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 meter/second, and stack diameter was assumed to be 1 meter. 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 stack diameters.

When 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 2.2 below.

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, published as Part C of [Analyses Performed for the Risk-Screening Environmental Indicators]). For several years, RSEI stack data combined data from the AIRS Facility Subsystem (AFS) database within the Aerometric Information Retrieval System (AIRS); (2) the National Emission Trends Database (NET); (3) EPRI; and (4) State-level databases from California, New York and Wisconsin. See Technical Appendix E from RSEI Version 2.3.4 for details. With the Version 2.3.5 update, the methodology was modified to rely solely on the National Emissions Inventory (NEI) and EPRI, and new data from the 2011 NEI were pulled and processed. That methodology is described in this Appendix.

2   General Derivation of Stack Parameter Values

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

2.1   National Emissions Inventory (NEI) Data

EPA's Emission Factor and Inventory Group (EFIG) 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 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 in some cases.

In 2011, for the first time, NEI required stack parameter data (stack height, velocity, and diameter, among other parameters) to be reported by all sources, as opposed to previous NEI versions for which where parameter reporting was not required, and default values were used in many cases to fill in unreported parameter values. Beginning with RSEI Version 2.3.5, RSEI uses NEI (2011) as the main source of stack data. The sections below describe the method by which facility-specific, industry-level, and overall parameter values were calculated.

2.1.1 Facility-specific parameter values from NEI

First, a database containing "eligible" stacks was constructed. NEI contains reporting information for criteria air pollutants like particulate matter and ozone, as well as TRI-reportable chemicals. To estimate stack parameter values for RSEI, EPA included only those stacks that were reported to release TRI-reportable chemicals, using a chemical crosswalk between NEI and TRI chemicals and chemical categories. Because a substantial number of stacks released trace amounts of TRI chemicals with large amounts of criteria pollutants, stacks were dropped if they released less than 5% of the total TRI-reportable chemicals reported to NEI for the facility.

EPA matched facilities between TRI and NEI using EPA's Facility Registry Service (FRS) linkages, which maintain the relationships between the various program identifiers at EPA (in this case, the TRI identifier and the NEI identifier). EPA then used the database of eligible stacks to calculate the median stack parameter values across eligible stacks for each facility.

2.1.2 Industry and overall median parameter values from NEI

All TRI facilities did not match into the NEI database for various reasons including 1) TRI covers facilities reporting from 1988-2015, while the 2011 NEI only covers facilities reporting in 2011; 2) EPA FRS linkages may be incomplete; 3) EPA attempts to match all TRI reporting facilities, although some may not report stack air releases; and 4) TRI and NEI reporting requirements differ.

For facilities that could not be matched, EPA used the constructed database of eligible stacks to calculated median parameter values for each 4-digit NAICS code reported by NEI facilities. Only 4-digit NAICS codes with five or more facilities reporting to NEI with eligible stacks were included in these calculations. The median parameter values across all eligible stacks were also calculated, and used as the overall default values assigned to a TRI-reporting facility if the

facility could not be matched and no valid NAICS-level stack parameters were available.

2.2 Electric Power Research Institute (EPRI)

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.

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.

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 NEI); 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.

2.2.1 Matching the EPRI Dataset to TRI Facilities

The EPRI facilities were originally matched to the TRI dataset in 1998, the first year of utility reporting. In that year, 604 facilities reported under utility SIC codes 4911, 4931, or 4939 (SIC codes were used at the time instead of NAICS codes). 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 in the RSEI "Facility" table by the source code 'EPRI fac' (meaning facility-specific).

After the match was done, some TRI facilities with in utility SIC codes 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. Table E-1 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-1
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

3 Results for Current RSEI Version (2.3.6)

The stack parameters were assigned to TRI reporting facilities based on the following hierarchy:
1. EPRI stack parameter values, based on facility match (electric utilities only).
2. NEI 2011 stack parameter values, based on facility match.
3. NEI 2005 stack parameter values, based on facility match.
4. EPRI stack parameter values, based on default median value for unmatched facilities (electric utilities only).
5. NEI 2011 stack parameter values, based on 4-digit NAICS match.
6. NEI 2011 stack parameter values, based on overall default value.

Table E-2 shows the stack parameter sources for the facilities reporting to TRI. Note that stack heights are assigned to all facilities, even if they currently have no reported stack releases.

Table E-2
Sources for Stack Parameters Used in RSEI Version 2.3.6,
RY 201

Stack Parameter Source	Number of Facilities
NEI 2011 4-Digit NAICS	40,333
NEI 2011 Facility Specific	9,825
NEI 2005 Facility Specific	8,590
NEI 2011 Overall Median Values	520
EPRI Facility Specific	419
EPRI Median Values	131

[revised 12/21/2017]