COMPILING THE 2002 POINT SOURCE NEI: WHOSE DATA GET PRIORITY?

Anne Pope
Emission Inventory Group, U.S. Environmental Protection Agency, Mail Drop D205, Research Triangle Park, NC 27711

Stephanie Finn and Darcy Wilson
Eastern Research Group, Inc., 1600 Perimeter Park, Morrisville, NC 27560

ABSTRACT

The 2002 point source National Emissions Inventory (NEI) for hazardous and criteria air pollutants (HAPs and CAPs) is composed of state, local, and tribal agency NEI Input Format (NIF) submittals, and data from the U.S. Environmental Protection Agency (EPA) Emission Standards Division (ESD), Clean Air Markets Division (CAMD), and Toxic Release Inventory (TRI). The data are compiled from these multiple sources to develop as complete an inventory as possible. This means, however, that there are duplicative estimates from one or more of these sources. The first step in the process of identifying duplicative estimates from different data sources is to accurately match the facilities and assign a common NEI Unique Facility ID. In the 2002 point source NEI Output Format (NOF), the field with this common Facility ID is the strNTISite ID (aka NEI Unique Facility ID). The Site ID is retained from the original data source (a facility can have multiple sites in the 2002 point source NEI). Once common facilities are identified, the data sources are selected based on the reported pollutants and priorities shown below:

1) Preferred ESD and CAMD data;
2) Tribal agency data;
3) Local agency data;
4) State data;
5) Other ESD-Maximum Available Control Technology data (non-preferred) and other (non-preferred) CAMD data; and
6) TRI data.

This paper discusses how this data priority scheme was implemented for the 2002 point source NEI. In the first round, two data selection passes are made for HAPs because overlapping compounds could be retained if different, individual pollutant codes are reported for the same pollutant category (e.g., mercury and compounds, lead and compounds, etc). For CAPs, one selection pass is made at the pollutant category level. A second seletion round deals only with Electric Generating Units (EGUs).

State, local and tribal agency reviewers should understand that the presence of two sites with the same name and address does not mean that estimates have been duplicated. Reviewers should look for duplicate sites that do NOT have a common NEI Unique Facility ID. When sites do share a common NEI Unique Facility ID, reviewers will want to focus on the estimates brought in from data sources other than their own during the data selection process.

INTRODUCTION

Emission inventories are critical for the efforts of state, local, and federal agencies to attain and maintain National Ambient Air Quality Standards that EPA has established for CAPs. Title 1, Section 110 of the Clean Air Act (CAA) requires states to submit emission inventories for CAPs as part of their State Implementation Plans (SIPs). The 1990 CAA Amendments established new periodic emission inventory preparation requirements for CAPs. In June 2002, the EPA promulgated the Consolidated Emissions Reporting Rule (CERR) to simplify reporting, offer options for data collection and exchange, and unify reporting dates for CAPs by state and local agencies and tribes. Using CAP emission inventory data reported by state and local agencies and tribes, the EPA compiles the National Emissions Inventory (NEI) for CAPs. The NEI for CAPs includes point, nonpoint, and mobile source estimates of CAP emissions. The NEI for CAPs is used in modeling to analyze potential regulations.

Title I, Section 112 of the CAA Amendments of 1990 requires that the EPA promulgate standards that require Maximum Achievable Control Technology (MACT) for sources emitting hazardous air pollutants (HAPs). In order to determine if the MACT program and other CAA programs are successful in reducing emissions and human health and environmental risk due to HAPs emissions, EPA compiles the NEI for HAPs. The NEI for HAPs was formerly known as the National Toxics Inventory (NTI). The NEI for HAPs includes point major and area, nonpoint area and other, and mobile source estimates of emissions. This requires national surveys of stationary major and area source facilities including MACT source categories emitting HAPs and an estimate of emissions associated therewith. Compiled every three years, the 1990 NTI, 1996 NTI and 1999 NEI for HAPs are currently available.

The EPA's Emission Inventory Group (EIG) is currently developing the 2002 NEI. For the first time, the goal is to compile a merged NEI for CAPs and HAPs. A number of steps are involved in the development of the point source NEI including the blending and merging of data from different sources to yield an integrated CAP and HAP point source inventory.

State and local agencies and tribes were asked to supply emission inventory data to the EPA. Inventory data and facility lists were also requested from ESD for MACT and Section 112(k) Area Source Standards categories. EIG also collected emission inventory data for EGUs from the Department of Energy's (DOE) Energy Information Agency (EIA) and CAMD.

To develop a complete point source NEI for HAPs, Toxics Release Inventory (TRI) data were also used. The purpose of appending TRI data to the tribal-, local-, state-, and ESD-combined databases was to make sure all emissions data for facilities that report to TRI are included in the NEI.[1] The TRI makes a significant contribution to the NEI for HAPs as more than 9,000 facilities in the HAP inventory only have TRI-provided estimates. This constitutes more than 10% of all the facilities in the inventory. In addition to the sources of data discussed above, the NEI does contain some data from the 1999 NEI.

This paper summarizes EPA's data selection methodology when more than one data source supplied estimates for a facility and discusses the complexities and issues that arise from data selection. It also presents steps state, regional, and tribal data reviewers can take to assess the data selection process.

Blend-Merging (aka Data Selection)

Because the NEI is composed of databases submitted from multiple sources, there can be overlapping estimates from one or more of these sources. The NEI blend-merge or data selection process attempts to eliminate duplication. It is important to note, however, that no estimate is actually deleted from EPA's "master" inventory. Estimates deemed as duplicative are simply "unselected" and do not appear in any of the draft output or summary files. This method allows EPA to track competing estimates and refine its merging or data selection routine over time using different rules of selection.

Prior to any blend-merging, EPA must first match the facilities from the multiple data sources and assign common IDs to facilities found in one or more datasets. In preparation for the compilation of the integrated 2002 NEI, EIG created a crosswalk of NEI HAP and CAP facilities from the 1999 NEI. EIG built this crosswalk by first matching HAP and CAP facilities to one another and assigning unique identifiers to every facility in this crosswalk—the NEI Unique Facility ID (strNTISiteID in tblPointSI). (The crosswalk is posted at: ftp://ftp.epa.gov/EmisInventory/draftnei2002/point/summaries/ and is labeled "2002 NEI Facility File.").[2] Facilities found in both the HAP and CAP inventories share the same NEI Unique Facility ID. When data providers use different Site IDs (strStateFacilityIdentifier in tblPointSI) for their CAP and HAP inventories, the different Site IDs are retained. The common NEI Unique Facility ID indicates that sites are at the same facility. Different site records supplied by different data sources, also share a common NEI Unique Facility ID when the plants are deemed to be the same (Table 1). Local identifiers are retained in this case as well in both the crosswalk and tblPointSI.

When state, local and tribal data submittals were received in June 2004, EPA compared facilities from these submittals to the crosswalk. When there was a name or local identifier match between the new data set and the crosswalk, EPA verified that other information such as state, county, address, zip code, TRI ID (or other type of ID), and latitude/longitude coordinates were the same or very similar. If so, both of the sites received the corresponding NEI Unique Facility ID. Facilities not found in the crosswalk were assigned a new NEI Unique Facility ID. More details on the NEI facility matching process can be found in the paper "2002 NEI for Point Sources: Integration of HAPs and CAPs".[3] After NEI Unique Facility IDs were assigned, data selection took place.

Table 1. Assignment of common NEI Unique Facility ID to site records submitted by multiple data sources

State County FIPs	Tribal Code	State/Local ID	NEI Unique Facility ID	Facility Name	Data Source
42003	000	4200300157	NEI13183	Orion Power Midwest, Cheswick Station	State
42003	000	EGU0978	NEI13183	Cheswick Power Plant	CAMD
42003	000	T$1432	NEI13183	Cheswick Power Plant	TRI

Data selection is divided into two distinct "Rounds," each with their own "selection passes." In summary, the process is divided as follows:

- Round One - HAP and non-EGU CAP
  - Selection Pass 1 - By pollutant code
  - Selection Pass 2 - By pollutant category

- Round Two - EGUs CAP and HAP
  - Selection Pass 1 - NO_X and SO_2
  - Selection Pass 2 - All other pollutants

The selection rounds and individual selection passes are described below.

Round One

In the first round for HAP emission estimates, the selection routine looks only at state (S), local (L), regional (R), tribal (B), TRI (T) and ESD (M or P) data and ranks the estimates. The highest ranked estimate is selected from among the duplicates in the specified grouping. Selection passes are made at two grouping levels for HAPs:

- Facility (NEI Unique Facility ID), pollutant code, data source (ranked).

- Facility (NEI Unique Facility ID), HAP category (ranked), data source (ranked).

The ranking of the data sources from highest to lowest is as follows:

Code	Definition
P	"Preferred" MACT
B	Tribal Data
L	Local Agency Data
S	State Data
M	MACT Data
T	TRI

MACT categories given the highest ranking ("P" codes) include: Small and Large Municipal Waste Combustors (MWCs), Medical Waste Incinerators, Brick and Structural Clay Products Manufacturing, and Coke Ovens. These data were supplied by ESD engineers and are preferred because they are based on extensive source test data, and/or a very complete inventory for the category.

The results of both passes are evaluated, and a final selection decision is made. Two passes are necessary, because if the selection is confined to specific pollutant codes or Chemical Abstract Service (CAS) numbers, then pollutants with different pollutant codes, where one is reported by CAS number and the other by HAP category, could be retained and result in double counting. For example, pollutant code 195 (lead and compounds) will not appear to duplicate pollutant code 133527 (lead oxide), and both pollutants will get through the pollutant-specific selection pass (Table 2). However, in the second pass, the less specific "lead and compounds" is de-selected since there is already a pollutant in this grouping from a higher-ranked data source. The result of both passes is that "lead and compounds" is de-selected from the final output.

The second selection pass, therefore, looks for duplication at the category level at a facility, so that only

one data source for the pollutant group is selected. Both pollutant specific and HAP category selection passes are necessary, since the HAP category pass (pass 2) would de-select specific pollutants in the same HAP category (e.g., chromium III vs. chromium VI). Note that, individual pollutants for the same compound group from the same data provider for the same facility, never compete against one another and all are retained.

For CAPS, only the second selection pass for round one was conducted. This approach was taken to avoid blending particulate matter (PM) data from different sources (because of the relationship among the PM fractions). Specifically, there are many fractions of PM (PM-PRI, PM-FIL, PM-CON, PM2.5-PRI, etc.) and for this selection process they are all categorized as PM.

Table 2. Round One - HAP data selection. Selection passes one and two

NEI Unique Facility ID	Pollutant Code	HAP Name	HAP Category	Emissions (TPY)	Pass 1	Pass 2	Result: Pass 1 and Pass 2	Data Source
NEI18334	195	Lead and Compounds	Lead and Compounds	.125	select	de-elect	**de-select**	M
NEI18334	1335257	Lead Oxide	Lead and Compounds	.0395	select	select	**select**	S
NEI10585	18540299	Chromium (VI)	Chromium and Compounds	.005	select	select	**select**	S
NEI10585	16065831	Chromium III	Chromium and Compounds	.04	select	de-elect	**select**	M

Round Two

In the second round of data selection, EGU data are added to the selection process. "De-selected" records from Round One are not entered into Round Two. The 2002 EGU data are a combination of data from CAMD, DOE[4], and EPA-approved emission factors. For details on how the EGU emission estimates were developed, see "Documentation for the 2002 Electric Generating Unit (EGU) National Emissions Inventory (NEI)".[5] The EPA's 2002 EGU emission estimates are flagged in the emission record with one of the following codes:

767/CAMD: Record in both 2002 Form EIA-767 and 2002 Emission tracking system/continuous emissions monitoring (ETS/CEM); with SO_2, NO_x, and heat input values from ETS/CEM; condensible PM, and primary PM_{10} and $PM_{2.5}$ are recalculated using CAMD heat input values.

CAMD: Record only in 2002 ETS/CEM for SO_2, NO_x, and heat input values; other emissions estimated.

767/CAMD1: Combined cycle record (heat recovery steam generators (HRSG) + combustion turbine) in 2002 ETS/CEM used; the HRSG record in 2002 Form EIA-767 eliminated. SO_2, NO_x, and heat input values from ETS/CEM; other emissions estimated.

767/CAMD2: Combined cycle record (steam turbine + combustion turbine) in 2002 ETS/CEM used; the steam record in 2002 Form EIA-767 eliminated. SO_2, NO_x, and heat input values from ETS/CEM; other emissions estimated.

The ETS/CEM data files contain heat input and emissions that are generally based on monitoring data, which is preferable to calculating values using EIA-767 data and *AP-42* emission factors. EPA requires that all coal units have continuous emission monitors (CEMs) to report hourly data. Oil and gas units in general may, but are not required to, have CEMs; for additional information about EPA 's requirements, see http://www.epa.gov/airmarket/monitoring/factsheet.html.[6]

Because the NO_x and SO_2 ETS/CEM emissions data are deemed "preferred" by EPA, these data were ranked higher in the data selection process than state, local, or tribal emissions estimates.

In Round Two, for the matched sites that have been assigned the same NEI Unique Facility ID (hence, presumably they are the same facility), these rules guide data selection:

- EPA's EGU NO_x and SO_2 data were retained over state/local/tribal NO_x and SO_2 data for electric generation, industrial, and commercial-institutional fuel combustion (SCCs of 101*, 102*, 201*, 202*, 103*, 203*);

- For other HAPs and CAPs at the these facilities, the state/local/tribal data were retained as preferred;

- EPA's EGU HAP and other CAP (not NO_x or SO_2) data were retained only if not represented at the state/local/tribal facility;

- Any data at the state, local, tribal facility that is at a non-boiler SCC are retained;

- Criteria pollutants are handled on a category basis only. This approach was taken to avoid blending PM data from different sources; and

- All sources of data were given priority over TRI data.

The selection process may have excluded NO_x and SO_2 fuel combustion state/local/tribal emission records that did not duplicate the ETS/CEM emissions data. This is due to the fact that state-based industrial, and commercial-institutional fuel combustion units (SCCs 102*, 103*, 202* and 203*) "competed" against EGU records which may not have been actual industrial, and commercial-institutional combustion units. As these SCCs were never applied to the EGU database, it is not clear which portion of the CAMD/DOE-provided data are actually industrial and commercial-institutional fuel combustion units (versus electric generating units). Thus, rather than assuming no overlap, we assumed that the CAMD/DOE database contained all combustion units.

It is important to note that the data merging or selection of EGU data is not conducted by matching individual units or SCCs and is not a true "match and merge" exercise. Rather, this data selection process results in NEI facilities (with a common NEI Unique Facility ID) containing a mixture of combustion units from different data sources, each with their own set of non-overlapping pollutants. The facility's NO_x and SO_2 estimates are derived from CAMD data, while all other pollutants are sourced from the state or local agency (or from EPA if no state or local data). In order to obtain a complete set of emissions for an EGU facility, the user must find the NEI Unique Facility ID (stored in "strNTISiteID" in the NEI Output (NOF) files) and query for all estimates associated with this ID.

Table 3 below provides the results of the selection procedure for an EGU facility with multiple data

sources that involve fuel combustion and industrial processes and both HAPs and CAPs.

Table 3. Example of EGU data selection process

Facility	SCC	Pollutant	EPA EGU Emissions (tons)	S/L/B Emissions (tons)	EPA EGU Data	S/L/B Data
NEI01	101*	NO_x	100	150	select	de-select
NEI01	101*	SO_2	200	203	select	de-select
NEI01	101*	VOC	50	55	de-select	select
NEI01	101*	PM	5	*NULL*	select	N/A
NEI01	101*	Pyrene	0.002	0.003	de-select	select
NEI01	101*	Acenaphthene	0.005	NULL	select	N/A
NEI01	301*	Cadmium	*NULL*	0.0006	N/A	select
NEI01	301*	NO_x	*NULL*	20	N/A	select
NEI01	301*	PM	*NULL*	20	N/A	select
NEI01	301*	SO_2	*NULL*	25	N/A	select

[a] S/L/B = state, local, tribal

What are Some of the Implications and Issues arising from Data Selection?

Data selection or blending of data from different sources has inherent uncertainties and contributes to a more complete inventory which might be more accurate. However, data selection in the NEI leads to other complexities. One of the biggest issues that confronts the reviewer of the draft inventory is that multiple identifiers have been retained for the same site, unit, process, and stack. While ultimately, emissions are not duplicated, estimates for the same site/unit/process or stack may be associated with different IDs which gives the appearance of duplication. However, this system has benefits, especially during the review period:

- Users can readily identify where each data point came from because local identifiers have been retained in the strStateFacilityIdentifier field. Since emissions from one data source have not been artificially allocated or assigned to units or stacks from another data source, the reviewer can check the new data that have been added to the state/local or tribal data set and indicate agreement or disagreement with this supplementation. Records can be readily marked for deletion or revision.

- Comparisons with earlier datasets are easier when local identifiers are not changed. The history of each estimate is readily discernible and linking on common identifiers through multiple iterations of the inventory is easier. The inventory is more *transparent*. According to the EPA's Information Quality Guidelines (IQG), transparency is an important goal of the data collection/compilation process.[7]

- Data selection involving unit, process and or stack-level matching is difficult and takes additional resources. Ultimately, a more complex data selection process may not yield more accurate results in terms of summarizing, analyzing trends, and modeling the data. Furthermore, judgments have to be made about which set of unit attributes and stack parameters are better. For example, if CAMD emissions estimates are deemed better for EGUs, but state stack parameters are considered better than EPA parameters, then retaining state unit-stack information means that EPA emissions have to be matched and assigned to the right units. If there is a mismatch in the original datasets (e.g., one has more units than the

other), artificial judgments must be made regarding the allocation or aggregation of emissions per each unit. If a facility is entirely absent or none of its units are likely matches, then emissions from one database may not be included at all.

It is important to note, that in the 1999 NEI for criteria pollutants, emissions for EGUs were assigned to existing state and local units, so the process followed for the draft 2002 NEI is a departure from the method employed in the development of prior inventories.[8]

Below we discuss some of the issues regarding the current data selection system about which data users need to be aware:

- *Complete facility estimates are obtained using the NEI Unique Facility ID.* Because the estimates for a facility in the NEI may be provided by multiple data sources, it is necessary to sum facility emissions using the NEI Unique Facility ID and not just the state and local identifier (Table 4). This is particularly important for EGUs where preference is given to EPA data for NO_x and SO_2 and no other pollutant. Thus, to obtain the full complement of CAP and HAP estimates for these facilities it is necessary to query on the NEI Unique Facility ID. Querying on just the state and local identifier may drop out all SO_2 and NO_x if EPA data were provided for the facility. Conversely, querying by EPA's local identifier alone may result in only NO_x and SO_2 emissions if state or local data were provided for the other pollutants.

Table 4. Facility level emissions for a facility with state and CAMD-provided data

strTribalCode	strStateCountyFIPs	strState FacilityIdentifier	strFacilityName	strNTISiteID	strPollutantCode	Emissions (TPY)	DATA SOURCE
000	01055	0002	Power Company	NEI12788	253	0.16	S
000	01055	0002	Power Company	NEI12788	7647010	130.9	S
000	01055	0002	Power Company	NEI12788	7664393	9	S
000	01055	0002	Power Company	NEI12788	CO	85.37	S
000	01055	0002	Power Company	NEI12788	PM10-PRI	85.77	S
000	01055	0002	Power Company	NEI12788	PM25-PRI	37.13	S
000	01055	0002	Power Company	NEI12788	PM-PRI	130.57	S
000	01055	0002	Power Company	NEI12788	VOC	10.41	S
000	01055	EGU0008	Power Company	NEI12788	208968	0.000034	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	218019	0.000013	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	3697243	0.0000029	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	92524	0.00023	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	98828	0.00072	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	98862	0.0020	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	NH3	4.50	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	NOX	1917.8	767/CAMD
000	01055	EGU0008	Power Company	NEI12788	SO2	8740.40	767/CAMD

- *Complete EGU unit specific estimates are obtained using the NEI Unique Facility ID and common unit identifiers.* If the user needs to obtain all estimates for specific combustion units at an EGU facility, he/she must be careful to query on NEI Unique Facility ID and the specific unit. Again, querying using just state and local unit identifiers may drop out all SO_2 and NO_x if EPA data were provided for the facility. Conversely, querying only by EPA's unit IDs or ORIS Boiler IDs may result in only SO_2 and NO_x data if state or local data were provided for

the other pollutants. Finding common unit IDs is problematic as different data sources do not use the same unit IDs and ORIS boiler IDs may not be filled in by some data providers. Manual review of all units (provided by the different data sources) is necessary to find all pollutants being emitted by the unit(s) of interest (Table 5).

Table 5. Unit level emissions for a facility with state and CAMD-provided data

strTribal Code	strState CountyFIPs	strState Facility Identifier	strFacility Name	strNTISiteID	strPollutant Code	strEmission UnitID	ORISBoilerID	Emissions (TPY)	DATA SOURCE
000	01033	0010	Winston	NEI11769	PM-PRI	001		113	S
000	01033	0010	Winston	NEI11769	CO	002		0.24	S
000	01033	0010	Winston	NEI11769	PM-PRI	002		1.16	S
000	01033	0010	Winston	NEI11769	VOC	002		0.22	S
000	01033	EGU0004	Winston	NEI11769	7440020	1	1	0.05	767/CAMD
000	01033	EGU0004	Winston	NEI11769	80626	1	1	0.003	767/CAMD
000	01033	EGU0004	Winston	NEI11769	SO_2	1	1	4304.90	767/CAMD
000	01033	EGU0004	Winston	NEI11769	NO_x	1	1	2039.23	767/CAMD
000	01033	EGU0004	Winston	NEI11769	NH_3	1	1	5.68	767/CAMD
000	01033	EGU0004	Winston	NEI11769	7440360	1	1	0.0033	767/CAMD
000	01033	EGU0004	Winston	NEI11769	78933	1	1	0.0723	767/CAMD
000	01033	EGU0004	Winston	NEI11769	75150	1	1	0.024	767/CAMD
000	01033	EGU0004	Winston	NEI11769	60344	2	2	0.046	767/CAMD
000	01033	EGU0004	Winston	NEI11769	NO_x	2	2	2922.95	767/CAMD
000	01033	EGU0004	Winston	NEI11769	NH_3	2	2	8.09	767/CAMD
000	01033	EGU0004	Winston	NEI11769	SO_2	2	2	6170.48	767/CAMD
000	01033	EGU0004	Winston	NEI11769	50328	2	2	0.00001	767/CAMD
000	01033	EGU0004	Winston	NEI11769	208968	2	2	0.000067	767/CAMD

- *Units appear to have more stacks than actually exist.* Because the original unit and stack IDs have been retained from each separate data source, this means that a unit which is represented twice, e.g, once for HAPs and once for NO_x and SO_2, will have multiple stack or emissions release point records. Each of these records actually represents the same stack. While the emissions are not duplicated and sum up properly for the facility, this method assigns the emissions to different stacks when in fact the emissions come from the same stack. This may be an issue for modeling if the stacks have different parameters (i.e., diameter, height, flowrate, etc.).

- *Elimination of non-duplicative HAP estimates*. Because HAP merging is done at a facility and not a process level, it is possible that emissions for the same pollutant from DIFFERENT processes or units from different data sources are eliminated. Facility level merging was chosen in preference to unit and/or process level merging because SCCs and unit IDs have to be reliably filled in and capable of being matched among multiple databases. If these data elements do not match or are not filled in, then this type of merging leads to a significant amount of duplication when estimates cannot be reliably matched up. (SCCs are not a mandatory field for the HAP inventory.) Furthermore, TRI and some state data are only collected on a facility level basis and cannot be

merged on a unit or process level basis. Thus, the conservative approach taken here, while avoiding pollutant duplication for the facility, potentially may lead to data loss.

- *Inadvertent supplementation from shutdowns.* If one database contains a non-zero value for a pollutant estimate, and the state/tribal or local agency did not submit an estimate because the agency knows the site, unit, or process is shutdown, than the non-zero value will be added back in during data selection. To avoid this incorrect supplementation, S/L/T should submit zero emissions for recent or temporary shutdowns.

What is the Best Way to Review the Selection Process?

State, local and tribal reviewers can better understand some of the issues outlined above by looking at the NOF state and other summary files provided by EPA. Reviewers can assess the results of facility matching and improve the results of data selection in the following ways:

1) Review the sites table (tblPointSI) included in your state NOF file. Look for sites that appear to be identical facilities that do NOT have the same NEI Unique Facility ID (stored in the strNTISiteID column). If address or other key information were missing, duplicates may not have been detected. Reviewers can mark the duplicate record with a "D" for delete in the strSubmittalFlag. (See the README for complete instructions on submitting revisions to the NEI ftp://ftp.epa.gov/EmisInventory/draftnei2002/point/data/readme_2002neidraft_ptsources.pdf).[9] States should also submit a list of these duplicates along with their revision files so that the NEI Facility File is corrected and future data selection errors prevented. Reviewers do not need to submit separate delete records for the associated units, processes, etc. as all deletions will be cascaded through all tables.

   If two sites DO have the same NEI Unique Facility ID, EPA has attempted to de-select duplicate estimates among these sites while retaining non-overlapping pollutants. For these facilities, you should focus your efforts on reviewing estimates from data sources outside your own (see Step #2).

2) Review the emissions estimates for sites sharing the same NEI Unique Facility. The ALLNEI Summary table is a good place to start as it contains the emissions estimates linked to their facility name, IDs and data sources (see ftp://ftp.epa.gov/EmisInventory/draftnei2002/point/summaries/).[10] The data source code indicates the origin of each emissions record. Review the pollutants we have added or selected from other data sources. If you know a unit has closed and is not non-emitting or a chosen value appears too high or low, you will want to submit "D" delete or "RA/RD" revision records for tblPointEM. When one of your values has been "deselected" you will need to see the Historical Emissions Report to view it in context with its "competitors" from other data sources (see Step #3).

3) To compare "competing" estimates from multiple data sources, review the "selected" versus "de-selected" estimates for each pollutant/facility in the Historical Emissions Report. This report contains not only the chosen value, but lists the "de-selected" values as well on a facility-pollutant basis. The data source and values for each competing

estimate are also included (see Attachment A). You can determine if another value was chosen over your own by taking the following steps:

a) Find the facility/pollutant of interest in the table.
b) Note the value in the "CurrDataSource Reported." If this column is filled in with S, L or T then state, local or tribal data were used.
c) Note the value in the "Curr TPY" for the current emissions value in the draft NEI. Read across the row to find the other values submitted but not chosen.

If you disagree with a value that has been chosen over your own, you can copy/paste the record(s) of concern from the Historical Emissions Report into its own table. Add a comment column to the record and document your suggested revision and any other observations regarding the affected record(s). These comments can be included along with your other revisions submitted via the Central Data Exchange (CDX).

CONCLUSIONS

This paper gave an overview of the steps involved in the data selection process used to generate the draft HAP and CAP 2002 NEI. EIG must process, compile, and merge multiple data sources without creating duplicate estimates. Multiple data sources are accepted into the NEI because different data sources vary in their completeness for both source and pollutant coverage. Furthermore, different data sources have estimates of varying certainty and some data sources are preferred for certain source categories. Ignoring one more or more data sources would reduce the overall completeness and accuracy of the inventory.

A stepwise process and data ranking hierarchy have been established to select data from one source over another. For non-EGU facilities, HAPs and CAPs are merged on a facility-pollutant (or pollutant category) basis. For EGUs, a more complex data selection strategy is implemented involving SCCs. However, EGU data selection does not involve pairing up units or processes and cannot be considered a unit or process level merge.

When reviewing the draft NEI reviewers should be aware that the original IDs are retained from all data sources. The local site ID is retained in the strStateFacilityIdentifier field in tblPointSI. This enhances transparency, facilitates comparisons, and makes it easier to revise the data. However, users of the data need to be aware that a facility can have more than one state and local identifier associated with it. Users need to consult all records associated with the same NEI Unique Facility ID when reviewing records or summing emissions data for a given plant and need to be aware that different data sources use different unit, process and stack IDs for the same entities at the facility.

State, local and tribal reviewers can comment and help improve upon the data selection process. Reviewers should look for undetected duplicates sites and pollutants that have been incorrectly added to the inventory. Reviewers can compare selected data to their own values and submit comments when they believe an aberrant value was chosen. Comments and revisions should be submitted using the guidance outlined in the READ ME and this paper.

REFERENCES

1. U.S. EPA. 2004. Public Data Release. 2004 Toxics Release Inventory. U.S. Environmental

Protection Agency, Office of Environmental Information, Washington, D.C. http://www.epa.gov/tri/

2. U.S. EPA. 2005. Summary Information for 2002 National Emissions Inventory (NEI). 2002 NEI Facility File. U.S. Environmental Protection Agency, Emission Inventory Group, Research Triangle Park, North Carolina. ftp://ftp.epa.gov/EmisInventory/draftnei2002/point/summaries/.

3. Pope, A., Finn, S., and Wilson, D. 2004. "2002 NEI for Point Sources: Integration of HAPs and CAPs." Presented at the 13th International Emission Inventory Conference "Working for Clean Air in Clearwater." June 8-10, Clearwater, Florida.

4. U.S. DOE. 2003. "Steam-Electric Plant Operation and Design Report," Form EIA-767, data files for 1985-2002, U.S. Department of Energy, Energy Information Administration, Washington, D.C.

5. U.S. EPA. 2004. Documentation for the 2002 Electric Generating Unit (EGU) National Emissions Inventory (NEI). U.S. Environmental Protection Agency, Emission Inventory Group, Research Triangle Park, North Carolina. ftp://ftp.epa.gov/EmisInentory/draftnei2002/point/documentation/egu2002doc.pdf

6. U.S. EPA. 2004. 2002 Emission Tracking System/Continuous Emissions Monitoring (ETS/CEM) Annual and Ozone Season Data Files, U.S. Environmental Protection Agency/Clean Air Markets Division (CAMD), Washington, D.C. http://www.epa.gov/airmarket/monitoring/factsheet.html.

7. Pope, A., Finn, S., and Wilson, D. "The Challenge of Meeting New EPA Data Standards and Information Quality Guidelines in the Development of the 2002 NEI Point Source Data for HAPs." Presented at the 13th International Emission Inventory Conference "Working for Clean Air in Clearwater." April 2003, San Diego, CA.

8. U.S. EPA. 2004. Documentation for the Final 1999 National Emissions Inventory (NEI) (Version 3) for Criteria Air Pollutants and Ammonia Point Sources. U.S. Environmental Protection Agency, Emission Inventory Group, Research Triangle Park, North Carolina. ftp://ftp.epa.gov/EmisInventory/finalnei99ver3/criteria/documentation/point/point_99nei_finalv3_0204.pdf

9. U.S. EPA. 2005. READ ME File for the 2002 NEI Point Source NEI for HAPs and CAPs. U.S. Environmental Protection Agency, Emission Inventory Group, Research Triangle Park, North Carolina. ftp://ftp.epa.gov/EmisInventory/draftnei2002/point/data/readme_2002neidraft_ptsources.pdf

10. U.S. EPA. 2005. Summary Information for 2002 National Emissions Inventory (NEI). ALLNEI CAP ANNUAL, ALLNEI HAP ANNUAL. U.S. Environmental Protection Agency, Emission Inventory Group, Research Triangle Park, North Carolina. ftp://ftp.epa.gov/EmisInventory/draftnei2002/point/summaries/

KEY WORDS
Point Source Inventory
National Emissions Inventory
Criteria Pollutant
Hazardous Air Pollutant
Electric Generating Utilities

Attachment A. Historical Emissions Report

State County FIPS	TribalCode	Facility Name	NTISiteID	Pollutant Code	Pollutant Name	CurrVersionDate	CurrTPY Reported	CurrData Source Reported	CurrS	CurrT	CurrEGU767
01001	000	AB Paper Company	NEI8560	100414	Ethyl Benzene	2/1/2005	16.96	S	16.96		4.98E-02
01001	000	AB Paper Company	NEI8560	100425	Styrene	2/1/2005	2.21	S	2.21		2.85
01001	000	AB Paper Company	NEI8560	107028	Acrolein	2/1/2005	6.28	S	6.28		6.02
01001	000	AB Paper Company	NEI8560	107062	Ethylene Dichloride	2/1/2005	0	S	0		4.54E-02
01001	000	AB Paper Company	NEI8560	108054	Vinyl Acetate	2/1/2005	0	S	0		0.00025
01001	000	AB Paper Company	NEI8560	108101	Methyl Isobutyl Ketone	2/1/2005	6.02	S	6.02		
01001	000	AB Paper Company	NEI8560	108883	Toluene	2/1/2005	4.24	S	4.24		1.443
01001	000	AB Paper Company	NEI8560	108907	Chlorobenzene	2/1/2005	0	S	0		5.067E-02
01001	000	AB Paper Company	NEI8560	108952	Phenol	2/1/2005	10.23	S	10.23	5.5	7.72E-02
01001	000	AB Paper Company	NEI8560	110543	Hexane	2/1/2005	0.89	S	0.89		1.577
01001	000	AB Paper Company	NEI8560	120821	1,2,4-Trichlorobenzene	2/1/2005	0.09	S	0.09		
01001	000	AB Paper Company	NEI8560	127184	Tetrachloroethylene	2/1/2005	0.07	S	0.07		7.149E-02
01001	000	AB Paper Company	NEI8560	130498292	PAH, total	2/1/2005	0.11	T		0.11	2.351E-03
01001	000	AB Paper Company	NEI8560	1319773	Cresol	2/1/2005	8.67	S	8.67	16.5	
01001	000	AB Paper Company	NEI8560	1330207	Xylenes (Mixture of o, m, and p Isomers)	2/1/2005	2.91	S	2.91		7.987E-03
01001	000	AB Paper Company	NEI8560	155	Dioxins	2/1/2005	0.00000081585	T		0.00000081585	
01001	000	AB Paper Company	NEI8560	195	Lead & Compounds	2/1/2005	0.17	T		0.17	
01001	000	AB Paper Company	NEI8560	198	Manganese & Compounds	2/1/2005	0.95	T		0.95	
01001	000	AB Paper Company	NEI8560	50000	Formaldehyde	2/1/2005	4.57	S	4.57	3.982	6.7118
01001	000	AB Paper Company	NEI8560	56235	Carbon Tetrachloride	2/1/2005	3.05	S	3.05		6.768E-02
01001	000	AB Paper Company	NEI8560	67561	Methanol	2/1/2005	307.65	S	307.65	352.85	
01001	000	AB Paper Company	NEI8560	67663	Chloroform	2/1/2005	1.89	S	1.89		4.421E-02
01001	000	AB Paper Company	NEI8560	71432	Benzene	2/1/2005	2.57	S	2.57		6.378
01001	000	AB Paper Company	NEI8560	71556	Methyl Chloroform	2/1/2005	0	S	0		4.718E-02
01001	000	AB Paper Company	NEI8560	74839	Methyl Bromide	2/1/2005	0.12	S	0.12		2.8004E-02
01001	000	AB Paper Company	NEI8560	74873	Methyl Chloride	2/1/2005	10.1	S	10.1		5.24E-02
01001	000	AB Paper Company	NEI8560	74884	Methyl Iodide	2/1/2005	0.11	S	0.11		

Attachment A. Historical Emissions Report (Continued)

State County FIPS	TribalCode	Facility Name	NTISiteID	Pollutant Code	Pollutant Name	CurrVersionDate	CurrTPY Reported	CurrData Source Reported	CurrS	CurrT	CurrEGU767
01001	000	AB Paper Company	NEI8560	75070	Acetaldehyde	2/1/2005	38.82	S	38.82	39	1.273
01001	000	AB Paper Company	NEI8560	75092	Methylene Chloride	2/1/2005	3.22	S	3.22		0.4460
01001	000	AB Paper Company	NEI8560	75150	Carbon Disulfide	2/1/2005	3.85	S	3.85		0.0045
01001	000	AB Paper Company	NEI8560	75354	Vinylidene Chloride	2/1/2005	0.53	S	0.53		
01001	000	AB Paper Company	NEI8560	7647010	Hydrochloric Acid	2/1/2005	163.69	S	163.69	170	69.438
01001	000	AB Paper Company	NEI8560	78933	Methyl Ethyl Ketone	2/1/2005	13.61	S	13.61	13.875	2.140E-02
01001	000	AB Paper Company	NEI8560	79005	1,1,2-Trichloroethane	2/1/2005	4.54	S	4.54		
01001	000	AB Paper Company	NEI8560	79016	Trichloroethylene	2/1/2005	2.99	S	2.99		4.984E-02
01001	000	AB Paper Company	NEI8560	84742	Dibutyl Phthalate	2/1/2005	0.17	S	0.17		
01001	000	AB Paper Company	NEI8560	91203	Naphthalene	2/1/2005	2.62	S	2.62		0.14893
01001	000	AB Paper Company	NEI8560	95476	o-Xylene	2/1/2005	0	S	0		4.032E-02

[1] Not all columns are shown. For complete tables structure see: READ ME File for the 2002 NEI Point Source NEI for HAPs and CAPs.