Presented at the 9th Annual CMAS Conference, Chapel Hill, NC, October 11-13, 2010

Emission Factor Uncertainty: Focusing on NOx Emissions from Electric Generating Units

Emily Wisner
EPA Summer Intern, 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 27711

1. INTRODUCTION

Emission factors are important for estimating and characterizing emission sources of air pollution. Emissions are being released into the air everyday from different sources and are monitored in various ways. Emission factors are generally estimated from an average of all available data. 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. These estimates were calculated by taking emission data from source categories and using these to make inferences about all other units with the same Source Classification Code (SCC). In many cases, the limited number of data points leads to increased uncertainty in the emission estimates. This uncertainty is catalogued by data quality indicators—letter grades A through E. The judgmental nature of these indicators does not allow users to make a quantitative assessment of uncertainty of emission inventories and air quality modeling applications. The objective of this study was to explore potential options to statistically and objectively quantify the uncertainty of emission factors. The focus of this study was on nitrogen oxide (NOx) emissions from electric generating units, which is probably the best tested and cate.

The EPA has compiled emission factors in a document entitled, Compilation of Air Pollutant Emission Factors, AP-42. These factors were basically averages from available source tests. In many cases, the available source test is from a very small sample set. The ideal situation would be to have numerous tests from a variety of sources. The minimal numbers of tests lead to uncertainty in the emissions estimates based on emission factors.

Currently, analyses of the uncertainty of emission factors from CEMS are not available. Characterizing the uncertainty of these and other emission sources would enable the scientific community to quantify the accuracy of emissions estimates across combustion and other sources. This study consisted of three main objectives: (1) compare the NOx emission factors from combustion sources with currently available continuous emission monitoring data; (2) develop quantitative uncertainty indicators for A through E rated data quality indicators for 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.

In order to analyze the variability of NOx emission factors from EGU sources, several databases of information needed to be combined. First, the CEMS monitoring data from the EPA's Clean Air Markets Division contains hourly NOx emission rates in lbs of NOx per million British thermal units (lbs/10^6Btu). Second, the DOE's Energy Information Administration has monthly fuel information for selected EGUs. This set of data includes the quantity for fuel consumed per month at a given plant and the heat content of the fuel (MMBtu/ton of fuel). Third, the National Emission Inventory contains plant information, including stack parameters and the Source Classification Codes. The years of data were 1997 to 2007. There were data for 52 different SCCs in the initial data.

2. METHODS

2.1 Phase I

To create a database where emission factors from different SCCs could be compared, all AP-42 values were standardized to lbs of NOx per million British thermal units (lbs/10^6Btu). 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.

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 an order of magnitude 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.

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

Between the years 2002 and 2007, some plants phased in controls between May 1st and September 30th through the various control strategies, as shown in figure 1. For this reason, data from these plants were removed from the analysis since only uncontrolled, or as combusted, emission factors were of interest, as shown in figure 2. 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.

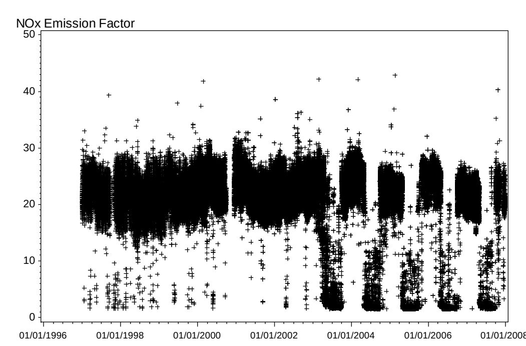

Figure 1. Time plot of NOx emission factors for SCC 10100203, plant 963.

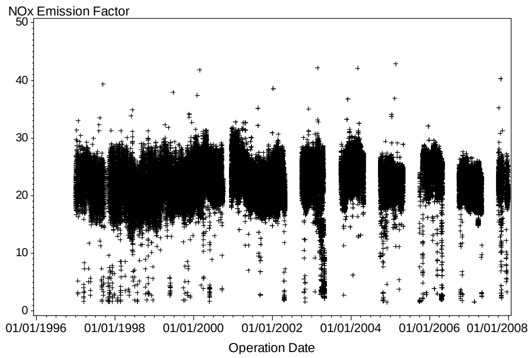

Figure 2. Time plot of NOx emission factors for SCC 10100203, plant 963.

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.

2.2 Phase II

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 2 shows the assumed sample sizes associated with each of the data quality ratings. (Note that many other considerations contribute to a data quality rating, but this analysis assumed sample size was the key attribute.)

Table 1. Letter grades and assumed associated sample sizes.

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

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:

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

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

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

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

where $\bar{x}$ refers to the sample mean; μ is the population mean; σ is the population standard deviation; and Z is the standard z-score. 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 eq 4 for each SCC.

2.3 Phase III

To determine the possibility 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 study by RTI International and included emission factor data for 44 different pollutant and source category combinations. 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.

3. RESULTS AND DISCUSSION

3.1 Phase I

After looking at the percent difference between the AP-42 emission factor and the mean NOx 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. 62% of SCCs had a percent difference between EPA's AP-42 emission factor and mean of continuous emissions data greater than ±25%. 29% of SCCs had a percent difference between EPA's AP-42 emission factor and mean of continuous emissions data greater than ±50%. 14% of SCCs had a percent difference between EPA's AP-42 emission factor and mean of continuous emissions data greater ±100%. Based on the analysis of Phase I, most 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.

3.2 Phase II

Table 2. Average uncertainties for A through E letter grades.

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

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 2 shows the rounded average uncertainties for the five letter grades of all 21 SCCs. 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.

3.3 Phase III

The second data set of 44 different pollutant and source category combinations 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. The uncertainties for each pollutant and source category combination were averaged and then rounded. These rounded averages were then combined with the NOx emission factor uncertainties to create uncertainty ranges for each letter grade that could possibly be applied to any pollutant, as shown in table 3. According to these calculated uncertainty ranges, an A 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 emission factors for any pollutant is taken, the probability that the sample mean is not within 10% of the true mean is between 25% and 50%.

Table 3. Uncertainty ranges for emission factor data quality indicators.

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

4. CONCLUSIONS

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. This means that substantial targeted and prioritized parametric source testing needs to be done on many source categories to provide a reliable database to develop new and better quality emission factors. Even though the AP-42 emission factor values did not match well with the CEMS data, the letter grades for each SCC found in the AP-42 were generally appropriate for the distribution shapes of the CEMS data and matched fairly well with subjective letter grades. 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* (true emission factor mean. The letter grade uncertainties were then averaged across SCC to calculate overall letter grade uncertainties for NOx emissions. Using the CV was then where *n* is the sample size associated with the different letter grades) will not be within 10% of the proposed as a way to possibly rank a sample of emission factors as either A, B, C, or D. For the majority of SCCs, the CV letter grades matched reasonably well with the AP-42 letter grades and the subjective letter grades. To determine the possibility of applying the letter grade uncertainties computed for NOx emissions to other pollutants, another data set with various combinations of pollutants and firing methods was analyzed. Uncertainties for each letter grade were calculated for the new data set and compared to those calculated from the continuous NOx emissions data. Uncertainty ranges were then computed based on the NOx emissions uncertainties and the uncertainty values from the second data set. These uncertainty ranges could possibly be applied to many different types of pollutants and source categories.

5. RECOMMENDATIONS

Being able to apply these findings to air quality modeling simulations to access the uncertainty of those results would be ideal. Also, more research likely needs to be done on using the coefficient of variation as a metric for assigning data quality ratings.

6. 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 for preparation of the initial NOx combustion data base. In addition, Barrett Parker of USEPA is acknowledged for his support of the project and for use of the data from other pollutants and source categories.

7. REFERENCES

Compilation of Air Pollutant Emission Factors. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Office of Air and Radiation: Research Triangle Park, NC, 1995. Available at http://www.epa.gov/ttn/chief/ap42 (accessed 2009).

Emissions Inventories. U.S. Environmental Protection Agency. Technology Transfer Network: Clearinghouse for Inventories & Emissions Factors. Available at http://www.epa.gov/ttn/chief/eiinformation.html (accessed 2009).

Continuous Emissions Monitoring Systems Data. U.S. Environmental Protection Agency. Clear Air Markets Division. Available at http://camddataandmaps.epa.gov (accessed 2009).

Form FERC-423 Database. Monthly Cost and Quality for Fuels and Electric Generating Units. U.S. Department of Energy. U.S. Energy Information Administration. Available at http://www.eia.doe.gov/cneaf/electricity/page/ferc423.html (accessed 2009).

SAS Institute Inc., SAS 9.2 Help and Documentation, Cary, NC: SAS Institute Inc., 2000-2004.

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 2009).

Procedures for Preparing Emission Factor Documents. U.S. Environmental Protection Agency. Office of Air Quality Planning and Standards. Office of Air and Radiation, 1997, p. 4-24. Available at http://www.epa.gov/ttn/chief/efdocs/procedur.pdf (accessed 2010).

Emissions Factor Uncertainty Assessment. RTI International, 2007, p. *4-5 – 4-6.*