variable is the natural log of fractional emissions, the intercept term gives the mercury removal at zero chlorine content according to the relationship $\text{Hg Removal}_{\text{zero chlorine}} = 1 - \exp(A)$, and the B term gives the sensitivity of emissions to chlorine content.

The use of the ICR Part III data for this analysis presents what is known as a "mixed effects" problem. The chlorine term is a "fixed effect", because it is systematic and controllable by the experimenter. Surrounding the systematic effect of chlorine is random variation among units and across repeat tests for individual units. The variability associated with units is termed a "random effect" because the value is not controllable by the experimenter, but is determined (prior to testing) at the time a unit and the test conditions are selected. Variability associated with repeat tests is a second random effect, which is nested within the unit effect. Further, the replication factor is not constant, because, after controlling for data quality, some units are represented by one or two tests, not three.

The usual regression approach – entering the individual test data for units into a pooled regression – has a key limitation with nested data because it assumes that each data point is independent. Here, the repeated tests of an individual unit are not fully independent, because each shares the same unit effect. If a unit is destined to exhibit high or low emissions on the day of the test, each of its repeat tests will also tend to be high or low. Because the repeat tests are not fully independent, the usual pooled approach will overstate the actual degrees of freedom in the problem and will imply greater statistical power than actually is present in the data.

Specialized software exists to estimate mixed effects regression problems with nesting, but it is seldom used, difficult for most audiences to understand, and not available in the short timeframe permitted for this study. In its absence, the study has been conducted on data that are the averaged values of stack emissions and coal mercury and chlorine content for each unit. As a result, each unit is represented by one data point. It should be noted that this approach is subject to caveats, because it does not make use of the part of the repeat tests that is independent, and therefore, does not make full use of the data. It has also been necessary to give equal weight to all units, even when some are represented by fewer than three tests.

Nevertheless, this approach should be favored over the usual pooled analysis because it gives a more accurate appraisal of the statistical power of the data and, therefore, the confidence limits of the correlations. A relatively small number of units (80) were tested in ICR Part III, and the sample sizes for individual technology groups are relatively small (4 to 16 units). Therefore, the data have relatively modest statistical power for the analysis of mercury removal performance by the top-performing technologies.

Having structured the analysis to use unit-average values of emissions and coal characteristics, the steps of the analysis are as follows:

1. A chlorine term is estimated for the technology groups whenever the data demonstrate a statistically significant effect of chlorine on emissions. In such cases, the resulting model is a correlation equation of the form given by Eq. 4-1, where the coefficients ln(A) and B are estimated by the regression.

2. When chlorine is shown not to be statistically significant, a fixed mercury removal rate has been estimated by fitting an intercept term (only) to the data. The fit takes place in natural log space and is equivalent to determining the mean logarithm of fractional emissions. When exponentiated, the result is referred to as a fixed mercury removal rate, but it has not been estimated as an arithmetic average of data.

3. All units of a given technology have been used to estimate the correlations, regardless of coal rank and blends. Statistical tests were applied to the residuals to confirm that differences in mercury removal by coal rank are fully accounted for. Separate correlations were estimated for each coal rank whenever statistically significant differences remained, as was the case for the ESPc technology group. A cap was imposed on correlation predictions for lignite units, as explained in the final section of this chapter, to assure that the analysis did not over-predict the mercury removal capability actually demonstrated in lignite units.

4. Statistical tests have been used to determine whether the mercury removal performance of similar or related technology groups can reliably be distinguished using the available data. Technology groups have been combined, and a composite correlation estimated for the combined group, when the data indicate that mercury removal performance cannot be distinguished.

5. Coal SO_2 and ash content were also tested for effects on mercury removal, but given the relatively modest statistical power of the available data, no such terms have been included in the correlations.

Confidence intervals have been determined for each correlation selected to represent a technology group. Two types of confidence intervals are formed at the 95 percent confidence level:

- 95 percent confidence interval for the regression
- 95 percent confidence interval for the population

Because the regression line represents the average unit, the confidence interval for the regression represents the uncertainty in the prediction of the average. The confidence interval for the population is wider and represents the range within which individual data points (averages for three tests of a single unit) may fall.[18]

Two components of the variance can be recognized in the data:

- $\sigma^2(U)$ = unit performance variability
- $\sigma^2(T)$ = repeat test variability

Individual data points in the regression that are the average of three repeat tests will be subject to a variance of $\sigma^2 = \sigma^2(U) + \sigma^2(T)/(3-1)$. Following completion of the correlation study for a technology group, the correlation equation (or the average mercury removal rate) was used to predict emission values for each of the individual tests that were performed on units in the group. A components of variance analysis[19] was used to decompose the observed residual variance into estimates of $\sigma^2(U)$ and $\sigma^2(T)$ for the group. These variance components will be used later in developing confidence bounds for the emissions levels expected from top-performing technologies under adverse conditions.

4.3 <u>Results of Correlation Analysis</u>

There are potentially 3* 4 = 12 technology groups formed from the combinations of three SO_2 control technologies (Not Scrubbed, Wet FGD and Dry FGD) with four PM control technologies (FF, ESPc, ESPh, and Mechanical PM). Not all combinations are represented in the data, and fewer combinations are represented with sufficient sample size to permit independent analysis. After examination of all technology groups, it was determined that two composite technology

---

[18] A treatment of confidence intervals for regression models can be found in J.S. Milton and J.C. Arnold, "Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences." (Irwin McGraw-Hill, Boston, MA1995). Section 12.4 (pp. 488-492)

groups – Fabric Filters (with or without scrubbing) and Wet Scrubbing with ESP (hot- or cold-side) – exhibited a sensitivity of mercury removal performance to coal chlorine content. The remaining technologies exhibited no sensitivity to chlorine content and were therefore represented by fixed mercury removal rates. These technologies were divided into five groups: Dry FGD with ESPc, ESPc with bituminous coal, ESPc with subbituminous or lignite coal, ESPh, and Mechanical PM controls. The chlorine-sensitive technology groups give the highest mercury removal rates on most coals, while among the other technologies, only ESPc with bituminous coal is able to achieve more than a *de minimis* removal rate.

4.3.1 Fabric Filter Technology Group

The database for the fabric filter (FF) technology group consists of test results for 16 units in total, of which eight are unscrubbed, two have a Wet FGD unit following the fabric filter, and six have Dry FGD followed by the fabric filter. When estimated using unit-average data, we can detect no differences in mercury removal performance among the configurations with and without scrubbing, once chlorine content is accounted for. The regression analysis is summarized in Table 4.1. The dependence of mercury removal is strong[20] and highly significant, and the regression indicates a mercury removal of 34 percent at zero chlorine content, although this is not statistically different from zero given the uncertainty in the regression. The regression based on chlorine content explains 61 percent of the observed variation among units in the group.

Figure 4.1 shows the unit-average data for this group in the semi-log space in which the regression is estimated. Note that the vertical axis is fractional emissions, rather than mercury removal. The middle line in the figure is the regression line itself, while the upper and lower lines are the 95 percent confidence interval for the population (meaning individual data points).

---

[19] A treatment of components of variance analysis can be found in G.W. Snedecor and W.G. Cochran, "Statistical Methods." (The Iowa State University Press, Ames, IA, 1967). Sections 10.16 through 10.19 (pp. 285-294).

[20] Coal chlorine contents measured in lbs/TBtu have values ranging from 1,000 lbs/TBtu to greater than 100,000 lbs/TBtu. Coefficient values for the chlorine term on the order of 1E-5 may appear to be numerically small, but represent a strong effect.

Table 4.1 Correlation for All Fabric Filters With or Without Scrubbing					
Mercury Removal = 1 – exp(A + B $\text{Chlorine}_{\text{lbs/TBtu}}$)					
	N Units	R2	A	B	Mercury Removal At Zero Chlorine Content
All Configurations	16	60%	-0.446 (t= 1.22)	-3.042 E-5 (t= 4.59)	36%

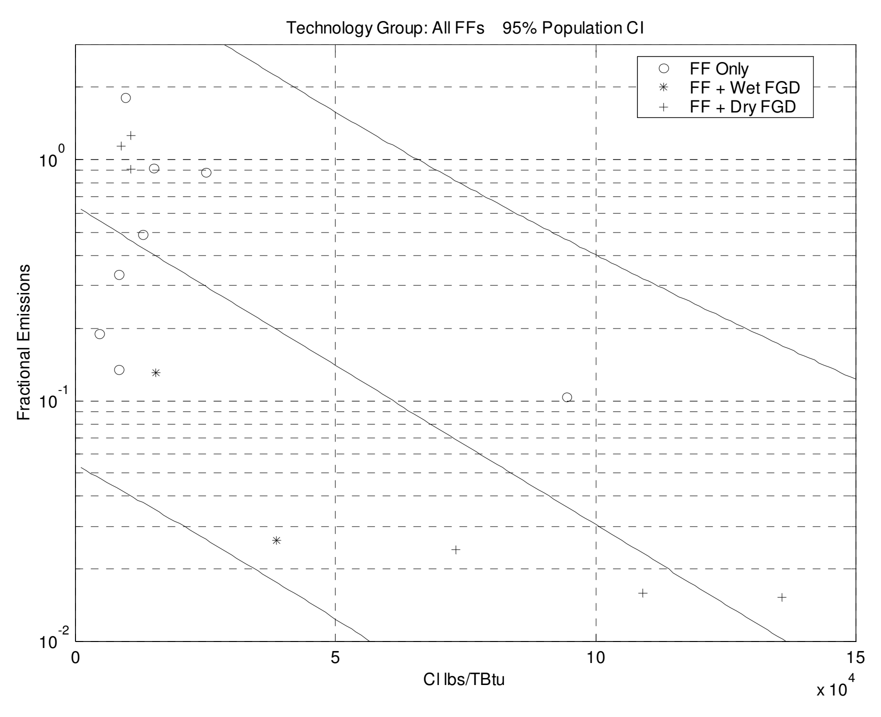

Figure 4.1 Mercury Removal Performance for Fabric Filter Technologies

One can see that the data for units with FF only (unscrubbed) and for units with FF combined with Wet or Dry FGD all lie within the 95 percent confidence bounds of the population. The FF only and FF + Dry FGD configurations are represented by points at both high and low chlorine values and by points that straddle both sides of the regression line. The FF + Dry FGD configuration is represented by only two data points (units); while these fall on the same side of the regression line, no statistical significance can be assigned to the result. Thus, we must conclude that the mercury removal performance of fabric filters alone cannot be distinguished from the performance of fabric filters combined with wet or dry scrubbers, at least within the modest statistical power of the available data to detect such an effect.

Statistical tests were conducted to determine whether the residuals from the correlation indicated the presence of statistically significant differences associated with coal rank. For this technology group, no differences by coal rank could be detected with statistical significance. However, this finding is, in large part, the result of the small sample sizes involved and not a conclusive demonstration that no such differences exist in fact. In particular, the two lignite units with fabric filters both show poorer mercury removal than predicted by the correlation line, and reports by operators of lignite units indicate both poor control and high variability in emissions. Thus, lignite units with fabric filters may not achieve in practice the level of mercury removal that is predicted by the correlations.

As seen in the preceding figure, the data are characterized by a relatively high variability in mercury removal at low chlorine contents. The average of three tests for individual units ranges from -80 percent removal (for a low mercury content coal) to as high as +85 percent removal. Regression studies that included coal ash and SO_2 content and stack gas temperature did not explain the variability at low chlorine levels. For selected units, a review of the utility test reports was made and discussions were held with unit operators in an effort to identify other factors that might explain the variability in performance. However, no additional factors were identified that could be used to improve the predictive power of the regression model.

The cause of the high variability at low chlorine content remains unknown. High levels of unburned carbon in the fly ash, which could act as an active surface for adsorption of mercury compounds, is a possible reason for some units to demonstrate better mercury removal than others in the absence of coal chlorine content. In field tests of advanced technologies, EPA has identified high carbon content of fly ash, as indicated by loss on ignition measurements, as a contributing cause for the atypically high removals at some units that were equipped with other controls (not fabric filters)[21]. Fly ash characteristics were not measured during the ICR Part III testing, and it is not possible to assess the potential contribution of fly ash characteristics to the

---

[21] See discussion of Brayton Point Unit 1 and Salem Harbor Unit 1 in the EPA paper referenced in footnote 18.

observed variation in mercury removal among units for the fabric filter or other control configurations.

The presence of high variability among units necessarily implies wide confidence bounds on the performance of fabric filters at low chlorine contents. Figure 4.2 shows the 95 percent confidence limits for the regression line – i.e., the uncertainty in predictions made by the correlations. Note that the vertical axis is mercury removal and the graph is linear (not logarithmic). The confidence bounds in the figure represent the range of uncertainty in conclusions that may be drawn in regard to the mercury removal performance of this technology.

Thus, the data support the conclusion (with 95 percent confidence) that fabric filters achieve between 75 percent and 90 percent removal (for the average of three tests) for coals with chlorine content of 50,000 lbs/TBtu . This result is applicable only to certain bituminous coals, because this chlorine level exceeds the highest level recorded for subbituminous or lignite coals in the ICR Part II data. At low chlorine contents, the average mercury removal performance is poorly characterized by existing data and may range from as low as –30 percent removal to as high as +70 percent removal with 95 percent confidence. Thus, for this technology, if adverse coals are defined as ones with chlorine content below approximately 10,000 lbs/TBtu, we cannot say based on present data, with 95 percent confidence, that a fabric filter will remove any of the mercury present in the coal, on average, when retrofitted to a population of units.

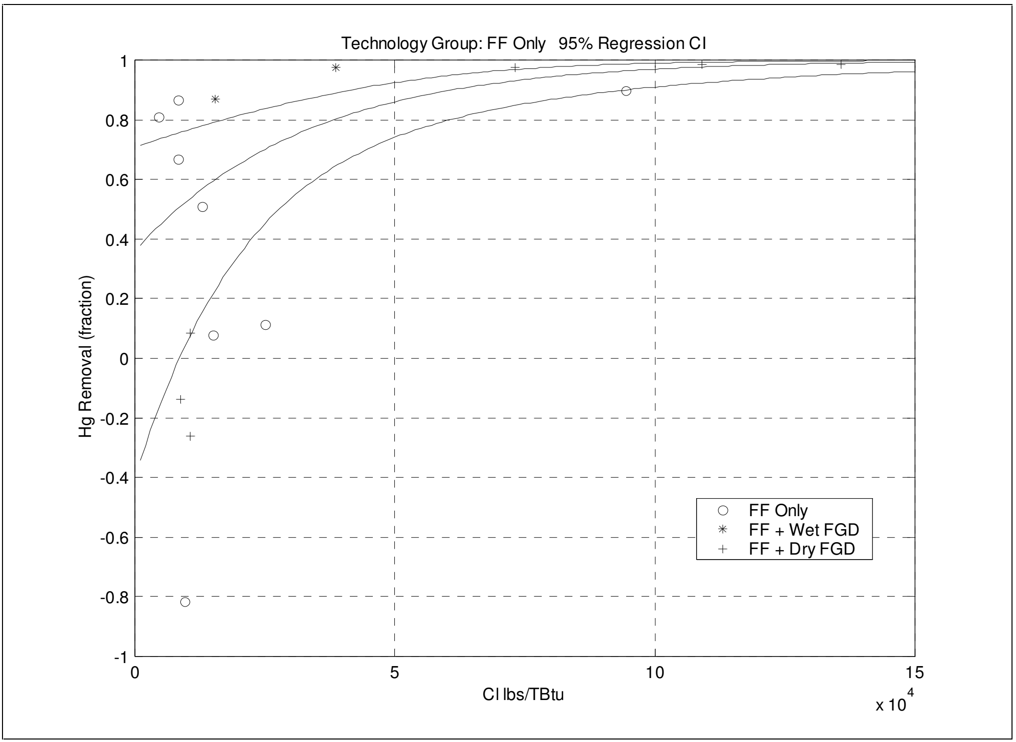

Figure 4.2 95 Percent Confidence Intervals on the Performance of Fabric Filter Technologies

4.3.2 Wet FGD with ESP

The Wet FGD with ESP technology group consists of 11 units in total, of which five have a cold-side ESP units and six have a hot-side ESP. When estimated using unit-average data, we can detect no differences in mercury removal performance among the two ESP configurations, once chlorine content is accounted for, and we find a statistically significant dependence of mercury removal on chlorine content that is about one-third less than that for fabric filters. The regression analysis is summarized in Table 4.2. The regression indicates a mercury removal of 3 percent at zero chlorine content, which is not statistically different from zero. Thus, mercury removal in Wet FGD units with ESP appears to be significantly associated with the concentration of chlorine in the coal and stack gases. The regression based on chlorine content explains 63 percent of the observed variation among units.

Table 4.2 Correlation for Wet FGD + ESP (Cold- or Hot-side)

Mercury Removal = 1 – exp(A + B Chlorine$_{\text{lbs/TBtu}}$)

	N Units	**R2**	**A**	**B**	**Mercury Removal At Zero Chlorine Content**
All Configurations	11	63%	-0.028 (t = 0.15)	-2.033 E-5 (t = 3.91)	3%

As for the fabric filter group, statistical tests were conducted to determine whether the residuals from the correlation indicated the presence of statistically significant differences associated with coal rank. For Wet FGD with ESP, no differences by coal rank could be detected with statistical significance, although as before, this finding is in large part a result of the small sample size and should not be interpreted as a conclusive demonstration that such differences do not exist. Only one lignite unit is present in this group, and its performance is poorer than predicted by the correlation line. Thus, lignite units with Wet FGD and ESP may not achieve in practice the level of mercury removal that is predicted by the correlation.

Figure 4.3 shows the unit-average data for this group in the semi-log space in which the regression is estimated, with confidence intervals for the population (data points) surrounding the regression line. Once chlorine content is accounted for, we see no difference in the mercury removal performance of cold- and hot-side ESP units, which fall on both sides of the regression line. Compared to the fabric filter technologies, the variability around the regression line is smaller, which translates into narrower confidence intervals, and is approximately the same at high and low chlorine levels.

Figure 4.4 shows the 95 percent confidence limits for the regression line in a linear form with mercury removal on the vertical axis. The somewhat irregular shape of the confidence limit envelope is an artifact of exponentiating the confidence limit parabolas in log space. The expected (average) mercury removal for Wet FGD with ESP technologies is nearly zero at zero chlorine content, and the confidence limits range from approximately –40 percent removal to +40 percent removal.

Fabric filters and Wet FGD + ESP are identified in later analysis as the best performing technologies for mercury removal from typical (average) coals. The confidence intervals around the correlations show, however, that for worst-case coals with low chlorine content, we have very little certainty that mercury removal will actually take place. The methods used in later chapters to estimate MACT floors and limits will account for all of the regression uncertainty represented by confidence limits in the figures, and make allowances for unit performance and repeat test variability that is appropriate to the form of the standard. Because of the uncertainty present in the analysis, the MACT floors and limits will not – and can not – reflect a high level of mercury removal.

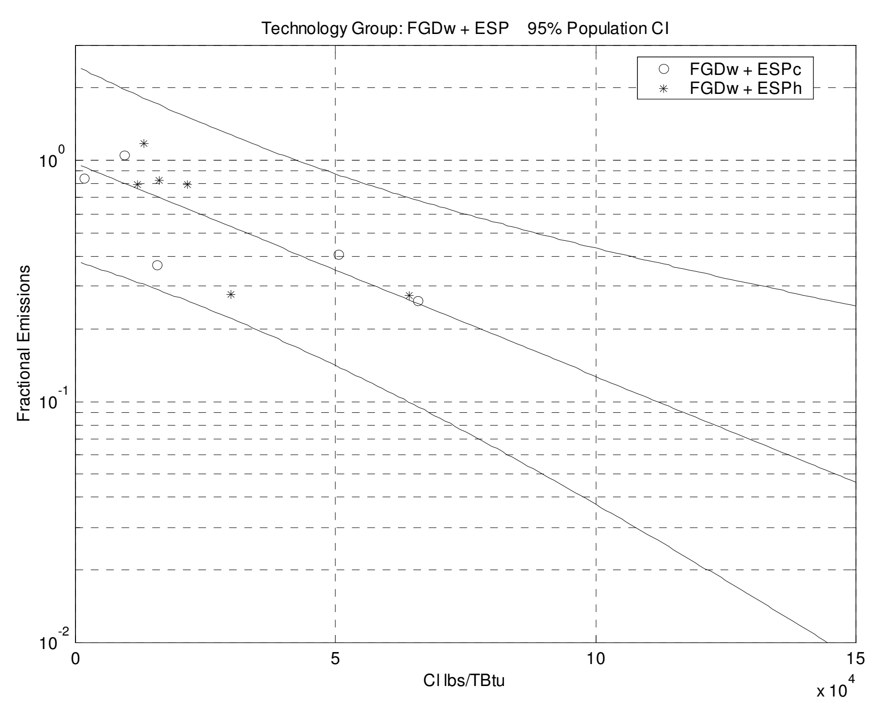

Figure 4.3 Mercury Removal Performance for Wet FGD + ESP Technologies

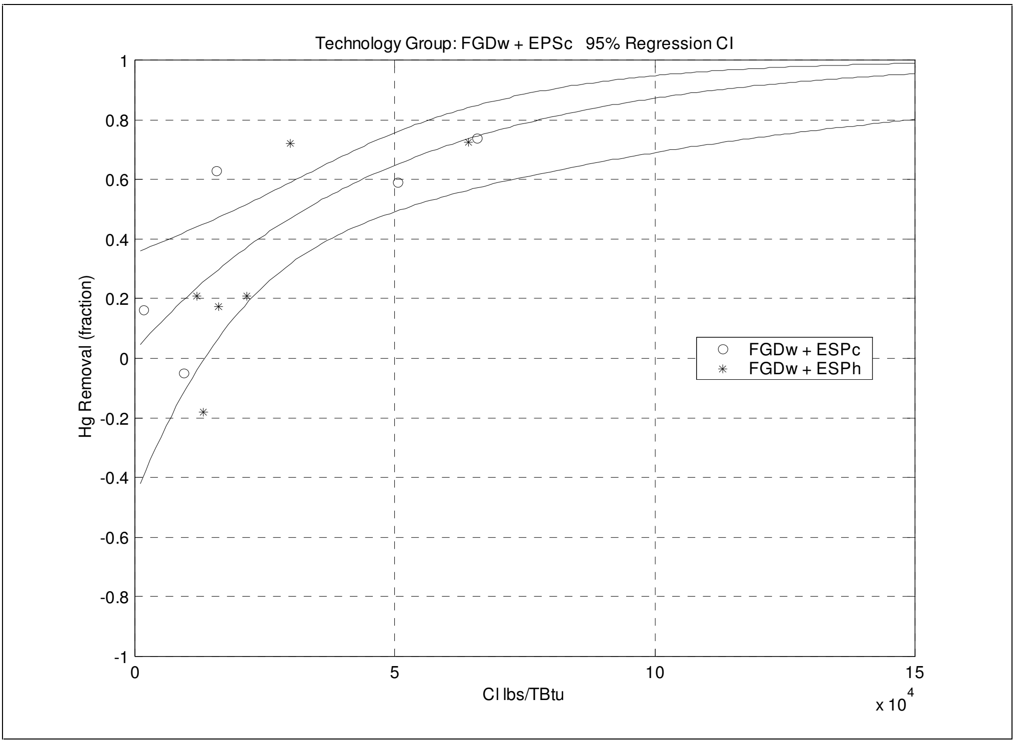

Figure 4.4 95 Percent Confidence Intervals on the Performance of Wet FGD with ESP Technology

4.3.3 Other Mercury Control Technologies

The examination of the other technology groups found no evidence of chlorine sensitivity. This result is more than a finding that chlorine is not a statistically significant predictor and extends to the apparent absence in the data of trends associated with coal chlorine content. Because sample sizes are small for these technologies, its remains possible that additional research could identify effects related to chlorine or other coal characteristics.

Because these technologies proved to be insensitive to chlorine content, their mercury removal performance is represented by fixed removal rates (see Table 4.3), which are estimated as intercept terms in natural log space. All detectable differences in coal rank are accounted for by the subdivision of the ESPc category into bituminous and non-bituminous groups. Of these technologies, only ESPc with bituminous coal achieves more than de minimis mercury removal.

Table 4.3 Summary of Other Technologies (Not Chlorine Sensitive)

Mercury Removal =1 – exp(A)

Technology	N Units	A	Fixed Mercury Removal
Dry FDG + Cold ESP	4	-0.077	7%
ESPc + Bituminous Coal	6	-0.621	46%
ESPc + Sub/Lig Coal	6	-0.149	14%
ESPh	8	-0.027	3%
Other PM Controls	8	-0.115	18%

4.4 Assessment of Unit and Test Variability

While the correlations are based on unit-average data, we may apply them to predict values for individual tests that were conducted on the units. Having done so for a technology group, the residual variability around the correlation line (or fixed removal rate) was decomposed into variance components for unit performance and repeat test variability. Because changes in coal chlorine content have been accounted for in the correlation line (or do not affect the mercury removal), the variance component estimated for repeat tests is not affected by variations in chlorine content among the tests.

Repeat test variability is easily understood as the variability of individual test results around the average value for a unit. Because the analysis has been conducted in log space, the test variability is estimated as a variance (and related standard deviation) of logarithm values. Exponentiating the log-space standard deviation and subtracting 1 gives a relative standard deviation (RSD) for the original data space; the RSD expresses the variation of test values as percentages of the unit mean value. This approach is also consistent with the ASTM standards for the Ontario Hydro test, which are stated in terms of RSD. In this form, the standard deviation of an emissions value (or fractional emissions estimate) is proportional to the value itself, so that the standard deviation is larger for large emissions (or fractional emissions) values.

Unit performance variability is a complex phenomenon and has been discussed at length in preceding chapters. Because of the short-term nature of the Ontario Hydro test used in the ICR Part III testing, unit performance variability represents the net effect of a constellation of factors. Among these are:

- Factors such as operational and environmental conditions that would cause a unit to test higher on one day and lower on another.
- Factors representing inherent differences among units having the same emission control configuration, which would cause one unit to exhibit higher or lower emissions than another unit, even when firing similar coals.

Because the variation in coal characteristics is observed primarily as differences in coals among units, and not as variations in coal among repeat tests of individual units, any coal-related characteristic that influences mercury emissions of units (and is not adequately represented by chlorine content as a surrogate variable) will also contribute to unit performance variability.

Table 4.4 summarizes the components estimated for unit performance and repeat test variability in each technology group. The tabulated values are log-space standard deviations that have been exponentiated and expressed in the form of relative standard deviations. The regression uncertainties for fabric filters and wet scrubbers + ESP, the technologies found to be sensitive to coal chlorine content, are cited at the midpoint of the data where the uncertainty will be smallest, while for other technologies the tabulated uncertainty reflects the standard error of the fixed removal rate estimated in place of a correlation line. The regression uncertainties indicate that we can measure the average removal performance of the best performing technologies only to within an error of ± 28 to 35 percent. The variability associated with the performance of units and with repeat tests is additive to this uncertainty. In conjunction with the regression uncertainty, these components of the unexplained variance will be used in the next chapter to estimate the emission level than can be achieved, with stated confidence, by a unit adopting any the best performing technologies.

Table 4.4 Components of Uncertainty in Emissions Estimates by Technology (Relative Standard Deviations, percent)

Technology Group	Regression Uncertainty	Unit Performance Variability	Repeat Test Variability
Fabric Filters	28%	179%	51%
Wet Scrubbers + ESP	35%	33%	38%
Dry FDG + Cold ESP	35%	117%	18%
ESPc + Bituminous Coal	33%	53%	23%
ESPc + Sub/Lig Coal	33%	53%	23%
ESPh	42%	51%	18%
Other PM Controls	42%	14%	40%

4.5 Adequacy of Correlations to Predict Performance of Lignite Units

The finding in the correlation analysis that lignite units consistently showed poorer performance than predicted by the correlation equations is of considerable concern. The two lignite units equipped with fabric filter controls showed average mercury removals across three tests of –9 percent and +8 percent, while the one lignite unit equipped with Wet FGD and ESP showed an average removal of –4 percent. The correlation equations would have predicted substantially greater removals for these units. Our concern in this regard was increased by the reports of unit operators that, in their experience, no lignite unit had achieved the predicted levels of mercury removal using existing controls.

With only three lignite units available in total for the analysis of these technology groups, the data have essentially no power for testing whether the poorer-than-predicted performance of the three units is due to chance or is indicative of an actual difference in the performance of lignite units compared to bituminous and subbituminous units with similar controls. Further, there is not enough data on lignite units in the ICR Part III testing to consider treating lignite units separately, either in the form of a correlation or as a fixed removal rate. Clearly, more data should be collected before EPA promulgates standards for lignite units.

Given the basis for concern that the correlation equations may over-predict the mercury removal that can be achieved in lignite units, and the inability of the available data to support a better characterization, we have elected to apply a cap to the predictions of correlations in later analysis. The cap is that predictions of the correlations, whether based on the correlation line alone or after accounting for uncertainty in the correlation, may not exceed the highest mercury removal that has been demonstrated by a lignite unit in the ICR Part III testing. This limit is provided by Stanton Station 10 (equipped with cold-side ESP controls only), which demonstrated 21.2 percent removal for the average of three tests. In the use of the correlation equations in the following chapters, predicted removals in excess of this cap were replaced by the 21.2 percent value and the replacement clearly noted for the reader.

5. DEVELOPMENT OF ALTERNATIVE MACT FLOORS FOR EXISTING UNITS

5.1 Introduction

In the Chapter 3 review of EPA's approach, recommendations were made for an improved methodology for determining MACT floors. The recommended methodology is used in this chapter to develop alternative MACT floors as follows:

1. Information on the frequency of emissions control configurations in the US population of generating units (used by EPA in designing the ICR Part III survey) is used to count ICR Part III units in a way that represents the top 12 percent of the entire US population for bituminous and subbituminous coals (each having many more than 30 units in the national inventory) and the equivalent of the top five units in the US population for lignite coal(having fewer than 30 units in the national inventory according to EPA data).

2. Units in the ICR Part III sample are ranked based on the annual emissions expected when firing an average coal of their rank. The choice of an average coal for ranking purposes is made to separate the performance of each unit from the coal it happens to fire and, by evaluating all units on the same basis, thereby determine which units exhibit better mercury removal performance, rather than merely the use of better raw material.

3. The adverse conditions for top-performing units are defined as firing a coal whose emissions would fall at the 90th percentile for the coals in each rank. The adverse coal is based on an actual annual coal supply received by a unit in 1999. Because the ICR Part III data do not sample emissions under conditions other than high load and steady-state operations, it is not possible to account for the contribution of adverse operating conditions to emissions.

4. Emissions of the top-performing units are estimated when firing the adverse coal based on the characteristics of the adverse coal and the correlation equations (or fixed removal rates) developed in Chapter 4. Emission estimates made at this stage are based on the predictions of the correlation equations and do not in any way account for the uncertainties in the correlation analysis or for the variability in unit performance or repeated tests than will occur; such estimates do not necessarily reflect the emissions performance that can actually and consistently be achieved by units.

5. To estimate MACT floors for existing units, the emissions of top-performing units are adjusted for the uncertainty in the correlation equations (or the fixed removal rates) and for the effects of unit performance and repeat test variability that are appropriate to the form of the standard. When, and only when, the adjustments for uncertainty and variability are made can the resulting estimates be taken as reflecting the emissions performance that can actually and consistently be achieved by units.

The emphasis in the analysis is the annual-average form of the standard, but results are given for a standard based on the average of three tests in order to illustrate the effect of variance attenuation in a long-term average. In both cases, the standards are consistent with the measurements of the Ontario Hydro test and have not been adjusted to account for the bias and variability that would be introduced by the use CEMs for compliance purposes. Results are presented for the MACT floors that can be met with 97.5 percent confidence by the top-performing units in the entire population.

5.2 Counting Top Performing Groups based on Population Statistics

The top performers for a subcategory having more than 30 units in the population are defined as the top 12 percent of units for which EPA has information. As has been argued in Chapter 3, the collection of the ICR Part III data through a stratified random sample requires that the method of counting top-performing units be based on sample units that represent the top 12 percent of the US population, not merely the top 12 percent of the sample as EPA has done. Any approach not based on population statistics will over-represent units that have more sophisticated emission controls compared to the actual distribution of controls in the population.

In this context, we have evaluated the performance of the top 12 percent of the entire population of bituminous and subbituminous units, and the top five units in the entire lignite population. To accomplish this, each ICR Part III unit has been assigned an expansion factor – i.e., the number of generating units in the national population that it represents. The expansion factor is based on EPA information on the number of units in each survey cell nationally[22] and the number of such units in the Part III database, after the exclusion of data as documented in Chapter 2.

Table 5.1 shows how the number of top-performing units in each category is determined from the national population of generating units, rather than the sample. For the bituminous subcategory, a number of ICR Part III units (not predetermined) will be selected in order of increasing emission rate, until the top-performing group represents a total of 66 units or 12 percent of the US bituminous population. Similarly, subbituminous units will be selected until the top-performing group represents a total of 27 units, or 12 percent of the US subbituminous

---

[22] The generating units counts used in the ICR Part III survey design differ in minor respects from other data sources on the inventory of generating plants in the US. The survey design data have been used to develop expansion factors to remain consistent with the ICR survey. Differences in the counts of units among sources do not materially affect the results of the analysis presented here.

population, and lignite units will be selected until the top-performing group represents 24 percent of the lignite population (the top 5). Population information is shown for completeness for FBC and IGCC units and for conventional units firing coal blends, but is not used in the analysis.

Table 5.1 Counting Top Performing Units based on Population Statistics

	Bit	**Sub**	**Lig**	**FBC**	**IGCC**	**Blends**	**Total**
US Population of Units (per EPA Sample Design)							
Total Units	661	228	26	38	3	130	1,086
Top Performers							
Units	79	27	5	5	3	n/a	n/a
Percent	12.0%	12.0%	19.2%	12.0%	100%	n/a	n/a
ICR Part III Sample							
Units Tested[1]	21	22	8	6	2	12	71
Units Represented							
Total	552	227	21	37	3	236	1,076
Top Performers							
Units	66	27	5	4	3	n/a	n/a
Percent	12%	12%	24%	12%	100%	n/a	n/a

[1] Units left after exclusion of data. Excludes 1 stoker unit not classified in table.

5.3 Ranking Units Based on Emissions when Firing an Average Coal

Units in the ICR Part III sample are ranked based on the emissions expected when firing an average coal to separate the performance of each unit from the coal it happened to fire on the day of the testing. For this purpose, the average coal has been defined as the (independently computed) Btu-weighted average of coal mercury and chlorine content in each rank, based on the ICR Part II information on coal characteristics. Table 5.2 summarizes the Btu-weighted average characteristics of coal by rank. Coal is a naturally-occurring product that exhibits varying characteristics depending on its specific origin. The characteristics cited in the table are averages only, and an individual coal can depart significantly from the average for its rank.

Table 5.2 Characteristics of Average Coal

	Mercury Content lbs/TBtu	Chlorine Content lbs/TBtu
Bituminous		
Mean value	7.6	79,000
(± 1 σ range)	(5.7-10.3)	(19,000-118,000)
Subbituminous		
Mean value	5.9	12,000
(± 1 σ range)	(3.2-8.3)	(4,000-14,000)
Lignite		
Mean value	12.9	19,900
(± 1 σ range)	(5.4-19.1)	(10,000-37,000)

Table 5.3 summarizes the result of ranking top-performing units based on population statistics and the emissions rate expected when firing average coal. For the bituminous subcategory, two technologies – fabric filters (with and without scrubbers) and Wet FGD with ESP – represent the 66 best bituminous units in the US and constitute the top-performing group. The top-performing technologies have an expected average stack emissions rate of 0.82 lbs/TBtu and a predicted mercury removal rate of 89.2 percent on average coal, based on the correlation analysis. These estimates do not account for uncertainty in the correlation prediction or the variability that is likely to occur among units and across tests. The estimates do not necessarily reflect the emissions performance that can consistently be achieved by units under actual operating conditions.

In the subbituminous category, the same two technologies (fabric filters, with and without scrubbers, and Wet FGD with ESP) are the top performers, and represent the 27 best subbituminous units in the US. The top-performing units have an expected average stack emissions rate of 2.68 lbs/TBtu and a mercury removal rate of 54.1 percent on average coal. For the lignite category, the best technology is again fabric filters (with or without scrubbing), and represents the top five units in the US.. The top-performing technology has an expected average stack emissions rate of 10.15 lbs/TBtu and a mercury removal rate of 21.2 percent on average coal; these values are based on the removal cap of 21.2 that is applied to predictions of the correlation equations for lignite units. Fabric filter technologies dominate the list of top-performing ICR Part III units in all subcategories, accounting for two-thirds or more of the top-performing units in the US population. These emission values also do not account for uncertainty or variability and do not necessarily reflect the emissions performance that can consistently be achieved by units in actual operation.

Table 5.3 Identification of Top Performing Units and their Control Technologies for Conventional Coal Subcategories (Without Consideration of Uncertainty or Variability)

			Estimated for Annual-Average Coal	
Technology	**Units in Top Performing Population**	**Population Weight**	**Predicted Removal Fraction**	**Predicted Emissions lbs/TBtu**
Bituminous				
FF (with or without scrubber)	42	7.6%	0.942	0.44
Wet Scrubber +ESP	24	4.4%	0.805	1.47
Average	66	12.0%	0.892	0.82
Subbituminous				
FF (with or without scrubber)	26	3.5%	0.556	2.60
Wet Scrubber +ESP	1	0.6%	0.238	4.46
Average	27	12.0%	0.541	2.68
Lignite				
FF (with or without scrubber)	5	23.8%	0.212*	10.15
Average	5	23.8%	0.212*	10.15

* The prediction of the correlation equation has been capped at the highest mercury removal demonstrated by a lignite unit in the ICR Part III data.

The MACT floors are based on the emission performance expected for the top-performing units under adverse circumstances that are reasonably foreseeable. Given the absence of sampling of adverse operating conditions in the ICR Part III data, adverse circumstances are taken to mean firing a coal with adverse emissions characteristics. The adverse coal for each rank is based on the actual annual coal supply in 1999 of a plant whose emissions would be at the 90th percentile of the population of plants even after adopting a best performing technology. Table 5.4

summarizes the characteristics of the adverse coals; as seen in the table, the coal chosen as representing adverse circumstances can vary depending on the technology being considered.

Table 5.4 Characteristics of the Adverse Coal by Rank

	Predicted Emissions[*] lbs/TBtu	**Mercury Content lbs/TBtu**	**Chlorine Content lbs/TBtu**
Bituminous Coal			
FF (with or without scrubbing)	2.35	6.51	18,900
Wet Scrubber + ESP (c or h)	4.31	6.51	18,900
Subbituminous Coal			
FF (with or without scrubbing)	4.08	7.27	4,300
Wet Scrubber + ESP (c or h)	6.60	8.06	8,400
Lignite Coal			
FF (with or without scrubbing)	24.63[**]	31.26	13,550

[*] Emissions estimates do not account for uncertainty or variability.

[**] The emission rate is based on the cap of 21.2 percent mercury removal for lignite units.

The adverse coals are generally characterized by low chlorine contents. The adverse bituminous coal for mercury emissions is one with low chlorine content, but below-average mercury content, reflecting the importance of chlorine content to mercury removal, as estimated in the correlation analysis. Further, it should be realized that other annual coal supplies, with higher chlorine content and above-average mercury content, would be consistent with the emission level estimated in the table for the selected adverse coal. For subbituminous coal, the adverse coal supply is one with below-average chlorine content and above-average mercury content. The adverse lignite coal is one in which chlorine content is modestly reduced compared to the average, but having a substantially higher mercury content.

Table 5.5 shows the emission rates expected for the top-performing units when firing adverse coals. In the bituminous subcategory, the low chlorine content of the adverse coal reduces the estimated mercury removal to 52.9 percent (from 89.2 percent on the average coal) and increases the estimated mercury emission rate to 3.06 lbs/TBtu. In the subbituminous category, the estimated mercury removal declines less dramatically, to 42.0 percent, when firing the adverse coal, while the emission rate increases to 4.20 lbs/TBtu. For the lignite category, the estimated mercury removal remains capped at 21.2 percent based on the highest removal demonstrated in the ICR Part III testing, and the emissions rate expected when firing the adverse coal is 24.63 lbs/Tbtu. As for the estimates based on average coal, these emission values do not account for uncertainty or variability and do not necessarily reflect the emissions performance that can consistently be achieved by units in actual operation.

5.4 Accounting for Uncertainty and Variability

The emission rates summarized in Table 5.5 are the *expected* values for the average of the top-performing units in each subcategory when firing adverse coal, based on the correlation equations developed for the two technologies. These emission rates are the best point estimates than can be made based on the available data and the emissions analysis presented in Chapter 4.

Table 5.5 Expected Emissions under Adverse Conditions for Top Performing Units (Without Consideration of Uncertainty or Variability)

			Estimated for Annual-Average Adverse Coal	
Technology	**Units in Top Performing Population**	**Population Weight**	**Predicted Removal Fraction**	**Predicted Emissions lbs/TBtu**
Bituminous				
FF (with or without scrubber)	42	7.6%	0.639	2.35
Wet Scrubber +ESP	24	4.4%	0.338	4.31
Average	66	12.0%	0.529	3.06
Subbituminous				
FF (with or without scrubber)	26	3.5%	0.439	4.08
Wet Scrubber +ESP	1	0.6%	0.181	6.60
Average	27	12.0%	0.420	4.20
Lignite				
FF (with or without scrubber)	5	23.8%	0.212*	24.63
Average	5	23.8%	0.212*	24.63

* The prediction of the correlation equation has been capped at the highest mercury removal demonstrated by a lignite unit in the ICR Part III data.

The presence of uncertainty and variability means that, if the top-performing units in the US population were subjected to comparable Ontario Hydro testing, the actual emission rates that would be measured could vary from the tabulated estimates to an extent that depends on the uncertainties in the analysis and the effects of unit performance and repeat test variability. To estimate MACT floors, these estimates must be adjusted to account for that type of variation.

Three components of the variance are applicable to the determination of MACT floors:

- $\sigma^2(R)$ = the uncertainty of the correlation equation
- $\sigma^2(U)$ = unit performance variability
- $\sigma^2(T)$ = repeat test variability

If compliance were based on the average of three Ontario Hydro tests conducted back-to-back, and one unit were selected at random from the group of top-performing units, the 95 percent confidence intervals associated with estimated emission levels will be:

$$CI_{95} = \pm\ t_{95,df} * \sigma^2(R, Cl) + t_{95} * [\ \sigma^2(U) + \sigma^2(T)/2\ ] \qquad (5\text{-}1)$$

where $t_{95,df} * \sigma^2(R,Cl)$ is the 95 percent confidence interval for the correlation equation at the point x = Cl on the regression line and $t_{95,df}$ is the two-sided t-statistic for 95 percent confidence and the degrees of freedom df in the regression, and where $t_{95} * [\ \sigma^2(U) + \sigma^2(T)/2\ ]$ is the 95 percent confidence interval for the variation expected in large samples for a single unit and the average of three tests (since variance is attenuated by averaging in proportion to 1 / N-1).

For an annual form of the standard, we must consider the variation that is expected when averages are computed over a long-time period. This is a straightforward matter for the repeat variability, which will be attenuated inversely in proportion to the number of tests that would be conducted to determine an annual average. If the Ontario Hydro test requires two hours for completion, then the repeat test variability would be attenuated in proportion to 1 / [12 * 365 –1] = 1 / 4379.

The treatment of unit performance variability is more complicated, because portions of it reflect the chance testing of some units on days when they will exhibit high emissions, and other units on days when they will exhibit low emissions. This portion of the variability will be attenuated in an annual average. Other portions reflect factors that pertain to inherent differences in units and that will not be attenuated in an annual average. Because the ICR Part III data reflect snapshots of unit performance in time, there is no basis in the data for assessing how the overall unit performance variability should be divided into these portions. For purposes of this analysis, we assume that the overall unit variability is evenly divided (50/50) into:

- σ^2(U1) = unit performance variability reflecting factors that would cause a unit to test with higher emissions on one day and lower emissions on another, which would be attenuated in an annual average
- σ^2(U2) = unit performance variability reflecting inherent differences among units that would not be attenuated in an annual average.

Given this partitioning, the 95 percent confidence interval that is applicable to an emissions estimate for an annual standard is given by:

$$CI_{95} = \pm\ t_{95,df} * \sigma^2(R, Cl) + t_{95} * [\ \sigma^2(U1) + \sigma^2(U2)/364 + \sigma^2(T)/4379\ ] \quad (5\text{-}2)$$

For each top-performing unit, the upper 97.5 percent confidence limit is computed for the emission rate when firing the adverse coal for its category. The computation applies Eq. 5-1 or Eq. 5-2 depending on the form of the standard. A corresponding lower confidence limit for the mercury removal rate of each unit is computed from the upper limit for emissions and the mercury content of the adverse coal. A one-sided 97.5 percent limit is the upper limit of a 95 percent confidence interval and, therefore, is the confidence level conventionally accepted for research purposes. It is also the confidence level selected by EPA for its analysis.

The analysis of the ICR data admits the statistical possibility that measured mercury removal rates can be negative at low chlorine levels because of the independent variation in stack emissions and coal mercury content tests. Negative removal is a physical impossibility, however, as long as mass balance is maintained. When emissions are averaged over long periods of time, only small negative removals are likely to occur because the independent variation is substantially attenuated. Therefore, we apply an additional limit that the associated lower confidence limit on mercury removal for any unit cannot be more negative than would be expected based on a zero mercury removal rate and the variability associated only with σ^2(U2) and σ^2(T) as appropriate to the form of the standard. That is, although the confidence intervals associated with the regression uncertainty σ^2(R) and the σ^2(U1) portion of the unit performance variability may admit the possibility of negative mercury removal rates in a particular circumstance, this result is not allowed to be more extreme (more negative) than the negative removal rate than could be measured in testing for a unit that actually achieves zero mercury removal.

MACT floors are computed by averaging the upper 97.5 percent confidence limit for the emission rate of each top-performing unit in a subcategory. The mercury removal rates are averaged to give an average removal rate that is associated with the MACT floor. The resulting MACT floors are the emissions rates that are achievable, on average, in adverse circumstances by the top-performing units. These emissions levels would be exceeded in practice only 2.5 percent of the time.

Table 5.6 presents the computation of MACT floors for the annual-average form of the standard. The MACT floors are estimated to be 6.9 lbs/TBtu for bituminous units, 7.9 lbs/TBtu for subbituminous units, and 34 lbs/TBtu for lignite units.

For the top-performing units in each subcategory, the uncertainty in the mercury removal performance of existing control technologies is sufficiently large that we can not say, with the selected confidence level, that emissions will be reduced below the mercury content of the coal. Thus, the MACT floor is largely determined by the mercury content of the adverse coal and the variability associated with σ^2(U2) and σ^2(T) that remains after attenuation in the annual average.

Table 5.7 summarizes the MACT floors that we estimate for existing, conventional coal units. If compliance were based on the average of three tests conducted back-to-back, the MACT floors needed to achieve 97.5 percent confidence are necessarily very high due to the variability that can occur in short-term testing. Large negative removals must also be admitted as a possible

Table 5.6 Determination of MACT Floors for the Annual-Average Form of the Standard

			Estimated for Annual-Average Adverse Coal	
Technology	**Units in Top Performing Population**	**Population Weight**	**Mercury Removal fraction**	**Emissions lbs/TBtu**
Bituminous				
FF (with or without scrubber)	42	7.6%	-0.08	7.0
Wet Scrubber +ESP	24	4.4%	-0.02	6.7
Average	66	12.0%	-0.06	6.9
Subbituminous				
FF (with or without scrubber)	26	3.5%	-0.08	7.8
Wet Scrubber +ESP	1	0.6%	-0.02	8.2
Average	27	12.0%	-0.08	7.9
Lignite				
FF (with or without scrubber)	5	23.8%	-0.08	34
Average	5	23.8%	-0.08	34

outcome of short-term testing, for the same reason that negative removals are observed in the ICR Part III data. For the 3-test form of the standard, the MACT floors represent marginal emissions limitations that could occur in the testing of top-performing units. The units could achieve lower average emission levels over a longer time as indicated by the MACT floors for the annual form of the standard, which are estimated to be 6.9 lbs/TBtu for bituminous units, 7.9 lbs/TBtu for subbituminous units, and 34 lbs/TBtu for lignite units. Given the uncertainties attending the characterization of emissions performance for lignite units, the actual performance of lignite units could be poorer than expected in this analysis, and the MACT floors for lignite units that provide 97.5 percent confidence of compliance could be higher than shown here.

Table 5.7 MACT Floors for Conventional Coal Units

Form of Standard	Mercury Removal (Percent)	Mercury Emissions lbs/TBtu
Average of 3 Tests		
Bituminous	-167%	17
Subbituminous	-350%	33
Lignite	-280%	119
Annual Average		
Bituminous	-6%	6.9
Subbituminous	-8%	7.9
Lignite	-8%	34

The uncertainty in the analysis of mercury removal performance for the best performing technology groups is sufficiently large that, at the present time and based on the ICR Part III data, one cannot say with acceptable confidence that the best performing technologies will reduce emissions below the mercury present in coal under the adverse circumstances of low chlorine content. Even in the annual form of the standard, in which the variability associated with unit performance and repeat tests is substantially attenuated, the uncertainty in the present knowledge of mercury removal is sufficiently large that one cannot conclusively expect existing controls to reduce emissions below the mercury content of adverse coals.

These results may appear surprising given the relatively more stringent standards proposed by EPA. EPA's analysis failed to reach this conclusion for two primary reasons. First and most importantly, EPA gives no consideration to the uncertainty in its assessment of the emissions performance of the best technologies. Second, EPA does not fully recognize or account for the extent of the unit performance variability that exists in the ICR Part III. As a result, EPA overstates the degree of certainty in the level of control that can be expected from the emission controls in the best performing units.

6. DEVELOPMENT OF MACT FLOORS FOR REGIONAL LIGNITE COALS

6.1 Introduction

The foregoing analysis developed MACT floors for subcategories of bituminous, subbituminous, and lignite coals. Coal is not a homogenous product, however, and substantial differences in characteristics exist within each rank depending on the specific origin of the coal. Such differences are acute for lignite coals, which are found in two primary geologic settings – the Fort Union and Gulf Coast formations – and are generally fired in minemouth generating units in areas of the Northern Great Plains and the US Gulf Coast. This chapter examines the regional differences in lignite coal characteristics and derives MACT floors for separate subcategories for units firing each of these types of lignite coal.

Differences in the characteristics of lignite coal from these two formations have been estimated using the limited information that was reported in the ICR Part II data for 1999. Users of coal from the Fort Union formation are represented by 6 plant sites in North Dakota and Montana, at which are located 10 of the 13 units that fire Fort Union lignite. Users of the Gulf Coast formation coal are represented by 4 plant sites in Texas, at which are located 6 of 15 units that fire Gulf Coast lignite[23].

Table 6.1 shows that Gulf Coast lignite has, on average, nearly twice the mercury content of Fort Union lignite (14.3 lbs/TBtu versus 8.6 lbs/TBtu, respectively) and that both Gulf Coast and Fort Union lignites are relatively low in chlorine content. While the chlorine content is somewhat higher in Gulf Coast lignite, the correlation equations indicate that the potential gain in mercury removal with chlorine-sensitive technologies is not sufficient to offset its higher mercury content. Other factors being equal, including emission control configuration and operating conditions, units that fire Gulf Coast lignite will tend to produce higher emissions than units firing Fort Union lignites.

---

[23] Four of the 15 lignite units fire blends of Gulf Coast lignite and subbituminous coal. These four units (Monticello 3 and Martin Lake 1, 2, and 3) are classified as firing coal blends and are excluded from the analysis presented here.

Table 6.1 Average Lignite Coal Characteristics Based on ICR Part II Shipments Data

	All Lignite Coal	Fort Union Lignite	Gulf Coast Lignite
Hg Content (lbs/TBtu)			
Mean value	12.9	8.6	14.3
(± 1 σ range)	(5.4-19)	(5.7-11.3)	(5.4-20)
Cl Content (lbs/TBtu)			
Mean value	19,900	14,100	21,900
(± 1 σ range)	(10,000-37,000)	(9,000-19,000)	(10,000-38,000)

Figure 6.1 shows the range in characteristics around the averages cited in the table. Fort Union lignite lies at the lower end of the ranges for the mercury and chlorine contents of Gulf Coast lignite, and Fort Union lignite is less variable than Gulf Coast lignite. Given the variation in characteristics among coals, the differences in the average characteristics of Fort Union and Gulf Coast lignite are sufficiently large to be statistically significant.

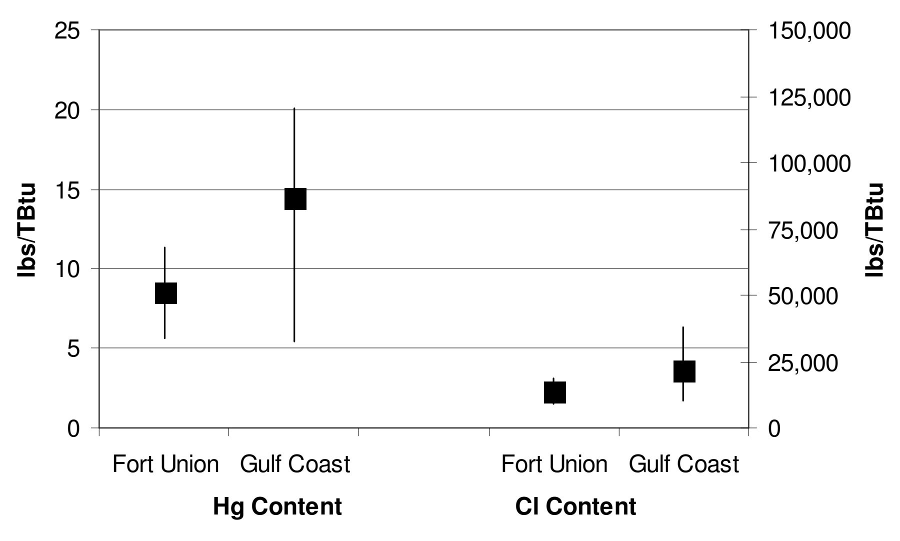

Figure 6.1 Distribution of Mercury and Chlorine Content for Regional Lignite Coal

In the following analysis, the methodology presented in Chapter 5 is used to develop separate MACT floors for the subgroups of Fort Union and Gulf Coast lignite. The analytical process is to:

- Identify the populations of units in the US that fire lignite coal of the two types.
- Determine the size of the top performing groups, which will consist of the top 5 units in each subgroup.
- Rank units based on annual emissions when firing an average coal, and identify the top performing control technologies used by the 5 best units in each subgroup.
- Estimate emissions of the top performing units reflecting emission control technologies when firing an adverse coal in each subgroup.
- Adjust emissions when firing an adverse coal for the uncertainty of the emissions correlations and for unit performance and repeat test variability as appropriate to the form of the standards.

The MACT floors for the subgroups are the emission levels of the 5 best units in each subgroup on an annual basis, as adjusted for uncertainty and variability appropriate to determining a standard they could meet with the prescribed confidence level.

6.2 Generating Unit Populations

The subpopulations of generating units firing lignite coals of each group have been identified using the database of US generating units from which EPA has estimated the mercury emissions inventory for 1999[24]. Units located in North Dakota and Montana were classified as firing Fort Union lignite, while units located in Texas were classified as firing Gulf Coast Lignite.

6.2.1 Fort Union Lignite Units

Table 6.2 summarizes the population of generating units that fire Fort Union lignite coals. A total of 13 units were identified, of which 12 are the conventional coal boilers considered by this analysis. The R. M. Heskett Station B2 unit is excluded because it is an FBC unit and processes coal in a fundamentally different way than the conventional boilers. The 5 top performing units in this subgroup represent the best 5 of 12 unit or the best 42 percent of the population.

---

[24] Contained in the Case by Case tool listed at http://www.epa.gov/ttn/atw/combust/utiltox/utoxpg.htm under the heading Relevant Technical Data (June 2001)

Table 6.2 Population of Fort Union Lignite Units

State	Generating Stations	Number of Units	ICR Part III Stack Testing	ICR Part II Coal Data
Conventional Boilers				
MT	Lewis & Clark B1	1	Unit B1	Yes
ND	Antelope Valley B1, B2	2	Unit B1*	Yes
ND	Coal Creak 1,2	2		No
ND	Coyote 1	1	Unit 1*	No
ND	Leland Olds 1, 2	2	Unit 2	Yes
ND	Milton R Young B1, B2	2		Yes
ND	Stanton 1, 10	2	Units 1, 10	Yes
Total		12		
Excluded (FBC)				
ND	R.M. Heskett Station B2	1	Unit B2	Yes

* Excluded from the analysis due to data quality.

Table 6.3 Population of Gulf Coast Lignite Units

State	Generating Stations	Number of Units	ICR Part III Stack Testing	ICR Part II Coal Data
Conventional Boilers				
LA	Dolet Hills 1	1		Yes
TX	Big Brown 1	1	Unit 1	No
TX	Limestone 1, 2	2	Unit 1	Yes
TX	Monticello 1, 2	2	Unit 1	No
TX	Pirkey 1	1		Yes
TX	San Miguel	1		No
TX	Sandow 4	1		No
Total		9		
Excluded (Blends, FBC)				
TX	Martin Lake 1-3 (Blend)	3		Sub Only
TX	Monticello 3 (Blend)	1	Unit 3	Sub Only
TX	TNP-One U1, U2 (FBC)	2	Unit U2	Yes

6.2.2 Gulf Coast Lignite Units

Table 6.3 summarizes the subpopulation of Gulf Coast lignite units. A total of 15 units were identified, of which 9 are conventional coal boilers firing only lignite coal. Six units are excluded from the analysis, including 4 units that fire blends of Gulf Coast lignite and sub-bituminous coal and 1 FBC unit. The 5 top performing units in this group are the best 5 of 9 or the best 56 percent of the population.

As described in Chapter 5, the best performing technologies are identified by ranking units on the annual emissions that are expected when firing an average coal typical of each subgroup. For the category of all lignite units, fabric filters (with or without scrubbing) were shown to be the emission control technology used by the top 5 units. For the subgroups for Fort Union and Gulf Coast lignite, fabric filters (with or without scrubbing) remain the best technology

6.3 Emissions of Top Performing Units under Adverse Circumstances

The MACT floors are based on the emissions performance expected for the top-performing units under the adverse circumstances that are reasonably foreseeable. As defined in Chapter 5, adverse circumstances are taken to mean firing a coal with adverse emissions characteristics. The adverse coal for each subgroup is based on the actual annual coal supply of a plant whose emissions would be at the 90th percentile of plants, even after adopting the best performing technology. Because of the small sample of units reporting lignite coal characteristics, the actual percentiles are less than the 90th percentile. The 1 plant site (out of 6) with highest emissions was chosen to represent the Fort Union group, while the 1 plant site (out of 4) with highest emissions was chosen to represent the Gulf Coast group. As seen in Table 6.4, the adverse coals for fabric filter controls are ones with mercury content close to the average for the subgroups, but with below-average chlorine content, reflecting the importance of chlorine to mercury control as estimated in the correlation analysis. Other annual coals supplies, with higher mercury and chlorine contents would be consistent with emission levels given in the table for adverse coals. Note that the adverse coal defined in Chapter 5 for the group of all lignite units is the same as the adverse coal for the Gulf Coast lignite presented here.

Table 6.4 Characteristics of Adverse Lignite Coals By Region

	Expected Mercury Emissions[*] lbs/TBtu	Mercury Content lbs/TBtu	Chlorine Content lbs/TBtu
Fort Union Lignite			
FF (with or without scrubbing)	6.63[**]	8.42	9,680
Gulf Coast Lignite			
FF (with or without scrubbing)	24.63[**]	31.26	13,550

[*] Emissions estimates do not account for uncertainty or variability.

[**] The emission rates are based on the cap of 21.2 percent mercury removal for lignite units.

Table 6.5 shows the emissions rates expected for the top-performing units when firing the adverse coals, applying the correlations derived in Chapter 4. In all cases, the mercury removal predicted by correlations have been capped at the 21.2 percent value demonstrated in the ICR Part III testing. For Fort Union lignite, the average mercury emission rate on adverse coal is estimated to be 6.63 lbs/TBtu. For Gulf Coast lignite, the average emission rate is estimated to be 24.63 lbs/TBtu. These rates do not account for uncertainties in the emissions analysis or for the variability in emissions and do not necessarily reflect the emissions performance that can consistently be achieved in units.

Table 6.5 Expected Emissions under Adverse Conditions for Regional Lignites Based on Correlations Curves for all Ranks (Without Consideration of Uncertainty or Variability)

			Estimated for Annual Average Adverse Coal	
Technology	**Units in Top Performing Population**	**Population Weight**	**Removal fraction**	**Emissions lbs/TBtu**
All Lignite				
FF (with or without scrubber)	5	24%	0.212	24.63
Average	5	24%	0.212	24.63
Fort Union Lignite				
FF (with or without scrubber)	5	42%	0.212	6.63
Average	5	42%	0.212	6.63
Gulf Coast Lignite				
FF (with or without scrubber)	5	56%	0.212	24.63
Average	5	56%	0.212	24.63

6.4 MACT Floors for Regional Lignite Coals

As explained in Chapter 5, a 97.5th percent confidence limit is computed for the emission rate of each top-performing unit in a subcategory, considering the effects of uncertainty in the emission analysis and the unit performance and repeat test variability that is appropriate to the form of the standard. These are the rates that are forecast to be achievable in adverse circumstances by each unit all but 2.5 percent of the time. The MACT floors are then defined as the average of the 97.5th percentile emission rates for the top-performing units.

Table 6.6 shows the calculation of MACT floors for the annual form of the standard. The MACT floors are estimated to be 9.1 lbs/TBtu for Fort Union lignite units and 34 lbs/TBtu for Gulf Coast lignite units. The MACT floor for all lignite units (developed in Chapter 5) is seen to be the same as that for Gulf Coast units, because the adverse coal and technology for best performing units for the all lignite subcategory analysis was the same as those for the Gulf Coast subcategory considered here.

Table 6.6 Determination of MACT Floors for Regional Lignite Coal Annual-Average Form of the Standard

			Estimated for Annual-Average Adverse Coal	
Technology	**Units in Top Performing Population**	**Population Weight**	**Removal fraction**	**Emissions lbs/TBtu**
All Lignite				
FF (with or without scrubber)	5	24%	-0.08	34
Average	5	24%	-0.08	34
Fort Union Lignite				
FF (with or without scrubber)	5	42%	-0.07	9.1
Average		42%	-0.07	9.1
Gulf Coast Lignite				
FF (with or without scrubber)	5	56%	-0.08	34
Average	5	56%	-0.08	34

Table 6.7 summarizes the MACT floors we estimate for existing, conventional lignite coal units. If compliance were based on the average of 3 tests conducted back-to-back, the MACT floors needed to achieve 97.5 percent confidence of compliance are necessarily very high due to the variability that can occur in short-term testing. And, as seen in Chapter 5 in the floors for the other coal ranks, large negative removals must also be admitted as a possible outcome of short term testing.

Table 6.7 MACT Floors for Regional Lignites

	Mercury Removal (Percent)	**Mercury Emissions (lbs/TBtu)**
Average of 3 Tests		
All Lignite	-280%	119
Fort Union Lignite	-290%	33
Gulf Coast Lignite	-280%	119
Annual Average		
All Lignite	-8%	34
Fort Union Lignite	-7%	9.1
Gulf Coast Lignite	-7%	34

The effect of uncertainty and variability is smaller for the annual form of the standard due to the attenuation of some components of variability in an annual average. At the low chlorine levels characteristic of the adverse coals, the uncertainty in the emissions analysis is sufficiently large that we cannot say, with 97.5 percent confidence, that the best performing technologies would reduce emissions below the mercury present in the coal. Much of the unit performance and repeat test variability is attenuated for an annual standard, and only modestly negative removals are likely to be measured on an annual basis. Thus, the MACT floors are largely determined by the mercury contents of the coals.

For the Fort Union lignite group, the MACT floor for an annual form of the standard is estimated to be 9.1 lbs/TBtu. This emission level is nearly the same as EPA's proposed 9.2 lb standard for all lignite units, for which all of the top performing units selected by EPA fired Fort Union coal. The substantially higher mercury content of adverse Gulf Coast lignite coal indicates the need for a separate and higher MACT floor, which is estimated here to be 34 lbs/TBtu.

7. ESTIMATING MACT LIMITS FOR NEW UNITS

7.1 Introduction

The emissions analysis and methodology for determining MACT floors developed in the prior chapters are used here to estimate MACT limits for new units. The MACT limits for new units are based upon the performance achieved by the single best unit in the population in each subcategory, rather than an average performance of the top 12 percent of units. Given the limitations of the ICR Part III data, however, it is not possible to identify the specific unit in the population that would have the lowest emissions rate on an annual basis. Instead, this analysis is based on the performance of the best performing control technology in the database – fabric filters (with or without scrubbing) – under the adverse conditions previously defined.

Because the new source limit under Section 112(d) is based on the performance of the best unit alone, it is not appropriate to make allowances for unit performance variability as was done in deriving the MACT floors for existing units. Therefore, all of the unit performance variability identified in the analysis in Chapter 4 has been eliminated in the determination, and the allowances made in assessing the Section 112(d) limit for new units are restricted to:

- The uncertainty in the correlation analysis
- The repeat test variability that pertains to the form of the standard.

Other factors used in the MACT determination remain unchanged from those used for deriving the MACT floor for existing units. The adverse coals are based on the annual coal supply actually fired in 1999 by the unit that would be at the 90th percentile of emissions. MACT limits are estimated for the 97.5 percent confidence level for two forms of the standard: the average of 3 tests and an annual-average standard.

7.2 New Unit MACT

Table 7.1 gives the MACT limits we estimate for new units. For a standard based on the average of three tests, the new unit MACT limits are estimated to be 7.1 lbs/TBtu for bituminous units, 10 lbs/TBtu for subbituminous units, and 9.7 lbs/TBtu and 31 lbs/TBtu for Fort Union and Gulf Coast lignite units, respectively. The corresponding floors for existing units, which account for the variability in performance among the top performing units, are 2 to 3 three times higher than for new units. The combination of uncertainty in the emissions analysis and repeat test variability means that the MACT limits for new units must admit the possibility of measured emission levels in excess of the mercury content of the coals (negative removals).

Table 7.1 Assessment of New Unit MACT

	New Units Limit		Existing Units
	Hg Removal (Percent)	MACT (lbs/TBtu)	MACT Floor (lbs/TBtu)
Average of 3 Tests			
Bituminous	-9%	7.1	16
Subbituminous	-38%	10	30
All Lignite	0%	31	109
Fort Union Lignite	-75%	15	33
Gulf Coast Lignite	-75%	55	109
Annual Average			
Bituminous	22%	5.1	6.9
Subbituminous	-1%	7.4	7.8
All Lignite	-1%	32	34
Fort Union Lignite	-1%	8.5	9.1
Gulf Coast Lignite	-1%	32	34

For the annual form of the standard, the new unit limits are reduced below the floors for existing units. The annual limits for new units are 5.1 lbs/TBtu for bituminous units, 7.4 lbs/TBtu for subbituminous units, and 8.5 lbs/TBtu and 32 lbs/TBtu for Fort Union and Gulf Coast lignite units, respectively.

With unit performance variability eliminated for new units, and repeat test variability attenuated in the annual form of the standard, the variability accounted for in the new unit MACT limits is dominated by the uncertainty of the correlation predictions. Therefore, the differences among ranks are strongly affected by the position of the adverse coal on the correlation line (determined by the chlorine content of the adverse coal) and by differences in the mercury content of the adverse coals. The MACT limits that can be achieved by new units with 97.5 percent confidence reflect a modest degree of mercury removal for bituminous units (22 percent), for which the adverse coal has a chlorine content of 18,900 lbs/TBtu. The MACT limits for sub-bituminous coal and for Fort Union and Gulf Coast lignites reflect no mercury removal due to the confidence bounds of the correlations at the low chlorine contents of the adverse coals and the uncertainties attending the emissions performance of lignite units. The MACT for new Gulf Coast lignite units reaches 32 lbs/TBtu because of the very high mercury content of the adverse coal (31 lbs/TBtu).

8. ABILITY OF VARIOUS COALS TO COMPLY WITH EPA's PROPOSED MACT STANDARDS

A major implication of EPA's proposed rule is that a large portion of the US coal supply may face significant difficulty in achieving compliance with the proposed MACT standards, even when fired in units that have adopted the best control technologies identified in the ICR database. This is in spite of EPA's recognition that mercury MACT regulation should not cause disruption in US coal markets and EPA's effort to allow for adverse coal characteristics in setting the proposed standards.

Using the analysis of emissions performance by technology in Chapter 4 and the treatment of uncertainty and variability developed in Chapter 5, we have evaluated the ability of US coals to comply with the proposed MACT floors. The methodology is based on the ICR Part II data on coal shipments received at US generating plants during 1999. For each shipment, we identify the technology that will achieve the greatest reduction in mercury emission considering the characteristics of the coal. An emissions rate limitation is computed that can be met by the shipment with a given statistical confidence. Upper confidence limits ranging from 50 percent to 97.5 percent were examined, assuming an annual-average form for the standard. The 97.5 percent confidence limit is one that would be exceeded 2.5 percent of the time.

The assessment is based on the expected performance of the selected best technology in an average unit, without consideration of the variability in emissions performance among individual units. The probability of compliance will be less if the coals are fired in units that, although adopting the best technology, prove to have poorer emissions performance than the average unit. The probability of compliance will be greater if the coals are fired in units that prove to have better emissions performance than the average unit.

8.1 <u>Existing Units</u>

As shown in Table 8.1, a large portion of the US coal supply will be unable to achieve compliance, with high statistical confidence, with the proposed standards for existing units, even when fired in units having the best performing technologies. Nearly one-half of the US bituminous coal supply (49 percent) will be unable to achieve compliance with 97.5 percent confidence. A somewhat lower proportion of the subbituminous coal supply (41 percent) will be unable to comply with the stated confidence, as will 62 percent of the lignite coal supply (71percent for Gulf Coast units and 37percent for Fort Union units).

Table 8.1 Ability of Coal to Achieve Compliance with EPA Proposed MACT Limits with 97.5% Confidence When Fired in Units with Top Performing Control Technology (Percent of Btus)

	EPA Proposed Standards	Percent Achieving Compliance	Percent Not Achieving Compliance
Bituminous	2.0 lbs	51%	49%
Subbituminous	5.8 lbs	59%	41%
Lignite			
All Lignite	9.2 lbs	38%	62%
Fort Union Lignite	9.2 lbs	63%	37%
Gulf Coast Lignite	9.2 lbs	29%	71%

Table 8.2 demonstrates that many coals will face low probabilities of compliance with EPA's standards, even when fired in top-performing units. Approximately 30 percent of the US bituminous and subbituminous coal supply, and 55 percent of the lignite supply (68 percent for Gulf Coast units and 18 percent for Fort Union units), will be unable to comply with at least 80 percent probability in top-performing units. A smaller proportion (6 to 15 percent for most coals) will be unable to comply with at least 50 percent probability. Lignite coal has an overall low probability of compliance at all confidence levels due to the difficulty faced by Gulf Coast lignite. The compliance difficulty is caused by the high mercury content of Gulf Coast lignite and the absence of evidence in the ICR Part III data that existing controls can achieve greater than 21 percent removal in lignite units. As a result, EPA's proposed mercury standards are likely to have significant impacts on US coal supply and the coal industry.

Table 8.2 Coal Not Achieving Compliance with EPA Proposed MACT Limits with Stated Confidence When Fired in Units with Top Performing Control Technology (Percent of Btus)

Probability of Compliance with EPA MACT Floors	Bituminous	Subbituminous	Lignite		
			All Lignite	Fort Union Lignite	Gulf Coast Lignite
97.5	49%	41%	62%	37%	71%
95	43%	40%	60%	37%	68%
90	36%	38%	59%	33%	68%
80	29%	31%	55%	18%	68%
70	24%	18%	52%	13%	66%
60	20%	10%	52%	12%	66%
50	15%	6%	52%	12%	66%

8.2 New Units

For new units, EPA has proposed output-based standards that are expressed as pounds of mercury emitted per megawatt hour. For purposes of converting the limits to an input basis, we have assumed a heat rate of 9,500 Btus/kwh hour for new units, which is 5 percent lower than the heat rate (10,000 Btus/kwh) that the proposed rule adopts for existing units. This rate is based on an assessment of a prototypical new coal-fired US power plant firing western coal[25]; by adopting an improved heat rate, the analysis will avoid understating the input-based limits that are equivalent to the output-based standards. The resulting equivalent input-based limits are given in Table 8.3. New bituminous units would be required to meet a standard equivalent to 0.63 lbs/TBtu, subbituminous units a standard equivalent to 2.1 lbs/TBtu, and lignite units a standard equivalent to 6.5 lbs/TBtu. Using the methods described for existing units, we have estimated the probability that US coals can comply with EPA's proposed new unit standards, in units that have adopted the best control technologies identified in the ICR database.

---

[25] "Feasibility of New Coal Fueled Power Plants," presentation by Black & Veatch Energy Services Group at the 9th Clean Fossil Energy Technical Seminar, APEC Energy Working Group, March 2002.

Table 8.4 shows that only a very small portion of the US coal supply can comply with the proposed limits for new units with high statistical confidence. Only 18 percent of the bituminous coal supply, 8 percent of the subbituminous coal supply, and 22-24 percent of the lignite coal supply are found to achieve compliance with 97.5 percent confidence. More than three-fourths of the lignite supply, more than 80 percent of the bituminous supply, and more than 90 percent of the subbituminous supply cannot achieve compliance with high statistical confidence.

Table 8.3 EPA's Proposed MACT Limits for New Units (Assuming Heat Rate of 9,500 Btu/kwh)

	Proposed Output Limit 10^{-6} lbs/Mwh	Equivalent Input Limit Lbs/TBtu
Bituminous	6	0.63
Subbituminous	20	2.1
Lignite	62	6.5

Table 8.4 Ability of Coal to Achieve Compliance with EPA Proposed MACT Limits with 97.5% Confidence When Fired in New Units with Top Performing Control Technology (Percent of Btus)

	EPA Proposed Standards	Percent Achieving Compliance	Percent Not Achieving Compliance
Bituminous	0.63 lbs	18%	82%
Subbituminous	2.11 lbs	8%	92%
Lignite			
All Lignite	6.5 lbs	23%	77%
Fort Union Lignite	6.5 lbs	24%	76%
Gulf Coast Lignite	6.5 lbs	22%	78%

US coals in all ranks will face low probabilities of compliance, even when fired in new units that use the best performing technologies. As Table 8.5 shows, 57 percent of the US bituminous supply will be unable to comply with at least 80 percent confidence, as will 89 percent of the subbituminous supply and 72 percent of the lignite supply (64 percent for Fort Union lignite and 74 percent for Gulf Coast lignite). Forty two percent of the bituminous supply will have no better than a 50 percent probability of compliance, as will 62 percent of the subbituminous supply and 70 percent of lignite supply (59 percent for Fort Union lignite and 74 percent for Gulf Coast lignite). That the difficulty in complying that is faced by Gulf Coast lignite coals is largely unaffected by the confidence level is a direct result of the high mercury content typical of these coals and the absence of evidence in the ICR Part III data that any control configuration can achieve more that about 21 percent removal in lignite units. As a result, EPA's proposed limits for new units are likely to have very significant impacts on the ability of US coals to contribute to meeting future needs for new electric generating capacity and on the efforts of the US electric industry to provide for these needs.

Table 8.5 Coal <u>Not</u> Achieving Compliance with EPA Proposed MACT Limits with Stated Confidence When Fired in New Units with Top Performing Control Technology (Percent of Btus)

Probability of Compliance with EPA MACT Floors	Bituminous	Subbituminous	Lignite		
			All Lignite	Fort Union Lignite	Gulf Coast Lignite
97.5	82%	92%	77%	76%	78%
95	77%	91%	75%	76%	75%
90	68%	90%	75%	75%	75%
80	57%	89%	72%	64%	74%
70	51%	83%	70%	59%	74%
60	46%	73%	70%	59%	74%
50	42%	62%	70%	59%	74%

9. CRITIQUE OF THE USE OF ICR PART III DATA IN RECENT EPA PAPER ON CONTROL OF MERCURY EMISSIONS

9.1 Introduction

EPA has released a white paper entitled "Control of Mercury Emissions from Coal-Fired Electric Utility Boilers"[26] that reviews the performance and status of technologies for controlling mercury emissions, including:

- Existing technologies for controlling PM and SO2 emissions that also reduce mercury emissions
- Technologies designed specifically for mercury control, such as activated carbon injection (ACI), and
- NOx control technologies such as SCR that appear to enhance mercury capture in existing SO_2 and PM controls.

In its review of R&D issues, the EPA paper uses a variety of statistics taken from the ICR Part III data to provide both an R&D context and to characterize the mercury control performance of existing control technologies. The use of these statistics leaves the unwarranted impression that existing controls have been demonstrated to achieve a substantial, consistent and dependable level of mercury control. And, unlike its discussion of R&D issues related to advanced control options, the paper does not disclose the limitations of the ICR data or qualify the inferences drawn from the data.

As presented earlier in this report, the ICR Part III data used by EPA embody a very limited number of tests on a small subset of the coal units in the US population. The test results are highly variable – both within the repeat tests conducted for each unit and across the average results for units with similar control configurations. Further, it is clear that the performance of controls in highly variable with the characteristics of the coal. The following discussion highlights areas in which the EPA paper presents misleading statistics, impressions, and conclusions based on the ICR data.

---

[26] "Control of Mercury Emissions from Coal-Fired Electric Utility Boilers". Air Pollution Prevention and Control Division. U.S. Environmental Protection Agency (Research Triangle Park, NC). 2004.

9.2 The Limitations of Small Samples

The ICR Part III testing measured stack emissions at a sample of 80 generating units. The sample was designed as a stratified survey that selected up to three units for testing in each of 36 different survey cells, defined by combinations of SO_2 and PM controls and the rank of coal fired. As a result, there is wide coverage of technology configurations in the ICR data, but only very limited depth for any particular technology, and one must exercise considerable caution in relying on statistics for individual technology configurations. Statistics for individual technologies can be unduly influenced by characteristics of the individual units tested, and by the conditions prevailing on the day of the test and may not be representative of the actual performance of the technology.

Table 9.1 gives the sample sizes that correspond to the average mercury capture rates reported in Table 1 of the EPA White Paper[27]. The sample size is very limited for every technology cited by EPA as achieving an appreciable capture rate, and this is specifically true for the fabric filter technologies which are identified as achieving the highest capture rates.

Fabric filter controls only (not scrubbed) are cited as achieving 90 percent mercury capture for bituminous coal and 72 percent for sub-bituminous coal, but this is based on samples of only 4 and 2 units, respectively. The results for individual tests vary widely. One bituminous fired unit (Valley 3) exhibited negative removals (stack emissions in excess of coal mercury content) on each of its three tests. For the three other units, individual test results ranged from a low of 35 percent removal to as high as 91 percent, and the units showed removals of 49 percent, 81 percent, and 90 percent for the average of the three tests. The data do not support a conclusion that fabric filters on units firing bituminous coal consistently achieve a control level at or near 90 percent.

---

[27] The data in the table reflect conventional coal units (pulverized coal and cyclone boilers) and exclude 6 FBC units, 2 IGCC units, and 1 stoker-fired boiler. All configurations of existing controls are tabulated for conventional units, including categories not covered in Table 1 of the EPA paper.

Table 9.1 Sample Sizes for Technology Groups Tested in ICR Part III

	Number of Units Tested in ICR Part III		
Configuration	**Bituminous Coal**	**Subbituminous Coal**	**Lignite Coal**
PM Control Only			
CS-ESP	8[28]	5	2
HS-ESP	3	6	Not tested
FF	4	2	Not tested
FF + CS-ESP	2	Not tested	Not tested
PS	2	Not tested	1
PM Control + Spray Dryer Adsorber			
SDA + CS-ESP	Not tested	3	Not tested
SDA + FF	3	3	3
Sorbent Injection + CS-ESP	1	Not tested	Not tested
PM Control and Wet FGD System			
FGD + CS-ESP	3	3	2
FGD + HS-ESP	3	3	Not tested
FGD + FF	2	Not tested	Not tested
FGD + PM	Not tested	6	Not tested

Ninety percent control may be achievable with fabric filters on some units, some of the time in short-term testing, but a lower level of control should be expected in a population of units, and no data has demonstrated a comparable, sustained level of control in long-term testing.

Fabric filter controls combined with dry scrubbing (spray drier adsorber, or SDA) are cited as achieving 98 percent capture (with or without SCR) on bituminous units, but this category is represented by only 3 units in the ICR Part III data. The units that were tested show consistently high mercury removals, but all were tested on coals with relatively high chlorine content – one unit was tested on coals with 70,000 to 75,000 lbs of chlorine per TBtu and the other two units on coals with 110,000 to 135,000 lbs/TBtu. High coal chlorine content is widely believed to enhance the capability of fabric filters to capture mercury from stack gases. That these units exhibited consistently high capture rates in the ICR Part III testing may be due more to their high coal chlorine contents than to the demonstrated performance capability of fabric filters combined with dry scrubbing. Substantially lower mercury capture rates would be expected on coals with lower chlorine content.

Fabric filter controls combined with wet scrubbing are cited as achieving 98 percent capture for bituminous units, but this category is represented by only 2 units in the ICR Part III data. The two bituminous units that were tested showed mercury removals ranging from 79 to 99 percent in individual tests, and the units averaged 86 percent and 97 percent for the average of 3 tests. We compute the average for the category as 91 percent, not EPA's 98 percent. One unit (Intermountain 2SGA) stands out from other units employing fabric filters in showing relatively high mercury removal on test coals that were very low chlorine content in comparison to typical bituminous coals. As EPA notes in its discussion of field test units, unit-specific configurations and factors – including the sizing of controls, lengths of duct runs, and characteristics of fly ash – can make individual units unrepresentative of the larger population employing similar controls.

The differing performance levels cited by EPA for the three configurations of fabric filter controls appear to be more the result of the interpretation of small samples in isolation from each other and without an effort to address the effects of coal chlorine content on the mercury removals. Further, the report makes no effort to clarify that the data base that was used limits the applicability of the reported results. EPA should have noted that the results were achieved in single short term tests under controlled conditions, the performance of the controls is known to be highly variable and that the levels reported have not been achieved consistently over a range of coals, operating conditions and boiler configurations.

Chapter 4 of this report presented a re-analysis of the ICR Part III data for the purpose of characterizing mercury removal performance by technology as a function of coal chlorine content. The results of that analysis demonstrated that, once the effect of coal chlorine content is accounted for, the performance of fabric filter controls does not appear to differ among the configurations with and without scrubbers. That is, firing a coal with high or low chlorine content was found to have a larger effect on mercury removal and, within the modest statistical

---

[28] One of the eight units, which was blowing soot during the first test sequence, was retested on a second occasion.

power of the small sample of data on these units, the presence or absence of a scrubber had no detectable effect.

Based on the analysis in Chapter 4, the ICR Part III data support the conclusion (with 95 percent confidence) that fabric filters achieve between 75 percent and 90 percent removal (for the average of three tests) for bituminous coal with a chlorine content of 50,000 lbs/TBtu. At lower chlorine contents, the average mercury removal performance is poorly characterized by existing test data and may range from as low as –30 percent removal to as high as +70 percent removal, with 95 percent confidence. Thus, the data that EPA has used will not allow one to say with 95 percent confidence that a fabric filter will remove, on average, any of the mercury present in low chlorine coals, regardless of rank. EPA is incorrect in its assessment based on the ICR Part III data that fabric filter controls, in any configuration, have been demonstrated to achieve high levels of mercury control on a sustained basis.

Whenever sample sizes are small, as in this case, the data provide little or no statistical confidence – i.e., statistics will have wide confidence bounds (uncertainties). A further danger of small sample sizes is that the performance, even if consistently measured in repeat tests, may be influenced or caused by unit-specific factors that may not be representative of the actual performance of a larger population of units having the technology. In samples of 2, 3, or 4, the average can be greatly affected if even one unit is atypical of the larger population in some respect, and it is not possible to test for "outliers" because there are too few units to compare against.

An example of the problems encountered in small samples can be found in the group of units with cold-side ESP for PM control only. Table 1 of the EPA paper cites an average mercury capture of 36 percent when firing bituminous coals, and this statistic is important because there are many units in the US population with this control configuration. Eight units with this control configuration were tested three times each, and one of the units (Gibson 3) was later tested again, for a total of 27 tests. Four of the 27 tests exhibited negative mercury removal. The one unit that was later retested showed negative removals on each of its three initial tests, with stack emissions that were three times the mercury content of the coal. Five of the remaining eight units gave average removals (over three tests) that were below the 36 percent removal average that EPA cites, while three units gave higher values.

Although EPA cites an appreciable removal value, the performance of this technology is highly variable. In fact, the emissions analysis presented in Chapter 4 estimates that cold-side ESP on bituminous units achieves a 46 percent mercury removal rate, but with a 95 percent confidence interval that ranges from –61 percent to 100 percent. That is, the confidence bounds are so wide that, based on the ICR Part III data, one cannot conclude with adequate statistical confidence that the technology is actually capable of reducing mercury emissions below the mercury content of the coal.

9.3 <u>Short-term Nature of Testing and Operating Bias</u>

As documented in the EPA paper, the processes of mercury liberation, capture, and emission are complex. Capture in existing control devices is thought to depend largely on chemical reactions in the stack gas to form mercury compounds that can be captured in PM or SO_2 controls. The composition of coals and the stack gas temperatures prior to the control devices play an important role in the stack gas chemistry. These factors, and therefore the mercury capture performance of control devices, can be expected to vary widely from one unit to the next and, for a given unit, to vary within the year and across a range of operating circumstances from full to partial load, steady-state to transient operations, and during maintenance events.

The ICR Part III data are the result of one-time testing of units under conditions that EPA admits sampled emissions only under full-load and steady-state operation. Three Ontario Hydro tests, lasting approximately 2 hours each, were conducted back-to-back on each unit, typically on the same day or shift. Therefore, the ICR data sample only about six hours of operation for each unit or about 0.1 percent of the hours of operation in a year.

While the available data on advanced mercury control technologies is described in the EPA paper as “limited” based on the short-term nature of testing, the testing periods (4-9 days) actually exceed by far the operating hours measured for any unit in the ICR data. The ICR data are at best “snapshots” of emissions performance at a given point in time and provide *no* *information* that can be used to identify the units or technologies that are capable of achieving significant mercury capture over an annual period (as expected for compliance demonstration).

The nature of the ICR data will be to understate the variability in emissions performance for each unit, because only the particular coal and operating characteristics that prevailed on the day of testing were sampled. Further, a high degree of variability will be displayed across units, because some units happened to be tested on days when coal and operating characteristics were favorable to mercury capture, while other units were tested on days when these factors were unfavorable. In small samples (as for the fabric filters configurations), it is possible to have units that happened to be tested on good days and that therefore lead to mistaken conclusions about the mercury capture capability of an entire technology class.

9.4 High Variability Present in the ICR Data

The ICR data are characterized by a high degree of variability in emissions performance. The Ontario Hydro method is a relatively complex test method and a number of other parameters (coal mercury content and calorific value, coal feed rates, and stack gas flow rates) must be measured in order to compute mercury emissions in lbs per Trillion Btu (lbs/TBtu) and mercury capture rates. The variability inherent in the test methods is compounded by the snapshot nature of the testing and by the potential for unidentified factors to have significant and unrecognized effects in the data. Further, the high variability is not adequately attenuated by the small number of tests that were conducted on each unit. The resulting high variability must translate into wide confidence bounds (uncertainties) for any statistics developed from the data. Nowhere does EPA present or discuss the uncertainties of the statistics on mercury capture for existing controls.

Key conclusions of the EPA paper are that a high level of mercury removal can be achieved for bituminous coal and that fabric filters are a top-performing technology, particularly when combined with dry scrubbers (spray dryer adsorbers). Of the 9 bituminous units with fabric filters that were tested, one unit was measured to emit more mercury than present in its coal as the average over 3 tests, another unit achieved 53 percent average removal, three units achieved 80-90 percent removal, and 4 units more than 90 percent removal. The fabric filter units achieving the highest mercury removals were all tested on high chlorine coals. High removal rates are not likely to be achieved when bituminous coals of low chlorine content are fired. The use of average statistics for small groups tends to give a simple and narrow view of the data, and the EPA paper gives no indication of the variability in the results.

While appreciable removal rates appear to be possible for bituminous units with fabric filters, the testing results are not uniform among units, and there is evidence that other factors (not accounted for in the simple statistics) may influence removal rates. EPA's own analysis identified in field testing that two of the units being used to evaluate advanced technologies produced high levels of unburned carbon in their fly ash and were likely to be unrepresentative of other units equipped with similar technologies. The ICR Part III data did not collect information on fly ash characteristics, so that the presence of unburned carbon in fly ash could contribute to mercury removal in fabric filter units in ways that cannot be identified and that will not be representative of larger populations of generating units.

9.5 Conclusion

The EPA paper presents a narrow and misleading view of the mercury capture performance of conventional SO2 and particulate control technologies. If the purpose of the paper was to communicate what is and is not known about mercury control, the paper should have discussed the limitations of the data from which conclusions were drawn, the variability and uncertainty of the results in that data, the performance that can be expected over a range of coal types, the confidence intervals for those estimates and what EPA is doing to improve the state of knowledge on the effectiveness of conventional as well as advanced control systems.

The basis is very weak for EPA conclusions to the effect that by 2015, control systems for NO_x and SO_2 will "… have the potential to achieve 90 to 95 percent control of mercury." Existing SO_2 and PM emissions controls clearly have potential for mercury removal, and that potential may be enhanced through the development of advanced technologies. However, the state of knowledge regarding mercury capture by existing controls – based largely on snapshot testing of a small number of units in each technology group – is more tentative than acknowledged in the EPA paper, and in fact, mercury control will be difficult to achieve in many existing units.