Development of Nonpoint Hazardous Air Pollutant Emissions Estimates from 1990 to 2010 Based on Historical Emissions Data

Michael Regan

The Potomac School
1301 Potomac School Rd
McLean, VA 22101

Abstract

The purpose of this study was to develop estimates for Hazardous Air Pollutant (HAP) emissions for 20 years worth of historical emissions estimates, based on the National Emissions Inventory (NEI), for long-term trend analysis. Released semi-annually by the Environmental Protection Agency (EPA) beginning in 1990, the NEI provides quantitative estimates of Criteria Air Pollutant (CAP) emissions, including carbon monoxide, sulfur dioxide, nitrogen dioxide, ammonia, and particulate matter, as well as volatile organic compounds (VOCs), based on county-level reports submitted by state environmental officials. However, due to the difficulty of measuring HAP emissions, pollutants known to cause adverse health effects, the NEI did not augment these until the 2011 report. VOCs emissions include the values of HAPs, thus, individual HAP emissions can be represented as a fraction of VOC emissions. Using RStudio as the analytical platform, HAP emissions estimates were developed using previously created annual emissions estimates which were based on the NEI reports and lacked HAP information from 1990 to 2010. To generate these HAP estimates, weighting factors for individual HAPs were calculated based on the 2016 release of NEI data, including their inventory source characteristics (geographical location, sector-specific identifiers, Source Classification Code [SCC], and annual emissions). Next, these weighting factors were applied to previous VOC entries with matching source characteristics in the historical dataset. This complete emissions dataset can assist in long-term climate studies, providing information on trends of VOC and HAP emissions, helping to facilitate climate-related analyses at a finer scale.

Acknowledgements

Dr. Bok-Haeng Baek (George Mason University), the supervising scientist of this research project, as well as Dr. Chi-Tsan Wang (George Mason University), provided the project idea and the dataset of annual historical emissions that was used for analysis. They provided assistance with questions regarding background research and participated in developing the procedure to find a potential topic for analysis, as well as giving revisions for the written report.

Table of Contents:

Abstract 1

Acknowledgements 1

1. Introduction 3

2. Materials and Methods 4

3. Results 6

4. Discussion 13

5. References 14

1. Introduction

Understanding the location and concentration of pollutants is critical to developing regulations that preserve human health by limiting exposure to harmful chemical compounds. Recent research by the WHO has determined that 99% of the world population did not have their air quality guidelines met for their location of residence in 2019, and that ambient air pollution caused an estimated "4.2 million premature deaths" in 2016.[1] Quantifying and tracking exposure to air contamination, specifically, poses an issue, due to the mobile and reactive nature of these chemicals and particles. Much has been done in the development of past estimates of emissions trends through a variety of methods, such as by establishing emissions factors for each source of pollution,[2,3] which connects an emissions-producing activity to the amount at which that activity emits pollutants.[4]

One category of pollutants that require emissions estimates is Volatile Organic Compounds (VOCs). VOCs are organic compounds that are typically found in the gaseous state at room temperature because of their ease of evaporation and are emitted from numerous sources, including fuel burning and exhaust, paints, and many more.[5] Some specific, more toxic VOCs are called Hazardous Air Pollutants (HAPs), such as benzene, styrene, toluene, xylene, and ethylbenzene. HAPs are a subset of VOCs determined by the Clean Air Act,[6] which are substances known to cause serious health risks, requiring the EPA to heavily regulate their emissions.[7] Exposure to HAPs have been found to cause negative health effects in both the short and long-term, thus, understanding the quantities of emissions of HAPs are critical.[8,9,10]

A tool used to estimate emissions is the National Emissions Inventory (NEI), a database developed by the EPA first published in 1990 and now released on a semi-annual basis. Estimates are divided into the following categories in the NEI: stationary, mobile, and event-based emissions sources.[6] The NEI offers emissions estimates for VOCs and Criteria Air Pollutants (CAPs) including ozone, particulate matter, carbon monoxide, lead, sulfur dioxide, and nitrogen dioxide, as well as estimates for HAPs, but these were not augmented until the 2011 report.[6] VOCs and HAPs are not CAPs, but they are significant precursors of two CAPs, ozone (O3) and secondary organic aerosol (SOA).

Although previous studies have recorded observational data and developed yearly historical estimates of CAP emissions utilizing data from multiple sources, including the NEI, historical HAP data is limited.[11] One way in which the EPA generates estimates for HAPs is through a database called SPECIATE. This repository contains profiles for specific emissions sources, including VOC emissions sources, that are developed through previous scientific literature as well as laboratory testing and surveys. Within these profiles are weighting factors for HAP emissions as a percentage of the entire reported amount of VOC emissions.[12] The NEI uses this database to generate HAP emissions by assigning these profiles to matched VOC sources in the current NEI release.[13] However, SPECIATE was initially released in 2002, and there were very limited profiles available for HAP emission generation, leaving many previous years without detailed emissions data for these harmful pollutants.[14] Thus, generating estimates for HAP emissions is critical to improving the understanding of their long-term impacts on the environment and humans.

This project aims to develop HAP emissions estimates for the years of NEI data lacking this information, using previously generated annual emissions estimate files for CAPs and VOCs that were based on NEI reports from 1990 to 2010.[11] The creation of these estimates will help to

facilitate long-term air quality analyses, providing a better understanding of emissions trends across the US to help in the refinement of policy.

2. Materials and Methods

Materials

The annual historical emissions estimates from 1990-2010 were provided by a collaborative library from previously developed data for a research publication.[11] RStudio (R version 4.1.2, released in 2021) was used for the development of all programs to manipulate and extract unknown values from the data. The desired data was extracted into Microsoft Excel 2016 (16.0.5254.1000) for visualization.

Methods

In order to develop HAP emissions estimates for the years that did not contain estimates (i.e. 1990-2010), we applied the Sparse Matrix Operator Kerner Emissions (SMOKE) modeling system to process the 2016 NEI data and generate the inventory report of the data. The 2016 inventory report was divided into sectors by SMOKE - different categories of sources of emissions - and were imported into RStudio. Each entry of these sector-level inventory reports includes the annual or monthly emissions from various pollutants, FIPS code, and source classification code (SCC) of each reported emissions source.[15] All entries are reported in tons/year.

After being imported, only reported entries with VOC (I_VOC_INV) values greater than 0 were selected for future analysis, while entries where total VOC values were not provided or where individual HAP values exceeded VOC values, were excluded. Next, surrogate profiles for multiple sectors were developed, outlined in Table 1. Each entry of the profile consisted of the aforementioned VOC values, each with a unique FIPS code and SCC combination, as well as the ratios of selected HAP emissions, with matching FIPS code and SCC, to the total VOC emission value, called weighting factors. These weighting factors were calculated by equation [1], where F represents the individual HAP weighting factor, E_VOC is the reported VOC emissions value for an entry, E represents any HAP emission value for an entry, i is the individual species of HAP, s is the SCC, and f is the FIPS code for each entry. Thus, $E_{i,s,f}$ represents the reported emissions in the inventory report of HAP i with the SCC of s and FIPS code of f, $E\_VOC_{s,f}$ represents the reported VOC emissions from the SCC s and FIPS code of f, and $F_{i,s,f}$ is the weighting factor of HAP i with the SCC of s and FIPS code of f.

$$F_{i,s,f} = \frac{E_{i,s,f}}{E\_VOC_{s,f}} \qquad [1]$$

The desired HAPs for each surrogate varied by sector, evidenced in Table 1. In this project, results are based on the combined surrogate profiles for the nonpoint (nonpt) sector, as well as the nonpoint source oil and gas (np_oilgas) sector. The nonpt sector is made up of emissions from sources that are at a fixed location, but "individually are too small in magnitude to report as point sources."[16] Some examples of emissions in the nonpt sector are sources of fuel combustion such as heaters, industrial processes, the use of solvents, and the storage and transportation of chemicals and waste.[6] The 2016 nonpt data is released on a monthly rather than annual basis, but changes between months are minuscule, thus, the January report for the nonpt sector was used for weighting factor calculation.

Table 1: 2016 NEI sectors for which this project developed profiles, and the selected HAPs in each sector for which weighting factors were calculated.[17]

Sector	HAPs with Calculated Weighting Factors
rwc (Residential wood combustion)	Formaldehyde, benzene, acetaldehyde, naphthalene, xylenes, butadiene, acrolein, toluene, chlorine
Ag (Agricultural)	Methanol, benzene, acetaldehyde, xylenes, toluene
Airport (Aircraft and ground support equipment)	Benzene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene
nonpt (Remaining nonpoint)	Formaldehyde, methanol, benzene, acetaldehyde, naphthalene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene, chlorine
nonroad (Nonroad)	Formaldehyde, benzene, acetaldehyde, naphthalene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene, hexane, propional, TRMEPN224
Np_oilgas (Nonpoint source oil and gas)	Formaldehyde, methanol, benzene, acetaldehyde, naphthalene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene, chlorine
othpt (Other point sources not from the NEI)	Acetaldehyde, propionaldehyde and higher aldehydes, ethene, ethanol, formaldehyde, internal olefin carbon bond (R-C=C-R), isoprene, methanol, non-volatile compounds, terminal olefin carbon bond (R-C=C), toluene and other monoalkyl aromatics, unreactive, benzene, ethane, acetone, ethyne, ketone groups, xylene and other polyalkyl aromatics, naphthalene
pt_oilgas (Point source oil and gas)	Benzene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene, chlorine, ACET
ptag_fire (Agricultural fires with point resolution)	Benzene, xylenes, butadiene, acrolein, toluene
ptegu (Point source electric generating units)	Benzene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene, chlorine
ptfire3D (Layered point source fire emissions)	Formaldehyde, methanol, benzene, acetaldehyde, naphthalene, xylenes, styrene, butadiene, acrolein, toluene
ptnonipm (Remaining non-EGU point sources)	Benzene, xylenes, ethylbenzene, styrene, butadiene, acrolein, toluene, chlorine
rail (Locomotives)	Formaldehyde, benzene, acetaldehyde, naphthalene, butadiene, acrolein

Initially, this combined surrogate profile was compared with the annual estimates of the nonpt sector from 1990-2010 (historical data) to identify matching entries between the current profile and previous data based on complete matches of both FIPS codes and SCCs. However, due to the limited matches of complete FIPS codes and SCCs between the profiles and the historical data, 6 total 2016 surrogate profiles with differing levels of aggregation for each entry's designated FIPS and SCC (schemas) were developed, as seen in Table 2. This was done to increase the number of matches between profiles and historical data, and thus, the number of entries in the historical data for which HAP emissions estimates could be generated.

To calculate weighting factors for the 5 remaining schemas, individual entries in the combined nonpt profile were summed based on the level of aggregation of that schema, then the weighting factors were developed by equation [2], which maintains the same values as equation

[1], except *s* now represents the SCC aggregation for each schema and *f* now represents the FIPS code aggregation for each schema. Thus, matching $E_{i,s,f}$ and $E\_VOC_{s,f}$ values are summed before calculating the resulting weighting factor.

$$F_{i,s,f} = \frac{\Sigma E_{i,s,f}}{\Sigma E\_VOC_{s,f}} \qquad [2]$$

In order to calculate the individual HAP emissions estimates for the historical nonpt data, matches in the historical data that contained VOC values greater than 0 (data_wVOC) to the 2016 surrogate based on the varying aggregations of each schema were identified (data_match). Then, through equation [3], where *H* represents any individual historical HAP emissions estimate for an entry, and *y* represents any year in the 1990-2010 timeframe, the weighting factors from each schema were multiplied by the historical VOC emissions value, *HE_VOC*, with matching *s* and *f* values to the weighting factor, generating HAP estimates for each entry over the timeframe.

$$H_{y,i,s,f} = HE\_VOC_{y,s,f} \times F_{i,s,f} \qquad [3]$$

The historical data which was used in this project contained relatively consistent SCCs throughout all 20 years.[11] Thus, based on the 2010 historical nonpt data, entries that did not have a match were separated and totaled for each of the 6 schemas, allowing for the determination of emission sources that are no longer in use as of 2016, as well as the most useful schema for analysis. Since data was not available for the year 1995, it was represented as the average emissions between 1994 and 1996. Data for all figures was rounded to the second decimal place.

3. Results

Overview of All Six Schemas

Long-term emissions estimates of HAPs have not been compiled before. In this paper, estimates for HAPs in the nonpoint sector of the NEI were generated through varying levels of aggregation and examined for implementation in historical studies. To examine the efficacy of each schema for identifying matches with the historical nonpt data, the number of matched entries in the 2010 historical nonpt data with their VOC emissions value (ANN_EMS) greater than zero were first tabulated for each level of aggregation. A matched entry in the historical data refers to an entry with matching FIPS code and SCC based on the schema aggregation.

Table 2: Aggregation levels of FIPS code and SCC in the 6 different schemas developed in this project as well as their abbreviated name.

Schema	Aggregation of FIPS code	Aggregation of SCC
FIPS/SCC	Entire FIPS code	Entire SCC
FIPS/8SCC	Entire FIPS code	First 8 digits of SCC
FIPS/6SCC	Entire FIPS code	First 6 digits of SCC
State/SCC	First 2 digits of FIPS code	Entire SCC
State/8SCC	First 2 digits of FIPS code	First 8 digits of SCC
State/6SCC	First 2 digits of FIPS code	First 6 digits of SCC

Table 3: Number of matched and unmatched VOC entries for all 6 schemas in the nonpt sector of the 2010 historical dataset compared to the combined 2016 NEI nonpt and np_oilgas profiles. Total entries in the 2010 historical data were only those where the pollutant code (POLCODE) was VOC and the ANN_EMS was greater than 0. Matches based on varying levels of aggregation of only the SCC were calculated as well.

Aggregation Level	Matched Entries	Unmatched Entries	Percent Matched (%)
FIPS/SCC	126,687	103,797	54.97
FIPS/8SCC	134,026	96,458	58.15
FIPS/6SCC	168,170	62,314	72.96
State/SCC	138,427	92,057	60.06
State/8SCC	146,238	84,246	63.45
State/6SCC	179,675	50,809	77.96
Full SCC	171,693	58,791	74.49
First 8 SCC digits	173,893	56,591	75.45
First 6 SCC digits	197,719	32,765	85.78

There were 230,848 unique entries that matched this ANN_EMS criteria in the 2010 data. Unfortunately, the SCCs were not completely consistent between the historical and 2016 NEI data, evidenced by the row in Table 3 with an aggregation level of Full SCC. In this row, 58,791 VOC entries were found without a match in the 2010 historical nonpt data, meaning that these entries either had SCCs that were retired or no longer reported on. Based on the schema consisting of only exact matches between the two years (FIPS/SCC), 54.97% of entries were matched in the 2010 historical nonpt data, the lowest percentage matched of all the schemas. Following this was the FIPS/8SCC, then State/SCC, State/8SCC, FIPS/6SCC, and finally, the

Table 4: Total amount of VOC emissions estimated for each of the 6 schemas compared to the total reported historical nonpt amount for 2010 (5,841,069 tons/year). Total VOC estimated was calculated by summing the total estimated emissions for each of the varying levels of aggregation.
*As these schemas only provide HAP estimates, the total VOC estimated also represents HAP emissions estimated.
**For reference, the 2016 NEI nonpt profile had 3,747,807 tons/year of VOC emissions, and 308,327 tons/year reported from the 12 selected HAPs in Table 1, representing 8.23% of total VOC emissions.

Schema	Total VOC Estimated (tons/year)*	Percent of Total Reported VOC (%)**
FIPS/SCC	205,188	3.51
FIPS/8SCC	211,389	3.62
FIPS/6SCC	284,316	4.87
State/SCC	212,224	3.63
State/8SCC	217,885	3.73
State/6SCC	273,663	4.69

State/6SCC, matching 77.96% of entries with the 2016 NEI combined nonpt profile. The actual quantity of VOC estimated, rather than the number of entries, follows a similar trend as well.

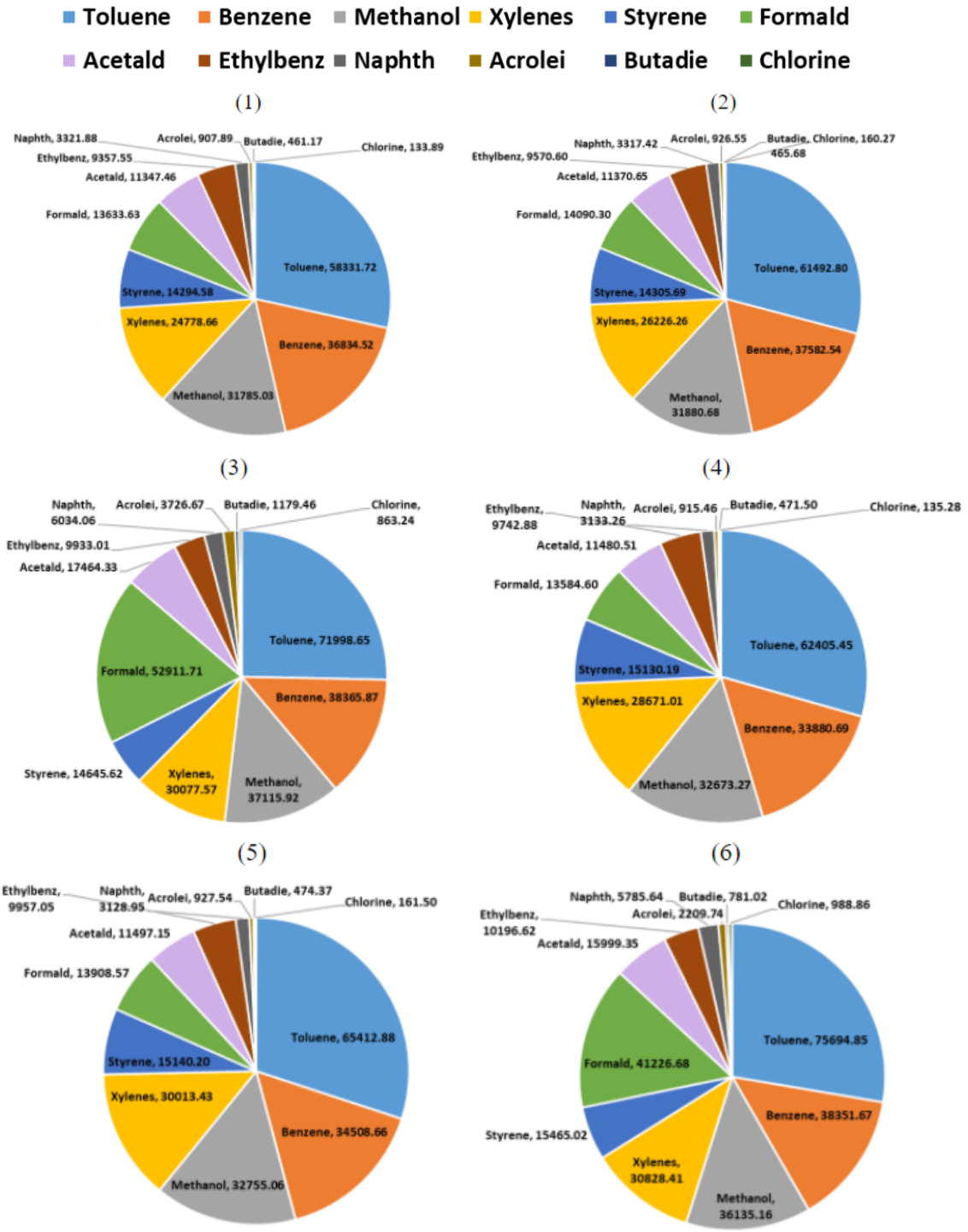

Figure 1: Individual estimated emissions distribution of HAPs for each of the 6 schemas in tons/year. Data for these figures was taken from the year 2010, and each HAP is shown with their annual emissions value as a fraction to the total estimated HAP emissions in 2010. The schemas are listed in the following order; (1): FIPS/SCC, (2): FIPS/8SCC, (3): FIPS/6SCC, (4): State/SCC, (5): State/8SCC, (6): State/6SCC.

As seen in Table 4, the lowest percentage of total VOC estimated of the 6 schemas is the FIPS/SCC schema, and it follows the same order of increase as Table 3 (excluding SCCs, as estimates were not generated for this level of aggregation), with the exception of the FIPS/6SCC exceeding the State/6SCC for the highest percentage of the total estimated (4.87% > 4.69%). This result is unexpected, as Table 3 shows the State/6SCC schema has roughly 11,000 more total matched entries than the FIPS/6SCC schema, yet the FIPS/6SCC schema developed more VOC estimates in terms of volume of emissions. While these overall percentages of estimated emissions appear small, many other factors that contribute to VOC emissions.[5]

Evidenced in Figure 1, throughout all 6 schemas, the most prominent HAP in the estimated estimations was toluene (cyan), which is used in the production of benzene, solvents, and polymers, and is found in gasoline as well, among many other uses.[18] Schemas with full SCCs or 8 digits of their SCCs for their aggregation share very similar distributions of HAP emissions (charts 1, 2, 4, 5), however, the charts with the first 6 digits of their SCCs as their aggregation (charts 3 and 6) drastically increase in their share of formaldehyde (green) that is estimated. Thus, for the figures containing in-depth trend analysis for one of the schemas, the FIPS/SCC schema was used, due to the similar percentages of total VOC estimated for the FIPS/SCC, FIPS/8SCC, State/SCC, and State/8SCC (3.51%, 3.62%, 3.63%, and 3.73%, respectively), and the more specific aggregation that the FIPS/SCC schema offered. Despite these dissimilar distributions of HAPs, the overall HAP estimations for all six schemas follow a similar trend to the historical amounts of total VOC emissions. Both the schemas and the actual VOC amount all decrease substantially in the 20-year timeframe, evidenced in Figure 2, and have similar shapes.

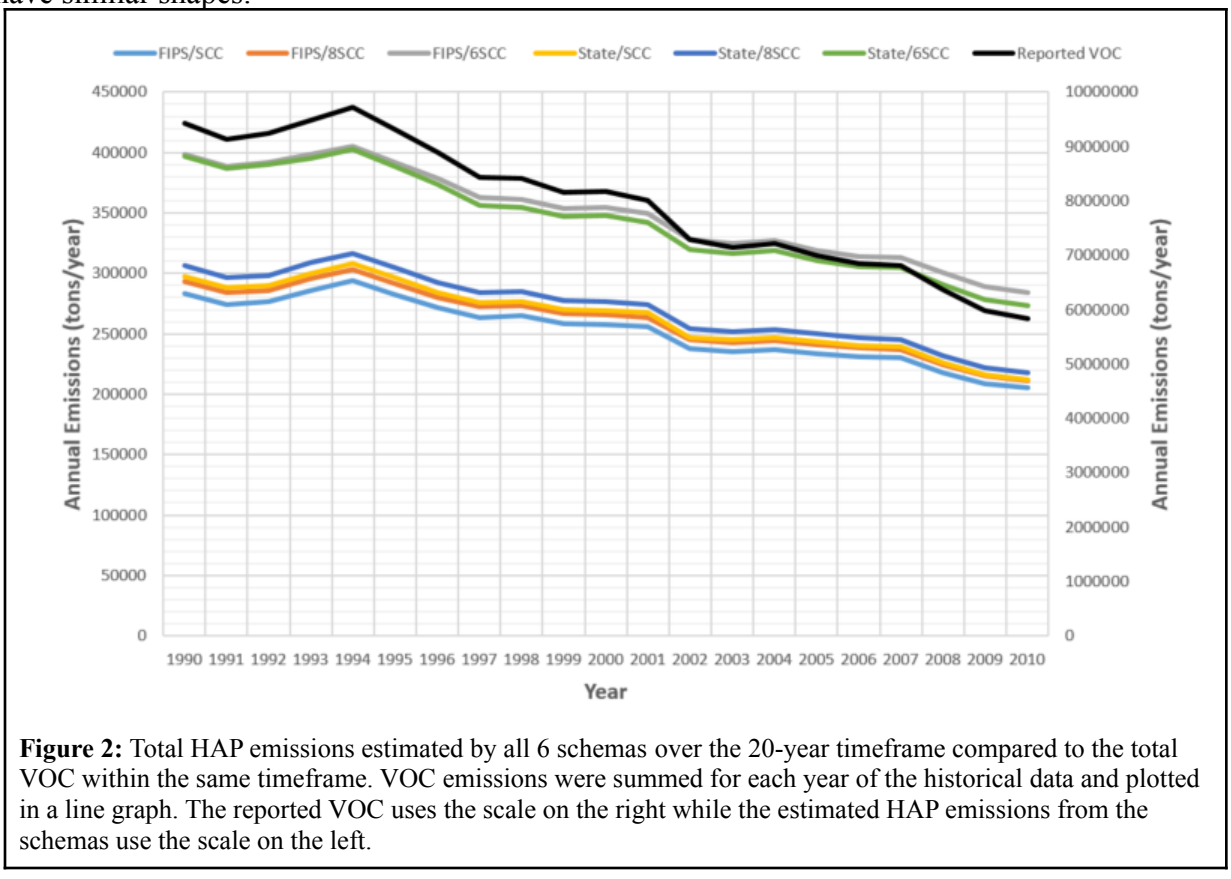

Figure 2: Total HAP emissions estimated by all 6 schemas over the 20-year timeframe compared to the total VOC within the same timeframe. VOC emissions were summed for each year of the historical data and plotted in a line graph. The reported VOC uses the scale on the right while the estimated HAP emissions from the schemas use the scale on the left.

Table 5: Change in estimated nonpt total HAP emissions from various tier 2 sectors from 1990 to 2010 based on the FIPS/SCC schema, compared to the historical VOC emissions from the nonpt sector for the same timeframe. The number in parenthesis represents the identifier for that specific tier in the SCC. All values were aggregated based on their tier 1 and tier 2 SCC identifiers.

		Total VOC Emissions (tons/year)			**FIPS/SCC nonpt Estimated HAP Emissions (tons/year)**		
Tier 1	**Tier 2**	**1990**	**2010**	**Δ**	**1990**	**2010**	**Δ**
Stationary Source Fuel Combustion (21)	Industrial (02)	11526.67	8700.72	-24.52%	361.50	369.15	2.12%
	Commercial/Institutional (03)	9256.95	7922.25	-14.42%	364.21	326.03	-10.49%
	Residential (04)	788638.88	392433.24	-50.24%	604.18	568.58	-5.89%
Industrial Processes (23)	Chemical Manufacturing (01)	308715.64	73550.59	-76.18%	21.34	5.09	-76.18%
	Food and Kindred Products (02)	39047.06	39047.06	-	4015.70	4015.70	-
	Mineral Processes (05)	930.52	930.52	-	0.52	0.52	-
	Petroleum Refining (06)	18386.55	18386.55	-	181.03	181.03	-
	Oil and Gas Exploration and Production (10)	855084.03	855084.03	-	19797.70	19797.70	-
Solvent Utilization (24)	Surface Coating (01)	1800932.59	1015205.82	-43.63%	71999.99	40587.20	-43.63%
	Misc. Industrial (40)	691099.03	389580.24	-43.63%	241.98	136.41	-43.63%
	Misc. non-Industrial - Consumer and Commercial (60)	799012.72	450412.39	-43.63%	52115.21	29377.93	-43.63%
	Misc. non-Industrial - Commercial (61)	497415.07	280398.43	-43.63%	14382.87	8107.78	-43.63%
Storage and Transport (25)	Petroleum and Petroleum Product Storage (01)	1181794.62	587452.20	-50.29%	32403.26	15981.69	-50.68%
	Petroleum and Petroleum Product Transport (05)	148065.16	74228.74	-49.87%	2155.02	1093.07	-49.28%
Waste Disposal, Treatment, Recovery (26)	Open Burning (10)	13752.65	13752.65	-	69769.97	69769.97	-
	Landfills (20)	26813.73	26813.73	-	2597.26	2597.26	-
	Wastewater Treatment (30)	121597.52	121597.52	-	11543.84	11543.84	-
Miscellaneous Area Sources (28)	Agriculture Production - Crops (01)	154519.36	185456.15	20.02%	33.78	25.91	-23.29%
	Other Combustion (10)	6392.80	6392.80	-	503.90	503.90	-
	Automotive Repair Shops (40)	1662.14	1662.14	-	199.23	199.23	-

Examination of FIPS/SCC Schema

The SCC for each entry can be broken down into four tiers. For example, the SCC "2102001000." The first and second digits (21) represent tier 1 (Stationary Source Fuel Combustion), the third and fourth digits (02) represent tier 2 (Industrial), the fifth, sixth, seventh, and eighth digits (0010) represent tier 3 (Anthracite Coal), and the final two digits (00) represent tier 4 (Total: All Boiler Types).

Within the FIPS/SCC schema of the nonpoint sector, there were 6 unique tier 1 sectors identified. Table 5 shows a breakdown of tier 1 nonpoint sectors by tier 2 separators, and their estimated change in the 20-year timeframe based on the FIPS/SCC schema, compared with the actual historical VOC emissions for those years. Many tier 2 categories exhibit equal percent changes across the two datasets, or no change at all in tier 1 sectors such as industrial processes or waste disposal, treatment, and recovery. This suggests that the entries were left unchanged in the historical dataset, or decreased by an equal proportion for all entries within that SCC tier. The breakdown of the tier 2 agricultural production in tier 1 miscellaneous area sources (2801) reveals an interesting difference between the historical VOC emissions and those from only HAPs, as historical VOC emissions increased by roughly 20%, while the estimated HAP emissions from the sector decreased by 23.29%. However, the majority of the selected sectors show a decrease in emissions over time, in accordance with results from Figure 2. The stationary source fuel combustion has different delta values for all three tier 2 sectors, which will be further examined.

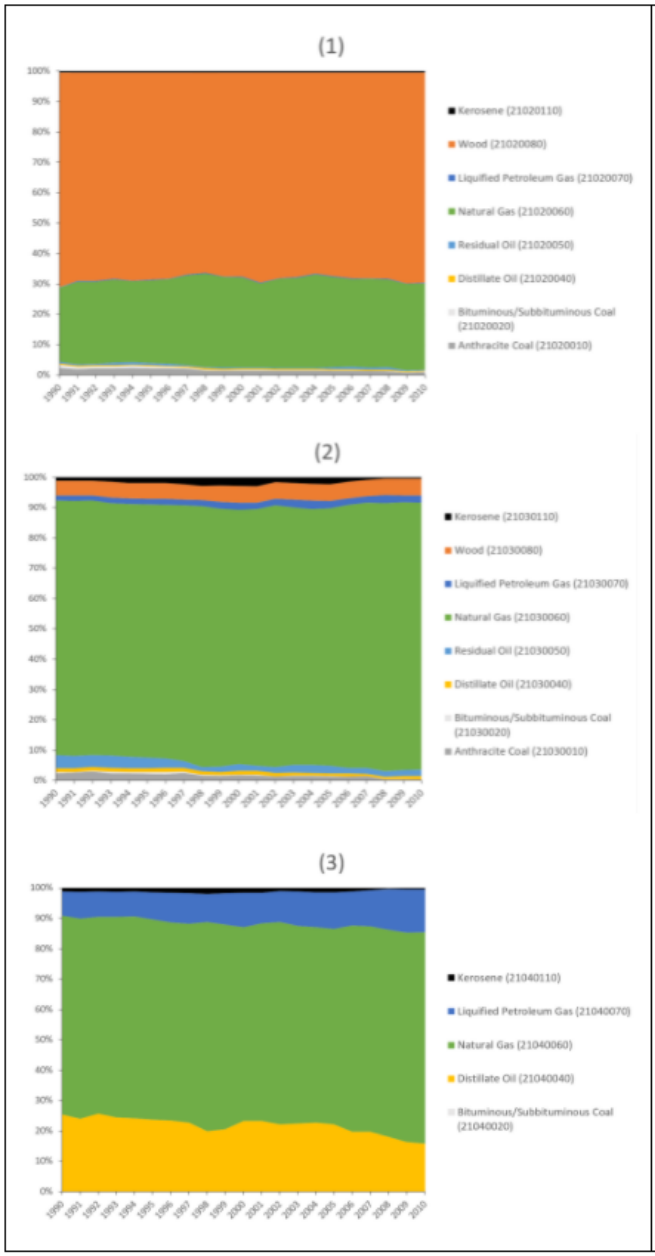

Figure 3: Changes in estimated HAP emissions from 1990-2010 using the FIPS/SCC schema estimations. Data was aggregated and summed by tier 3 SCC separators from the tier 1 sector (stationary source fuel combustion) for each year, then plotted in a stacked area chart as a fraction of total estimated HAP emissions for each of the identified tier 2 sectors within stationary source fuel combustion. The number in parenthesis represents the SCC identifier for all three tiers. Each separate graph represents a different tier 2 identifier; (1): Industrial, (2): Commercial/ Institutional, (3): Residential.

Figure 3 shows the percent distribution of estimated emissions in the stationary source fuel combustion sector (21) in the industrial, commercial/institutional, and residential areas based on the FIPS/SCC schema. In the industrial tier 2 identifier (02), the majority of HAP emissions come from wood and natural gas usage, whereas the commercial/institutional sector (03) is almost entirely natural gas. Residential (04) is also majority natural gas, with some distillate oil, however, there is a visible trend in the decrease in distillate oil HAP emissions over time in residential areas, representing roughly 25% of the total residential HAP estimated emissions in 1990 to less than 20% in 2010.

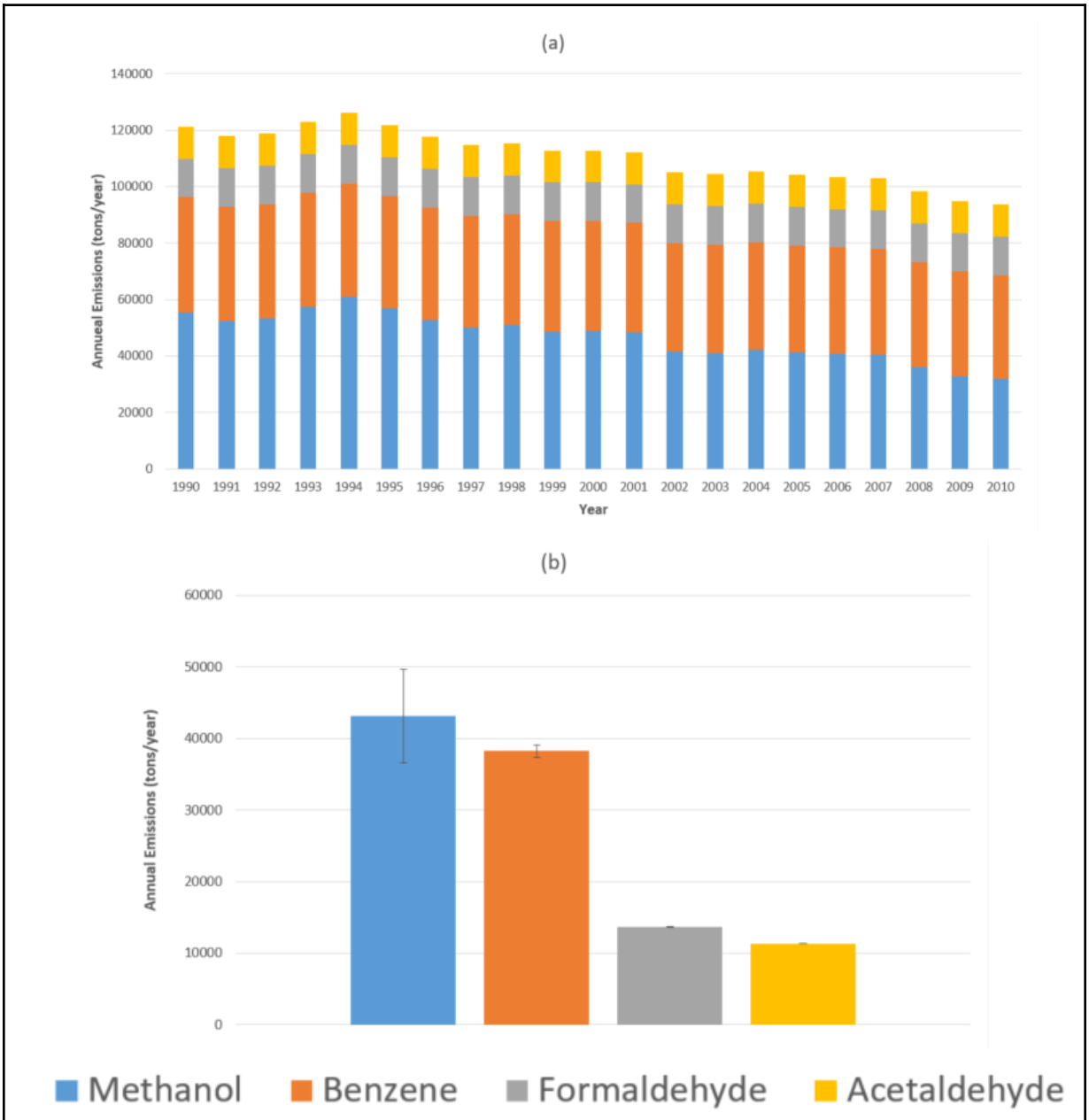

Figure 4: (a) 20-year estimated emissions trends of 4 major primary reactive species of HAPs (benzene, acetaldehyde, formaldehyde, methanol) based on FIPS/SCC schema estimated totals. Estimated emissions were summed based on matching POLCODE for each year, then plotted in the stacked bar graph. **(b)** Mean of annual emissions values for each individual HAP from 1990-2010, with error bars showing standard deviation.

Figure 4 offers another interpretation of the FIPS/SCC schema, looking at the data from the perspective of estimated individual HAP trends over the 20-year period rather than their sources of emission. Looking at the estimated emissions trends of 4 major primary reactive HAP species (benzene, acetaldehyde, formaldehyde, methanol) in Figure 4a, there is a consistent trend of an overall decrease in the data that has been prevalent throughout the other figures in tables. However based on Figure 4b, the majority of this decrease in emissions is due to the decrease in estimated Methanol production throughout the 20-year timeframe, as the standard deviation of Methanol is much higher than that of all other pollutants, and it also represents the largest share of emissions out of the four HAPs. Methanol experiences the largest decrease during the given timeframe, from 55393.28 tons/year to 31785 tons/year, a 43% decrease in emissions, followed by a 9.8% decrease in Benzene. Formaldehyde and acetaldehyde remained basically unchanged over the period of time under investigation (less than a 1.0% decrease).

4. Discussion

Overall, this project developed an original sense of long-term HAP emissions trends in the United States by modifying the data from historical emissions reports that lacked in-depth HAP emissions data. HAP weighting factor profiles were developed for 13 sectors in the 2016 NEI report that can be applied to other years of NEI profiles lacking data, allowing for new long-term analyses of historical emissions trends. In this project, the 6 surrogate profiles developed for the nonpt sector were utilized to generate estimates for 12 HAPs from 1990 to 2010. For all of the levels of aggregation, the nonpt profile successfully identified and provided estimates for over half of the entries in any given year, despite the fact that many SCCs are no longer reported on, visible in Table 3. Figure 1 shows that toluene was the most prominent HAP in terms of volume of emissions in 2010 in the entire United States. All 6 levels of aggregation demonstrated similar emissions trends for HAPs compared to the actual historical value of total VOCs, all decreasing over the 20-year time frame. These emissions profiles can be used to approach long-term HAP analyses in multiple ways, for example, the data can be approached by analyzing the changes of total HAP emissions in different sectors, such as in Figure 3, or by examining changes in individual HAPs similar to the approach in Figure 4.

As past NEI releases did not have strict requirements regarding HAP reporting, these profiles contribute an element of consistency throughout the data and contribute to the knowledge base of HAP emissions as a whole. Further, previous studies that suffered from a lack of specific historical information could benefit from these more precise estimates.[11] Although many studies have been performed examining past emissions trends, there has been much less work done in projecting the future for HAPs, specifically.[7] Thus, the weighting factors developed in this project's profiles could have applications in studies regarding projections of future HAP emissions as well.

There are many aspects of this project that could be examined further. Although the profiles have been developed for 13 sectors, complete estimates were only developed for the combined nonpt sector (representing both nonpt and np_oilgas in the 2016 NEI data). Thus, emissions estimates could be generated for the remaining sectors. Additionally, the results of Table 5 suggest that the emissions for some sectors had constant reporting for the 20-year timeframe in the historical data, meaning there is room for refinement in the estimates. One unexpected result was the State/6SCC schema matching roughly 11,000 more entries than the FIPS/6SCC schema, yet the latter estimated more tons/year worth of HAP emissions, leaving

room for future analysis as well. The number of unmatched entries for all 6 schemas also suggests a lack of clarity on the overall emissions picture in the US for this timeframe. Thus, additional research could be done with present NEI data and available HAP emissions (i.e. 2011 and 2014 reports) to compare the accuracy of the emissions estimates generated by this project's profiles with an existing dataset.

5. References

1. "Ambient (Outdoor) Air Pollution." World Health Organization, World Health Organization, 22 Sept. 2021, https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

2. McDonald, Brian C., et al. "Long-Term Trends in Nitrogen Oxide Emissions from Motor Vehicles at National, State, and Air Basin Scales." Journal of Geophysical Research: Atmospheres, vol. 117, no. D21, 2012, https://doi.org/10.1029/2012jd018304.

3. Dallmann, Timothy R., and Robert A. Harley. "Evaluation of Mobile Source Emission Trends in the United States." Journal of Geophysical Research, vol. 115, no. D14, 2010, https://doi.org/10.1029/2010jd013862.

4. "Basic Information of Air Emissions Factors and Quantification." EPA, Environmental Protection Agency, https://www.epa.gov/air-emissions-factors-and-quantification/basic-information-air-emissions-factors-and-quantification.

5. Anand, S.S., et al. "Volatile Organic Compounds." Encyclopedia of Toxicology, 2014, pp. 967–970., https://doi.org/10.1016/b978-0-12-386454-3.00358-4.

6. "2017 National Emissions Inventory: January 2021 Updated Release, Technical Support Document." U.S. Environmental Protection Agency, Feb. 2021.

7. Huang, Kaixin, and Matthew J. Eckelman. "Estimating Future Industrial Emissions of Hazardous Air Pollutants in the United States Using the National Energy Modeling System (NEMS)." Resources, Conservation and Recycling, vol. 169, 2021, p. 105465., https://doi.org/10.1016/j.resconrec.2021.105465.

8. Rajabi, Hamid, et al. "Emissions of Volatile Organic Compounds from Crude Oil Processing – Global Emission Inventory and Environmental Release." Science of The Total Environment, vol. 727, 2020, p. 138654., https://doi.org/10.1016/j.scitotenv.2020.138654.

9. Miller, R. R., et al. "Styrene Production, Use, and Human Exposure." Critical Reviews in Toxicology, vol. 24, no. sup1, 1994, https://doi.org/10.3109/10408449409020137.

10. Strafella, Elisabetta, et al. “Occupational Styrene Exposure Induces Stress-Responsive Genes Involved in Cytoprotective and Cytotoxic Activities.” PLoS ONE, vol. 8, no. 9, 2013, https://doi.org/10.1371/journal.pone.0075401.

11. Xing, J., et al. “Historical Gaseous and Primary Aerosol Emissions in the United States from 1990 to 2010.” Atmospheric Chemistry and Physics, vol. 13, no. 15, 2013, pp. 7531–7549., https://doi.org/10.5194/acp-13-7531-2013.

12. Bray, Casey D., et al. “An Assessment of Important Speciate Profiles in the EPA Emissions Modeling Platform and Current Data Gaps.” Atmospheric Environment, vol. 207, 15 June 2019, pp. 93–104., https://doi.org/10.1016/j.atmosenv.2019.03.013.

13. Simon, Heather, et al. “The Development and Uses of EPA’s Speciate Database.” Atmospheric Pollution Research, vol. 1, no. 4, 2010, pp. 196–206., https://doi.org/10.5094/apr.2010.026.

14. Hsu, Ying, et al. “SPECIATE Version 5.0.” U.S. Environmental Protection Agency, June 2019.

15. “SMOKE v4.8.1 User’s Manual.” Smoke V4.8.1 User's Manual, The University of North Carolina at Chapel Hill, 2020, https://www.cmascenter.org/smoke/documentation/4.8.1/html/index.html.

16. “National Emissions Inventory (NEI).” EPA, Environmental Protection Agency, https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei.

17. “Technical Support Document (TSD) Preparation of Emissions Inventories for 2016v1 North American Emissions Modeling Platform.” U.S. Environmental Protection Agency, Sept. 2020.

18. “Toluene.” U.S. Environmental Protection Agency, 2016.