In: Proceedings, Annual Meeting of the Air & Waste Management Association, held June 24-28, 2001 in Orlando, FL, and published by A&WMA, Pittsburgh, PA.

Quantification of Variability and Uncertainty in Emission Inventories: A Prototype Software Tool with Application to Utility NO_x Emissions

2001-Abstract #588

H. Christopher Frey
Department of Civil Engineering, North Carolina State University, Raleigh, NC 27695

Junyu Zheng
Department of Civil Engineering, North Carolina State University, Raleigh, NC 27695

ABSTRACT

The quality of stationary source emission factors is typically described using data quality ratings, as in AP-42, "Compilation of Air Pollutant Emission Factors." Such ratings are qualitative and provide no indication of the precision of the emission factor for an average emission source, nor of the variability in emissions from one source to another within a category. Advances in methodology and computing power enable the application of a quantitative approach to characterizing both variability and uncertainty in emission factors. Variability refers to actual differences in emissions from one source to another due to differences in feedstock composition, design, maintenance, and operation. Uncertainty refers to lack of knowledge regarding the true emissions because of measurement errors (both random and systematic), limited sample sizes (statistical random sampling error), and non-representativeness (which can introduce additional errors, including systematic errors). The set of numerical methods generically known as bootstrap simulation are a powerful tool for characterization of both variability and random sampling error. In this paper, we demonstrate the use of bootstrap simulation and related techniques for the quantification of variability and uncertainty for a selected example of NO_x emissions from coal-fired power plants. We have developed a prototype software tool that enables a user to display data sets for emission factors and activity factors for selected power plant technology groups. The user can select a parametric distribution to fit to the data. The user enters information regarding the number of power plant units in the inventory, and can display a variety of results regarding both variability and uncertainty in the inputs to the inventory, as well as uncertainty in various outputs of the inventory. While our example is focused upon emission factors for a selected criteria pollutant, the same methodology can be applied to other pollutants (e.g., hazardous air pollutants, greenhouse gases). The policy relevance of probabilistic inventories will be discussed.

Key Words: Variability, Uncertainty, Emissions, Emission Inventories, Emission Factors, Activity Factors, Monte Carlo simulation, probabilistic modeling, bootstrap simulation, nitrogen oxides, power plants, data quality, quality assurance

1.0 INTRODUCTION

Emission Inventories (EIs) are a vital component of environmental decision making. For example, emission inventories are used at federal, state, and local governments and private corporations for: (a) characterization of temporal emission trends; (b) emissions budgeting for regulatory and compliance purposes; and (c) prediction of ambient pollutant concentrations using air quality models. If random errors and biases in the EIs are not quantified, they can lead to erroneous conclusions regarding trends in emissions, source apportionment, compliance, and the relationship between emissions and ambient air quality.

Emission inventory work should include characterization and evaluation of the quality of data used to develop the inventory. For example, the National Research Council has recently recommended that quantifiable uncertainties be addressed in estimating mobile source emission factors, and in the past has addressed the need for understanding of uncertainties used in air quality modeling.[1,2] In this project, we demonstrate a quantitative approach to the characterization of both variability and uncertainty as an important foundation for conveying the quality of estimates to analysts and decision makers. The approach is illustrated by example for the case of emissions of NO_x from electric utility power plants. The example is conveyed via the Analysis of Uncertainty and Variability in Emissions Estimation (AUVEE) prototype software tool that has been developed for proof-of-concept purposes.

The AUVEE software takes into account both variability and uncertainty in the process of developing a probabilistic emission inventory. Variability is the heterogeneity of values with respect to time, space, or a population. Uncertainty arises due to lack of knowledge regarding the true value of a quantity. Variability in emissions arises from factors such as: (a) variation in feedstock (e.g., fuel) compositions; (b) inter-plant variability in design, operation, and maintenance; and (c) intra-plant variability in operation and maintenance. Uncertainty typically arises due to statistical sampling error, measurement errors, and systematic errors. In most cases, emissions estimates are both variable and uncertain. Therefore, we employ a methodology for simultaneous characterization of both variability and uncertainty based upon previous work in emissions estimation, exposure assessment, and risk assessment.[3-8] The method features the use of Monte Carlo and bootstrap simulation.

The specifics of the methodology used by the AUVEE software are documented in Frey and Zheng.[9] A previous report by Frey, Bharvirkar, and Zheng illustrates the application of similar methods to three case studies.[10] Probabilistic methods have previously been demonstrated in the context of air toxics emissions estimation, highway vehicle emission factors, and utility emissions.[3,5,10-18]

The objectives of this project are to:

(1) Demonstrate a general probabilistic approach for quantification of variability and uncertainty in emission factors and emission inventories;
(2) Demonstrate the insights obtained from the general probabilistic approach regarding the ranges of variability and uncertainty in both emissions factors and emission inventories;
(3) Demonstrate how probabilistic analysis can be used to identify key sources of variability and uncertainty in an inventory for purposes of targeting additional work to improve the quality of the inventory;

(4) Develop a prototype software tool for calculation of variability and uncertainty in statewide inventories for a selected emission source and pollutant; and
(5) Facilitate the transfer of the general approach and prototype software tool to federal, state or local governments or other recipients via development of appropriate technical and software documentation of the approach and the prototype software.

The AUVEE prototype software tool was developed to satisfy these five objectives. The specific example selected, power plant NO_x emissions, was chosen because power plant emissions represent a large contribution to national NO_x emissions. NO_x emissions are a significant concern because of their contribution to local and regional ozone formation.

It should be noted that the perspective of this uncertainty analysis is with respect to trying to estimate future emissions. Clearly, with the prevalence of continuous emission monitoring (CEM) equipment for measuring hourly NO_x emissions from a large number of power plants in the U.S., it is possible in many cases to characterize recent emissions of these plants with a comparative high degree of accuracy (e.g., perhaps precise to within approximately plus or minus 3 percent -- see Frey and Tran[14]). However, when making estimates of emissions any time into the future, it is more difficult to make a precise prediction. This is because there is underlying variability in the emissions of a single unit from one time period to another, even if the unit load is similar. Therefore, the purpose of the AUVEE prototype software tool is to assist in developing probabilistic estimates of future emission inventories based upon statistical analysis of representative CEMs data.

The prototype software tool was developed to demonstrate a methodology. The general approach employed to quantify variability and uncertainty in an emission inventory includes the following major steps:

1. Assemble and evaluate a database
2. Visualize data by developing empirical cumulative distribution functions for individual variables and scatter plots to evaluate dependencies among pairs of variables
3. Select, fit, and critique alternative parametric probability distribution models for representing variability in activity and emissions factors
4. Characterize uncertainty in the distributions for variability
5. Evaluate the effect of averaging, over both time and space, on variability and uncertainty
6. Propagate uncertainty and variability in activity and emissions factors to estimate uncertainty in statewide emissions.
7. Identify key sources of uncertainty in the emission inventory.

Because the main focus of this work was to demonstrate a methodology, this project was not intended to be comprehensive in terms of scope of coverage of all possible power plant technologies. To illustrate the methodology, five "technology groups" have been selected for characterization. A "technology group" is a combination of power plant unit furnace technology and of NO_x control technology (e.g., tangential-fired furnace with combustion-based NO_x control). The methods used to characterize variability and uncertainty in the emissions associated with these five technology groups can be extended later to include other technology

groups. Furthermore, the methods can be extended to other source categories and other pollutants.

In developing emission inventories, it is important to keep in mind the averaging time associated with the inventory. For example, in the prototype version of the AUVEE software tool, we include two different averaging times for power plant NO_x emissions. One is a 6-month averaging time, which is inclusive of the 2nd and 3rd quarters of the year. This 6-month period, therefore, includes the summer months which constitute the peak of the "ozone season." The other averaging time is a 12-month average, which would be useful for developing estimates of uncertainty in annual emission inventories. The prototype AUVEE software tool does not currently have a provision for calculating emission inventories for any other averaging time. Because the range of uncertainty in emission inventories is a function of the averaging time used in the inventory, the results of the uncertainty analyses from the prototype AUVEE software should not be applied to other averaging times without appropriate adjustments.

Although the methodology used in the AUVEE prototype software tool is one that can be widely applied, the results generated by the program are specific to the technology groups, averaging times, user input assumptions (e.g., number of units of each technology group and their sizes), data sets, and probabilistic assumptions (e.g., selection of parametric distributions) used in applying the software. Therefore, when reporting results from the use of the AUVEE software tool, we recommend that the user carefully document all of the assumptions used in a given case study so that another user could reproduce the same results.

This paper includes a discussion of the database developed for this work, some aspects of the development and implementation of the AUVEE system, and a case study demonstrating the development of a probabilistic emission inventory and the insights obtained from the inventory.

2.0 DEVELOPMENT OF A DATABASE

The methodology for probabilistic analysis is applied to a case study of variability and uncertainty in electric utility coal-fired power plant NO_x emissions. In this section, the focus is primarily on the development of the database

The data used for the case study is based upon Continuous Emission Monitoring (CEM) for individual power plant units obtained through the U.S. Environmental Protection Agency. The data are from the "Preliminary Summary Emissions Reports" of the Acid Rain Program of the U.S. Environmental Protection Agency (EPA). These files contain summary emissions information for electric utilities regulated by the EPA's Acid Rain Program. Each power plant unit subject to the Acid Rain Program regulations is required to report hourly data, describing emissions and operation, to EPA at the end of each calendar quarter. In this project, only the quarterly data files are used.

In the case studies of this project, two averaging times are considered: (1) 6-month; and (2) 12-month. The purpose of the 6-month averaging time is to characterize emissions that include the "ozone season." The purpose of the 12-month averaging time is to be able to characterize annual emissions for emissions budgeting and other purposes.

To develop the data necessary for these case studies requires combining data from two or more quarters and calculation of activity and emissions for the desired averaging times. The 6-month time period is intended to be inclusive of summer months. Therefore, the 6-month averages are based upon combining data from the 2nd and 3rd quarters of the year, including the months from April through September. The 12-month averages are based upon the entire year, and include the months from January through December. At the time that the data collection effort was made, quarterly data were available for the 1st quarter of 1997 through the 2nd quarter of 1999. Therefore, complete datasets of four quarters were available only for 1997 and 1998. Furthermore, data sets needed to characterize the 6-month period inclusive of the summer were available only for 1997 and 1998. In order to combine data from multiple quarters into a single data base, "macros" were developed using Visual Basic in Microsoft Excel™.

After the data combination and screening processes were completed, the final database was ready for statistical analysis. In this database, each record represents a power plant unit or stack. Each record contains the following information:

Unit/Stack Identification (Unit ID and ORISPL identifier)
General Information (State, Region)
Technology Group (Boiler Type, NO_x Control Technology)
Operation Data (Capacity, Operating Time)
Ten Quarters of NOx Emission Data

This database is used as a basis for the internal database of the prototype AUVEE software.

In developing an emission inventory, both activity and emission factors are needed. An emission factor characterizes the amount of emissions produced per unit of activity. For example, for a power plant, emission factors are often reported as mass of pollutant produced per unit of fuel consumed. The activity factor, therefore, is the amount of fuel consumed. To estimate fuel consumption for a power plant, one method is to use the power plant electrical generation, which is accurately measured, and the power plant efficiency in order to calculate the fuel input. Power plant efficiency is typically reported as a "heat rate", which is the ratio of fuel input with respect to electricity generation, in units of BTU of fuel input per kWh of electricity generated. Power plant load is often summarized using the capacity factor.

Four quantities are calculated from the combined database developed in this project. These quantities are: (1) unit/stack heat rate (BTU/kWh); (2) unit/stack capacity factor (actual kWh generated/maximum possible kWh); (3) NO_x emission rate on a fuel input basis (g/GJ); and NO_x emission rate on an energy output basis (g/GJ). Data from the final database are used to calculate the average emission factors and activity factors for each unit or stack. The averaging time includes 12-month averages and 6-month averages.

The emissions and activity data are calculated for selected technology groups. Four of the technology groups were selected based upon the most prevalent types of units in the data base. These include: (1) dry bottom, wall-fired boilers with no NO_x control; (2) dry bottom, wall-fired boilers with low NO_x burners (LNB); (3) tangential-fired boilers no NO_x controls; and (4) tangential-fired boilers with low NO_x burners and overfire air option 1, referred to as LNC1. The number of data points for these four technology groups ranges from 36 to 136, depending upon the technology group and the averaging time used. In addition, one other technology group was

selected that has a small sample size. The reason for selecting this group was to demonstrate that the probabilistic method for developing estimates of variability and uncertainty in emission inventories is able to deal with small data sets. The category for dry bottom, turbo-fired boilers with overfire air has only six data points and was selected for inclusion in the database.
In order to simplify the development of a database for use in case studies, possible statistical dependencies within the database were evaluated. To simplify the database as much as possible, it is desirable to be able to select data for one representative year. For both 1997 and 1998, there are four quarters of data. There were only two quarters of data available for 1999 at the time that this work was done. Therefore, the data for 1997 and 1998 were compared to identify similarities and differences between them. The data for the two years were similar, implying that data for either year could be used as the basis for analysis. The more recent 1998 data were selected. In addition, possible dependencies between activity and emission factors were evaluated. No significant dependencies were found. Therefore, it was not necessary to attempt to simulate statistical dependencies among emission and activity factors.

3.0 AUVEE SYSTEM DEVELOPMENT AND IMPLEMENTATION

The probabilistic methodology for emission inventory estimation was implemented in a prototype software, AUVEE. Here, we briefly introduce the functional design of AUVEE, the composition of the main modules and the relationships among them.

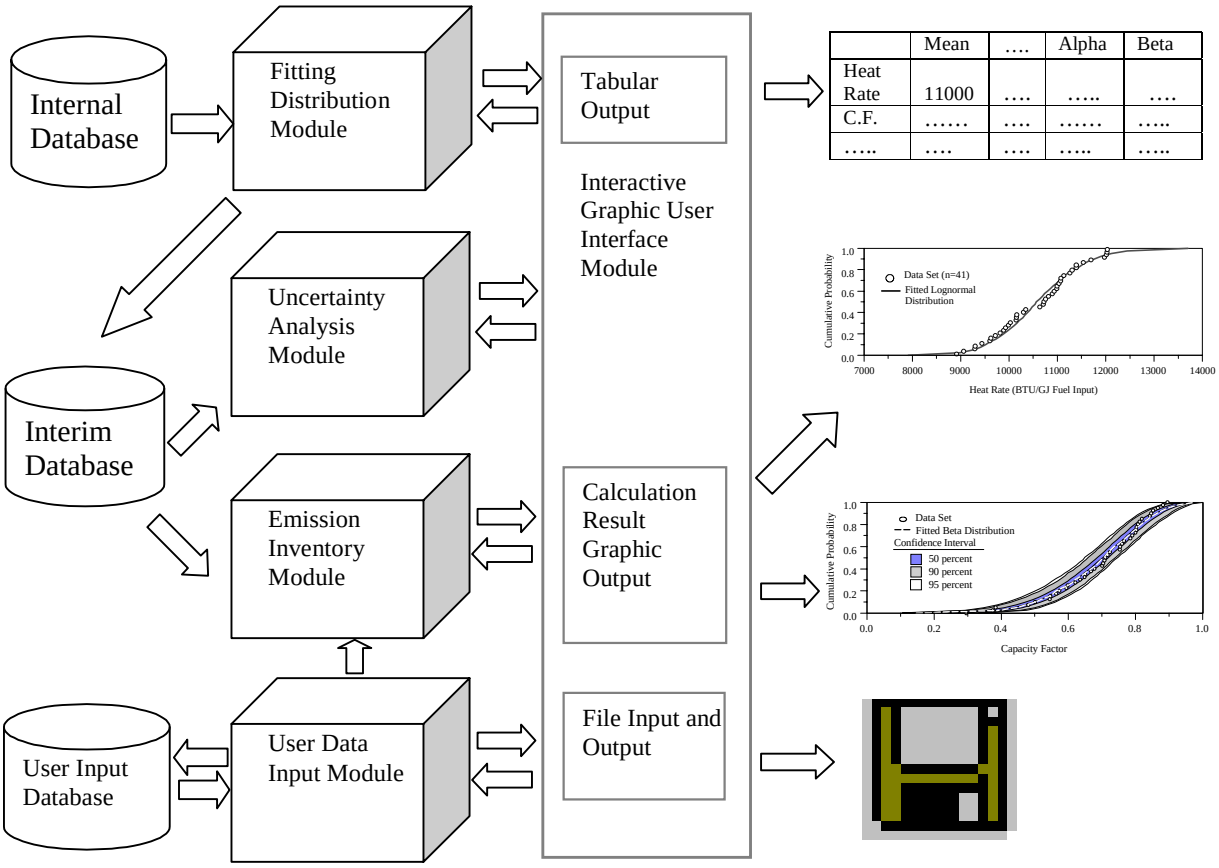
In AUVEE, user sets up a project. The project contains information on the choice of an internal emission factor and activity factors database, project name, project comments, and user data regarding the number of power plant units included in the inventory, the boiler and emissions control technology for each unit, and the capacity of each unit.

Figure 1 shows the conceptual design of AUVEE. AUVEE is composed of 3 databases, which include an internal database, a user input database and an interim database. In addition, AUVEE includes four main modules: (1) fitting distributions; (2) characterizing uncertainty; (3) calculating emission inventories; and (4) user data input. AUVEE features an interactive Graphical User Interface (GUI).

The internal database for AUVEE includes emission and activity factors obtained from CEMS data, and was described in detail in the previous section. The user may select either a 6-month average or a 12-month average database as the basis for developing either a 6-month or 12-month emission inventory, respectively. The internal database cannot be modified by the user in the prototype version of the software. The user input database stores data that the user provides regarding the number of power plant units, the boiler and emission control technology for each unit, and the capacity of each unit. This database can be edited by the user via the user data input module shown in Figure 1.

The interim database in AUVEE is used to store the results from the fitting distribution module and to store project information. The interim database provides fitted distribution information needed by the uncertainty analysis and emission inventory modules shown in Figure 1. A default interim database is provided so that the user can proceed to calculate emission inventory results even without making a new selection of parametric distributions to represent each input to the emission inventory.

Figure 1. Conceptual Design of the Analysis of Uncertainty and Variability in Emissions Inventories (AUVEE) Prototype Software System



The fitting distribution module implements all calculations for fitting parametric distributions to emission factor and activity factor data. This module provides graphs comparing fitted distributions to the data, allowing the user to evaluate the goodness of fit of parametric distributions fitted to datasets from the internal database. The user has the option, via a pull-down menu, to select alternative parametric distributions for fit to the data. When the user exits the fitting distribution model, the current set of fitted distributions are saved to the interim database for use by other modules in the program.

The characterizing uncertainty module implements the function of characterizing uncertainty in emission factors or activity factors based upon the EPA database and based upon the number of units of each technology group that are in the internal database. The characterizing uncertainty module uses data from the interim database to get distribution information including distribution type and the parameters of the fitted distributions for emission and activity factors. Uncertainty estimates of the mean emission and activity factors, and other statistics, are calculated using the numerical method of bootstrap simulation. The results of the uncertainty analysis are displayed in the GUI. Because this module uses data from the internal database, which may contain a relatively large number of power plant units compared to an individual state emission inventory, the estimates of uncertainty in the mean and in other statistics are typically a lower bound on the

range of uncertainty in the same statistic applicable to an emission inventory that includes a smaller number of power plant units.

The emission inventory module has the following functions: (1) it allows the user to visit the user database and append, modify or delete user input data; (2) it characterizes the uncertainty in emission factors and activity factors based on user project data; (3) it calculates uncertainty in the emission inventory; and (4) it calculates the key sources of uncertainty from among the different technology groups. It is via the emission inventory module that the user has access to the user data input module. The estimates of uncertainty in the emission inventory module are based upon the number of power plant units of each technology group specified by the user. For example, although there may be 36 power plant units of a given type in the internal database, the user may have only 10 units of that type in the emission inventory of interest. The uncertainty in the emission and activity factors for that technology group will be estimated based upon a sample size of 10, not 36.

The GUI is actually a general control module in AUVEE, and it makes all independent modules, platforms and databases work together. In addition, the GUI is a bridge which links user input to internal implementation within AUVEE, and provides model output to the user. Through the GUI, the user can build or open a project, enter a database of emission sources, implement user's choice of parametric distributions, view or save all calculation results, and manage the message passing between the different modules.

4.0 CASE STUDY: A PROBABILISTIC EMISSION INVENTORY FOR A SINGLE STATE

The approach for developing a probabilistic emission inventory using AUVEE is illustrated here using a case study. The case study is based on the number of units of each selected technology group within the state of North Carolina. This case study was selected because the number of units representing each of four power plant technologies is dissimilar. Thus, there are different amounts of uncertainty, based on random sampling error, associated with the emissions estimates for each of the technologies. Specifically, the number of units ranges from only three, in the case of dry-bottom wall-fired boilers using low NO_x burners, to 19, in the case of uncontrolled tangential-fired boilers. No units are available for the dry bottom turbo-fired with overfire air technology group in the state of North Carolina. In addition, not all technology groups in North Carolina are included in the case study.

The uncertainty in the emission inventory can be characterized by the propagation of probabilistic model inputs through the emission inventory model. For a power plant, model inputs in the emission inventory model include activity factors and emission factors. Activity factors include heat rate (BTU/kWh), capacity factor, and capacity (MW) for individual units. In this project, heat rate and capacity factor were probabilistically characterized. Capacity was assumed to be fixed without uncertainty and variability. However, the approach could be extended to treat capacities probabilistically if there were reasons to believe that the reported capacities were in error. Compared to variability and uncertainty in heat rate and capacity factor, it is unlikely that uncertainty or variability regarding true plant capacity would play a significant role in most cases, other than due to data recording errors. All emission factors were characterized probabilistically.

4.1 Fitting Distributions to Data to Represent Inter-Unit Variability

The case study is based upon a 6-month period, inclusive of summer months. From the internal database of the model, 6-month average data obtained from the EPA CEMS database were analyzed via the AUVEE user interface. Parametric probability distributions were fit to each activity and emission factor required for the inventory. The parameters of the distributions were estimated by AUVEE using Maximum Likelihood Estimation (MLE). Examples of the fitted distributions for the example of one technology group are shown in Figures 2, 3, and 4 for an emission factor, a capacity factor, and a heat rate, respectively. The fitted distributions are a compact means for representing inter-unit variability. The goodness-of-fit can be evaluated qualitatively by comparing the fitted distribution with the data. For example, the Lognormal distribution fitted to the emission factor data agrees with the tails of the distribution of the data and with the central tendency of the data. There are some deviations of the fitted distribution from the data in the regions of the 25th and 75th percentiles, indicating that the fit is not particularly good. In contrast, the Beta distribution fitted to the capacity factor data agrees very well with the data, as does the Lognormal distribution fitted to the heat rate data.

4.2 Quantifying Uncertainty in Statistics of the Fitted Distributions

Bootstrap simulation is used to quantify uncertainty in the inputs to the emission inventory. Bootstrap simulation was introduced by Efron as a means for calculating confidence intervals for statistics in a general manner for situations in which analytical solutions are not available.[19] A probabilistic framework for calculating uncertainty in emissions estimation using Bootstrap simulation is described in detail elsewhere.[3-5,9,10] Bootstrap simulation is a numerical method for simulating random sampling error. In bootstrap simulation, a probability distribution is assumed to be a best estimate of the true but unknown population distribution for a quantity. For example, the parametric distributions fit to datasets for inter-unit variability in emissions and activity data are assumed to be the best estimates of the true but unknown population distribution for inter-unit variability of these quantities.

Using a random sampling technique, synthetic data sets of the same sample size as the original observed data set are simulated from the assumed population distribution. The random sampling technique employed is Monte Carlo simulation. A synthetic random sample of the same sample size as the original data is referred to as a *bootstrap sample*. For each bootstrap sample, a new value of the statistic(s) of interest, such as the mean, standard deviation, distribution parameters, or fractiles of the distribution, are calculated. An estimate of a statistic calculated from a bootstrap sample is referred to as a *bootstrap replicate* of the statistic.

To obtain an estimate of uncertainty for the selected statistic(s), bootstrap samples are drawn repeatedly from the assumed population distribution. For example, if the original data set contained 36 data points, perhaps 500 random samples of 36 data points would be simulated, and for each of these 500 bootstrap samples, a bootstrap replicate of the mean is calculated. The 500 bootstrap means describe a *sampling distribution* for the mean. A sampling distribution is a probability distribution for a statistic. From the sampling distribution, probability ranges can be inferred. For example, the 95 percent probability range for the mean can be estimated.

Figure 2. Comparison of Fitted Lognormal Distribution and Six-Month Average NO_x Emission Factor Data for Tangential-Fired Boilers with NO_x Control.

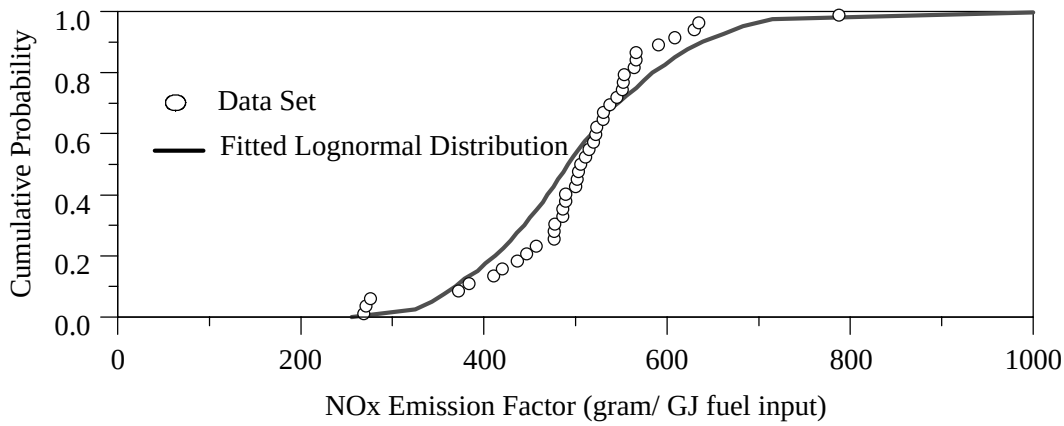


Figure 3. Comparison of Fitted Beta Distribution and Six-Month Average Capacity Factor Data for Tangential-Fired Boilers with NO_x Control.

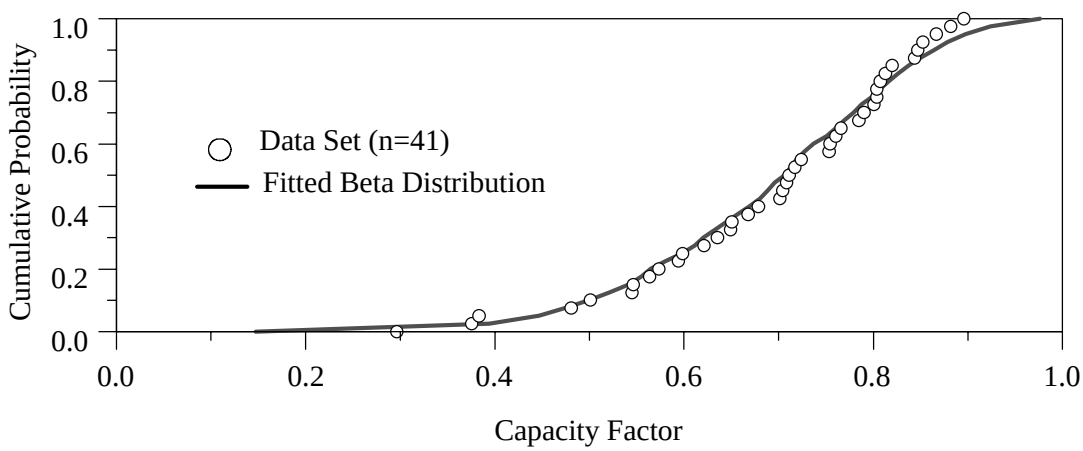
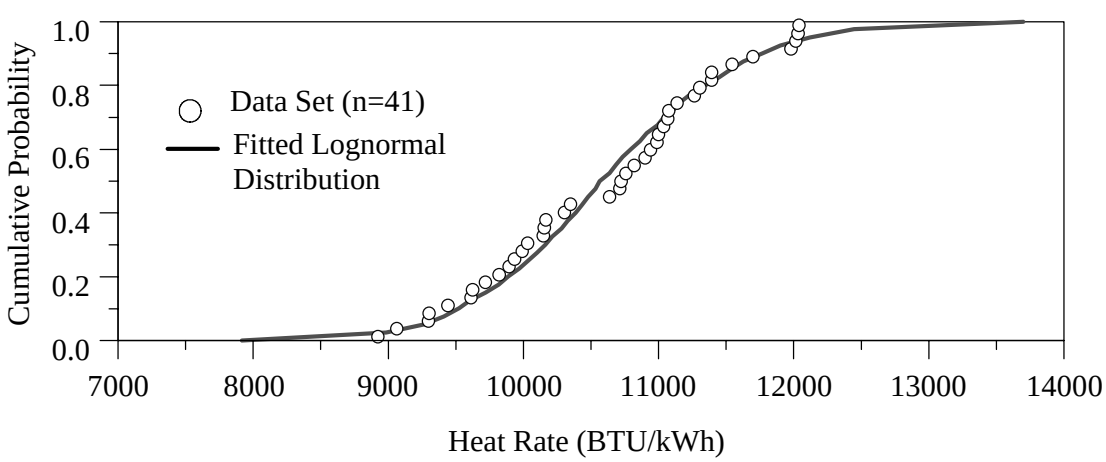


Figure 4. Comparison of Fitted Lognormal Distribution and Six-Month Average Heat Rate Data for Tangential-Fired Boilers with NO_x Control.



4.3 Evaluating Goodness-of-Fit Using Bootstrap Results

Bootstrap simulation can be used to help evaluate the goodness of a fit of a distribution with respect to the original data. Confidence intervals for the fitted distribution can be estimated and compared with the original data.

For example, Figures 5, 6, and 7 show a comparison of confidence intervals for the fitted distribution with the datasets for the emission factor, capacity factor, and heat rate, respectively, for one technology group. In Figure 5, it appears that most of the data are contained within the 95 percent confidence interval; however, few of the data are contained within the 50 percent confidence interval. Thus, it appears that the Lognormal distribution may adequately describe the inter-unit variability in emissions for some data quality criteria, but perhaps not for others. Later, we will return to consider whether this particular input was important to the overall estimate of uncertainty in the inventory. For the other two cases, the fitted distributions agree very well with the data. For example, more than half of the data are enclosed by the 50 percent confidence intervals, and all but one or two data points out of 41 are contained within the 95 percent confidence intervals. Thus, the fits in these two cases are reasonably good ones. From these comparisons, which the user may view via the AUVEE GUI, one may conclude that the fitted distributions adequately characterize inter-unit variability.

4.4 Quantifying Uncertainty in the Inputs to an Emission Inventory

After the user has entered data regarding the number of units of each technology group that are included in the inventory, a simulation of uncertainty specific to the particular inventory may be performed. For example, in the example inventory, there are only 11 units of the specific technology group represented in Figures 5, 6, and 7. Thus, although there are a total of 41 such units represented in the database, the uncertainty estimate specific to the example inventory must account for the fact that there are only 11 units in the inventory. An assumption is that the 11 units are a random sample of the population of all units of the same technology group. The uncertainty in the mean emission rate among 11 units should be based upon a sample size of 11 and not a sample size of 41. If the 11 units are a random sample from the population, then the sampling distribution for the mean of the 11 units must reflect stochastic variation in the mean for a random sample of only 11. Therefore, bootstrap simulation with bootstrap samples of 11 synthetic data points is used to quantify uncertainty in the distribution used to describe inter-unit variability in emissions for a sample of 11 units.

An example of results for uncertainty based upon the number of units actually in the inventory is shown in Figure 8 for the case of the capacity factor. In comparing Figure 8 with Figure 6, it is apparent that the confidence intervals are much wider in Figure 8. The increased width of the confidence intervals in Figure 8 corresponds to the smaller sample size of 11 versus 41, the latter of which is the basis for the bootstrap simulation results shown in Figure 6. With a random sample of only 11, there is more random fluctuation in the mean, median, standard deviation, parameter values, fractiles, and other statistics that may be calculated from the bootstrap samples. With a smaller number of units, the range of uncertainty is larger.

Figure 5. Probability Band for Distribution Fitted to Example NO_x Emission Factor Data

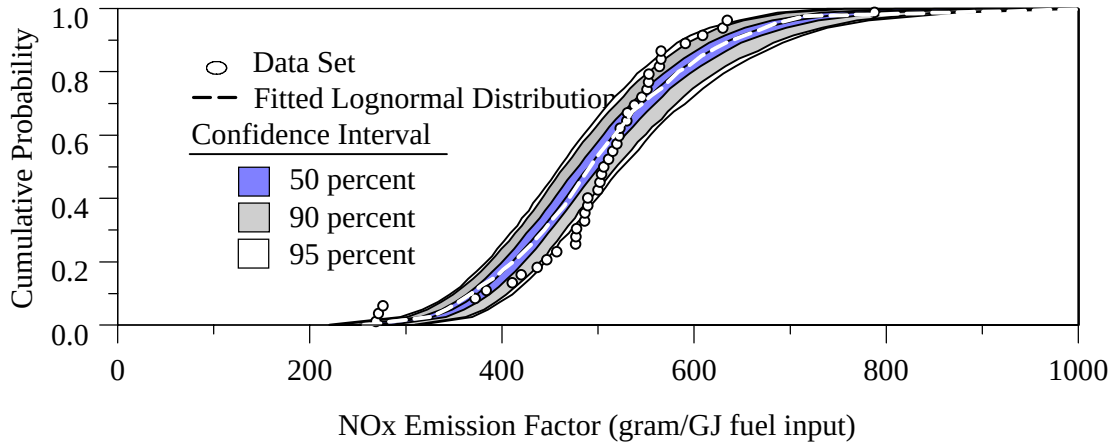


Figure 6. Probability Band for Distribution Fitted to Example Capacity Factor Data.

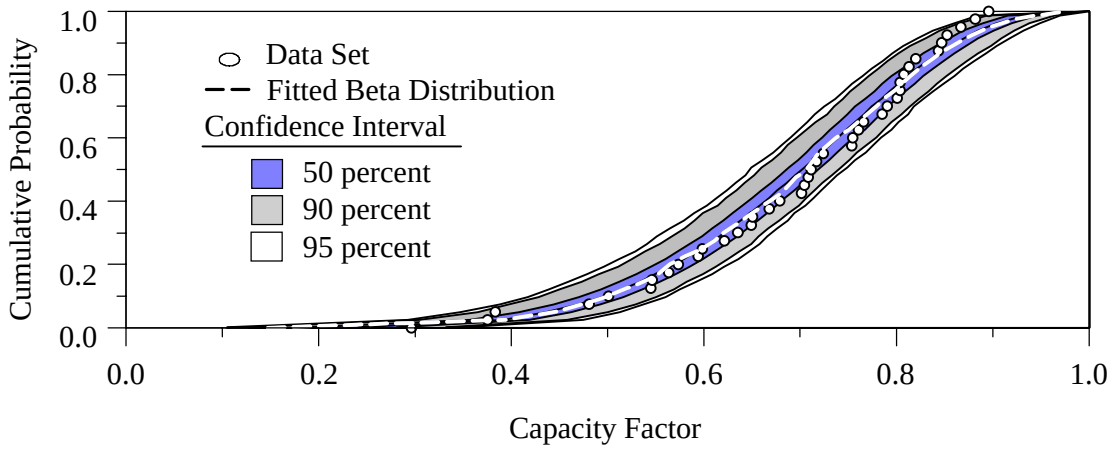


Figure 7. Probability Band for Distribution Fitted to Example Heat Rate Data

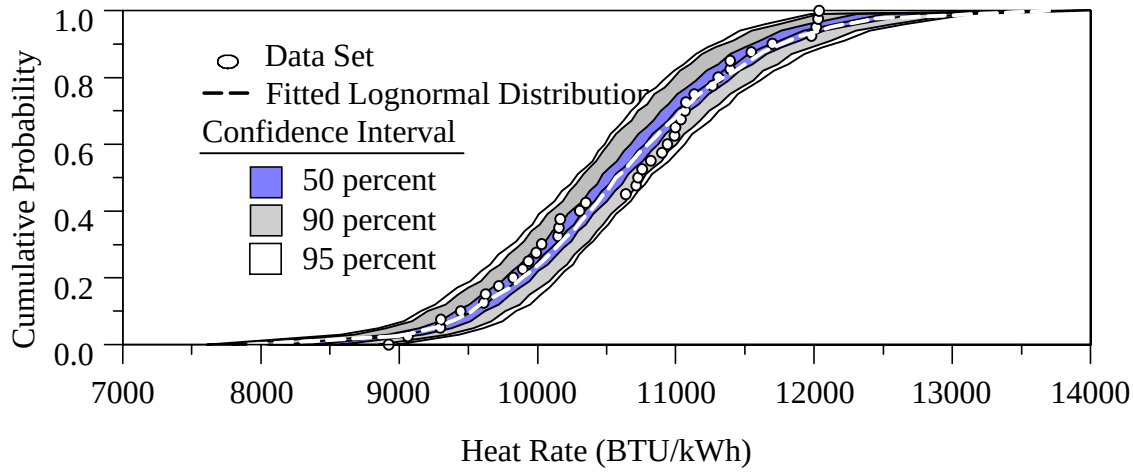
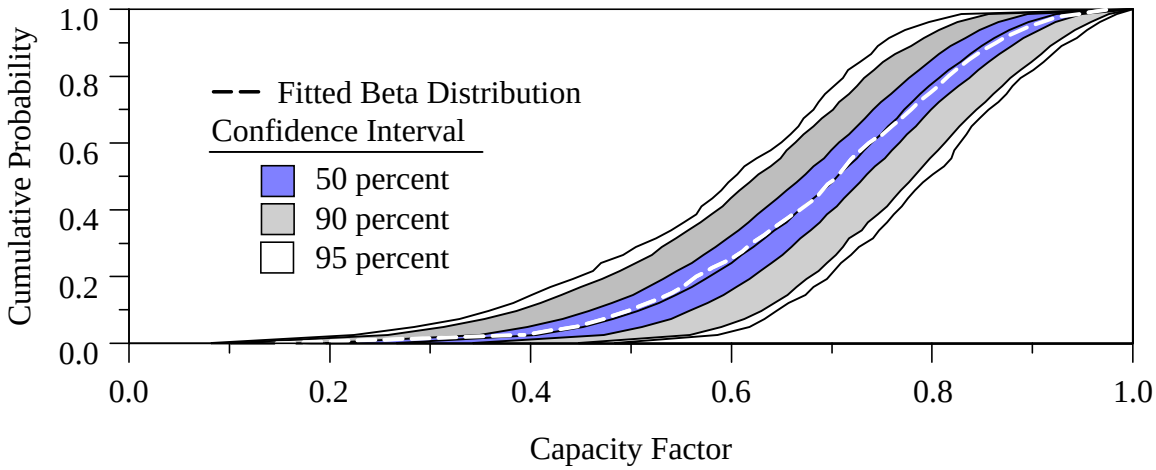


Figure 8. Probability Bands Based Upon Number of Units in the Emission Inventory (n=11) for the Example of Capacity Factor.



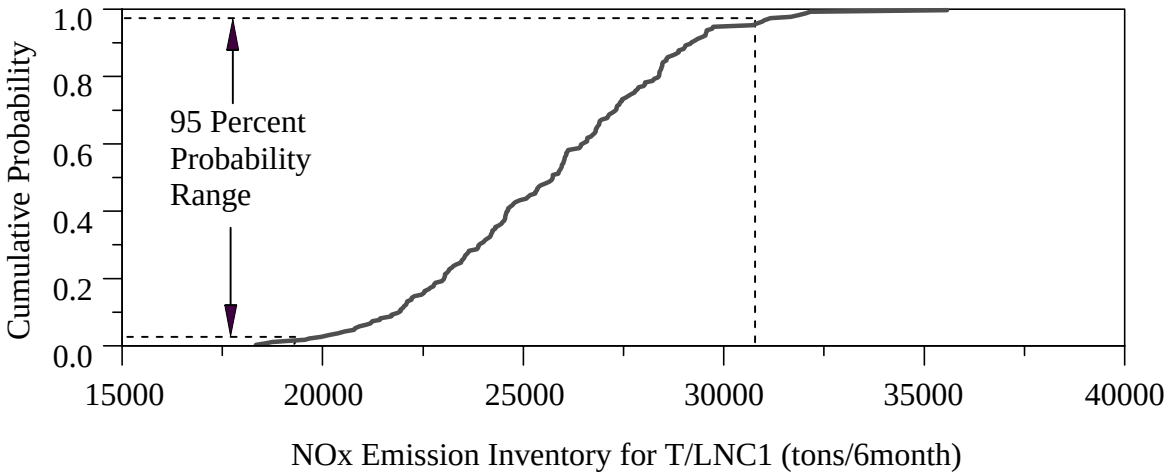
4.5 Propagating Uncertainty in Emission Inventory Inputs to Predict Uncertainty in Emission Inventory Outputs

To estimate uncertainty in the total emissions for the inventory, emissions for individual units of each technology group are simulated. For example, if there are 11 units in a technology group, then 11 random samples are simulated from the fitted distributions for emission factor, capacity, and heat rate. Each of these 11 values are paired with one of the 11 user-entered values for unit capacities. Eleven values of emissions for each of the 11 units are calculated and summed to represent one possible realization of total emissions for the technology group. This process is repeated many times for the technology group to develop hundreds or thousands of estimates of total emissions within the group. The distribution of values of the total emissions for the group represents uncertainty in the total emissions. This process is repeated for all technology groups in the inventory. The uncertainty in the inventory, inclusive of all technology groups, is calculated by summing the emissions from each technology group for each of the hundreds or thousands of realizations of uncertainty, to create hundreds or thousands of alternative random estimates of the emission inventory.

Figure 9 illustrates an uncertainty estimate for the total emissions from one technology group. In this case, the emissions are from 11 units. The mean value of the emissions is 25,200 tons of NO_x emitted over a six month period. The 2.5th percentile of the distribution of uncertainty in emissions is 19,800 tons of NO_x emitted over a six month period. The 97.5th percentile is 31,100 tons of NO_x emitted over a six month period. The 2.5th and 97.5th percentiles enclose the 95 percent probability range. Expressed on a relative basis, the 95 percent probability range for uncertainty is minus 21 percent to plus 23 percent with respect to the mean value.

The range of uncertainty in the emissions for the example technology group is slightly asymmetric, reflecting the fact that many of the inputs to the emission inventory have skewed distributions (e.g., as in the case of the Lognormal distribution fit to the emission factor data) and

Figure 9. Uncertainty in a Six-Month NO_x Emission Inventory for an Individual Technology Group Comprised of 11 Units.



small sample sizes (e.g., n=11). The range of uncertainty reflects the large amount of inter-unit variability in the inputs to the inventory.

To illustrate the amount of inter-unit variability in the inputs to the emission inventory, refer to Figures 2, 3, and 4. In Figure 2, the highest value of the emission factor is approximately three times larger than the lowest value of the emission factor. Capacity factor varies from approximately 0.3 to 0.9, as indicated in Figure 3. The heat rate varies from approximately 9,000 BTU/kWh to approximately 12,000 BTU/kWh. The wide range of variation in performance and operation of these types of units is reflected in the comparatively wide range of uncertainty for the total emissions of this technology group.

The overall uncertainty in the emission inventory, inclusive of all four technology groups considered, is shown in Figure 10. The estimated mean emission rate is 84,800 tons of NO_x emitted in a six month period. The 95 percent probability range is enclosed by emissions of 71,800 tons and 99,900 tons. This is a range of -13,000 tons to +15,100 tons, or -15 percent to +18 percent, with respect to the mean. The asymmetry of the 95 percent probability range is a result of skewness in many of the input assumptions among the four technology groups.

A summary of the uncertainty results for the entire emission inventory is given in Table 1. Although the absolute range of uncertainty for the total inventory is greater than the absolute range of uncertainty for the selected technology group, the relative range of uncertainty is smaller. While this result may seem counter-intuitive, the result occurs because the uncertainty in emissions for each technology group is assumed to be statistically independent of the other technology groups. There is no compelling reason to assume, for example, that if emissions are high at a particular tangential-fired boiler, that they must also be high at a dry-bottom boiler also located in the region of the inventory.

Figure 10. Uncertainty in a Six-Month NO_x Emission Inventory Inclusive of Four Technology Groups.

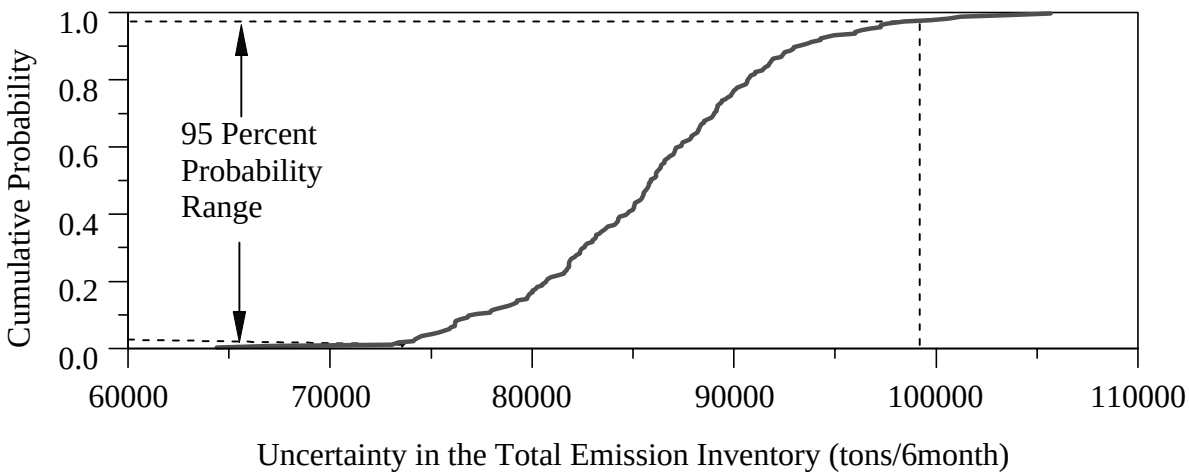


Table 1. Summary of Uncertainty Results for the Emission Inventory Case Study

Technology Group	2.5th Percent	Mean	97.5th Percentile	*Random Error (%)*[a]	
				Negative	Positive
DB/U	21,700	31,100	40,100	30	29
DB/LNB	5,600	8,100	11,400	31	39
T/U	15,300	20,400	28,600	25	40
T/LNC1	19,800	25,200	31,100	21	23
Total	71,800	84,800	99,900	15	18

[a]Results shown are the relative uncertainty ranges for a 95 percent probability range, given with respect to the mean value.

A property of probabilistic simulations is that, in general, it is not possible to sum the values of selected percentiles of each model input to obtain an estimate of the same percentile of the model output. For example, the 2.5th percentile of the total emission inventory, which is 71,800 tons, does not correspond to a sum of the 2.5th percentile of each of the four technology groups. However, for linear models, the sum of the means is usually the same as the mean of the sum, unless there is a correlation among the model inputs.[6]

4.6 Identifying Key Sources of Uncertainty in the Inventory

A method for identifying which technology groups contribute the most to uncertainty in the overall emission inventory is included in AUVEE. The method is based upon calculating the correlation between the uncertainty in emissions from an individual group and the uncertainty in total emissions. The correlation is a measure of the linear covariation of the two uncertainty distributions. The larger the magnitude of the correlation, the stronger the linear dependence between the two. Of the four technology groups, the dry-bottom, uncontrolled (DB/U) group has the strongest correlation with uncertainty in the total emission inventory, with a correlation coefficient of approximately 0.7. In contrast, the controlled tangential boiler group used as the

basis for the examples in Figures 2 through 8 has a correlation of approximately 0.45, and was only the third most important of the four groups in contributing to uncertainty in the total inventory.

As noted earlier, the fitted distribution for the controlled tangential boiler group emission factor was not a particularly good fit to the data. However, given that this particular group is only the third most important contributor to uncertainty in the total inventory, the discrepancies in the fit are not likely to contribute substantially to errors in the overall estimate of uncertainty in the inventory.

5.0 CONCLUSIONS

This project has demonstrated a prototype software environment for calculation of probabilistic emission inventories. The prototype enables a user to visualize, in the form of empirical probability distributions, the data used to develop the inventory. Therefore, the user is able to observe the range of variability in the data. This is sharp contrast from typical emission inventory work, in which point estimate values of emission factors are used to calculate a single estimate of the inventory. The range of variability in the example datasets was shown to be large. For example, the range of inter-unit variability in emission factors for one technology group was a factor of approximately three from the smallest to the largest value in the dataset.

Although it is not possible to quantify all sources of uncertainty, it is important to quantify as many sources of uncertainty as is practical. The example case study demonstrates the the range of uncertainty attributable to random sampling error is substantial. For individual technology groups, the range of uncertainty is as large as approximately plus or minus 30 percent, and for the total inventory the range of uncertainty is approximately plus or minus 15 percent. These ranges of uncertainty are likely to be substantially larger than measurement errors in the data. The case study is based upon a relatively large sample of continuous emission monitoring data. Therefore, it is likely that the data used in the case study are reasonably representative of actual emissions among the population of units for the technology groups studied. For the case study here, it is likely that random sampling error is the most important contributor to overall uncertainty.

The estimates of uncertainty reflect the lack of information than an emissions estimator would have regarding future emissions for the selected source category. As noted early in the paper, it is now possible to have a high degree of uncertainty regarding recent actual emissions at power plants equipped with CEM equipment. However, given the inherent variability in emissions from one unit to another, and at a single unit over time, it is not possible to have certainty regarding what the emissions will be at a future time, whether in the near or distant future. In estimating distant future emissions, an additional refinement that may be needed in the case study would be to consider changes in capacity factor and the effects of capacity expansion. For relatively short term future estimates (e.g., a year or two into the future), the methodology employed as is may provide a reasonable estimate of absolute emissions. However, the relative range of uncertainty estimated using the methods presented here are likely to be indicative of the relative range of uncertainty in a future emission inventory, unless there is a large shift in the relative contributions of different technology groups to the total inventory.

In addition to quantifying the substantial range of uncertainty in the inventory, the case study demonstrates the capability to identify key sources of uncertainty in the inventory. As noted, the largest contribution to uncertainty comes from one technology group. Therefore, if it were an objective to reduce uncertainty in the overall inventory, resources could be focused on collecting more or better data for the most sensitive technology group. Knowledge of key sources of uncertainty can also aid in identifying where it is not necessary to target additional data collection. For example, even though there were some discrepancies in the fit of a parametric distributions to one of the emission factors, that particular emission factor does not contribute substantially to uncertainty in the overall inventory. Therefore, there would not be a large benefit associated with improving the characterization of uncertainty for that particular input.

The project has demonstrated a probabilistic approach for development of emission inventories. Because of the widespread use of inventories for policy making, planning, and research purposes, it is important that the quality of the inventories be known and that any shortcomings in the inventories be identified and prioritized for improvement. The method illustrated here enables quantification of the variability and uncertainty in each input to an inventory, quantification of the precision of the inventory, and identification of key sources of uncertainty that can be targeted for reduction via additional data collection and research. The latter is especially a critical concern when allocating scarce dollars to potentially expensive field studies or surveys.

The quantification of uncertainty has many important implications for decisions. For example, it enables analysts and decision makers to evaluate whether time series trends are statistically significant or not. It enables decision makers to determine the likelihood that an emissions budget will be met. Inventory uncertainties can be used as input to air quality models to estimate uncertainty in predicted ambient concentrations, which in turn can be compared to ambient air quality standards to determine the likelihood that a particular control strategy will be effective in meeting the standards. In addition, using probabilistic methods, it is possible to compare the uncertainty reduction benefits of alternative emission inventory development methods, such as those based upon generic versus more site-specific data. Thus, the methods presented here allow decision makers to assess the quality of their decisions and to decide on whether and how to reduce the uncertainties that most significantly affect those decisions.

6.0 ACKNOWLEDGMENTS

The authors acknowledge the support of the Office of Air Quality Planning and Standards (OAQPS) of the U.S. Environmental Protection Agency, which funded most of this work. The authors appreciate the guidance and encouragement of Mr. Steve Rhomberg, formerly with U.S. EPA, and Ms. Rhonda Thompson of U.S. EPA.

Some support for the methodological components of this work was also provided via U.S. EPA Science to Achieve Results (STAR) Grants Nos. R826766 and R826790.

Although the research described in this paper has been funded wholly or in part by the U.S. EPA, including the STAR grant program, this paper has not been subject to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

The authors also thank Mr. Zhen Xie for his contributions to the development of the internal database used in the AUVEE prototype software.

7.0 REFERENCES

1. NRC, *Modeling Mobile Source Emissions*, National Academy Press, Washington, D.C., 2000.

2. National Research Council. *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, National Academy Press: Washington, DC, 1991.

3. Frey, H.C. and Rhodes, D.S.; "Characterizing, simulating and analyzing variability and uncertainty: An illustration of methods using an air toxics emissions example", *Human and Ecological Risk Assessment.* 2(4): 762-797 (1996).

4. Rhodes, D.S., and H.C. Frey, "Quantification of Variability and Uncertainty in AP-42 Emission Factors Using Bootstrap Simulation," Emission Inventory: Planning for the Future (held October 28-30 in Research Triangle Park, NC), Air and Waste Management Association: Pittsburgh, Pennsylvania, October 1997, pp. 147-161.

5. Frey, H.C., and D.S. Rhodes, "Characterization and Simulation of Uncertain Frequency Distributions: Effects of Distribution Choice, Variability, Uncertainty, and Parameter Dependence," *Human and Ecological Risk Assessment*, 4(2):423-468 (April 1998).

6. Cullen, A.C., and Frey, H.C.; *Use of Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Plenum Press: New York, 1999.

7. U.S. EPA, Guiding Principles for Monte Carlo Analysis, EPA/630/R-97/001, U.S. Environmental Protection Agency, Washington, DC, May, 1997.

8. U.S. EPA, Report of the Workshop on Selecting Input Distributions for Probabilistic Assessment, EPA/630/R-98/004, U.S. Environmental Protection Agency, Washington, DC, January 1999.

9. Frey, H.C, J. Zheng, "Quantitative Analysis of Variability and Uncertainty in Emissions Estimation: Development of a Database, Prototype Software and Example Case Study," Draft, Prepared by North Carolina State University for U.S. Environmental Protection Agency, Research Triangle Park, NC, January 2001.

10. Frey, H.C., R. Bharvirkar, J. Zheng; "Quantitative Analysis of Variability and Uncertainty in Emissions Estimation"; Final Report, Prepared by North Carolina State University for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, pp 1-10,1999.

11. Kini, M.D., and H.C. Frey, Probabilistic Evaluation of Mobile Source Air Pollution: Volume 1, Probabilistic Modeling of Exhaust Emissions from Light Duty Gasoline Vehicles, Prepared by North Carolina State University for Center for Transportation and the Environment: Raleigh, December 1997.

12. Dickson, R.J., Hobbs A. D.; “Evaluation of Emission Inventory Uncertainty Estimation Procedures.” Paper No. 89-24.8, In *82nd Annual Meeting*, Air & Waste Management Association: Anaheim, CA 1989.

13. Balentine, H. W.; Dickson, R. J. “Development of Uncertainty Estimates for the Grand Canyon Visibility Transport Commission Emission Inventory.” In *Conference on the Emission Inventory*, Air & Waste Management Association: Research Triangle Park, NC, 1995.

14. Frey, H.C., and L.K. Tran, *Quantitative Analysis of Variability and Uncertainty in Environmental Data and Models: Volume 2. Performance, Emissions, and Cost of Combustion-Based NO_x Controls for Wall and Tangential Furnace Coal-Fired Power Plants*, Report No. DOE/ER/30250--Vol. 2, Prepared by North Carolina State University for the U.S. Department of Energy, Germantown, MD, April 1999.

15. Pollack, A.K., P. Bhave, J. Heiken, K. Lee, S. Shepard, C. Tran, G. Yarwood, R.F. Sawyer, and B.A. Joy, "Investigation of Emission Factors in the California EMFAC7G Model. PB99-149718INZ, Prepared by ENVIRON International Corp, Novato, CA, for Coordinating Research Council, Atlanta, GA. 1999.

16. Frey, H.C., and S. Bammi; "Quantification Of Variability and Uncertainty in Lawn And Garden Equipment NO_x and Total Hydrocarbon Emission Factors," Proceedings of the Annual Meeting of the Air & Waste Management Association, Orlando, FL, June 2001 (in press)

17. Frey, H.C., and S. Li; "Quantification of Variability and Uncertainty in Natural Gas-fueled Internal Combustion Engine NO_x and Total Organic Compounds Emission Factors," Proceedings of the Annual Meeting of the Air & Waste Mangement Association, Orlando, FL, June 2001 (in press)

18. Radian, *Evaluating the Uncertainty of Emission Estimates, Final Report*, Prepared by Radian Corporation for the Emission Inventory Improvement Program, State and Territorial Air Pollution Control Officers’ Association, Association of Local Air Pollution Control Officers, and U.S. Environmental Protection Agency, Research Triangle Park, NC, 1996.

19. Efron, B.; Tibshirani, R. J., *An Introduction to Bootstrap*; Chapman and Hall: New York 1993.