Technical Appendix E

Derivation of Stack Parameter Data

Table of Contents

1 Introduction

In July 1997, EPA's Science Advisory Board (SAB) reviewed and commented on the methodology used in the Risk-Screening Environmental Indicators (RSEI) model developed by EPA. In response to one of SAB's comments, EPA sought to improve the estimate of facility stack height used in modeling air emissions of Toxics Release Inventory (TRI) chemicals. The sensitivity analysis of the air emission modeling used in the RSEI model demonstrated that stack height has the greatest impact on predicted concentrations of air pollutants. At the time of SAB's review, all stacks in the model were assumed to be 10 meters high. Also at that time, all exit gas velocities, which represent the second most important variable impacting air emissions modeling, were assumed to be 0.01 m/sec, and stack diameter was assumed to be 1 m. As EPA began improving the accuracy of stack height estimates, the Agency determined that it could also readily improve the estimation of exit gas velocities and diameters. This Appendix describes the Agency's improvements to the accuracy of the RSEI Model through two types of changes: 1) the incorporation of facility-specific median stack heights and median exit gas velocities where available; and, 2) the estimation of median values for stack heights and exit gas velocities by Standard Industrial Classification (SIC) codes.[1] These SIC estimates are assigned to facilities without facility-specific data.

To obtain facility-specific stack heights and exit gas velocities as well as estimates of stack heights and exit gas velocities by SIC code, the Agency relied on (1) the AIRS Facility Subsystem (AFS) database within the Aerometric Information Retrieval System (AIRS); (2) the National Emission Trends Database (NET); and (3) databases from three states (California, New York, and Wisconsin), as described in Section 2 below. When electric utilities (coal- or oil-burning facilities in SIC codes 4911, 4931, and 4939) were added for Reporting Year 1998, additional work was done to accurately characterize these stacks, as they were presumed to be generally taller than other facilities' stacks. Data were obtained from the Electric Power Research Institute (EPRI) for these facilities, as described in Section 3 below. Beginning with Version 2.1.5, facility-specific data from the National Emissions Inventory (NEI) was added.

The first analysis and construction of a stack height database was performed in early 1999, as fully described in *Estimates of Stack Heights and Exit Gas Velocities for TRI Facilities in OPPT's Risk-Screening Environmental Indicators Model* (June 1999). This Appendix summarizes the information contained in that report, and also presents results from a subsequent analysis performed for Reporting Year 1998 when Electric Utilities were first required to report to TRI. This appendix presents the results from the initial analyses; however, the data extraction

[1] Beginning with RY 2006, TRI-reporting facilities are required to use North American Industry Classification System (NAICS) codes instead of SIC codes. Because data on stack parameters (among other data) is indexed by SIC code, RSEI maintains historically-reported SIC codes for each facility. For facilities reporting for the first time in RY 2006 (and therefore with no historically-reported SIC codes), EPA looks up their reported NAICS codes in the Census Bureau's crosswalk to obtain the appropriate SIC codes.

and processing from AIRS, NET, and NEI has been repeated several times. In years when AIRS and NET data is not extracted and processed, data from the previous year is applied to each facility. New facilities that have not reported in previous years (and hence have no facility-specific data) are assigned the SIC-code level default or the overall default values. Beginning in May 2007, facility-specific data were obtained from EPA's National Emissions Inventory (NEI) database. New data were not obtained from AFS or NET, but the SIC-code level defaults based on previously-pulled AFS/NET were retained. The final section of this appendix summaries the final parameters that were used, based on the initial analyses and the annual updating.

2 Derivation of Primary Stack Data

RSEI uses stack parameter data from several sources. The following sections describe each source and how it is used in the model.

2.1 AFS Overview

AFS originally was a component of AIRS, which is a computerized database management system for airborne pollution administered by EPA's Office of Air Quality Planning and Standards (OAQPS). AFS contains emissions, compliance, and enforcement data on stationary sources of air pollution. Included sources cover the spectrum from large industrial facilities to relatively small operations such as dry cleaners, although facilities must meet certain threshold requirements to be included in AFS. These threshold requirements vary by pollutant.

In general, facilities collect emissions data in compliance with their permits and send the data to their state environmental agencies. Some emissions data are based on actual measurements; others are based on estimation methods. Sometimes inspectors collect emissions data. Most facilities prepare emissions inventories once every five years. Each year, States consolidate the data received from facilities reporting in that year and send it to the EPA Regional Offices, where it is entered into AFS. Since the stack data for RSEI was first obtained, AFS has been superseded by NET and NEI as a repository of emissions data, and is no longer a component of AIRS. Historical data that was obtained previously is still used for facility-specific and SIC-code level defaults, but no new data from AFS was available for RSEI Version 2.1.6.

2.1.1 Pollutants Included in AFS

AFS includes data on a total of 52 specific pollutants or pollutant classes (not counting fugitive emissions, visible emissions, coke oven emissions, fugitive dust, odors, and other). These data include release estimates for the following five air pollutants:

1. particulate matter smaller than ten microns (PM_{10});
2. sulfur oxides, with sulfur dioxide (SO_2) as a marker for all SO_x;
3. nitrogen oxides, with nitrogen dioxide (NO_2) as a marker for all NO_x;

4. carbon monoxide (CO); and
5. lead (Pb).

These are the "criteria" pollutants for which EPA's OAQPS has set National Ambient Air Quality Standards.[2]

Of the 52 pollutant and pollutant classes in AFS, 39 are either TRI chemicals or likely to contain TRI chemicals, presented in Table E-1.

Table E-1
Pollutant and Pollutant Classes in AFS that are TRI Chemicals or Assumed to Include TRI Chemicals

acetylene	cadmium compounds *	lead compounds *
aldehydes	chlorofluorocarbons	manganese compounds *
ammonia	chlorophenols	mercury
antimony compounds *	chromium compounds *	mercury compounds *
aromatics	cobalt compounds *	nickel compounds *
arsenic	copper compounds	olefins
arsenic compounds *	cyanide compounds *	organic acids
asbestos *	Fluorides	polybrominated biphenyls
barium compounds	glycol ethers *	polynuclear aromatics
benzene *	hydrochloric acid *	selenium compounds *
beryllium	hydrofluoric acid *	vinyl chloride *
beryllium compounds *	ketones	VOCs
cadmium	lead	zinc

* Indicates that chemical or chemical class is classified as a Hazardous Air Pollutant (HAP).

2.1.2 Emission and Stack Height Data in AFS

AFS tracks data in a hierarchy with four levels: (1) facilities; (2) stacks, the locations at which emissions are introduced into the atmosphere; (3) points, the processes that produce pollutant emissions; and (4) segments, which are components of the processes. For the criteria pollutants, estimated emissions are available in pounds per year at the facility level. For the HAPs,

[2] Although PM_{10} is the current particulate criterion pollutant, total particulate mass (PT) was the previous criterion for particulates. Depending upon the vintage of a given facility's data, PT may be listed in place of PM_{10}.

emissions may be estimated using "emissions factors" for specific production processes at the segment level. These processes are categorized by Source Classification Codes (SCCs), six-character identifications of the specific production processes.

Each facility in AFS has a primary SIC code, recorded at the four-digit level. The primary SIC code reflects the principal product or service generated by the facility. Within a facility, each stack is assigned a stack identification number. For each stack, the rate of emission in mass per time of each stack pollutant (identified by CAS number or other chemical identification number) is provided, along with the non-zero height of the stack measured in feet.

2.1.3 Analyses of Stack Height Data in AFS

To use facility-specific stack height data in the RSEI model wherever possible, the Agency attempted to identify TRI facilities in AFS for those States that reported to AFS. The match was performed as follows. For the reporting facilities, the AFS database includes an EPA ID, the only facility identifier common to both the TRI and AFS databases. On the TRI Form R, a facility is asked to report up to four EPA IDs associated with the facility. EPA identified TRI forms with non-zero stack releases, obtained all EPA IDs reported by those facilities on their forms, and matched the TRI facilities with the AFS facilities by EPA ID. For the 1995 TRI reporting year, which, at the onset of this analysis, was the most recent year with TRI data available, there are 41,528 Form Rs with non-zero stack releases, submitted by 13,204 facilities. These 13,204 facilities map to 12,106 EPA IDs.[3]

EPA identified 4,813 facilities in AFS that have primary 4-digit SIC codes in the range 2011 through 3999, not including Federal facilities, and that have stacks with non-zero stack height. EPA was able to link the 12,106 TRI EPA IDs to 1,231 AFS EPA IDs, albeit with some overlap, due to some TRI facilities having more than one EPA ID, and other TRI facilities sharing EPA IDs. After completing this analysis, EPA found 1,212 EPA IDs which represent 1,209 unique TRI facilities with non-zero stack heights in common to both AFS and TRI. In other words, about a quarter of the AFS facilities in the SIC code range required to report to TRI and with non-zero stack height can be found in TRI. Only about nine percent of TRI facilities with non-zero stack releases (1,209 of 13,204) are found in AFS with non-zero stack height. The low percent of matches can be explained by the following reasons:

- AFS data are not fully representative of all States;
- AFS reporting thresholds may exceed the threshold for reporting to TRI; and,
- AFS only covers 39 pollutant and pollutant classes that are either TRI chemicals or likely to contain TRI chemicals.

3 Some TRI facilities do not have or do not report an EPA ID; others have more than one EPA ID. It is also possible for one EPA ID to match to more than one TRI ID.

2.1.4 Analysis of Stack Height Data by Chemicals Emitted and SIC Code

After identifying facilities in common to both AFS and TRI, EPA began investigating ways of estimating stack heights for TRI facilities *not* in AFS or in the three State databases. First, EPA identified 37,390 unique stacks in AFS associated with the 4,813 facilities listing their primary facility SIC code in the range 2011 through 3999, not including Federal facilities. The mean height of these stacks is 46.7 feet (14.2 meters). Based on the pollutants recorded in AFS as being emitted from these stacks, the Agency classified each of the 37,390 stacks as either "emitting a possible TRI chemical" or "not emitting a possible TRI chemical." The set of AFS pollutants that are classified as possible TRI chemicals for the purpose of this analysis were shown in Table E-1. It is important to note that the VOCs and other chemical classes may contain more than just TRI chemicals. If at least one pollutant emitted from a stack was considered a possible TRI chemical, then the stack was designated as "emitting a possible TRI chemical". If none of the emitted pollutants were considered possible TRI chemicals, then the stack was designated as "not emitting a possible TRI chemical".

EPA then investigated the possibility that stack height varied by whether the stack emitted possible TRI chemicals or not. If stacks that do not emit possible TRI chemicals have different heights than stacks emitting possible TRI chemicals, then to include stacks that do not emit possible TRI chemicals in further analyses could bias the stack height results. Of the 37,390 stacks present, 16,889 (45.2%) emit pollutants considered as possible TRI chemicals. The remaining 20,501 emit only chemicals that are not considered as possible TRI chemicals from the AFS database. The mean height of those stacks emitting possible TRI chemicals is 46.9 feet (14.3 meters), with a standard deviation of 41.4 feet (12.6 meters). The mean height of the remaining stacks is 46.5 feet (14.2 meters), with a standard deviation of 35.4 feet (10.8 meters). The difference in the mean heights of these two groups of stacks is not statistically significant, as determined by using a Student's t-test to compare the means.[4]

Because the Agency noticed substantial variability in stack height across primary SIC codes of facilities in AFS, consideration was given to estimating stack height as a function of the SIC code of the facility. For 2-digit, 3-digit, and 4-digit SIC codes, EPA evaluated the mean stack heights for the two groups of stacks -- those emitting possible TRI chemicals and those that do not -- by testing the equality of the means using a Student's t-test at the five percent level of significance.[5] For each SIC group, EPA used an F-test to check whether the variances of the two stack groups were different. If the variances were equal, EPA assumed the two groups were drawn from the same population, and a Student's t-test was used to compare the means. If the variances were not equal, EPA assumed the two groups were from two different populations and therefore used a

---

4 The Agency compared means, rather than medians, because the test of means is a more powerful statistical test than the test of medians. The more powerful test is better able to differentiate dissimilar groups.

5 The significance level refers to the probability of rejecting the null hypothesis that the means are equal when actually it should not be rejected; this is the probability of committing a Type I error.

modified Student's t-test, accounting for the unequal variances, to compare the means. At the two-digit SIC code level, 14 SIC code groups indicated significant height differences between the two groups of stacks and six did not. At the 3-digit level, 55 SIC code groups indicated significant height differences between the two groups of stacks and 74 did not. At the four-digit level, 109 SIC groups indicated significant height differences between the two groups of stacks and 303 did not.

2.2 Overview of Stack Height Data in National Emission Trends Database (NET)

EPA's National Emission Trends (NET) database became available to OPPT early in 1998, well after relevant data for the project were obtained from AFS. EPA decided to use stack height data from NET to augment the AFS data because some States not included in AFS were included in NET. The NET database provides information on stack height measured in feet, and the annual emission rates of five criteria pollutants: VOCs, NO_x, CO, SO_2, and PM_{10}. To prevent double-counting of stacks from facilities in both AFS and NET, facilities present in both databases were identified based on the AFS ID.[6] If stack height data for a given AFS ID were present in both databases, the data in AFS were kept for further analyses, and the data in NET were removed from further consideration. The NET database does not include an EPA ID for facilities, and thus specific facilities in common to TRI and NET cannot be identified, nor can the number of facilities in common be estimated.

2.2.1 Analyses of Stack Height Data in NET

As with AFS, EPA evaluated the possibility that stack heights within NET varied by whether the stack emitted possible TRI chemicals. Unlike AFS, NET does not record specific pollutants emitted from each stack. NET does, however, record annual VOC emissions from each stack. EPA identified 90,167 unique stacks in NET associated with 16,682 facilities listing their primary facility SIC code in the range 2011 through 3999, not including Federal facilities. The mean height of these stacks is 49.9 feet (15.2 meters). For the purposes of this analysis, the Agency labeled any stack with non-zero VOC emissions as a stack emitting possible TRI chemicals. Based on this definition, of the 90,167 stacks used in the analysis, 62,245 (69.0%) are classified as emitting possible TRI chemicals. The mean stack height of those stacks emitting possible TRI chemicals is 46.7 feet (14.2 meters), with a standard deviation of 47.8 feet (14.6 meters). The mean height of the remaining stacks is 57.0 feet (17.4 meters), with a standard deviation of 51.0 feet (15.6 meters). The difference in the mean heights of these two groups of stacks is statistically significant, as determined by using a Student's t-test to compare the means'.[7]

---

6 From NET, EPA took the State Federal Information Processing Standard (FIPS) code, county FIPS code, and plant ID and concatenated them to form an identification number equivalent to an AFS ID.

7 Recall that for AFS data, the comparable analysis found that the difference in the mean heights of the two

2.3 Overview of National Emissions Inventory (NEI) Data

The Emission Factor and Inventory Group (EFIG) in the Environmental Protection Agency (EPA) compiles the National Emission Inventory (NEI) for hazardous air pollutants (HAPs) and criteria air pollutants (CAPs). The NEI for HAPs is compiled in order to support air the EPA air toxics programs and to quantify the success of the Clean Air Act (CAA) programs in reducing emissions and human health and environmental risk due to HAPs emissions. Title I, Section 110 of the CAA requires states to submit emission inventories for CAPs as part of their State Implementation Plans. The NEI contains estimates of facility-specific HAP and CAP emissions and their source-specific parameters necessary for modeling such as location and facility characteristics (stack height, exit velocity, temperature, etc.). Because complete source category coverage is needed, the NEI contains estimates of emissions from stationary point and nonpoint and mobile source categories. EPA performs numerous quality assurance checks on the NEI data, and estimates missing data or uses default values.[8]

Beginning with RSEI Version 2.1.5, and repeated for Version 2.1.6, NEI data were only used in cases where TRI facilities could be matched to NEI facilities. First, all NEI facilities with original (i.e., not estimated or default) data were matched to the TRI database of facilities using EPA's FRS ID, which is available for both datasets. Facilities with facility-specific EPRI data were not included in the set of TRI facilities; those data were given precedence over the NEI data. However, facilities with EPRI median parameters were assigned facility-specific NEI parameters if possible. For any other TRI facility that could be matched, the NEI stack parameters were adopted. The SIC-code and overall default parameters were not revised, and any facility that could not be matched to NEI was assigned whatever parameters had been used in the previous version of RSEI.

2.4 State Data

For three states not included in AFS (California, New York, and Wisconsin), EPA was able to obtain facility-specific data on stack heights. For California, 98 facilities matched TRI facilities; for New York, 279 facilities matched TRI facilities; and for Wisconsin, 44 facilities matched TRI facilities. Not all of these facilities contributed stack height data to the analysis, however, as not all facilities reported non-zero stack air releases for 1995. Again, note that although these facilities may also be present in the NET database, they cannot be identified as TRI facilities in NET because NET does not include an EPA ID for facilities.

---

groups of stacks was <u>not</u> statistically significant.

[8] Description taken from conference paper, "Truth or Dare: Data Augmentation in the Point Source 2002 NEI." Ann Pope, Madeline Strum, U.S. EPA and Stephanie Finn, Eastern Research Group, Inc. (no date given). Available at www.epa.gov/ttn/chief/conference/ei13/qaqc/strum.pdf

2.5 Results

2.5.1 Facility-specific Stack Heights

For the 421 California, New York, and Wisconsin facilities and the 1,209 facilities in common to the TRI and AFS databases, a representative stack height for each facility was estimated by calculating the median height for all of a facility's stacks with non-zero height. The median stack height was chosen rather than the mean because stack heights may not be normally distributed. No matter how the stack heights are distributed, the median is the appropriate measure of central tendency. For a facility with symmetrically-distributed stack heights, the median equals the mean. Therefore, for a given facility, the median of its stack heights was used as that facility's stack height in the RSEI model.

2.5.2 Estimated Stack Height by SIC Code

For the remaining TRI facilities with non-zero stack releases for which facility-specific data were not available, stack heights were estimated from AFS and NET based on facility SIC codes. EPA decided that the 3-digit SIC code was the appropriate level at which to analyze and use stack height data. At the 2-digit level, differences between stacks emitting TRI chemicals and stacks not emitting TRI chemicals are often masked because the variance in each population is so large. From a practical standpoint, 2-digit SIC codes represent too gross a level of aggregation for purposes of estimating stack height. At the other extreme, 4-digit SIC codes offer too fine a level of disaggregation; not only might one not expect much difference in stack height between, say, a facility manufacturing creamery butter and a facility manufacturing natural, processed, and imitation cheese, but the number of observations at the 4-digit level are often too few to make a meaningful comparison of the two stack groups. Thus, the remaining TRI facilities were classified into 3-digit SIC code groups by the assigned primary SIC code in the TRI database (i.e., the leading three digits of the first 4-digit SIC code listed). Of the 13,204 TRI facilities reporting non-zero air releases in 1995, 84% reported only one unique 3-digit SIC code; 12% reported two unique 3-digit SIC codes; 3% reported three, 0.8% reported four, and 0.2% reported five.

EPA determined that of the 37,390 stacks being analyzed from AFS and the 90,167 stacks being analyzed from NET, there were 18,967 stacks in common to the two databases. To avoid double-counting these stacks in the analysis, the Agency used the stack height data from AFS for these stacks, and removed the corresponding NET data from further consideration. Augmenting the stacks from AFS with the non-duplicative stacks from NET resulted in a total of 108,590 stacks (37,390 from AFS and 71,200 from NET).

Each TRI facility within a 3-digit SIC code group was assigned the median stack height of the AFS and NET stacks within that 3-digit SIC group according to the following hierarchy:

1. If the combined AFS and NET stack height data for that 3-digit SIC code group indicated no statistically significant difference between the mean height of stacks emitting possible

TRI chemicals and the mean height of stacks emitting non-TRI chemicals, then the median was estimated over all stacks in that group, regardless of whether the stack emitted possible TRI chemicals. This median height was then used as the estimated stack height for all TRI facilities in the 3-digit SIC code group that did not have facility-specific data in AFS or in the three State databases.

2. If the AFS and NET stack height data for that 3-digit SIC code group *did* indicate a statistically significant difference between the mean height of stacks emitting possible TRI chemicals and the mean height of stacks emitting non-TRI chemicals, then the median for *only* those stacks emitting possible TRI chemicals was used as the estimated stack height for all TRI facilities in that 3-digit SIC code group.

In both approaches, the stack heights of facilities that occur in both TRI and AFS (i.e., facility-specific data) are included in the calculation of the median height of their 3-digit SIC code groups. State data are not included in these analyses because of the potential of double-counting with NET data, which includes data from California, New York, and Wisconsin.[9] Table E-2 presents the number of 3-digit SIC codes with median stack heights falling in particular stack height ranges for 139 of the 140 unique 3-digit SIC codes in the range 201 to 399.[10] Note that the majority of SIC codes have median stack heights between 9.0 and 11.9 m; only one SIC code falls into each of the two highest ranges of stack heights.

---

9 Recall that NET facilities cannot be matched to TRI facilities because there is no facility identifier in common.

10 No estimates of stack heights were available for facilities in SIC code 316, luggage manufacturing.

Table E-2
Median Stack Heights by SIC Code

Range of Stack Heights (meters)	Number of 3-Digit SIC Codes with Median Stack Height in Range
6.0 to 6.9 m	7
7.0 to 7.9 m	13
8.0 to 8.9 m	13
9.0 to 9.9 m	37
10.0 to 10.9 m	25
11.0 to 11.9 m	11
12.0 to 12.9 m	14
13.0 to 13.9 m	2
14.0 to 14.9 m	2
15.0 to 15.9 m	3
16.0 to 16.9 m	2
17.0 to 17.9 m	0
18.0 to 18.9 m	2
19.0 to 19.9 m	2
20.0 to 24.9 m	4
25.0 to 29.9 m	1
30.0 to 39.9 m	1
TOTAL: 6.0 to 39.9 m	139

2.5.3 Estimation of Stack Heights for TRI Facilities with Missing or Invalid 3-digit SIC Codes

Of the 13,204 TRI facilities with non-zero stack air releases reported in 1995, stack heights were estimated as described above for 13,021 facilities. The estimation approaches used included: 1,209 facilities estimated directly from AFS; 69 facilities estimated from California State data; 192 facilities estimated from New York State data; 37 facilities estimated from Wisconsin State data; and 11,514 estimated based on the facilities' 3-digit SIC code. The remaining 183 facilities (13,204 facilities minus 13,021 facilities) reported SIC codes outside the range of 201 to 399, at the 3-digit level, or reported no SIC code.[11] For these 183 facilities, a stack height was assigned based on either the 2-digit SIC code (if a valid one was available) or on the median stack height for all 108,590 stacks from AFS and NET. The median stack height for all 108,590 stacks from AFS and NET is 10.67 m (35.0 ft). This median stack height of 10.67 m for *stacks* should not be confused with the median height of 9.14 m for all TRI *facilities*, which is based on AFS, NET, and State data. The median stack height at the 2-digit SIC code level was calculated according to the hierarchy used for the 3-digit SIC code analysis, presented in Section 2.4.2. Stack heights were estimated at the 2-digit SIC code level for 27 facilities. The stack heights for the remaining 156 facilities were estimated using the median stack height of all 108,590 stacks (10.67 m).

2.6 Analyses of Exit Gas Velocities

2.6.1 Facility-specific Exit Gas Velocities

An analysis similar to that performed for stack heights was conducted for exit gas velocities. Exit gas velocity data were available from AFS, NET, NEI and the New York and Wisconsin databases. (Data from California did not include exit gas velocities.) For the 216 New York and Wisconsin facilities and the 850 facilities in common to the TRI and AFS databases with non-zero exit gas velocities, a representative exit gas velocity for each facility was estimated by calculating the median exit gas velocity for all of a facility's stacks with non-zero height and non-zero exit gas velocity. Similar to the methodology used for the median exit gas velocity was chosen rather than the mean because exit gas velocities may not be normally distributed. No matter how the exit gas velocities are distributed, the median is the appropriate measure of central tendency. Therefore, for a given facility, the median of its exit gas velocities was used as that facility's exit gas velocity in the RSEI model. As with the stack height analysis, not all facilities provided by New York and Wisconsin could be matched to TRI facilities with non-zero stack air releases.

2.6.2 Estimated Exit Gas Velocities

For the remaining TRI facilities with non-zero stack releases and non-zero stack heights for which facility-specific data were not available, exit gas velocities were estimated from AFS and NET based on facility 3-digit SIC codes. As previously mentioned, EPA determined that of the 37,390 stacks being analyzed from AFS and the 90,167 stacks being analyzed from NET, there were 18,967 stacks in common to the two databases. To avoid double-counting these stacks in the analysis, the Agency used the exit gas velocity data from AFS for these stacks and removed the exit gas velocity data in NET from further consideration. Therefore, augmenting the stacks

---

[11] As noted previously, not all data provided by California, New York and Wisconsin were useable, because not all facilities reported non-zero stack air releases in 1995.

from AFS with the non-duplicative stacks from NET resulted in 108,590 stacks (37,390 from AFS and 71,200 from NET).

Each TRI facility within a 3-digit SIC code group was assigned the median exit gas velocity of the AFS and NET stacks within that 3-digit SIC group according to the following hierarchy:

1. If the combined AFS and NET stack height data for that 3-digit SIC code group indicated no statistically significant difference between the mean exit gas velocity of stacks emitting possible TRI chemicals and the mean exit gas velocity of stacks emitting non-TRI chemicals, then the median was estimated over all stacks in that group, regardless of whether the stack emitted possible TRI chemicals. This median exit gas velocity was then used as the estimated exit gas velocity for all TRI facilities in the 3-digit SIC code group that did not have facility-specific data in AFS or in the New York and Wisconsin databases.

2. If the AFS and NET exit gas velocity data for that 3-digit SIC code group *did* indicate a statistically significant difference between the mean exit gas velocity of stacks emitting possible TRI chemicals and the mean exit gas velocity of stacks emitting non-TRI chemicals, then the median for *only* those stacks emitting possible TRI chemicals was used as the estimated exit gas velocity for all TRI facilities in that 3-digit SIC code group.

In both approaches, the exit gas velocities of facilities that occur in both TRI and AFS (i.e., facility-specific data) are included in the calculation of the median exit gas velocity of their 3-digit SIC code groups. State data are not included in these analyses because of the potential of double-counting with NET data, which includes data from New York and Wisconsin.[12] Table E-3 presents the number of 3-digit SIC codes with median exit gas velocities falling in a particular exit gas velocity range for 137 of the 140 unique 3-digit SIC codes reported in TRI.[13] Note that for all 3-digit SIC codes in the range of 201 to 399, the median exit gas velocity is greater than or equal to 4.0 m/sec.

---

[12] Recall that NET facilities cannot be matched to TRI facilities because there is no facility identifier in common.

[13] No estimates of exit gas velocities were available for facilities in SIC codes 236 (girls', children's, and infants' outerwear), 316 (luggage manufacturing), and 317 (handbags and other personal leather goods).

Table E-3
Median Exit Gas Velocities by SIC Code

Range of Exit Gas Velocities (m/sec)	Number of 3-Digit SIC Codes with Median Exit Gas Velocity in Range
4.0 to 4.9 m/sec	3
5.0 to 5.9 m/sec	4
6.0 to 6.9 m/sec	4
7.0 to 7.9 m/sec	12
8.0 to 8.9 m/sec	44
9.0 to 9.9 m/sec	26
10.0 to 10.9 m/sec	26
11.0 to 11.9 m/sec	8
12.0 to 12.9 m/sec	7
13.0 to 13.9 m/sec	1
14.0 to 14.9 m/sec	2
TOTAL:	137

2.6.3 Estimation of Exit Gas Velocities for TRI Facilities with Missing or Invalid 3-digit SIC Codes

Of the 13,204 TRI facilities with non-zero stack air releases reported in 1995, exit gas velocities were estimated for 13,016 facilities. The estimation approaches used included: 850 facilities estimated directly from AFS; 192 facilities estimated from New York State data; 24 facilities estimated from Wisconsin State data; and 11,950 estimated based on the facilities' 3-digit SIC code. The remaining 188 facilities (13,204 facilities minus 13,016 facilities) reported SIC codes outside the range of 201 to 399, at the 3-digit level, or reported no SIC code. For these facilities, an exit gas velocity was assigned based on either the 2-digit SIC code (if a valid one was available) or on the median exit gas velocity for all 108,590 stacks. The median exit gas velocity for all 108,590 stacks from AFS and NET is 8.80 m/sec (28.9 ft/sec). This median exit gas velocity of 8.80 m/sec for *stacks* should not be confused with the median exit gas velocity of 8.90 m/sec for all TRI *facilities*. The median exit gas velocity at the 2-digit SIC code level was calculated according to the hierarchy used for the 3-digit SIC code analysis. Two significant figures are used for all exit gas velocities in the RSEI model.

3 Derivation of Stack Data for Electric Utilities

This section presents the method by which stack parameters for electric utilities were estimated from data provided by the Electric Power Research Institute (EPRI). Reporting Year 1998 was the first year that the electric utilities have been included in the TRI inventory and the process is repeated annually with each new release of TRI data. Since electric utilities have inherently different characteristics from other TRI facilities and may significantly contribute to risk estimates, it is important be as accurate as possible in representing the parameters for these facilities. A sensitivity analysis of RSEI's air modeling has demonstrated that stack height has the greatest impact on predicted concentrations of air pollutants, so special attention has been paid to this parameter.

EPA received two electric utility stack data files from EPRI:

1. Stk599.xls- containing information on *all* of the electric utilities selling electricity; and
2. Corrected final stack file.xls (cfsf)- containing more recent information on all of the *coal-fired* electric utilities.

The two EPRI files were combined, and as many facilities as possible were matched to the TRI facility database. For TRI facilities in the electric utilities SIC codes that could not be matched to a specific facility in the EPRI database, median parameters of all the relevant unmatched facilities in the EPRI database were assigned. These steps are described in more detail in the sections below.

3.1 Combining the EPRI Files

Each EPRI file contains stack parameter data, including height, diameter, velocity, chemical emitted, temperature, and flow, as well as facility data including plant name, owner name, and latitude/longitude. In these files, unique records are comprised of unique plant-boiler-stack combinations (similar to AFS); consequently, there are many records for each facility. The original file contains 3,275 records, and the corrected file of just coal-fired utilities contains 869 records.

First, stacks with zero chemical emissions were eliminated from both datasets. Since TRI only requires coal- or oil-fired utilities to report, facilities that used only gas were also eliminated from stk599.xls (gas-fired utilities were not included in the second file at all). The two files were combined, with data from the second file used whenever there was valid data on the same facility in both files. This resulted in a dataset of almost 1200 records, consisting of 575 unique facilities.

3.2 Matching the EPRI Dataset to TRI Facilities

The set of TRI facilities was comprised of those facilities in the 1998 data with the first listed SIC code of 4911, 4931, or 4939. This resulted in a set of 604 electric utilities.

Because there was no unique identifier between the two datasets, the matches were performed by considering plant name, state and latitude/longitude. Much of the matching was done by hand. Ultimately, 414 facilities were matched. For these facilities, the median value of each facility's stacks for stack height, diameter, and velocity were entered into the model's facility database. These facilities can be identified by the source code 'EPRI fac' (meaning facility-specific).

After the match was done, there were 161 TRI facilities that could not be matched to specific facilities in the EPRI dataset. For these facilities, median values of all stacks from the unmatched EPRI facilities were used. One facility in this group already had facility-specific data from AFS, so those stack parameters were retained. Table E-4 shows the results of this exercise. The numbers in the last column in bold under 'Median EPRI data of all stacks' show the median values that are used as the default in the RSEI model for unmatched electric utilities.

Table E-4
Results of facility matching between EPRI and TRI Datasets

Stack parameter	**All Stk599 data**	**Matched Stk599 and TRI facilities**	**Unmatched Stk599 and TRI facilities**
Number of stack-boiler pathways	2,869	2,309	560
Number of facilities	575	414	161
Average number of stack-boiler pathways per facility	5	6	3
Median EPRI data of all stacks			
Median stack height (m)	117	128	59
Median stack diameter (m)	5	5	4
Median stack velocity (mps)	23	23	17
Median stack temperature (°F)	290	290	288
Median stack flow (cmps)	26,272	29,934	12,955

4 Final Stack Parameters for Version 2.3.2

The following table shows the final sources for facility stack data used in RSEI Version 2.3.2.

Table E-5
Sources for Stack Parameters Used in RSEI Version 2.3.2, RY 2011

Source	Stack Height		Stack Velocity		Stack Diameter	
	# facilities	% of total	# facilities	% of total	# facilities	% of total
Facility-Specific Data from AFS, NET or NEI	15,879	28%	15,589	28%	15,873	28%
Facility-Specific Data from EPRI	424	1%	423	1%	424	1%
Facility-Specific Data from State Database	153	<1%	125	<1%	-	-
Median Value from EPRI	192	<1%	192	<1%	192	<1%
Median Value from 3-Digit SIC Code (from AFS/NET)	36,584	65%	36,903	66%	36,743	66%
Overall Median Value	2,818	5%	2,818	5%	2,818	5%

5 References

Bouwes, Sr., N.W., and S.M. Hassur. 1998. Ground-Truthing of the Air Pathway Component of OPPT's Risk-Screening Environmental Indicators Model. U.S. EPA, Office of Pollution Prevention and Toxics, Economics, Exposure, and Technology Division. October. Draft.