Exploration of Emissions Factor Adjustments for Using Emissions Factors in Noninventory Applications

Roy Neulicht, Breda Munoz, and Karen Schaffner
RTI International,* 3040 Cornwallis Road, Research Triangle Park, NC 27709
rmneulicht@rti.org

Barrett Parker
U.S. Environmental Protection Agency, MD C339-02, Research Triangle Park, NC 27711
parker.barrett@epa.gov

ABSTRACT

This paper summarizes the results of a study funded by and conducted for the U.S. Environmental Protection Agency to evaluate and develop adjustments to account for the uncertainty associated with using emissions factors in noninventory applications. The adjustments will enable users to apply the AP-42 emissions factors, which are based on mean values, in applications where other values (e.g., maximum or minimum values) may be more appropriate. For example, in estimating emissions data inputs for risk assessment analysis or for determining the applicability of a rule to a single source, an upper boundary emissions factor might be more appropriate than an average emissions factor to calculate the emissions from a single source.

During the study, the data sets for 44 A-rated AP-42 emissions factors from four industries were analyzed and the following pollutants were evaluated: particulate matter (PM); sulfur dioxide (SO_2); nitrogen oxides (NO_x); carbon monoxide (CO); and hazardous air pollutants (HAPs). Examination of the data for each of the emissions factor data sets indicates that the data are either log-normal or Weibull distributed. A statistical analyses based on the Monte Carlo technique was conducted on each of the emissions factor data sets to simulate the hypothetical population density of the emissions factor for the specific pollutant. Using the hypothetical population, default emissions factor adjustments were developed. The emissions factors analyzed, the statistical procedures used, and the default emissions factor adjustments developed are presented and discussed.

INTRODUCTION

The U.S. Environmental Protection Agency (EPA) and its predecessors have used emissions factors since 1968 to estimate emissions from point and area sources. Because of program priorities and goals, the emissions factor program primarily supported the development of the national trends emissions inventory and other inventories used for state and regional implementation plans. Over the last 10 years, however, the number of programs that use emissions factors has increased beyond the intended and supported national emissions inventory program use. In 2003, EPA began a complete re-evaluation of the emissions factor program. Part of this re-evaluation includes identifying ways to make the program more responsive to the broad and diverse range of emissions factors users. This paper presents the results of a study funded by and conducted for EPA to evaluate and develop adjustments for using emissions factors in noninventory applications.[1] The emissions factors analyzed, the statistical approach and procedures used to determine default emissions factor adjustments, the results, and the composite default adjustment factors are presented and discussed. The adjustments will enable users to apply the AP-42 emissions factors, which are based on mean values, in applications where other values (e.g., maximum or minimum values) would be more appropriate.[2] For noninventory applications of emissions factors, the AP-42 emissions factor would be multiplied by the composite default adjustment to estimate the emissions factor at the pertinent statistic (e.g., the 90th percentile). The true emissions from a single source of interest may fall anywhere within the range of emissions data. Although the

approach of applying an adjustment is not a fix for all situations, it is an improvement over the current use of emissions factors from AP-42 for all applications.

The characterization of the individual emissions factor data sets, the statistical procedures used to simulate a distribution of the population (from the sample made up of the AP-42 emissions factor data), and the statistical procedures used to calculate default adjustments and normalized Monte Carlo sampling distributions of the mean are discussed.

Technical Approach Overview

The overall technical approach consists of first selecting A-rated emissions factor data sets for analysis and using exploratory data analysis techniques to visualize and characterize these data sets. Then, statistical techniques are applied to each of the emissions factor data sets selected to determine preliminary emissions factor adjustments. Finally, default emissions factor adjustments are calculated for combined data sets. During the course of the project, several different statistical analyses were explored or conducted. The technical approach established for this project comprises the following steps:

1. Select and prepare initial emissions factor data sets for analysis.

2. Establish the statistical procedures.

3. Conduct statistical analyses of an emissions factor data set for an industry and calculate preliminary default emissions factor adjustments.

4. Review the initial results and refine the analytical approach.

5. Conduct statistical analyses of additional representative emissions factor data sets.

6. Calculate composite default emissions factor adjustments for the combined data sets.

7. Consider alternative approaches and compare results to default emissions factor adjustments for noninventory use.

Selecting Emissions Factor Data

AP-42 emissions factor data sets were selected for statistical analysis. These AP-42 emissions factors are representative of an industry average, and the supporting emissions data used to develop the emissions factors are also publicly available in the background documentation for each AP-42 industry-specific section. The rationale for selecting which data sets for the analysis was based on several criteria, including the following:

- The quality rating of the emissions factor
- The quantity of emissions data used to develop the factors (i.e., number of emissions tests)
- The number of pollutants included
- The accessibility of the supporting emissions data.

Based on these criteria, data sets from the AP-42 background documentation were selected for four source categories or industries. Each of these AP-42 sections provided the requisite supporting

background documentation and detailed data sets used to develop the emissions factors. The data sets were compiled for statistical analysis from the following AP-42 chapters:

- Wood Residue Combustion in Boilers (External Combustion Sources), Chapter 1.6
- Refuse Combustion (Solid Waste Disposal), Chapter 2.1
- Waferboard/Oriented Strandboard Manufacturing (Wood Products Industry), Chapter 10.6.1
- Hot Mix Asphalt Plants (Mineral Products Industry), Chapter 11.1.

These data sets included supporting emissions data for the following pollutants:

- Particulate matter (PM), including filterable, condensible, and total
- Sulfur dioxide (SO_2)
- Nitrogen oxides (NO_x)
- Carbon monoxide (CO)
- Hazardous air pollutants (HAPs), including acetaldehyde, arsenic, benzene, cadmium, chromium, formaldehyde, hydrogen chloride, lead, mercury, and nickel.

Data analyses have been completed for a total of 44 A-rated data sets. Each of these data sets included at least 15 emissions tests (*n*) to calculate the emissions factor.

Statistical Approach Overview

Several statistical approaches to developing default emissions factors adjustments for using emissions factors in noninventory applications were considered. These approaches are as follows:

- Target boundary statistics of the hypothetical population (is appropriate for noninventory uses of emissions factors)

- Normalized distributions for estimating the confidence interval about the mean of the hypothetical population (i.e., estimate uncertainty about the mean of the population) (is more appropriate for inventory uses of emission factors)

- Bayesian approach to account for uncertainty associated with the unknown portion of the population

- Variability approach that accounts for the uncertainty for three sources of variability, including skewness, the number of tests, and the number of process units.

Primary Statistical Approach

The primary approach selected for developing the emissions factor adjustments is designed to target selected boundary statistics of the population of emissions data. These adjustment factors will enable users to apply the AP-42 emissions factors, which are based on mean values, in applications where other population values (e.g., 90th percentile, 95th percentile, maximum, or minimum values) would be more appropriate. The true emissions from a single source of interest may fall anywhere within the range of emissions data.

Figure 1 presents a flow diagram of the statistical procedure used to calculate the emissions factor adjustments. The statistical analysis includes the following major steps:

1. Conduct data visualization techniques. Perform exploratory data analyses using summary statistics and histograms.

2. Fit probability density functions. Analyze empirical cumulative distribution functions (CDFs). Identify parametric theoretical probability density functions to model the data and estimate the parameters of the density function based on the data. Perform the Kolmogorov-Smirnov (KS) goodness-of-fit test to assess how well the model fits the data. Using maximization approaches implemented in the statistic software Splus® 7.0 for Microsoft Windows, obtain probability density function parameter estimates.

3. Simulate population. For each data set, use Monte Carlo techniques and the parameter estimates obtained in Step 2 to simulate the hypothetical population density of the emissions factor for the specific pollutant. For each simulated hypothetical population, calculate the following statistics: minimum, 1st percentile, 5th percentile, 10th percentile, 15th percentile, 20th percentile, 25th percentile, median, mean, 75th percentile, 80th percentile, 85th percentile, 90th percentile, 95th percentile, 99th percentile, and maximum.

4. For each hypothetical population, select 10,000 random samples of a specified size. Calculate the sample mean for each of the 10,000 samples. Repeat for samples (n = number of tests) of size 1, 3, 5, 10, 15, 20, and 25.

5a. Obtain adjustments. For each distribution of 10,000 means based on n samples, calculate the ratio of the population statistics (obtained in Step 3) and the sample mean. Because the sample mean converges in probability to the population mean, the distribution of this ratio will approach 1 as the sample size increases. The distribution of ratios characterizes the distribution of the adjustment for the emissions factor.

RTI International (RTI) conducted the statistical analysis for each of the 44 A-rated data sets identified from AP-42. For this analysis, RTI assumed that the data available were a representative sample of the population of interest. This is reasonable for A-rated emissions factors. Furthermore, RTI disregarded any precision concerns regarding the difference in number of sample runs comprising each emissions test value used to calculate the emissions factor. Typically, each emissions test is comprised of multiple sample runs (usually a minimum of three sample runs comprises an emissions test).

From the statistical analysis, a number of observations regarding the data sets were made. Each of the data sets was positively skewed. The Weibull, log-normal, and gamma functions were considered (Step 2), and these data sets were either log-normal or Weibull distributions. Figure 2 shows an example histogram (left graph) and empirical CDF (continuous line on right graph) for uncontrolled carbon monoxide data for wood residue combustion in boilers.

Figure 1. Schematic of statistical approach.

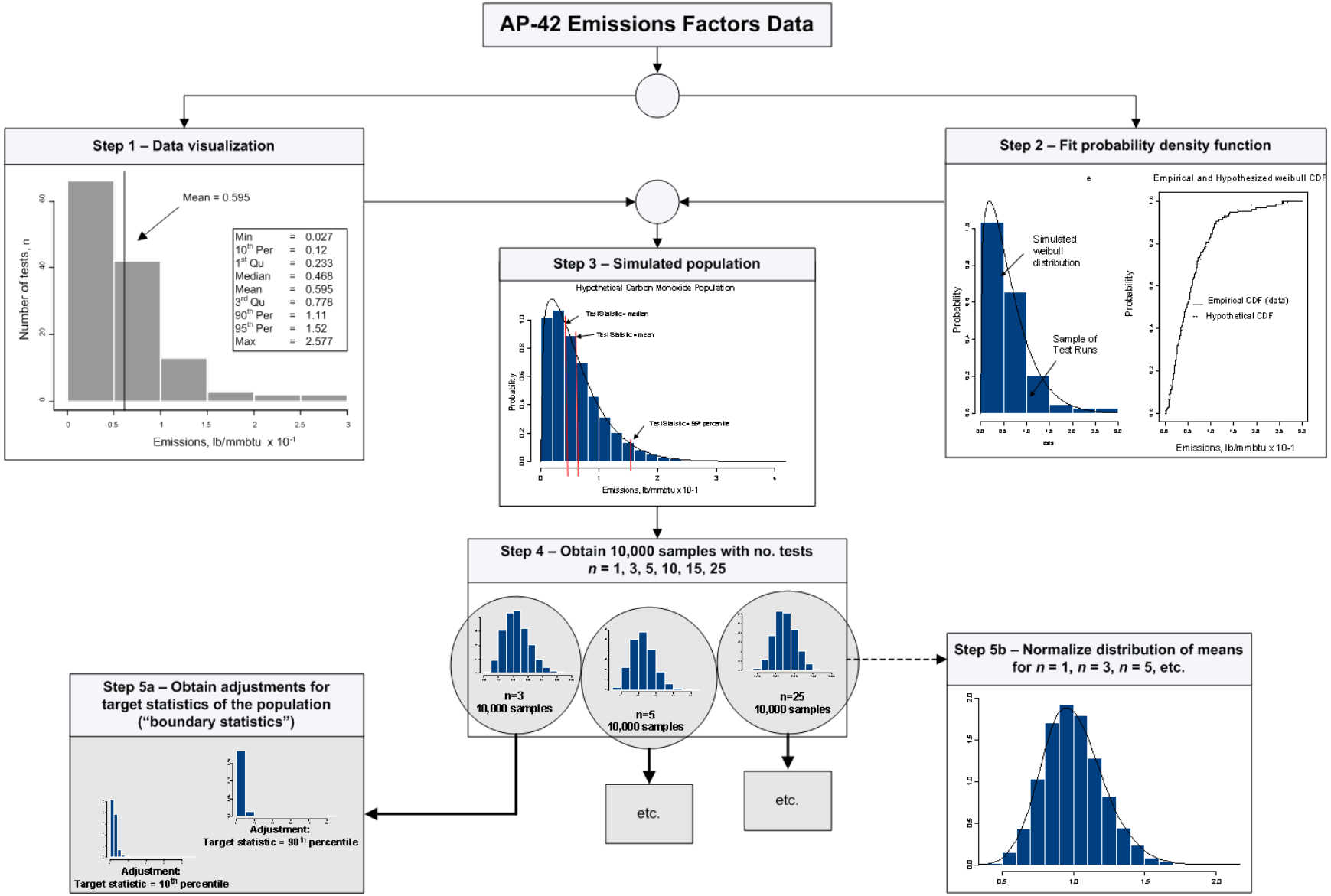
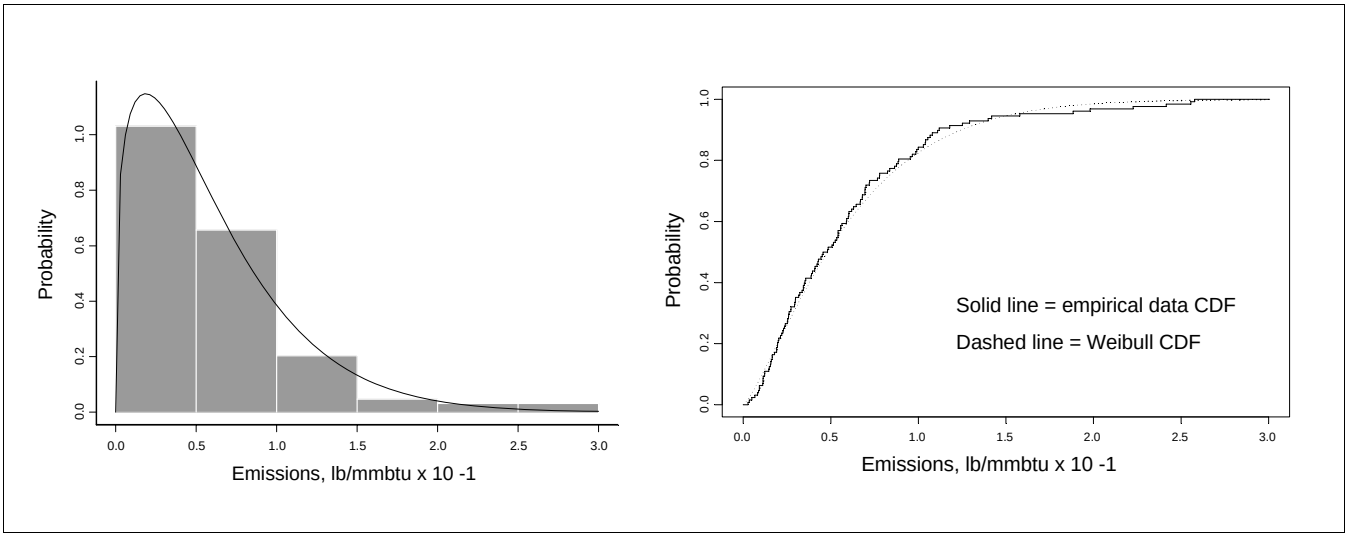


Figure 2. Graphical display of goodness-of-fit of the carbon monoxide data for wood residue combustion in boilers.



For the carbon monoxide example, ten thousand values were randomly drawn from the Weibull distribution with scale parameter 0.64 and shape parameter 1.26; the scale and shape parameters were calculated based on best fit of the data. This collection of 10,000 values will be referred to hereinafter as the hypothetical Weibull population or hypothetical distribution. The following population values (statistics) were calculated from the hypothetical population: minimum, maximum, mean, median, and the 1st, 5th, 10th, 15th, 20th, 25th, 75th, 80th, 85th, 90th, 95th, and 99th percentiles. Figure 2 also shows an example density of the hypothetical Weibull distribution superimposed on the histogram (left graph) for the carbon monoxide data set. An example plot of the hypothetical Weibull CDF (dotted line on the right graph) is also presented in Figure 2 for the carbon monoxide data. The similarity between the two lines observed in the CDF plot suggests a good fit was achieved for the data set.

Monte Carlo simulations refer to the repeated sampling of the hypothetical distribution to make conclusions about the data obtained from the population. Ten thousand samples of sizes 1, 3 to 5, 10, 15, 20, and 25 were randomly drawn from the hypothetical populations using a Monte Carlo approach. For each sample size, the mean was calculated for each sample.

The default adjustments were calculated for several pertinent statistics of interest, including the minimum, maximum, mean, median, and the 1st, 5th, 10th, 15th, 20th, 25th, 75th, 80th, 85th, 90th, 95th, and 99th percentiles of the data distribution. The default adjustments are a function of the number of tests, *n*, on which the emissions factor is based. In general, as the sample size increases, the magnitude of the adjustment decreases. Example results of the default adjustments for uncontrolled carbon monoxide for the wood residue combustion are shown in Table 1.

Referring to the example in Table 1, the Monte Carlo adjustment needed to estimate the 5th and 10th percentiles is about 0.2 and 0.3, respectively, for all number of tests. To target the population median, a factor of 0.9 is needed when the number of tests is greater than 3. To target the 90th percentile, an adjustment of 2 seems appropriate for all sample sizes. To target the 95th percentile, an adjustment of 2.7 is needed for an emissions factor based on 1 test and an adjustment of 2.4 is needed for an emissions factor based on 3 or more tests.

Table 1. Default emissions factor adjustments for carbon monoxide (uncontrolled), wood residue combustion, by number of tests (*n*) and target statistic.

	Target Statistic					
n	**5th Percentile**	**10th Percentile**	**Median**	**Mean**	**90th Percentile**	**95th Percentile**
1	0.19	0.30	1.0	1.2	2.2	2.7
3	0.17	0.27	0.91	1.0	2.0	2.4
5	0.16	0.26	0.89	1.0	1.9	2.3
10	0.16	0.26	0.88	1.0	1.9	2.3
15	0.16	0.26	0.88	1.0	1.9	2.3
20	0.16	0.26	0.88	1.0	1.9	2.3
25	0.16	0.26	0.88	1.0	1.9	2.3

An adjustment for the emissions factor based on the Monte Carlo simulation is defined as the ratio between the target population statistic (minimum, maximum, mean, median, and the 1st, 5th, 10th, 15th, 20th, 25th, 75th, 80th, 85th, 90th, 95th, and 99th percentiles) and the emissions factor, as shown in Equation 1.

$$EF_{target\ statistic} = \left(Adj\right) \times \left(EF\right) \quad (1)$$

where:

$EF_{target\ statistic}$ = Target population value of the emissions distribution, hereafter referred to as the target statistic (e.g., 95th percentile), in units of the AP-42 emissions factor
Adj = Default emissions factor adjustment, unitless
EF = Emissions factor, as presented in AP-42, in units of the AP-42 emissions factor.

Consideration of Alternative Statistical Techniques–Alternative Approach 1

Like the statistical analyses described in the primary approach above, the statistical analyses described in this section were conducted on the AP-42 emissions factor data. The first four steps of the analyses are the same as those described in the primary approach; however, the fifth step differed. The distribution of means obtained in Step 4 for each value of *n* (for *n* = 1 to 30) is known in the statistical literature as the "sampling distribution" of the mean. The spread of the sampling distribution of the mean decreases as the sample size increases, but the mean of the distribution, which approaches the mean of the hypothetical distribution, is not affected by the sample size; therefore, all sampling distributions are centered in the population mean. In Step 5b, all sampling distributions obtained in Step 4 were "normalized." In other words, each value of each sampling distribution was divided by the mean of the corresponding sampling distribution. As a result, all 30 normalized sampling distributions have a mean equal to 1.

As an example, Figure 3 presents the normalized sampling distribution of carbon monoxide, where *n* = 15. Observe that each normalized sampling distribution can be considered as the sampling distribution of the adjustment statistic if the goal is to target the population mean. The sampling distribution shows the probability of observing the different values of this adjustment statistic. With this "sampling distribution," it is possible to predict the 68 percent, 95 percent, and 99 percent confidence intervals for the population mean based on sample size of a specified size *n*. The 95 percent confidence interval around the mean of this sampling distribution has endpoints equal to the 2.5th and 97.5th percentiles of the normalized sampling distribution. The 95 percent confidence intervals are centered

on 1. These endpoints will define the lower and upper values for the adjustment factor if the goal is to target the mean of the hypothetical distribution. As a result, a 95 percent confidence interval for this adjustment factor produces an approximated 95 percent confidence interval for the mean. As an example, Table 2 presents selected sampling distribution percentiles for selected *n*-values for uncontrolled carbon monoxide for wood residue combustion in boilers.

Composite Default Emissions Factor Adjustments

Development of composite default emissions factor adjustments from the adjustment values determined for the individual emissions factor categories was explored. Two approaches to developing composite default adjustments were considered, including

- Clustering (categorizing) the individual adjustments based on the similarity of the distribution of the emissions factor data as measured by some statistical parameters (e.g., adjustments for emissions factors where the data exhibit a similar degree of skewness would be clustered)

- Clustering the individual adjustments based on an engineering/scientific property related to the emissions factor (e.g., similar pollutants [gaseous versus PM], controlled versus uncontrolled emissions, type of process).

The default emissions factor adjustments calculated for each data set were clustered by type of pollutant and control (e.g., controlled versus uncontrolled). Specifically the adjustments were clustered as follows:

- HAP, controlled
- HAP, uncontrolled
- PM-condensible
- PM-filterable, controlled
- PM-filterable, uncontrolled
- Gaseous criteria pollutants.

Composite default emissions factor adjustments were determined for each category of pollutant by calculating the mean value. The composite default adjustment depends on the number of emissions tests, *n*, used to calculate the emissions factor being adjusted. An emissions test consists of multiple sample runs (typically at least three valid sample runs.) Throughout the analyses conducted for this study, adjustments were calculated for the following values of *n*: 1, 3, 5, 10, 15, 20, and 25. The composite default adjustments were calculated for each value of *n*. Examination of the composite default values indicates that for each of the pollutant categories, the adjustment values begin to stabilize when *n* is 10 or greater. Furthermore, the composite adjustments for $n = 5$ and $n = 10$ are similar. Consequently, to further simplify the application of default emissions factor adjustments, we recommend providing fewer adjustment factors, for $n < 3$, $3 \leq n < 10$, $10 \leq n < 25$, and $n \geq 25$ for each pollutant category. Table 3 presents the recommended composite default adjustments, based on boundary statistics, for HAP, controlled; HAP, uncontrolled; PM-condensible; PM-filterable, controlled; PM-filterable, uncontrolled; and gaseous criteria pollutants. Table 4 presents the composite default adjustments, based on normalized sampling distributions (i.e., the Alternative Approach 1), for HAP, controlled; HAP, uncontrolled; PM-condensible; PM-filterable, controlled; PM-filterable, uncontrolled; and gaseous criteria pollutants.

Figure 3. Normalized Monte Carlo sampling distribution of the mean (n = 15) of carbon monoxide for wood residue combustion in boilers.

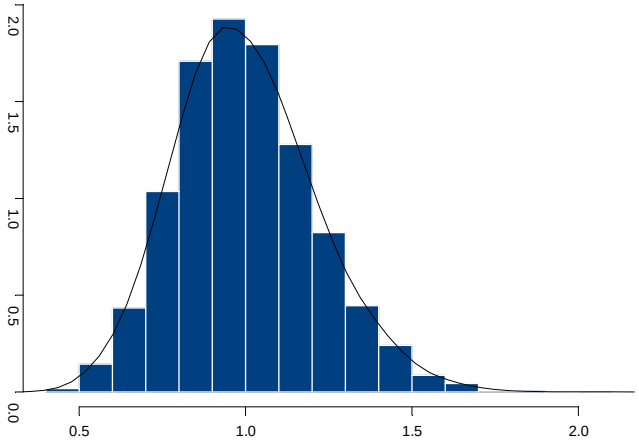


Table 2. Normalized Monte Carlo sampling distribution of emissions factors (means) from the population: wood residue combustion, carbon monoxide, uncontrolled.

EF Sample Size	1st	5th	10th	50th (Median)	90th	95th	99th	Mean
n = 1	0.03	0.10	0.19	0.81	2.04	2.55	3.69	1.00
n = 2	0.13	0.27	0.37	0.90	1.75	2.07	2.72	1.00
n = 3	0.23	0.37	0.47	0.94	1.60	1.86	2.35	1.00
n = 4	0.29	0.45	0.54	0.95	1.54	1.73	2.13	1.00
n = 5	0.35	0.49	0.58	0.96	1.47	1.64	1.98	1.00
n = 6	0.40	0.53	0.61	0.97	1.42	1.57	1.88	1.00
n = 7	0.43	0.56	0.64	0.97	1.40	1.54	1.79	1.00
n = 8	0.46	0.58	0.66	0.97	1.37	1.50	1.74	1.00
n = 9	0.47	0.60	0.68	0.98	1.34	1.46	1.70	1.00
n = 10	0.51	0.62	0.69	0.98	1.33	1.44	1.67	1.00
n = 11	0.52	0.65	0.71	0.98	1.31	1.41	1.63	1.00
n = 12	0.55	0.65	0.72	0.99	1.30	1.40	1.61	1.00
n = 13	0.55	0.66	0.73	0.99	1.28	1.38	1.57	1.00
n = 14	0.57	0.68	0.74	0.99	1.28	1.36	1.56	1.00
n = 15	0.58	0.69	0.75	0.99	1.26	1.35	1.52	1.00
n = 16	0.60	0.70	0.76	0.99	1.26	1.34	1.51	1.00
n = 17	0.60	0.71	0.77	0.99	1.25	1.33	1.50	1.00
n = 18	0.61	0.71	0.77	0.99	1.24	1.32	1.48	1.00
n = 19	0.63	0.71	0.77	0.99	1.24	1.31	1.48	1.00
n = 20	0.63	0.72	0.78	0.99	1.23	1.31	1.46	1.00
n = 21	0.64	0.73	0.78	0.99	1.23	1.30	1.43	1.00
n = 22	0.64	0.74	0.79	0.99	1.22	1.29	1.43	1.00
n = 23	0.66	0.75	0.80	0.99	1.21	1.28	1.42	1.00
n = 24	0.66	0.75	0.80	0.99	1.21	1.28	1.40	1.00
n = 25	0.67	0.76	0.81	0.99	1.21	1.27	1.40	1.00
n = 26	0.66	0.75	0.80	1.00	1.20	1.27	1.39	1.00
n = 27	0.68	0.77	0.81	0.99	1.19	1.26	1.38	1.00
n = 28	0.68	0.76	0.81	0.99	1.19	1.26	1.38	1.00
n = 29	0.69	0.77	0.81	0.99	1.19	1.25	1.38	1.00
n = 30	0.69	0.77	0.82	0.99	1.19	1.25	1.36	1

Table 3. Composite default emissions factor adjustments based on boundary statistics.

Pollutant	Target Statistic	Number of Emissions Tests Used to Determine AP-42 Emissions Factor			
		$n < 3$	$3 \leq n < 10$	$10 \leq n < 25$	$n \geq 25$
HAP, Controlled	10th Percentile	0.3	0.2	0.2	0.2
	25th Percentile	0.5	0.4	0.4	0.3
	Median	1.0	0.8	0.7	0.7
	75th Percentile	1.9	1.5	1.3	1.3
	90th Percentile	3.4	2.6	2.3	2.2
	95th Percentile	4.7	3.6	3.2	3.1
HAP, Uncontrolled	10th Percentile	0.1	0.1	0.1	0.1
	25th Percentile	0.3	0.2	0.2	0.2
	Median	1.0	0.5	0.4	0.4
	75th Percentile	3.4	1.6	1.2	1.0
	90th Percentile	9.8	4.1	2.9	2.6
	95th Percentile	19.1	7.8	5.1	4.5
PM-Condensible	10th Percentile	0.2	0.2	0.2	0.1
	25th Percentile	0.5	0.3	0.3	0.3
	Median	1.0	0.7	0.6	0.6
	75th Percentile	2.2	1.5	1.3	1.2
	90th Percentile	4.4	3	2.5	2.4
	95th Percentile	6.9	4.7	3.9	3.6
PM-Filterable, Controlled	10th Percentile	0.4	0.3	0.3	0.3
	25th Percentile	0.6	0.5	0.5	0.5
	Median	1.0	0.8	0.8	0.8
	75th Percentile	1.7	1.4	1.3	1.2
	90th Percentile	2.8	2.3	2.1	2.0
	95th Percentile	3.9	3.1	2.8	2.7
PM-Filterable, Uncontrolled	10th Percentile	0.5	0.5	0.4	0.4
	25th Percentile	0.7	0.6	0.6	0.6
	Median	1.0	0.9	0.9	0.9
	75th Percentile	1.5	1.3	1.3	1.2
	90th Percentile	2.2	1.9	1.8	1.8
	95th Percentile	2.7	2.3	2.2	2.2
Gaseous Criteria Pollutants	10th Percentile	0.3	0.3	0.3	0.3
	25th Percentile	0.6	0.5	0.5	0.5
	Median	1.0	0.8	0.8	0.8
	75th Percentile	1.9	1.4	1.3	1.2
	90th Percentile	3.5	2.5	2.1	2.0
	95th Percentile	5.4	3.6	3.0	2.8

HAP = hazardous air pollutant
PM = particulate matter

Table 4. Composite default emissions factor adjustments based on normalized sampling distribution of emissions factor (mean).

Pollutant	Distribution Statistic	Number of Emissions Tests Used to Determine AP-42 Emissions Factor			
		$n < 3$	$3 \leq n < 10$	$10 \leq n < 25$	$n \geq 25$
HAP, Controlled	10th Percentile	0.2	0.4	0.6	0.8
	25th Percentile	0.4	0.6	0.8	0.8
	Median	0.7	0.8	0.9	1.0
	Mean	1.0	1.0	1.0	1.0
	75th Percentile	1.2	1.2	1.2	1.1
	90th Percentile	2.1	1.7	1.4	1.3
	95th Percentile	2.9	2.1	1.6	1.4
HAP, Uncontrolled	10th Percentile	0.1	0.2	0.4	0.5
	25th Percentile	0.1	0.3	0.5	0.7
	Median	0.3	0.6	0.8	0.9
	Mean	1.0	1.0	1.0	1.0
	75th Percentile	0.9	1.1	1.2	1.2
	90th Percentile	2.2	2.1	1.8	1.6
	95th Percentile	3.8	3.1	2.3	1.9
PM-Condensible	10th Percentile	0.1	0.3	0.5	0.6
	25th Percentile	0.3	0.4	0.6	0.8
	Median	0.5	0.7	0.9	0.9
	Mean	1.0	1.0	1.0	1.0
	75th Percentile	1.1	1.2	1.2	1.2
	90th Percentile	2.2	1.9	1.6	1.4
	95th Percentile	3.3	2.6	2.0	1.6
PM-Filterable, Controlled	10th Percentile	0.3	0.5	0.6	0.7
	25th Percentile	0.4	0.6	0.8	0.8
	Median	0.7	0.8	0.9	1.0
	Mean	1.0	1.0	1.0	1.0
	75th Percentile	1.2	1.2	1.1	1.1
	90th Percentile	2.0	1.7	1.4	1.3
	95th Percentile	2.8	2.1	1.7	1.4
PM-Filterable, Uncontrolled	10th Percentile	0.5	0.6	0.8	0.8
	25th Percentile	0.6	0.7	0.9	0.9
	Median	0.8	0.9	1.0	1.0
	Mean	1.0	1.0	1.0	1.0
	75th Percentile	1.2	1.2	1.1	1.1
	90th Percentile	1.7	1.5	1.3	1.2
	95th Percentile	2.2	1.7	1.4	1.2
Gaseous Criteria Pollutants	10th Percentile	0.4	0.5	0.7	0.8
	25th Percentile	0.5	0.7	0.8	0.9
	Median	0.8	0.9	0.9	1.0
	Mean	1.0	1.0	1.0	1.0
	75th Percentile	1.2	1.2	1.1	1.1
	90th Percentile	1.8	1.5	1.3	1.2
	95th Percentile	2.3	1.9	1.5	1.3

HAP = hazardous air pollutant
PM = particulate matter

Comparison of the Two Statistical Techniques

Two different statistical analyses to develop default emissions factor adjustments were conducted. The first, or primary, statistical approach targets boundary statistics of the hypothetical population and may be better suited for noninventory uses of emissions factors. The second, or alternative, statistical approach examined estimates of uncertainty about the mean of the population and may be better suited for inventory uses of emission factors. As expected, adjustment factors calculated using the first approach are greater because they target a boundary statistic of the hypothetical population for a single measurement, whereas the second approach calculates an adjustment that provides uncertainty measures (confidence intervals) for the mean. The adjustments based on the two approaches can be considered to represent endpoints of a continuum; the adjustments calculated by the first approach target a boundary statistic for application to a single source, while the adjustments calculated by the second approach target the uncertainty about the mean value of a large sample. Some situations do not fall perfectly into one of these two categories, making it unclear which adjustment is most appropriate. This is particularly true when estimating emissions from a small number of similar sources for noninventory applications, for example, estimating the total emissions from a facility with three similar boilers. As suggested earlier, the default adjustments calculated by the two approaches represent endpoints; consequently, the adjustment values for the situations involving multiple sources should fall between the two endpoints. One approach for addressing the multiple-source situation is to start with the noninventory adjustment factor and apply a correction to reduce the adjustment factor applied because emissions from multiple emissions units are being estimated. A practical, nonstatistical approach based on a linear interpolation of the difference in the adjustments from the two statistical approaches was used to develop correction factors when applying the adjustments to up to 10 sources. For 11 or more sources, the emissions factor adjustment is equivalent to the adjustment determined by the normalized sample distribution about the mean.

Consideration of Alternative Statistical Techniques

An alternative approach designed to account for the uncertainty induced by three sources of variability was reviewed; however, we did not conduct any analyses using this approach. The three sources of variability considered in this approach were (1) the skewness of the distribution of emissions data, (2) the number of tests comprising the emissions factor, and (3) the number of process units for which emissions are being estimated. This analysis was based on hypothetical populations and did not use the actual AP-42 emissions data.

The basic approach is to sample each of the hypothetical populations to develop separate data sets; one data set to replicate emissions factors ("calculated" emissions factor values) and the other to represent actual emissions from the process units. The results were compared to determine how well the calculated emissions factors represent the actual emissions of the process units; the calculated emissions factor value was subtracted from the actual process unit value. The sampling that simulates emissions factor process units was replicated nine times to represent nine different facilities having from one to nine similar process units. The sampling that simulates the development of the calculated emissions factors values was replicated 20 times to represent emissions factors developed from 1 to 20 tests (i.e., n = 1 to 20). All combinations of differences between the 9 sampling distributions representing actual emissions factor process units and the 20 sampling distributions representing calculated emissions factors were calculated. This approach recreates the distribution of the differences of means based on different sample sizes. The adjustment for the emissions factor was defined as (1 + differences). Selected percentiles of the distribution of adjustment produced upper bounds for the mean.

A relevant aspect of this approach is the incorporation of the uncertainty due to the differences between the number of tests the emissions factor was based on and the number of process units. Also, this approach is based on the assumption that it is possible to model all the pollutants using one probability distribution with few varying parameters, which in some sense follows the finding from this project, that the pollutants considered were either Weibull or log-normally distributed. As expected, when the number of process units increases and the number of tests (*n*) used to calculate the emissions factor increases, the difference between the means will tend to zero.

CONCLUSIONS

Default adjustment factors based on boundary (target) statistics of the population have been developed; these emissions factor adjustments are appropriate to use for noninventory applications. Default adjustment factors were developed for numerous target statistics and values of *n*. To simplify presentation of the data and application of the results, default adjustments for selected target statistics are reported. Similarly, adjustments were calculated for numerous values of *n*; to simplify, composite adjustments for selected ranges of *n* are proposed. A more extensive presentation of additional target statistics can be developed for use in an electronic database or lookup table to provide a broader set of options, if needed.

For applications where the target statistic of interest is the *mean*, such as inventories, the appropriate adjustment to use is a selected confidence interval (upper and/or lower confidence limits). Although the primary focus of this study was on non-inventory applications of emissions factors using statistics other than the mean, adjustments based on the normalized sampling distribution of the mean for $n = 1$ to 30 were also calculated as part of the study. Multiplying selected adjustments (percentiles of the normalized distribution) by the AP-42 emissions factor will produce the desired confidence interval for the mean. A more extensive presentation of additional confidence limits about the mean can be presented in an electronic database or lookup table implemented in a user-friendly Java applet to provide a broader set of options, if needed.

The following general conclusions result from the analyses:

- All of the emissions factor data sets examined are either Weibull or log-normally distributed.

- A consistent pattern is shown for all of the pollutants: as the number of tests, *n*, increases, the value of the default emissions factor adjustment decreases. This pattern holds for all of the pollutants, regardless of the number of tests available from the supporting emissions data set or the control status (e.g., controlled versus uncontrolled).

- For each of the pollutant categories, the adjustment values nearly stabilize when *n* is 10 or greater.

- There are some differences from pollutant to pollutant regarding the range of the default emissions factor adjustments as a function of *n*. For some pollutants, regardless of the *n* value, the default adjustment does not change significantly (e.g., PM-filterable, uncontrolled). For other emissions factor data sets, the default adjustment varies more widely depending on the value of *n*.

- The HAP emissions factor data exhibit the highest degree of variability and result in the largest emissions factor adjustments. The majority of the emissions factors for HAP contain at least one data point identified as a problematic outlier.

REFERENCES

1. Neulicht, R., B. Munoz, K. Schaffner, L. Barr, Using Emissions Factors in Noninventory Applications–Options Paper for Default Emissions Factor Adjustments. Prepared for U.S. Environmental Protection Agency by RTI International, Research Triangle Park, NC. February 2006.

2. *Compilation of Air Pollutant Emissions Factors, AP-42, Fifth Edition, Volume I: Stationary Point and Area Sources*. U.S. Environmental Protection Agency, 1995.

KEY WORDS

emissions factor adjustments
adjustments
emissions factors
emissions factor uncertainty