United States Environmental Protection Agency

Emissions Uncertainty: NOx Emissions from Electric Generating Units

Emily Wisner and Bill Hunt
Statistics Department,
North Carolina State University, Raleigh, NC

David Mobley and George Pouliot
Atmospheric Modeling and Analysis Division,
National Exposure Research Laboratory,
Environmental Protection Agency,
Research Triangle Park, NC

19th International Emission Inventory Conferenc
San Antonio, TX September 30, 2010

Outline

- Introduction
- Phase I—Compare CEMS data and AP-42 Emission Factors
  - Approach and Results
- Phase II—Estimate NOx/Combustion emission factor uncertainties
  - Approach and Results
- Phase III—Estimate emission factor uncertainties for all pollutant and source categories
  - Approach and Results
- Conclusions and Recommendations

Introduction

- Emission factors (representative values that attempts to relate the quantity of a pollutant released to the atmosphere with an activity associated with the release of that pollutant) are important for estimating and characterizing emission sources of air pollution.

- EPA has estimates of emission factors for different Source Classification Codes (SCC) in AP-42(EPA's official documentation and dissemination of emission factors), which are usually an average of samples

- A limited number of tests lead to uncertainty in estimating the emission factors. Cost of testing limits the number of tests that can be performed

Slide 3

JHS5 Want to add point about why? Each test can cost 10's, even 100'a of thousands of dollars and the budget and priorities have not come forth to allow testing on a large scale for parametrically planned tests under strict control of the tester.

JimS, 7/22/2010

Introduction
Problem

- Currently, no analyses of the uncertainty of emission factors from Continuous Emission Monitors (CEMS) are available.

- The scientific community is unable to quantify the accuracy of emissions estimates across combustion sources.

Slide 4

JHS6 I believe I recall reading that CEMS data are within 5%, though they are not emission factors. If you have CEMS, why would you want to use emission factors?
JimS, 7/22/2010

JHS7 Never has, never will!
JimS, 7/22/2010

United States
Environmental Protection
Agency

Introduction
Objectives

1. Compare the NOx emission factors from combustion sources to currently available continuous emission monitoring data. JHS8
2. Develop quantitative uncertainty indicators for A through E data quality rated emission factors based on NOx emissions from combustion sources.
3. Determine the feasibility of applying these quantitative uncertainty indicators to other pollutants and source types.
4. *Ultimate Goal: Develop a modeling emission inventory with some estimate of the range of uncertainty that can be carried through the modeling process* JHS9

Slide 5

JHS8 Beyond the academics, why?
JimS, 7/22/2010

JHS9 I believe the CEMS were only, or primarily, on EGU's and there were well over 100 stack tests, which is highly unusual in database terms for the remainder of the source categories which may have as few as a couple of tests; totally different set of conditions.
JimS, 7/22/2010

Introduction
Database

- The CEMS data set is a measure by hour of the nitrogen oxide (NOx) emissions from combustion sources.

- The data are split into three major fuel types for combustion; coal, oil, and natural gas.

Slide 6

JHS10 Data "ARE" always plural. Ancient but true.
JimS, 7/22/2010

Introduction
Database Continued

The database was a combination of information from:

- The CEMS data reported to the EPA for the Acid Rain Cap and Trade program
- The fuel use and fuel quality data reported to DOE for electric generating units
- The boiler/fuel characteristics reported to the States and EPA for the National Emission Inventory (2005 NEI used in this analysis)
- Emission factors from AP-42 via WebFIRE (EPA's online database that catalogs descriptions of SCCs and their emission factors)

The first three databases were joined into one to create hourly NOx emission factors from ~2000 sources across the United States from the years 1997 through 2007 for 52 different SCCs and then compared to EPA's reported Emission Factors

Phase I
Approach

Data quality issues:

- After careful inspection of the data, it appeared there were issues with the quality as a result of some CEMS having multiple SCCs (e.g., multiple fuels).
- WJHS11his was the case, the most dominant SCC was kept and the others were thrown out of the analysis.
- Deciding which SCC was dominant:
  - If the SCC had the most hours of operation.
  - If it had an order of magnitude greater in emissions than any other SCC for a particular boiler.
- As a result, a total of 13 SCCs were entirely removed from the database.

Slide 8

JHS11 This denies reality. The better option would have been to treat the separate data set as another "shadow" source and made it independent. The reality really happened!

JimS, 7/22/2010

Phase I
Approach Continued

Extraneous values:

- Boxplots of all the SCCs revealed that some individual plants in each SCC clearly had very extraneous emission factor values. JHS12
- To resolve this problem and to reduce variability in the data, the dataset was trimmed. JHS13
- The highest 2% of emission factor values and the lowest 2% of emission factor values were removed from each SCC.

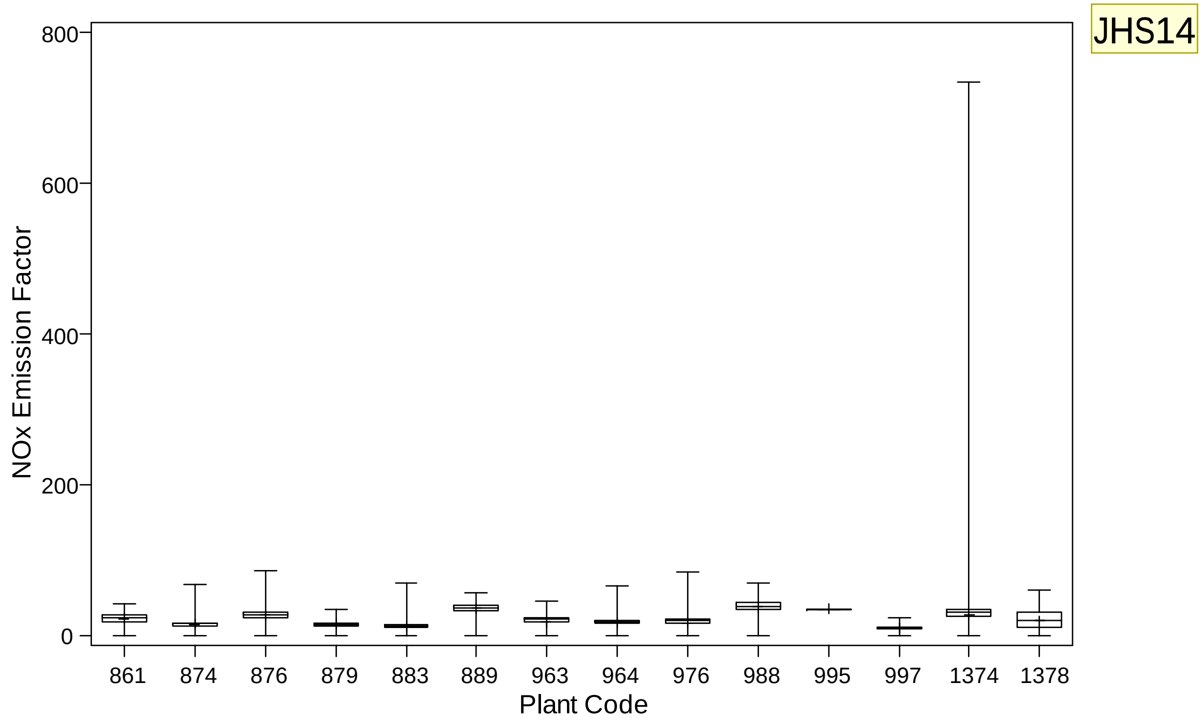
Slide 9

JHS12 Were any of these from the group that was rejected for having multiple fuels/SCC's?
JimS, 7/22/2010

JHS13 Is this a case of not liking the results, so you cange the data to agree?
JimS, 7/22/2010

Phase I
Approach Continued

SCC 10100203 (coal, cyclone furnace) boxplots by plants before 2% trim

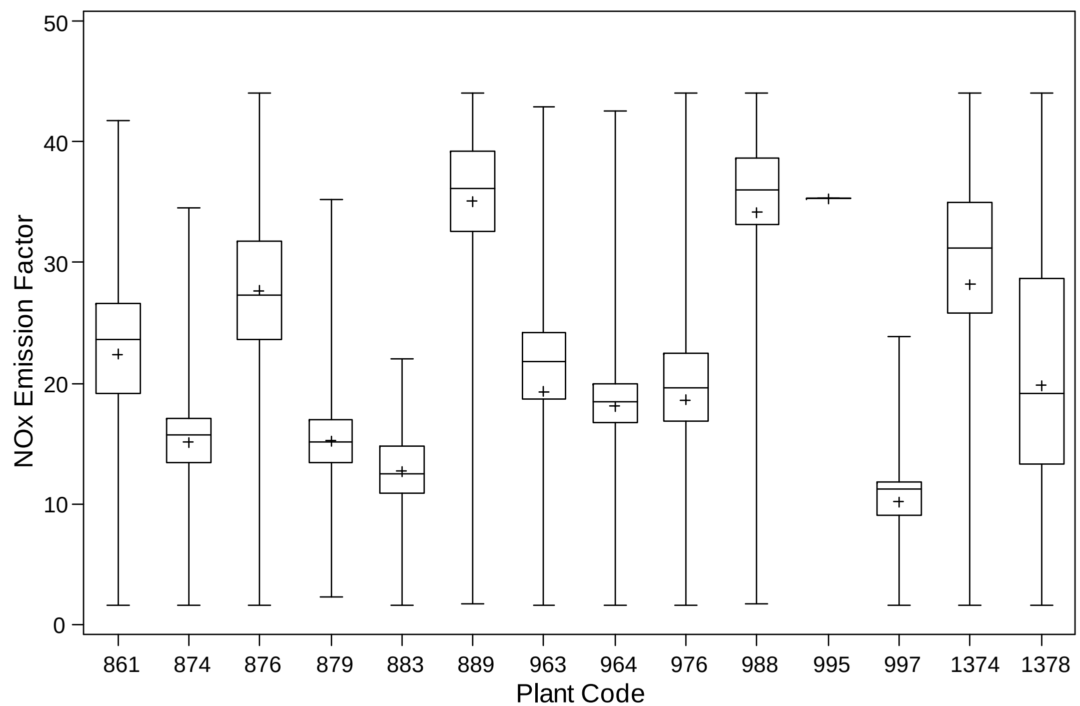


Slide 10

JHS14 If the long whisker is really happening, it can overwhelm emissions from the better performing sources in a very short time.
JimS, 7/22/2010

Phase I
Approach Continued

SCC 10100203 (coal, cyclone furnace) boxplots by plants after 2% trim



United States
Environmental Protection
Agency

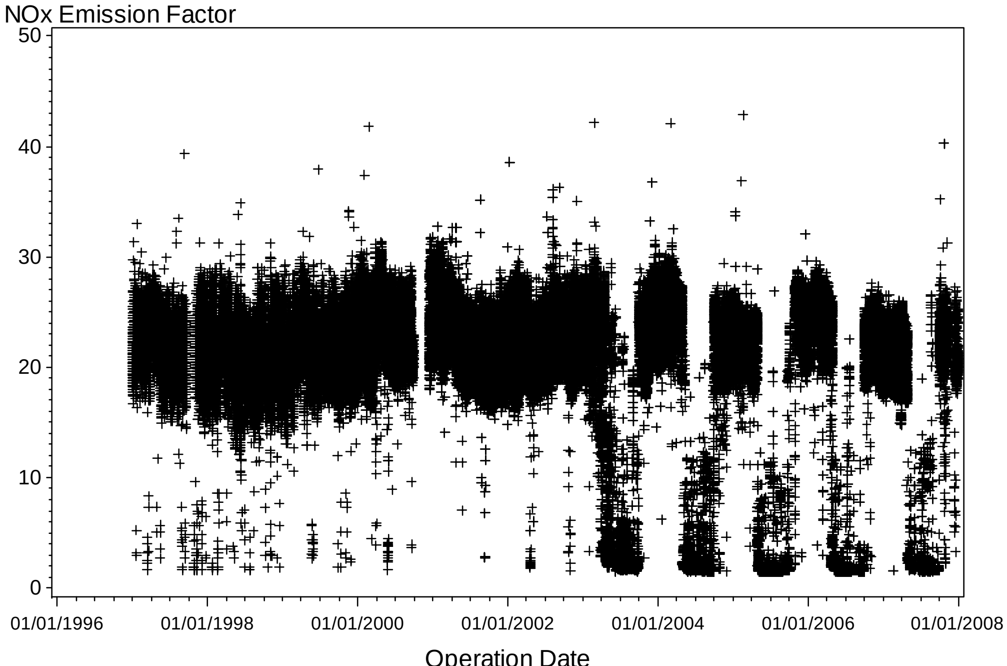
Phase I
Approach Continued

- Between the years 2002 and 2007, some plants phased in controls between May 1st and September 30th through the State Implementation Plan.

- For this reason, data from these plants were removed from the analysis, since only uncontrolled (as combusted) emission factors were of interest.

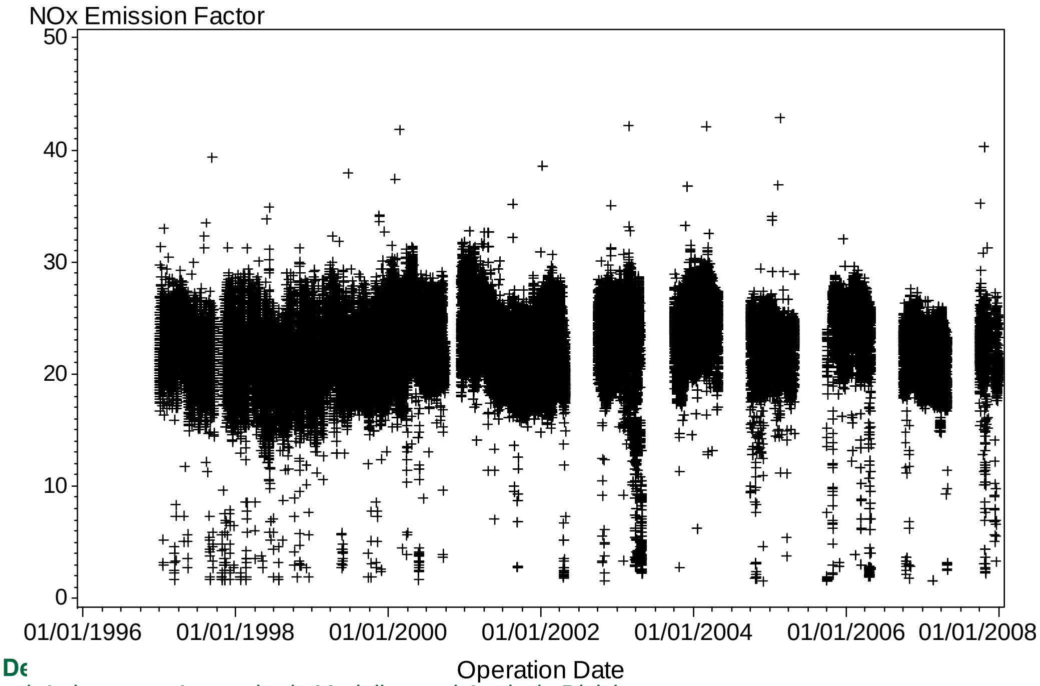
Phase I
Approach Continued

Time plot of NOx emission factors for SCC 10100203, plant 963 (coal, cyclone furnace) with control period included



Phase I
Approach Continued

Time plot of NOx emission factors for SCC 10100203, plant 963 (coal, cyclone furnace) with control period not included



Phase I
Approach Continued

- 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.

  – This excluded units with post combustion NOx controls such as selective catalytic reduction of NOx with ammonia.

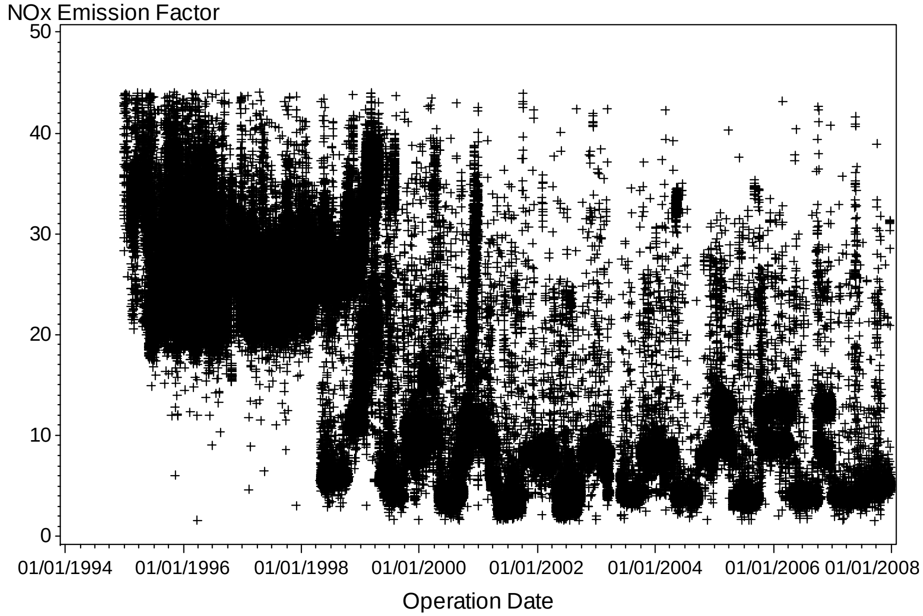
Slide 15

JHS15 It should be easy to verify which had controls.
JimS, 7/22/2010

JHS16 However, I contend that modified combustion units should be treated as a separate SCC, not as a control device.
JimS, 7/22/2010

Phase I
Approach Continued

Time plot of NOx emission factors for SCC 10100203, plant 2364 (coal, cyclone furnace) with control period included



For this plant, data was removed beginning on 04/01/1998

Phase I
Approach Continued

- 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.

  – The analysis began with 52 SCCs.

Phase I
Approach Continued

- After the data were properly formatted and appropriate dates were removed from each plant, the SAS analysis program was used to compute the mean emission factor for each SCC. JHS17

- 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.

Slide 18

JHS17 Does the SAS program have a name?
JimS, 7/22/2010

Phase I
Results

JHS18

Summary statistics for coal SCCs

SCC	Fuel Type	Mechanism	AP-42	Mean	% Diff.
10100201	Coal	Pulvarized Coal: Wet Bottom	31	15.4	**101.0%**
10100202	Coal	Pulvarized Coal: Wet Bottom	12	11.5	**4.7%**
10100203	Coal	Cyclone Furnace	33	22.3	**48.1%**
10100204	Coal	Spreader Stoker	11	8.7	**26.8%**
10100212	Coal	Pulverized Coal: Dry Bottom, Tangential	10	10.3	**-2.7%**
10100221	Coal	Pulverized Coal: Wet Bottom	24	7.5	**220.6%**
10100222	Coal	Pulverized Coal: Dry Bottom	7.4	7.3	**1.4%**
10100223	Coal	Cyclone Furnace	17	13.1	**29.5%**
10100226	Coal	Pulverized Coal: Dry Bottom, Tangential	7.2	5.7	**27.2%**
10100301	Lignite	Pulverized Coal: Dry Bottom, Wall Fired	6.3	4.6	**37.8%**
10100302	Lignite	Pulverized Coal: Dry Bottom, Tangential Fired	7.1	4.5	**59.0%**
10100303	Lignite	Cyclone Furnace	15	9.5	**58.6%**

Slide 19

JHS18 More importantly, what would be the difference in using the AP-42 factors and the new derived factors as applied to each source and summed up for a sum of the National emissions estimates? This would factor back in some of the realities omitted by trimming the data to those that you like.

JimS, 7/22/2010

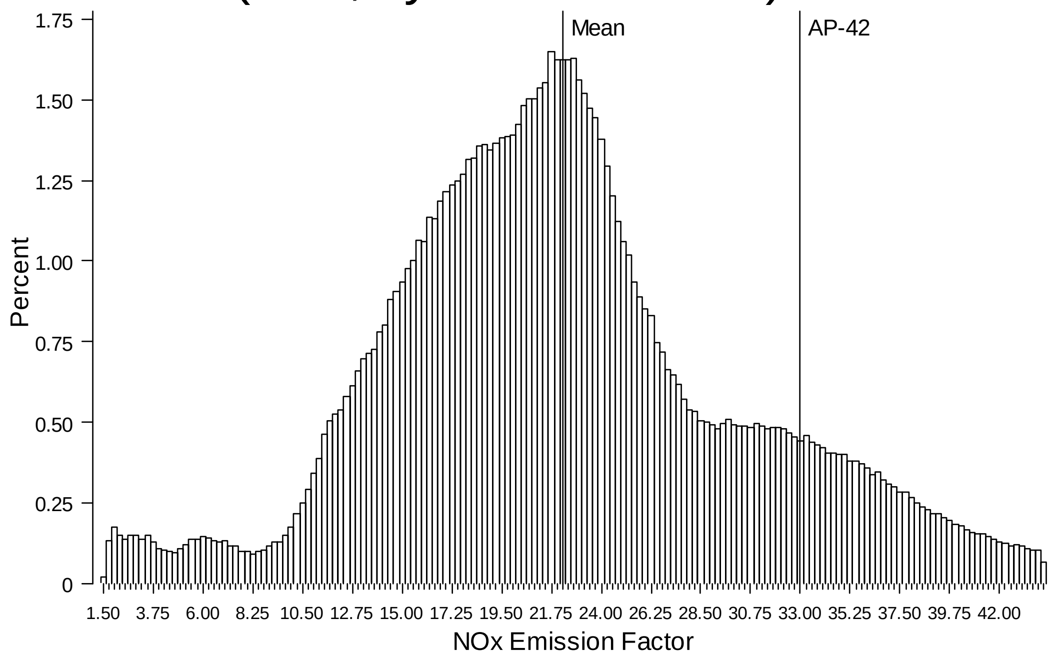
Phase I
Results Continued

Summary statistics for oil and natural gas SCCs

SCC	Fuel Type	Mechanism	AP-42	Mean	% Diff.
10100401	Residual Oil	Grade 6 Oil: Normal Firing	1.97	1.5	**30.1%**
10100404	Residual Oil	Grade 6 Oil: Tangential Firing	1.34	1.8	**-24.9%**
10100501	Distillate Oil	Grades 1 and 2 Oil	1.01	2.0	**-50.0%**
10100601	Natural Gas	Boilers > 100 Million Btu/hr except Tangential	0.19	0.2	**-6.9%**
10100602	Natural Gas	Boilers < 100 Million Btu/hr except Tangential	0.1	0.3	**-66.2%**
10100604	Natural Gas	Tangentially Fired Units	0.17	0.1	**18.9%**
10200601	Natural Gas	Boilers > 100 Million Btu/hr	0.19	0.2	**10.9%**
20100201	Natural Gas	Turbine	0.32	0.1	**286.8%**
20200201	Natural Gas	Turbine	0.32	0.3	**0.5%**

Phase I
Results Continued

Distribution of NOx emission factors for SCC 10100203
(coal, cyclone furnace)



Phase I
Results Continued

- Based on the analysis of Phase I, many of the AP-42 emission factor values for the 21 SCCs in this study likely need to be updated to reflect the currently available continuous NOx emissions data. JHS19
  - 62% of SCCs had % difference between AP-42 and mean of continuous emissions data > ±25%.
  - 29% of SCCs had % difference between AP-42 and mean of continuous emissions data > ±50%.
  - 14% of SCCs had % difference between AP-42 and mean of continuous emissions data > ±100%.

Slide 22

JHS19 From memory, but should be in the background documents; I believe Roy Huntley and his contractor took the existing CEMS data into account and used them in the updates done for the 5th edition.
But again, if you have the CEM results, why would you want to use emission factors?
JimS, 7/22/2010

United States Environmental Protection Agency

Phase II
Approach

- Goal: develop a quantitative measure of uncertainty for each of the EPA's qualitative data quality ratings currently being used. JHS20

- 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.

- This research assumed the following sample sizes were associated with each of the letter grades:

Letter Grade	Sample Size (*n*)
A	25
B	10
C	5
D	3
E	1

- Many other considerations contribute to a data quality rating, but this analysis assumed sample size was the key attribute.

JHS20 Very little connection. This work does not really answer that question.
JimS, 7/22/2010

Phase II
Approach Continued

- 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 emissions mean.

Phase II

Approach Continued

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

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

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

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

- $\bar{x}$ = sample mean
- μ = population mean
- σ = population standard deviation
- Z = standard z-score (the number of standard deviations an observation is away from the actual mean)
- n = sample size

United States Environmental Protection Agency

Phase II
Approach Continued

- To check the theoretical calculations, bootstrap methods through SAS programming wJHS21used for some of the very non-normal SCCs.
- 10,000 samples for each of the 5 sample sizes were simulated for the selected SCCs and the means were calculated. JHS22
- Uncertainty = 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).

Slide 26

JHS21 Again, the name of the SAS routine should be identified and explained.
JimS, 7/22/2010

JHS22 What made the 10,000 samples differ? In other words, why were they not all identical values? This seems to be a key unobserved variable that is key to making a better estimate of emissions - if you are going to reduce this to an emission factor question.
JimS, 7/22/2010

United States Environmental Protection Agency

Phase II

Approach Continued

- Creating a standardized way of ranking an SCC by letter grade was also of interest.

- Before establishing a standardized ranking variable, histograms of each SCC were observed and were subjectively ranked by letter grade.

- The standardized ranking variable chosen was the coefficient of variation (CV).

- The CV is a normalized measure of dispersion of a probability distribution compared to the mean.

$$CV = \frac{\sigma}{\mu}$$

Phase II
Approach Continued

- The CV statistic is free of any scale, and therefore is typically useful for comparing different sets of data.

- In a previous research study on emissions uncertainty by RTI International, CV values were found to correlate very well with the uncertainty ratios they calculated.

- CV intervals were determined to correspond to either an A data quality rating, or B, C, or D based on what the CV values were for histograms that were normal looking versus histograms that were very non-normal.

Phase II
Results

Coal SCCs and their corresponding uncertainties for A through E letter grades

SCC	Fuel Type	Mechanism	A (n=2)	B (n=10)	C (n=5)	D (n=3)	E (n=1)
10100201	Coal	Pulvarized Coal: Wet Bottom	30%	51%	64%	72%	84%
10100202	Coal	Pulvarized Coal: Wet Bottom	21%	43%	57%	66%	80%
10100203	Coal	Cyclone Furnace	16%	37%	53%	63%	78%
10100204	Coal	Spreader Stoker	25%	47%	61%	69%	82%
10100212	Coal	Pulverized Coal: Dry Bottom, Tangential	13%	34%	50%	60%	76%
10100221	Coal	Pulverized Coal: Wet Bottom	3%	18%	34%	46%	67%
10100222	Coal	Pulverized Coal: Dry Bottom	25%	46%	60%	69%	82%
10100223	Coal	Cyclone Furnace	12%	33%	49%	59%	76%
10100226	Coal	Pulverized Coal: Dry Bottom, Tangential	23%	44%	59%	68%	81%
10100301	Lignite	Pulverized Coal: Dry Bottom, Wall Fired	23%	45%	59%	68%	81%
10100302	Lignite	Pulverized Coal: Dry Bottom, Tangential	5%	22%	39%	50%	70%
10100303	Lignite	Cyclone Furnace	8%	27%	43%	54%	73%

Phase II
Results Continued

Oil and natural gas 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)
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	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	25%	47%	61%	69%	82%
10200601	Natural Gas	Boilers > 100 Million Btu/hr	39%	59%	70%	77%	86%
20100201	Natural Gas	Turbine	69%	80%	86%	89%	94%
20200201	Natural Gas	Turbine	37%	57%	69%	76%	86%

Phase II
Results Continued JHS23

Averaged SCC uncertainty values for NOx emissions from combustion sources

	A Uncertainty (n=25)	B Uncertainty (n=10)	C Uncertainty (n=5)	D Uncertainty (n=3)	E Uncertainty (n=1)
Average:	**25%**	**45%**	**60%**	**65%**	**80%**

Slide 31

JHS23 As a general rule of thumb, it has been said that one should have about 30 or more good tests before one can make a good estimate of the population. This seems to confirm that rule of thumb, but it does not provide the 30+ data points that are needed.
JimS, 7/22/2010

Phase II

Results Continued

- UJHS24tainty 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.

- This means, for example, that if an SCC received an A data quality 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 is not within 10% of the true emission factor mean.

Slide 32

JHS24 Seems to me that you have proven that it is absolutely certain that you will never be within 10% of the mean, unless you have huge numbers of test samples.
JimS, 7/22/2010

Phase II
Results Continued

Letter grades and possible corresponding CV intervals

	A	B	C	D
CV Ranking	**0 ≤ CV < 0.4**	**0.4 ≤ CV < 0.42**	**0.42 ≤ CV < 0.44**	**0.44 ≤ CV**

Phase II
Results Continued

Subjective, AP-42, and CV letter grades for coal SCCs

SCC	Fuel Type	Mechanism	AP-42 Grade	Subj. Grad	CV Grade	CV
10100201	Coal	Pulvarized Coal: Wet Bottom	D	C	D	0.482
10100202	Coal	Pulvarized Coal: Wet Bottom	A	A	A	0.398
10100203	Coal	Cyclone Furnace	A	A	A	0.355
10100204	Coal	Spreader Stoker	C	D	C	0.439
10100212	Coal	Pulverized Coal: Dry Bottom, Tangential	A	A	A	0.333
10100221	Coal	Pulverized Coal: Wet Bottom	E	A	A	0.235
10100222	Coal	Pulverized Coal: Dry Bottom	A	B	C	0.432
10100223	Coal	Cyclone Furnace	C	B	A	0.322
10100226	Coal	Pulverized Coal: Dry Bottom, Tangential	A	B	B	0.413
10100301	Lignite	Pulverized Coal: Dry Bottom, Wall Fired	C	C	C	0.420
10100302	Lignite	Pulverized Coal: Dry Bottom, Tangential	A	A	A	0.259
10100303	Lignite	Cyclone Furnace	C	C	A	0.286

Phase II
Results Continued

Subjective, AP-42, and CV letter grades for oil and natural gas SCCs

SCC	Fuel Type	Mechanism	Subj. Grade	AP-42 Grade	CV Grade	CV
10100401	Residual Oil	Grade 6 Oil: Normal Firing	**A**	**A**	A	0.378
10100404	Residual Oil	Grade 6 Oil: Tangential Firing	**A**	**A**	A	0.363
10100501	Distillate Oil	Grades 1 and 2 Oil	**B**	**D**	D	0.480
10100601	Natural Gas	Boilers > 100 Million Btu/hr except Tangential	**A**	**A**	D	0.619
10100602	Natural Gas	Boilers < 100 Million Btu/hr except Tangential	**D**	**B**	D	0.461
10100604	Natural Gas	Tangentially Fired Units	**A**	**A**	C	0.437
10200601	Natural Gas	Boilers > 100 Million Btu/hr	**A**	**A**	D	0.580
20100201	Natural Gas	Turbine	**D**	**A**	D	1.236
20200201	Natural Gas	Turbine	**D**	**A**	D	0.562

Phase II
Results Continued

- CV grade vs. subjective letter grade
  - Grades were the same for 12 SCCS.
  - Grades were different for 9 SCCs.
  - Consistent for about 57% of the SCCs.

- CV grade vs. AP-42 grade
  - Grades were the same for 10 SCCs.
  - Grades were different for 11 SCCs.
  - Consistent for about 48% of the SCCs.

- 76% of the SCCs were either the same or within 1 grade for the subjective rating.

United States Environmental Protection Agency

Phase II
Results Continued

- The CV is affected by whether a distribution is skewed left or skewed right.
- If a distribution is left skewed, the bulk of the data and the mean will be around a higher value, leading to a smaller CV.
  - This is because the CV is equal to the standard deviation divided by the mean, so having a larger denominator leads to a smaller fraction.
- If a distribution is right skewed, the bulk of the data and the mean will be around a smaller value, leading to a larger CV.

Phase II

Results Continued

- Although the CV grades do not compare that well to the subjective and AP-42 data quality ratings, the CV values correlate very well with the calculated uncertainty values. JHS25
  - This is a result of the equation used for calculating the uncertainties being a function of the CV.
- SCCs with high CV values (and lower CV grades as a result) also have higher uncertainty values.
- This consistency makes sense since uncertainty was calculated as the probability of a sample mean not being within 10% of the true mean.
- If a distribution is right skewed and the true mean is small, 10% of the mean will also be small, making it more unlikely to obtain a sample with a mean within the small range of ±10% of the true mean.

Slide 38

JHS25 Of course.
JimS, 7/22/2010

Phase III
Approach

- To determine the limitations of applying the uncertainties associated with the different letter grades for NOx emissions to other pollutants, another data set consisting of various pollutants was analyzed.

- This new data set is from the previous study by RTI International and included emission factor data for 44 different pollutant and source category combinations.

Phase III
Approach Continued

- The uncertainty values for the five letter grades were calculated for each of these pollutant and source category combinations as done in 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 NOx 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.

United States Environmental Protection Agency

Phase III

Results JHS26

EF uncertainties for the different pollutant and source category combinations

Source Category/Pollutant	A (n=25)	B (n=10)	C (n=5)	D (n=3)	E (n=1)
Hot Mix Batch Filterable PM JHS27	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%

Slide 41

JHS26 I would think you would need n= 25-30 to really justify even a C rating.
JimS, 7/22/2010

JHS27 These source categories have almost, if not absolutely, no CEM data, so how can you extrapolate from the CEMs comparison into a totally unrelated regeime? Deterioriates to a pure numbers game with out connection to reality.
JimS, 7/22/2010

Phase III

Results Continued

EF uncertainties for the different pollutant and source category combinations cont.

Source Category/Pollutant	A (n=25)	B (n=10)	C (n=5)	D (n=3)	E (n=1)
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%

Phase III
Results Continued

EF uncertainties for the different pollutant and source category combinations cont.

Source Cateogy/Pollutant	A (n=25)	B (n=10)	C (n=5)	D (n=3)	E (n=1)
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%

Phase III
Results Continued

Averaged uncertainty values for the 44 pollutant and source category combinations

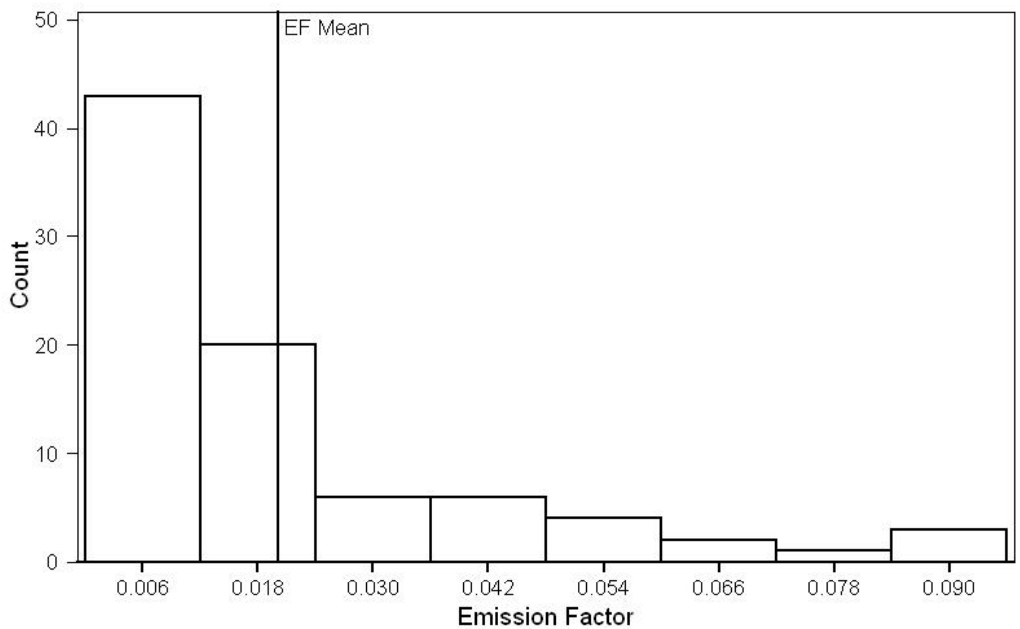
	A Uncertainty (n=25)	B Uncertainty (n=10)	C Uncertainty (n=5)	D Uncertainty (n=3)	E Uncertainty (n=1)
Average:	**50%**	**65%**	**75%**	**80%**	**90%**

Phase III
Results Continued

- 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 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.

Phase III
Results Continued

Hot mix batch filterable PM—Log-normally distributed



United States Environmental Protection Agency

ATMOSPHERIC MODELING AND ANALYSIS DIVISION • MOVING SCIENCE INTO APPLICATIONS •

Phase III
Results Continued

Uncertainty ranges from all pollutants and source categories

A Uncertainty (n=25)	B Uncertainty (n=10)	C Uncertainty (n=5)	D Uncertainty (n=3)	E Uncertainty (n=1)
25-50%	45-65%	60-75%	65-80%	80-90%

If these uncertainty ranges can be generalized to any pollutant…

- An A data quality rated sample of emission factors, assuming the sample size was 25, would have between 25% and 50% uncertainty associated with it.

- In other words, if a sample of size 25 data points for any pollutant is taken, the probability that the sample mean is not within 10% of the true mean would be between 25% and 50%.

Phase III
Results Continued

- CV values were also calculated for the 44 pollutant and source category combinations, and CV letter grades were assigned.
- Since most of these distributions were log-normal, the CV values were very large, resulting in most of the pollutant and source category combinations receiving Ds.
  - However, most actually received As in EPA’s AP-42.
- Thus, our work on using the CV is inconclusive at this time.

Summary and Conclusions

- **Phase I:** The inconsistency between the CEMS data and the AP-42 for most SCCs suggests the AP-42 needs to be updated to reflect the continuous emissions data now available.
  - 62% of SCCs had % difference between AP-42 and mean of continuous emissions data > ±25%.
  - 29% of SCCs had % difference between AP-42 and mean of continuous emissions data > ±50%.
  - 14% of SCCs had % difference between AP-42 and mean of continuous emissions data > ±100%.

United States Environmental Protection Agency

Summary and Conclusions

Continued

- **Phase II:** Uncertainty values for NOx emissions were calculated as the probability that a sample mean from a sample of size *n* (where *n* is the sample size associated with the different letter grades) will not be within 10% of the true emission factor mean.

	A Uncertainty (n=25)	B Uncertainty (n=10)	C Uncertainty (n=5)	D Uncertainty (n=3)	E Uncertainty (n=1)
Average:	25%	45%	60%	65%	80%

JHS28

United States Environmental Protection Agency

Summary and Conclusions
Continued

- **Phase III:** Uncertainty ranges were calculated for each letter grade using the NOx emissions uncertainty and a data set consisting of various pollutant and source category combinations that could possibly be applied to any pollutant.

A Uncertainty (n=25)	B Uncertainty (n=10)	C Uncertainty (n=5)	D Uncertainty (n=3)	E Uncertainty (n=1)
25-50%	45-65%	60-75%	65-80%	80-90%

Slide 51

JHS28 Demonstrates how mixing of statistics to explain unrelated associations is purely playing with numbers.
JimS, 7/22/2010

Areas for Future Research

- Apply these findings to air quality modeling simulations to access the uncertainty of those results.

- More research likely needs to be done on using the coefficient of variation as a metric for assigning data quality ratings.

Acknowledgements

The authors acknowledge Professor William Hunt Jr., Alissa Anderson, Colin Geisenhoffer, Brody Heffner, and Michael Shaw from North Carolina State University for their previous work on this project in ST 495, Environmental Data Analysis Class. They also acknowledge Jenise Swall and Kristen Foley of USEPA for their statistical advice and input. Further, CSC is acknowledged for their support in preparation of the initial NOx combustion data base. In addition, Barrett Parker of USEPA and Roy Neulicht of RTI are acknowledged for use of the data from other pollutants and source categories.