

# **Quantification of Emission Factor Uncertainty**

**George Pouliot**

Atmospheric Modeling and Analysis Division, National Exposure Research Laboratory,  
Environmental Protection Agency, Research Triangle Park, NC 27711

**Emily Wisner**

Statistics Department, North Carolina State University, Raleigh NC

**David Mobley**

Atmospheric Modeling and Analysis Division, National Exposure Research Laboratory,  
Environmental Protection Agency, Research Triangle Park, NC 27711

**William Hunt, Jr**

Statistics Department, North Carolina State University, Raleigh NC

## INTRODUCTION

Emission factors are important for estimating and characterizing emission sources of air pollution. Emissions are being released into the air every day from different sources and are monitored in various ways. Emission factors are generally estimated from an average of all available data<sup>1</sup>. However, the majority of emissions factors are based off estimates created by the US Environmental Protection Agency (EPA) in years past, using data of various quality and quantities. The EPA has compiled emission factors in a document entitled, *Compilation of Air Pollutant Emission Factors*<sup>1</sup> (AP-42). These factors are basically averages from available source tests. In many cases, the available source tests are from a very small sample set. Since we don't have numerous tests from a variety of sources to estimate an emission factor, the minimal numbers of available tests leads to uncertainty in the emission factors. A letter rating has been given for each emission factor estimated by the EPA to represent a general indication of the reliability or robustness of the emission factor and is assigned based on the estimated reliability of the tests used to develop the factor and on both the quantity and representativeness of the data. Different sources and pollutants that have the same robustness and reliability in the measured emission factor and in the representativeness of the measured values are given the same rating. The AP-42 emission factor rating is based on two factors. First, the quality of the test data used to create the emission factor is given a letter rating from A to D. These are defined in Table 1. Second, the ability of the factor to represent a national average emission factor for that category is assessed. These two pieces of information are then used to define the A to E letter rating of AP-42 which are defined in Table 2. The AP-42 emission factor estimates represent an average estimate of the emission factors from all units of a given Source Classification Code (SCC)<sup>2</sup>. In many cases, the limited number of reliable data points leads to increased uncertainty in the emission estimates. The nature of the letter rating indicator does not provide a quantitative assessment of uncertainty of emission inventories used in air quality modeling applications. An EPA report from the office of the inspector general<sup>3</sup> specifically recommended that EPA establish a rating system that provides the quantitative range of uncertainty for emissions factors for both inventory and non-inventory purposes. The objective of this study was to explore potential options to statistically and objectively quantify the uncertainty of emission factors for inventory purposes. The focus of the first part of this study was on nitrogen oxide (NO<sub>x</sub>) emissions from electric generating units (EGUs), which is probably the best tested and

characterized of any pollutant and source category combinations. The EGUs of interest in this study that produce NO<sub>x</sub> use coal, oil, or natural gas as a fuel in an external combustion boiler to generate electricity.

### **Key Terms for this Study**

- Emissions Uncertainty – The uncertainty associated with the emissions calculated for a pollutant and source category.
- CEMS – Continuous Emission Monitoring Systems, a method for continuously monitoring emissions and collecting data averaged over intervals of a few minutes.
- Emission Factor – a representative value that attempts to relate the quantity of a pollutant released to the atmosphere with an activity associated with the release of that pollutant.
- AP- 42 – this publication series is the principal means by which the Environmental Protection Agency documents and distributes its emission factors.
- SCC – Source Classification Code, an eight digit number, or 10 digits for area sources, which the EPA uses to classify various source categories. It provides a way to categorize emission sources in a consistent way when creating an emission inventory.
- WebFIRE – the EPA’s online database that attempts to catalog descriptions of SCCs and their emission factors.
- Standard z-score – the number of standard deviations an observation (or mean of observations) is away from the actual mean; often used in statistics for calculating probabilities of seeing extreme values.

### **Problem**

Currently, statistical quantification of the uncertainty of emission factors provided in AP-42 is not available. Quantifying the uncertainty of emission factors would enable the scientific community to apply them more meaningfully. Using a quantitative characterization of emission factor uncertainty, scientists will be able to more accurately characterize the uncertainty associated with air quality modeling and emissions inventories. A quantitative measure of uncertainty will also give decision makers the ability to determine the confidence which should be placed in the analysis of the data being used for policy decisions.

## **Objectives and Assumptions**

This study consisted of three main objectives and a separate Phase of the study addressed each objective:

1. Compare the NO<sub>x</sub> emission factors from combustion sources with currently available continuous emission monitoring data. CEMS data has been shown to be accurate to within 5%<sup>2</sup>.
2. Develop quantitative uncertainty indicators for A through E rated data quality indicators for emission factors based on NO<sub>x</sub> emissions from combustion sources.
3. Estimate a range of uncertainty associated with the A through E ratings by applying the methods from phase 2 to additional sources and pollutants.

Although there are different emission factors for different pollutants and different source categories, there are some “common” characteristics of the emissions factors that have been assigned the same “letter rating” for their data quality indicator. This letter rating represents a general indication of the reliability or robustness of the emission factor and is assigned based on the estimated reliability of the tests used to develop the factor and on both the quantity and representativeness of the data. Different sources and pollutants that have the same robustness and reliability in the measured emission factor and in the representativeness of the measured values are assumed to have a similar quantifiable uncertainty. For the purposes of comparison, we are assuming that the emission factor estimates from source categories with the same letter rating have enough robustness and consistency that we can quantify the uncertainty of these common emission factors based on the qualitative indication of data quality which is known for almost all factors. Thus, we will try to infer an uncertainty for all emission factors based on their qualitative data quality indicator by the quantitative analysis of a subset of the data. Although this assumption can be challenged, this paper still provides a framework for the comparison of emission factors from multiple sources and pollutants. These results could be refined if additional data become available.

## **Database**

In order to analyze the variability of NO<sub>x</sub> emission factors from EGU point sources, several databases of information needed to be combined. First, the CEMS monitoring data from the EPA's Clean Air Markets Division contains hourly NO<sub>x</sub> emission rates in lbs of NO<sub>x</sub> per million British thermal units (lbs/10<sup>6</sup>Btu)<sup>4</sup>. Second, the Department of Energy's Energy Information Administration has monthly fuel information for selected point source EGUs<sup>5</sup>. This set of data includes the quantity for fuel consumed per month at a given plant and the heat content of the fuel (10<sup>6</sup> Btu/ton of fuel). The EPA's National Emission Inventory contains plant information, including stack parameters and the Source Classification Codes for these point sources<sup>6</sup>. All three databases use the Department of Energy's Office of Regulatory Information Systems (ORIS) identification code, a unique robust primary key, to identify specific EGU plants by individual boiler. By using this common identifier to join the databases, a NO<sub>x</sub> emission factor (in tons of NO<sub>x</sub> per ton of fuel consumed) was calculated on an hourly basis for all plant/boilers that are common to the three databases. Hours in which the plant was operating for only a fraction of an hour (start-up and shutdown) were discarded. In addition, only CEMS records marked 'measured' were included. All estimated values in the CEMS database were discarded. Thus, a new database containing hourly computed NO<sub>x</sub> emission rates comparable to the AP-42 emission factors for all the facilities in the United States was created. The relevant sections of AP-42 are chapters 1.1 through 1.4. The actual emission factors from AP-42 were extracted from EPA's WebFIRE, rather than the actual text of AP-42. These emission factors are for "uncontrolled" sources. Uncertainties may exist in the measurements in the CEMS database. However, the CEMS data is considered to be reliable to within 5%<sup>2</sup>, so this analysis did not take any possible uncertainties associated with the CEMS database into concern. The years of data were 1997 to 2007. There were data for 52 different SCCs in the initial data set.

## **PHASE I**

### **Approach**

To create a database where emission factors from different SCCs could be compared, all AP-42 values were standardized to lbs of NO<sub>x</sub> per million Btu (lbs/10<sup>6</sup> Btu). This standardization was done in order to ensure that values of the same units were being compared during the analysis. Once the data were properly formatted, SAS (Statistical Analysis Software) programming was

used throughout the duration of the project for most of the analyses. Table 3 summarizes the SCCs used in this study. The top 2% of NO<sub>x</sub> emission factor values from each SCC were removed. The decision to do this trim was based on a previous study of volatile organic compounds emissions from petrochemical plants, in which the researchers attributed extraneous ambient air quality values to equipment malfunctions or some abnormal operation<sup>7</sup>. The top 2% of emission factor values from each SCC were removed rather than the top 2% of values from each plant, because only a few plants in each SCC appeared to have extraneous values. The bottom 2% of emission factor values were also removed from the database because of a large number of 0 values, which were attributed to errors or plant shutdowns. Figures 1 and 2 show the box plots of all the plants for one SCC before and after trimming the data. Only one plant appeared to have any extraneous values for this SCC. This example shows why the top and bottom 2% were removed from the datasets.

After careful inspection of the data, it appeared there were issues with the quality as a result of some EGUs having multiple SCCs (e.g., multiple fuels). When this was the case, the most dominant SCC was kept and the others were thrown out of the analysis. Deciding which SCC was dominant was based on whether it had the most hours of operation and if it had at least an order of magnitude (i.e. a factor of 10) greater in emissions than any other SCC for a particular boiler. In the process of cleaning up the data set, the influence of starter fuels and duplicate values were removed. As a result, a total of 13 SCCs were entirely removed from the database.

Between the years 2002 and 2007, some plants phased in controls between May 1<sup>st</sup> and September 30<sup>th</sup> through the various control strategies. For this reason, data from these plants were removed from the analysis since only uncontrolled, or as combusted, emission factors were of interest since this is the basis for the AP-42 emission factors. To determine which plants were in fact phasing in controls, time plots of every plant of every SCC were produced using SAS programming and were carefully observed. Figure 3 is a time plot of NO<sub>x</sub> emission factors for plant 963 of SCC 10100203, where periods of controls clearly began in 2003. Figure 4 shows the results of removing the controlled data. Figures 3 and 4 are likely examples of plants which employed either non catalytic or selective catalytic reduction of NO<sub>x</sub> with ammonia to achieve

the emissions reductions. Since the facilities were required to reduce emissions only during specified periods, they reduced their overall cost by not operating the equipment at all times.

Upon looking through each individual plant of each SCC, some plants appeared to have controls in during every month of the year starting at varying dates around the year 2000. Plants that exhibited this trend had data removed starting at the dates in which controls clearly looked present. Figure 5 is a time plot of one plant in which controls looked present throughout the year beginning in April, 1998. These screening techniques eliminated data from facilities which may have changed their operating characteristics during the study period. A subsequent study of the efficiency of post-combustion control equipment may be desirable, but this study was limited to as combusted emission factors for comparison to AP-42 values

Although some SCCs had over 100 plants, others had as few as 1 or 2 plants. SCCs with only 1 or 2 plants were removed from the analysis due to insufficient amounts of data. With this removal of SCCs from the analysis as well as SCCs previously being removed due to starter fuels and other issues, the final analysis only consisted of 21 different SCCs. However, the number of observations for the 21 SCCs used in this analysis accounted for about 77% of the total number of observations in the original database. Of these 21 SCCs, 12 were coal, 3 were oil, and 6 were natural gas.

After the data were properly formatted and appropriate data were removed from each plant, SAS programming was used to compute the mean emission factor for each SCC. The percent difference between the mean emission factor and AP-42 value were then computed to determine how well the values in AP-42 compared to the continuous emissions data. The fundamental approach here is to compare existing AP-42 emission factors with CEMS emission factors. AP-42 emission factors are "uncontrolled or as combusted" by definition. The efficiency of control devices is determined by other means and applied after the emission factor.

## **Results and Discussion**

After looking at the percent difference between the AP-42 emission factor and the mean NO<sub>x</sub> emission factor for each SCC, based on this analysis of CEMS data, it is clear that many of the

AP-42 values were significantly different from the CEMS values. Although 13 of the 21 SCCs in this study received AP-42 letter grades of A, the majority of the percent differences between the AP-42 emission factor values and the means were substantially large (greater than 50%), as shown in Table 3. The discrepancy between the two is likely due to the fact that some of the NO<sub>x</sub> emission factors from the AP-42 were last updated in September 1998 and other factors were last updated even earlier, and technology has improved and changed since then<sup>1</sup>.

Many of the AP-42 emission factors need updating to reflect the CEMS emission measurements. There were 62% of SCCs which had a percent difference between EPA's AP-42 emission factor and mean of continuous emissions data greater than  $\pm 25\%$ . There were 29% of SCCs which had a percent difference between EPA's AP-42 emission factor and mean of continuous emissions data greater than  $\pm 50\%$ . There were 14% of SCCs had a percent difference between EPA's AP-42 emission factor and mean of continuous emissions data greater  $\pm 100\%$ . For example, SCC 10100203, as shown in figure 6, received an A data quality rating for the AP-42 letter grade. For this SCC, however, the AP-42 emission factor is 48% above the mean emission factor of the CEMS data. The continuous NO<sub>x</sub> emission factor data for this SCC being reasonably normally distributed with the AP-42 emission factor far above the mean suggests the AP-42 value needs to be updated for this particular SCC. Based on the analysis of Phase I, several of the AP-42 emission factor values for the 21 SCCs in this study need to be updated to reflect the currently available continuous NO<sub>x</sub> emissions data.

## **PHASE II**

### **Approach**

*Calculating Uncertainty Values.* The goal of Phase II of this research was to develop a quantitative measure of uncertainty for each of the EPA's qualitative letter grades currently being used as data quality indicators. In order to do this, a few assumptions had to be made about what characterized an AP-42 emission factor as either an A, B, C, D, or E data quality rating. Table 4 shows the assumed sample sizes associated with each of the data quality ratings<sup>8</sup>. We included the sample size assumptions that were made because they were not explicitly stated in the AP-42 documentation. Different assumed sample sizes associated with each letter grade would result in a different analysis. For this analysis, as noted earlier there are two key factors



associated with the data quality rating: (1) quality of the testing method (2) representativeness of the factor for a national emission inventory. Since these data are from multiple facilities, we feel that our large data sample is representative for a national emissions inventory, satisfying factor (2). All the data used was from CEMS, implying that the quality of the data was consistent for the entire dataset. We used a varying sample size as indicated in Table 4 to simulate the AP-42 emission factor rating of A through E. The level of uncertainty for each of the 5 sample sizes,  $n$ , for each SCC was calculated to be the probability that a sample mean of a sample of size  $n$  will not be within 10% of the population mean:

$$\text{Uncertainty} = P(|\bar{x} - \mu| > 0.1\mu) \quad (1)$$

$$= P(\bar{x} - \mu > 0.1\mu) + P(\bar{x} - \mu < -0.1\mu) \quad (2)$$

$$= P(\bar{x} > 0.1\mu + \mu) + P(\bar{x} < -0.1\mu + \mu) \quad (3)$$

$$= 2P(Z > \frac{0.1\mu}{\sigma/\sqrt{n}}) \quad (4)$$

where  $\bar{x}$  refers to the sample mean;  $\mu$  is the population mean;  $\sigma$  is the population standard deviation; and  $Z$  is the standard z-score. We use this definition of uncertainty rather than one based only on the standard deviation because this definition is dimensionless and thus easier to interpret. The population mean and standard deviation for each SCC was assumed to be the calculated emission factor mean and standard deviation of the entire SCC, since the CEMS data consisted of such an enormous number of observations. SAS programming was used to compute the probabilities from equation 4 for each SCC.

This approach, however, does depend on normality. Calculating probabilities about sample means from the standard normal distribution assumes the sampling distribution is normal. A sampling distribution will be normal if the population distribution is normal, or if the sample size is large, no matter what the population distribution. An A data quality rating is assumed to have a sample size of 25, which can be considered large enough for a representative characterization of the source category. However, D and E ratings are assumed to have sample sizes of 3 and 1, respectively. These small sample sizes pose some problems, since not all of the SCCs were normally distributed. To check the theoretical calculations, bootstrap methods through SAS programming were used for some of the very non-normal SCCs. Ten thousand samples for each

of the 5 sample sizes were simulated for the selected SCCs and the means were calculated. For each sample size for these SCCs, the level of uncertainty was calculated to be the percentage of sample means out of 10,000 that did not fall within 10% of the population mean (the mean of the entire SCC). These uncertainty values matched up extremely well with the theoretically calculated uncertainty values, even for very small sample sizes. It was concluded that using the theoretically calculated uncertainties as opposed to using bootstrap methods for every SCC would make no difference in this study, particularly since rounding was to be done. The letter grade uncertainties for each SCC were then averaged to create overall uncertainties for the five letter grades and then rounded.

## **Results and Discussion**

Uncertainty was defined as the probability that a sample mean of a sample of size  $n$ , where  $n$  is 25, 10, 5, 3, or 1, will not be within 10% of the true mean. Table 5 shows the uncertainties for the five letter grades for all 21 SCCs. These provide quantitative metrics for the letter grades. The rounded average uncertainty for an A rating is 25%. This means that if an SCC received an A rating (assuming an A rating means a sample of size 25 was taken to compute the AP-42 emission factor), there is about a 25% chance the sample mean will not be within 10% of the true emission factor mean. On the other hand, if an SCC receives an E rating (assuming an E rating means a sample size of only 1 was taken), there is about an 80% chance the sample mean will not be within 10% of the true emission factor mean.

## **PHASE III**

### **Approach**

To determine the possibility of applying the uncertainties associated with the different letter grades for NO<sub>x</sub> emissions to other pollutants, another data set consisting of various pollutants and source categories was analyzed. This new data set is from the study by RTI International for EPA and included emission factor data for 44 different pollutant and source category combinations<sup>9</sup>. The uncertainty values for the five letter grades were calculated for each of these pollutant and source category combinations as described under Phase II. The uncertainties for each of the letter grades were averaged across pollutant and source category combination. These

letter grade uncertainty averages were then combined with the uncertainties calculated in Phase II to construct overall uncertainty ranges for each of the five letter grades that could possibly be applied to any pollutant.

### **Results and Discussion**

Table 6 shows the 44 different pollutant and source category combinations used in the previous study along with the calculated uncertainties for each letter grade. These sets of data yielded higher uncertainty values than the previous data set, which is due to most of the pollutant and source category combination distributions being log-normal. A log-normal distribution, as shown in figure 7, is skewed right with the bulk of the data, as well as the mean, to the left. As described earlier, this leads to larger uncertainties. The uncertainties for each pollutant and source category from phases two and three were combined into a single table (the data in tables 5 and 6). The interquartile range of the uncertainty for each letter grade from this table was used to estimate the overall range of uncertainty associated with that letter grade. These overall ranges of uncertainty are shown in Table 7. According to these calculated uncertainty ranges, an A rated sample of emission factors, assuming the sample size was 25, would have between 25% and 62% uncertainty associated with it. In other words, if a sample of size 25 emission test data sets for any pollutant is taken, the probability that the sample mean is not within 10% of the true mean is between 25% and 62%.

### **SUMMARY AND CONCLUSIONS**

NO<sub>x</sub> emission factors from EGUs were computed from CEMS data as well as EPA and DOE data sets and compared to AP-42 emission factors. The inconsistency between the emission factors from CEMS data and the AP-42 emission factors for several SCCs suggests that AP-42 needs to be updated to reflect the NO<sub>x</sub> continuous emissions data now available for EGUs. This also indicates that substantial targeted and prioritized parametric source testing needs to be done on many source categories and pollutants to provide a reliable database to develop new and better quality emission factors. Uncertainty values were calculated for each letter grade for each SCC, under the assumption that certain sample sizes were associated with the letter grades.

Uncertainty was calculated as the probability that a sample mean from a sample of size  $n$  will not be within 10% of the true emission factor mean. To determine the possibility of applying the letter grade uncertainties computed for NO<sub>x</sub> emissions from EGUs to other pollutants and source categories, another data set with various combinations of pollutants and source categories was analyzed. Uncertainties for each letter grade were calculated for the new data set and compared to those calculated from the continuous NO<sub>x</sub> emissions data from EGUs. Uncertainty ranges were then computed based on analysis of the two data sets.

For the AP-42 data quality ratings, the quantified uncertainty ranges were 25-62% for A rated emission factors, 46-75% for B rated emission factors, 60-82% for C rated emission factors, and 69-86% for D rated emission factors, and 82-92% for E rated emission factors. These uncertainty ranges represent a typical range of that could be used for other sources and categories when only a letter rating is available. Using these uncertainty estimates for emissions will enable quantification of uncertainty estimates for air quality modeling and emission inventories. This quantitative measure of uncertainty will also give decision makers the ability to determine the confidence which should be placed in the analysis of the data being used for policy decisions. Chapter 2 of the 2006 EPA report from the Office of the Inspector General<sup>3</sup> explains why use of unreliable emission factors adversely impacts key environmental decisions.

## **ACKNOWLEDGEMENTS**

The authors acknowledge Alissa Anderson, Colin Geisenhoffer, Brody Heffner, and Michael Shaw from North Carolina State University for their previous work on this project in ST 495 and ST 498 Environmental Data Analysis Class in Fall 2009 and Spring 2010. They also acknowledge Jenise Swall and Kristen Foley of USEPA for their statistical advice and input. Further, CSC is acknowledged for their support for preparation of the initial NO<sub>x</sub> combustion data base. In addition, Barrett Parker of USEPA is acknowledged for his support of the project and for use of the data from RTI for other pollutants and source categories.

## **REFERENCES**

1. U.S. Environmental Protection Agency. *Compilation of Air Pollutant Emission Factors*. Research Triangle Park, NC, 1995. Available at <http://www.epa.gov/ttn/chief/ap42> (accessed November 2009).
2. Gluck, S.; Glenn C., Logan T., Vu B., Walsh M., Williams P., Evaluation of NO<sub>x</sub> Flue Gas Analyzers for Accuracy and Their Applicability for Low-Concentration Measurements, *Journal of Air and Waste Management Association*, **2003**, 53,749–758.
3. U.S. Environmental Protection Agency, *Evaluation Report: EPA Can Improve Emissions Factors Development and Management*, 2006-P-00017, Office of Inspector General, U.S. Environmental Protection Agency, Washington, DC., March 22, 2006. Available at <http://www.epa.gov/oig/reports/2006/20060322-2006-P-00017.pdf> (accessed November 2011)
4. U.S. Environmental Protection Agency *Continuous Emissions Monitoring Systems Data*. Available at <http://camddataandmaps.epa.gov> (accessed November 2009).
5. U.S. Department of Energy *Form FERC-423 Database. Monthly Cost and Quality for Fuels and Electric Generating Units*. Available at <http://www.eia.doe.gov/cneaf/electricity/page/ferc423.html> (accessed November 2009).
6. U.S. Environmental Protection Agency *Emissions Inventories. Technology Transfer Network: Clearinghouse for Inventories & Emissions Factors*. Available at <http://www.epa.gov/ttn/chief/eiinformation.html> (accessed November 2009).
7. Louise Camalier, Brendan Yoshimoto and Brian Stines, “A Statistical Method to Corroborate VOC Emission Inventories Using Air Quality Data—Applied to Houston and Atlanta.” *Undergraduate Research Journal of North Carolina State University*, First Edition, 2004. Available at <http://www.ncsu.edu/undergrad-research/urj/> (accessed November 2009).
8. U.S. Environmental Protection Agency *Procedures for Preparing Emission Factor Documents*. 1997, p. 4-24. Available at <http://www.epa.gov/ttn/chief/efdocs/procedur.pdf> (accessed April 2010).
9. *Emissions Factor Uncertainty Assessment*. RTI International, 2007, p. 4-5 – 4-6. Available at <http://www.epa.gov/ttn/chief/efpac/uncertainty.html> (accessed April 2010)

About the authors:

George Pouliot (corresponding author) is a Physical Scientist in the Atmospheric Modeling and Analysis Division in the National Exposure Research Laboratory at the Environmental Protection Agency. He can be reached at [pouliot.george@epa.gov](mailto:pouliot.george@epa.gov)

Emily Wisner is member of the Statistics Department at North Carolina State University in Raleigh NC and was a summer intern at the EPA in 2010.

David Mobley is Deputy Director of the Atmospheric Modeling and Analysis Division in the National Exposure Research Laboratory at the Environmental Protection Agency

William Hunt is a Professor of Statistics at North Carolina State University

**Table 1.** Definition of ratings used for Test data quality Rating:

Test Data Quality Rating Letter	Description
A	Tests are performed by a sound methodology and are reported in enough detail for adequate validation.
B	Tests are performed by a generally sound methodology, but lacking enough detail for adequate validation.
C	Tests are based on an unproven or new methodology, or are lacking a significant amount of background information
D	Tests are based on a generally unacceptable method, but the method may provide an order-of-magnitude value for the source.

**Table 2:** Definition of AP-42 Emission Factor Ratings

Letter Rating	Description
A — Excellent	Factor is developed from A- and B-rated source test data taken from many randomly chosen facilities in the industry population. The source category population is sufficiently specific to minimize variability.
B — Above average	Factor is developed from A- or B-rated test data from a "reasonable number" of facilities. Although no specific bias is evident, it is not clear if the facilities tested represent a random sample of the industry. As with an A rating, the source category population is sufficiently specific to minimize variability.
C — Average	Factor is developed from A-, B-, and/or C-rated test data from a reasonable number of facilities. Although no specific bias is evident, it is not clear if the facilities tested represent a random sample of the industry. As with the A rating, the source category population is sufficiently specific to minimize variability.
D — Below average	Factor is developed from A-, B- and/or C-rated test data from a small number of facilities, and there may be reason to suspect that these facilities do not represent a random sample of the industry. There also may be evidence of variability within the source population.
E — Poor	Factor is developed from C- and D-rated test data, and there may be reason to suspect that the facilities tested do not represent a random sample of the industry. There also may be evidence of variability within the source category population.

**Table 3.** Summary statistics of the 21 different SCCs

SCC	Fuel Type	Mechanism	AP-42 Grade	AP-42 EF (Lb/Tons)	CEMS Mean (Lb/Tons)	% Diff.
10100201	Bituminous/ Subbituminous Coal	Pulverized Coal: Wet Bottom (Bituminous Coal)	D	31	15.43	101.0%
10100202	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom (Bituminous Coal)	A	12	11.46	4.7%
10100203	Bituminous/ Subbituminous Coal	Cyclone Furnace (Bituminous Coal)	A	33	22.28	48.1%
10100204	Bituminous/ Subbituminous Coal	Spreader Stoker	C	11	8.68	26.8%
10100212	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom, Tangential (Bituminous Coal)	A	10	10.27	-2.7%
10100221	Bituminous/ Subbituminous Coal	Pulverized Coal: Wet Bottom (Subbituminous Coal)	E	24	7.49	220.6%
10100222	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom (Bituminous Coal)	A	7.4	7.30	1.4%
10100223	Bituminous/ Subbituminous Coal	Cyclone Furnace (Subbituminous Coal)	C	17	13.13	29.5%
10100226	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom, Tangential (Subbituminous Coal)	A	7.2	5.66	27.2%
10100301	Lignite	Pulverized Coal: Dry Bottom, Wall Fired	C	6.3	4.57	37.8%
10100302	Lignite	Pulverized Coal: Dry Bottom, Tangential Fired	A	7.1	4.46	59.0%
10100303	Lignite	Cyclone Furnace	C	15	9.46	58.6%
10100401	Residual Oil	Grade 6 Oil: Normal Firing	A	1.97	1.51	30.1%
10100404	Residual Oil	Grade 6 Oil: Tangential Firing	A	1.34	1.79	-24.9%
10100501	Distillate Oil	Grades 1 and 2 Oil	D	1.01	2.02	-50.0%
10100601	Natural Gas	Electric Generation, Boilers > 100 Million Btu/hr except Tangential	A	0.19	0.20	-6.9%
10100602	Natural Gas	Boilers < 100 Million Btu/hr except Tangential	B	0.1	0.30	-66.2%



10100604	Natural Gas	Tangentially Fired Units	A	0.17	0.14	18.9%
10200601	Natural Gas	Industrial, Boilers > 100 Million Btu/hr	A	0.19	0.17	10.9%
20100201	Natural Gas	Electric Generation, Turbine	A	0.32	0.08	286.8%
20200201	Natural Gas	Industrial, Turbine	A	0.32	0.32	0.5%

*Note:* EF = emission factor; AP-42 Grade = the letter grade found in the AP-42; *AP-42 EF* = emission factor as found in WebFIRE; %Diff. = AP-42 emission factor minus the mean emission factor divided by the AP-42 emission factor.

**Table 4.** Letter grades and assumed associated sample sizes.

<b>Letter Grade</b>	<b>Sample Size (<i>n</i>)</b>
A	25
B	10
C	5
D	3
E	1

**Table 5.** SCCs and their corresponding uncertainties for A through E letter grades

SCC	Fuel_Type	Mechanism	A (n=25)	B (n=10)	C (n=5)	D (n=3)	E (n=1)
10100201	Bituminous/ Subbituminous Coal	Pulverized Coal: Wet Bottom (Bituminous Coal)	30%	51%	64%	72%	84%
10100202	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom (Bituminous Coal)	21%	43%	57%	66%	80%
10100203	Bituminous/ Subbituminous Coal	Cyclone Furnace (Bituminous Coal)	16%	37%	53%	63%	78%
10100204	Bituminous/ Subbituminous Coal	Spreader Stoker	25%	47%	61%	69%	82%
10100212	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom, Tangential (Bituminous Coal)	13%	34%	50%	60%	76%
10100221	Bituminous/ Subbituminous Coal	Pulverized Coal: Wet Bottom (Subbituminous Coal)	3%	18%	34%	46%	67%
10100222	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom (Bituminous Coal)	25%	46%	60%	69%	82%
10100223	Bituminous/ Subbituminous Coal	Cyclone Furnace (Subbituminous Coal)	12%	33%	49%	59%	76%
10100226	Bituminous/ Subbituminous Coal	Pulverized Coal: Dry Bottom, Tangential (Subbituminous Coal)	23%	44%	59%	68%	81%
10100301	Lignite	Pulverized Coal: Dry Bottom, Wall Fired	24%	45%	59%	68%	81%
10100302	Lignite	Pulverized Coal: Dry Bottom, Tangential Fired	5%	22%	39%	50%	70%
10100303	Lignite	Cyclone Furnace	8%	27%	43%	54%	73%
10100401	Residual Oil	Grade 6 Oil: Normal Firing	19%	40%	55%	65%	79%
10100404	Residual Oil	Grade 6 Oil: Tangential Firing	17%	38%	54%	63%	78%
10100501	Distillate Oil	Grades 1 and 2 Oil	30%	51%	64%	72%	84%
10100601	Natural Gas	Electric Generation, Boilers > 100 Million Btu/hr except Tangential	42%	61%	72%	78%	87%
10100602	Natural Gas	Boilers < 100 Million Btu/hr except Tangential	28%	49%	63%	71%	83%
10100604	Natural Gas	Tangentially Fired Units	26%	47%	61%	69%	82%

10200601	Natural Gas	Industrial, Boilers > 100 Million Btu/hr	39%	59%	70%	77%	86%
20100201	Natural Gas	Electric Generation, Turbine	69%	80%	86%	89%	94%
20200201	Natural Gas	Industrial, Turbine	37%	57%	69%	76%	86%

**Table 6.** Emission factor data quality rating uncertainties for the different pollutant and source category combinations.

<b>Source Category/Pollutant</b>	<b>A (n=25)</b>	<b>B (n=10)</b>	<b>C (n=5)</b>	<b>D (n=3)</b>	<b>E (n=1)</b>
Hot Mix Batch Filterable PM	64%	77%	83%	87%	93%
Hot Mix Batch Inorganic Cond	63%	76%	83%	87%	92%
Hot Mix Batch Organic Cond	62%	75%	82%	86%	92%
Hot Mix Benzene	49%	66%	76%	81%	89%
Hot Mix Drum Filterable PM	47%	65%	75%	80%	88%
Hot Mix Drum Inorganic Cond	56%	71%	79%	84%	91%
Hot Mix Drum Organic Cond PM	66%	78%	85%	88%	93%
Hot Mix Formaldehyde	59%	73%	81%	85%	91%
OSB-HP PM Filterable	55%	70%	79%	83%	90%
Refuse Arsenic	48%	66%	75%	81%	89%
Refuse Cadmium	58%	72%	80%	85%	91%
Refuse CO	45%	63%	74%	79%	88%
Refuse HCL FF	62%	75%	82%	86%	92%
Refuse HCL Uncontrolled	25%	47%	61%	69%	82%
Refuse Lead	51%	68%	77%	82%	90%
Refuse Mercury	58%	73%	81%	85%	91%
Refuse Nickel	44%	63%	73%	79%	88%
Refuse NOx Uncontrolled	1%	11%	26%	38%	61%
Refuse PM DS FF	38%	58%	69%	76%	86%
Refuse PM ESP	58%	72%	80%	85%	91%
Refuse PM ESP RDF	68%	79%	85%	89%	93%
Refuse PM RDF Uncontrolled	18%	40%	55%	64%	79%
Refuse PM SD ESP	30%	51%	64%	72%	83%
Refuse PM SD FF	55%	71%	79%	84%	91%
Refuse PM Uncontrolled	28%	49%	63%	71%	83%
Refuse SO2	32%	53%	66%	73%	84%
Wood Comb Acetaldehyde	67%	79%	85%	88%	93%
Wood Comb Arsenic	78%	86%	90%	92%	96%
Wood Comb Benzene	71%	82%	87%	90%	94%
Wood Comb Cadmium	67%	79%	85%	88%	93%
Wood Comb Chromium	70%	81%	86%	89%	94%
Wood Comb CO	51%	68%	77%	82%	90%
Wood Comb Formaldehyde	71%	81%	87%	90%	94%
Wood Comb Lead	74%	84%	88%	91%	95%
Wood Comb Mercury	69%	80%	86%	89%	94%
Wood Comb Nickel	74%	83%	88%	91%	95%
Wood Comb Nox	15%	36%	52%	62%	77%
Wood Comb PM Cond	53%	69%	78%	83%	90%
Wood Comb PM Filter Drywood	27%	49%	62%	70%	83%
Wood Comb PM Filter MC Drywood	24%	45%	60%	68%	81%

Wood Comb PM Filter MC					
Wetwood	59%	74%	81%	85%	92%
Wood Comb PM Filter Wetwood	36%	57%	68%	75%	86%
Wood Comb PM Filter WS					
Wetwood	11%	32%	48%	58%	75%
Wood Comb SO2	67%	79%	85%	88%	93%

**Table 7.** Uncertainty ranges (the interquartile range) for emission factor data quality indicators.

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>Uncertainty</b>	<b>Uncertainty</b>	<b>Uncertainty</b>	<b>Uncertainty</b>	<b>Uncertainty</b>
<b>(n=25)</b>	<b>(n=10)</b>	<b>(n=5)</b>	<b>(n=3)</b>	<b>(n=1)</b>
25 – 62%	46 – 75%	60 – 82%	69 – 86%	82 – 92%

## FIGURE CAPTIONS

1. SCC 10100203 (bituminous coal, cyclone furnace) boxplots by plants before 2% trim. The NO<sub>x</sub> emission factor is in units of lbs/ton.
2. SCC 10100203 (bituminous coal, cyclone furnace) boxplots by plants after 2% trim. The NO<sub>x</sub> emission factor is in units of lbs/ton.
3. Time plot of NO<sub>x</sub> emission factors for SCC 10100203, plant 963 (bituminous coal, cyclone furnace) with control period included in lbs/ton.
4. Time plot of NO<sub>x</sub> emission factors for SCC 10100203, plant 963 (bituminous coal, cyclone furnace) with control period not included in lbs/ton.
5. Time plot of NO<sub>x</sub> emission factors for SCC 10100203, plant 2364 (bituminous coal, cyclone furnace) with control period included in lbs/ton. Data was removed from this plant beginning on 04/01/1998
6. Distribution of NO<sub>x</sub> emission factors for SCC 10100203 (bituminous coal, cyclone furnace) with the mean emission factor and AP-42 emission factor labeled. Units are in lbs/ton.
7. Hot mix batch filterable PM—example of a log-normal distribution. The Emission factor units are in lbs/ton.