

The development and uses of EPA's SPECIATE database

Heather Simon¹, Lee Beck², Prakash V. Bhawe¹, Frank Divita³, Ying Hsu³, Deborah Luecken¹, J. David Mobley¹, George A. Pouliot¹, Adam Reff⁴, Golam Sarwar¹, Madeleine Strum⁴

¹ US EPA, National Exposure Research Laboratory, Atmospheric Modeling and Analysis Division, 109 TW Alexander Drive, Research Triangle Park, NC 27711, USA

² US EPA, National Risk Management Research Laboratory, 109 TW Alexander Drive, Research Triangle Park, NC 27711, USA

³ E.H. Pechan & Associates, 5528B Hempstead Way, Springfield, VA 22151, USA

⁴ US EPA, Office of Air Quality Planning and Standards, 109 TW Alexander Drive, Research Triangle Park, NC 27711, USA

ABSTRACT

SPECIATE is the U.S. Environmental Protection Agency's (EPA) repository of volatile organic compounds (VOCs) and particulate matter (PM) speciation profiles of air pollution sources. These source profiles can be used to (1) provide input to chemical mass balance (CMB) receptor models; (2) verify profiles derived from ambient measurements by multivariate receptor models (e.g., factor analysis and positive matrix factorization); (3) interpret ambient measurement data; and (4) create speciated emission inventories for regional haze, climate, and photochemical air quality modeling. This paper describes the SPECIATE v4.2 database, provides specific examples of its use, and makes recommendations for future improvements.

Keywords:

SPECIATE

Emissions

VOC

Particulate matter

Speciation profiles

Article History:

Received: 01 April 2010

Revised: 27 May 2010

Accepted: 27 May 2010

Corresponding Author:

Heather Simon

Tel: +1-919-541-1803

Fax: +1-919-541-1379

E-mail: simon.heather@epa.gov

© Author(s) 2010. This work is distributed under the Creative Commons Attribution 3.0 License.

doi: 10.5094/APR.2010.026

1. Introduction

The SPECIATE database is an important product of the U.S. Environmental Protection Agency (EPA) that serves as the repository for source category-specific emission speciation profiles. The profiles contain weight fractions of chemical species of both volatile organic compounds (VOCs) and particulate matter (PM). The profiles are necessary since emission inventories typically contain reactive VOCs and total PM; however, many applications require information on the individual species and compounds which comprise the VOCs and PM mass. The weight fractions of individual VOC species are used in atmospheric chemistry mechanisms within air quality models to predict ozone concentrations in the atmosphere. The profiles of PM species weight fractions are specific to particle size ranges and are used to support air quality modeling for PM and visibility. SPECIATE supports air toxics assessments and assessments of compounds important to global climate (e.g., methane and black carbon), and is essential for source-receptor modeling applications. Data used to create these profiles come from a variety of sources including peer-reviewed journal articles and emissions testing conducted primarily by the EPA. The original source of the data used in each profile is documented in the SPECIATE database. Because the data come from a variety of sources, their quality varies greatly. The SPECIATE database includes a quality indicator field that ranks the quality of specific profiles.

SPECIATE was computerized in 1988 and the first electronic version was distributed to the user community in 1993. Although accessibility to the SPECIATE database has been sustained through EPA's Clearing House for Inventories and Emission Factors (CHIEF) website, updates to SPECIATE languished in the mid-1990s due to decreasing budgets. The U.S. National Research Council in its report on Research Priorities for Airborne Particulate Matter (NRC, 2004), the Clean Air Act Advisory Committee in its report of the Air Quality Management Working Group (Clean Air Act Advisory Committee, 2004), the North American Research Strategy for Tropospheric Ozone in its Emission Inventory Assessment (NARSTO, 2005), and other groups have recommended that the database be extensively updated and maintained in a dynamic manner.

Given the importance of SPECIATE to the process of air quality management, an EPA SPECIATE Workgroup was organized in 2005 to update SPECIATE. The scope of work undertaken by the Workgroup is to:

- Update the SPECIATE databases to capture recent and scientifically-meritorious VOC, total organic gas (TOG), and PM speciation profile data available from EPA, state agencies, peer-reviewed literature, and other relevant data sources;
- Modify the structure of the SPECIATE database to allow for storage of important information underlying each profile (meta-

data such as sampling and analysis methods, overall subjective profile quality ratings);

- Link the new profiles to Source Classification Codes (SCCs) in the National Emissions Inventory (NEI); and
- Assign species to photochemical reactivity classes.

An overview of the latest SPECIATE database is provided followed by examples of potential applications for this tool: expansion of our understanding of pollutant sources, interpretation of ambient data, creation of inputs necessary for air quality models, and new estimations of compounds important to radiative forcing.

2. SPECIATE v4.2 Overview

The latest SPECIATE database, version 4.2, was published in a June 2009 report (U.S. EPA, 2009) that summarizes the development and provides guidance on use of the database. In addition, the report documents revisions to auxiliary data tables including the VOC-to-TOG conversion table and the SCC-to-SPECIATE profile cross-reference table. One major update over previous SPECIATE versions is the addition of a new pollutant category called "other gases". This category contains speciated mercury, nitrogen oxides (NO_x), and semi-volatile organic compounds (SVOCs) that do not fall into VOC and PM profile categories. The new table entitled "SVOC Splitting Factors" provides suggested SVOC partitioning factors in PM and gaseous phases based on Schauer et al. (1999). Note that the partitioning factor of each SVOC species is not universal, it depends on sampling conditions such as temperature and dilution. There are 237 other gases profiles incorporated into SPECIATE v4.2.

Another SPECIATE v4.2 major update includes the development of a web-based application to allow access to the SPECIATE database through a web interface. This web-based application facilitates direct download of SPECIATE source profiles (i.e., MS Access is not needed). The application allows data searches by pollutant, keyword, and category. The SPECIATE Web Browser can be accessed at <http://cfpub.epa.gov/si/speciate/>. To date, the initiative to update SPECIATE has produced in version 4.2:

- 3 326 PM source profiles,
- 1 624 organic gas source profiles,
- 237 other gases source profiles,
- A total of 2 207 unique species,
- Composite profiles for 58 (47 PM and 11 VOCs) source categories,
 - An updated SCC-to-SPECIATE profile cross-reference table,
 - VOC-to-TOG conversion factors for applicable gas profiles,
 - A protocol for expansion of the database,
 - Suggested partitioning factors for SVOCs in gas and PM phases,
 - A mapping of VOCs not previously in the SPECIATE database into model species commonly found in atmospheric chemical mechanisms.

2.1. Using SPECIATE to calculate source-specific emissions of individual compounds

Because the SPECIATE database covers a large range of pollutants and sources and is difficult to "view" per se, here we present summaries of the database that demonstrate how it might affect some of the major applications for which it is used. One of the most important applications of this database is to drive air quality modeling studies of ozone and PM. For these applications, the database would be used in conjunction with an inventory of VOC and PM, such as the National Emissions Inventory (NEI, 2010), in order to provide spatial-, temporal- and source-resolved emission

estimates of individual VOCs, PM components, and other modeled species. The NEI reports emissions of specific pollutants (such as benzene) as well as classes of pollutants (such as PM, TOG, and NO_x) that may be determined through source-specific emission measurements, mass balance, source-specific models, emission models such as the Motor Vehicles Emissions Simulator (MOVES), or emission factors. In this last case, emissions are calculated by multiplying an emissions factor by an activity rate [Equation (1)]. For instance, emissions of NO_x from an industrial boiler may be calculated by multiplying the total cubic feet of natural gas burned per year (activity rate) by the average NO_x produced per cubic foot of combusted fuel for that boiler type (emissions factor). If emissions are routed through a control device, then emission estimates are multiplied by a control factor to determine the total amount of pollutant emitted into the atmosphere.

$$E = (\text{Activity Rate}) \times (\text{Emissions Factor}) \times (\text{Control Factor}) \quad (1)$$

Reff et al. (2009) calculated source-specific emissions of particulate trace metals by multiplying emissions of fine particulate matter (PM_{2.5}) from each source by an appropriate chemical speciation profile. This technique requires first grouping NEI emissions by source type and then determining the most representative profile for each category. Here we apply a similar technique to organic gasses: We determine total emissions of specific gaseous organic species by multiplying emissions of TOG for an individual source type by the weight fraction of the specific compound reported in the speciation profile for that same source type [Equation (2)],

$$E_{ij} = E_{\text{TOG},j} X_{ij} \quad (2)$$

where E is the mass emissions of compound i from source type j , and X is the weight fraction of compound i with respect to total TOG from source type j . For instance, if source tests have determined that ethylene makes up X (%) of tailpipe TOG emissions from on-road gasoline vehicles, we multiply the total TOG emissions from those vehicles by $X/100$ to obtain ethylene emissions from on-road gasoline vehicles.

In order to illustrate the SPECIATE database and its significance for ozone and toxic air pollution studies, we use the above methodology to generate information on emissions for specific VOC compounds by source category. To accomplish this, we use the complete set of TOG profiles in the SPECIATE database and link them with VOC emissions through an emission processing system. Emissions and SPECIATE profiles used in this study are consistent with EPA's 2005-based modeling platform version 4. Details about the emissions processing and chemical speciation profiles can be found elsewhere (U.S. EPA, 2010a). Using a set of SCC-to-SPECIATE profile cross references and VOC-to-TOG profile-based factors, we create a SCC-to-SPECIATE compound table. This master table is an easy way to view the VOC NEI inventory using the lens of the SPECIATE database and allows us to quickly analyze which source sectors are major contributors to nationwide emissions of specific VOC compounds that are important for ozone formation and toxics exposure. Since the master table is a very large sparse matrix, only selected parts of it can be visualized at any time. For the rest of this section, we will view different subsets of the master table to highlight various aspects of the VOC inventory from an air quality modeling viewpoint.

2.2. Sources of VOCs ranked by ozone-formation potential

Figure 1 shows total U.S. emissions for the 50 VOC species estimated to contribute the most mass on a typical day in July 2005. These 50 species comprise approximately 80% of the total mass emissions of VOCs. However, not all VOCs are equal in their

ability to produce ozone: one gram of ethene (C₂H₄), for example can produce 33 times more ozone than one gram of ethane (C₂H₆) under the same conditions (Carter, 2010). In Figure 2 we rank the VOC species according to their potential for ozone formation, by displaying the species that make up 80% of the total mass-weighted reactivity towards ozone formation. Reactivity is determined by multiplying the total emissions of each VOC by its Maximum Incremental Reactivity (MIR) value (Carter, 1994; Carter, 2010), which provides a way to compare quantitatively how different types of VOCs react differently in producing ozone. While this does not predict the actual amount of ozone formed, which depends on ambient atmospheric conditions, the MIR is a widely-used scale that allows us to perform a relative comparison of the importance of different VOCs (Hales, 2007; Luecken and Mebust, 2008) so that VOC emission reductions can focus on those VOCs which are most effective in reducing ozone. Some compounds are ranked high in both Figures 1 and 2, such as xylene (reported here as the sum of all 3 isomers). Another compound, isopentane, has a low MIR (1.28) and a high emissions (5th largest by mass) and therefore still has the potential to form high amounts of ozone

(17th largest). Even compounds that are exempted from EPA's definition of reactive VOC because they form little ozone on a unit basis (40 Code of Federal Regulations 51.100), have the potential to contribute to ozone because their emissions are high. For example, acetone is found to be 44th largest in terms of ozone formation out of approximately 1 000 VOCs stored in SPECIATE.

The two figures highlight the potential importance of mobile source VOC emissions in ozone production, both by mass and reactivity. Solvent utilization is also an important source of the ozone precursors: xylene, mineral spirits, acetone and alcohols. This method provides insight into the contribution of different source categories to the VOCs that are most important to ozone formation. This type of analysis can help to identify major sources of ozone precursors and direct focus for more in-depth studies but should not be used alone to make policy decisions because the actual impact of emitted pollutants from these sources depends on many factors including the co-emitted NO_x, biogenic emissions, and the meteorological conditions.

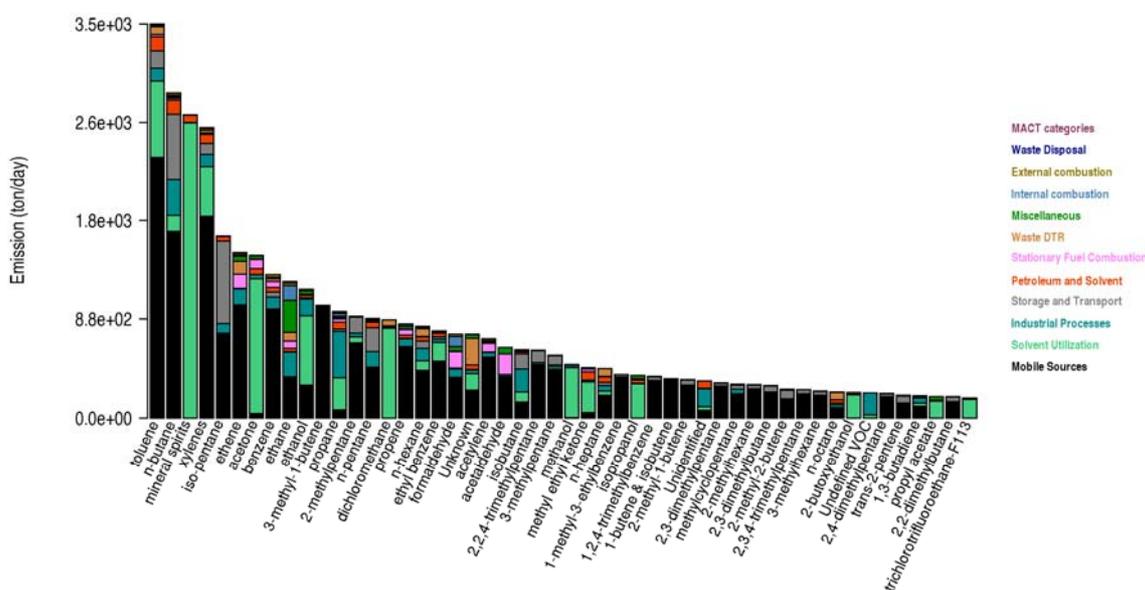


Figure 1. 2005 Emissions (metric tons day⁻¹) for a typical July day of the largest 50 VOCs using the SPECIATE profiles and the NEI inventory [Equation (2)].

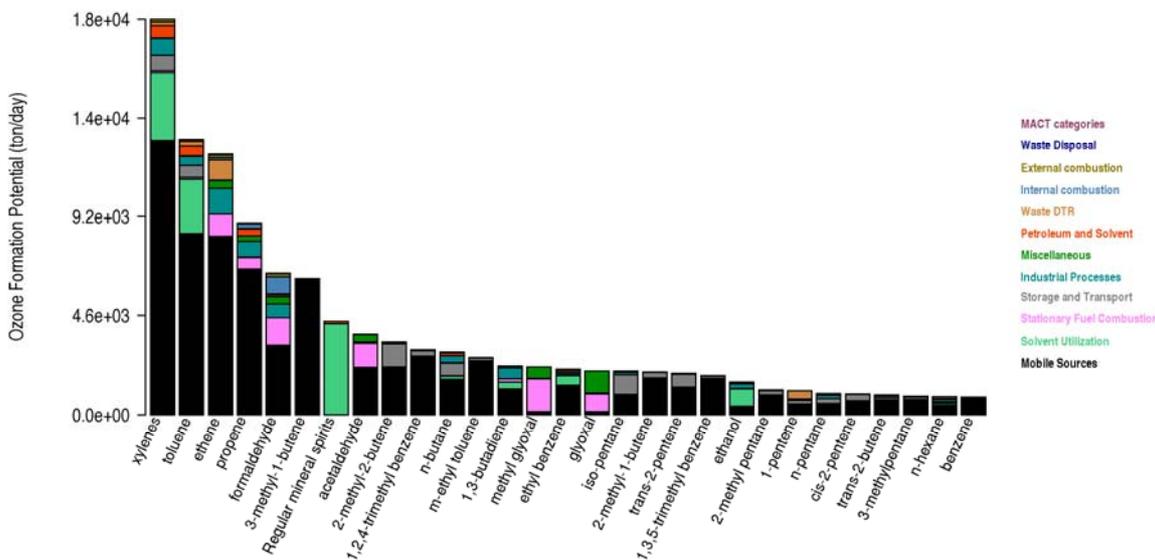


Figure 2. Ranking of the chemicals comprising the top 80% of total potential ozone reactivity, created by multiplying emissions from Figure 1 with the compound-specific MIR values.

2.3. Sources of toxic VOCs

Many of the species in Figures 1 and 2 are also of interest because EPA classifies them in Section 112(b) of the Clean Air Act as hazardous air pollutants (HAPs). The Clean Air Act has identified 187 explicit compounds as HAPs (U.S. EPA, 2010b), because they are known or suspected to cause cancer or other serious health effects. We now focus on 14 of the most toxic VOC HAPs, identified in the National Air Toxics Assessment (U.S. EPA, 2010c) as national and regional drivers or contributors to human health risk. In Figure 3, these 14 compounds are ranked by their relative toxicity in terms of non-cancer risk. The ranks were determined by dividing the total emissions by the reference concentration (RfC), defined as the concentration below which no effects are noted. Lower RfCs indicate pollutants with higher relative toxicity. Acrolein has a reference concentration of $2.0 \times 10^{-5} \text{ mg m}^{-3}$ that is at least two orders of magnitude lower than any of the other gas-phase HAPs, with the exception of toluene diisocyanate (RfC of $7.0 \times 10^{-5} \text{ mg m}^{-3}$). This low RfC, combined with a moderately high emission rate (130th highest mass out of 1 000 matched VOC species in the SPECIATE database), makes acrolein an important species to address in terms of toxicity. While toluene diisocyanate has a similar RfC, its mass emissions are only 13% of acrolein emissions, and is therefore less important.

Figure 4 shows the relative importance of these HAPs for cancer risk. The ranking is calculated by weighting the HAP emissions by the Unit Risk Estimate (URE), which quantifies the number of cancer cases over a lifetime for a unit concentration of each pollutant—the higher the URE, the higher the cancer risk of the pollutant. These rankings are more closely related to the total emissions than the non-cancer rankings, with benzene having both high emissions (7th largest) and a moderately high URE ($7.8 \times 10^{-6} \mu\text{g}^{-1} \text{ m}^3$). As in Figures 1 and 2, formaldehyde and 1,3-butadiene are also identified as potentially important. Because naphthalene has a high URE ($3.0 \times 10^{-5} \mu\text{g}^{-1} \text{ m}^3$) even though only moderately high emissions (99th), it also shows up as important.

Figures 3 and 4 show that mobile sources and to a smaller extent, solvent utilization sources contribute substantially to the mass-weighted cancer and non-cancer potential of VOC emissions. The actual risk of each source category to human exposure will also depend on a variety of factors including timing and proximity of both sources and human populations, meteorological conditions, local terrain, exposure route and duration, and many others.

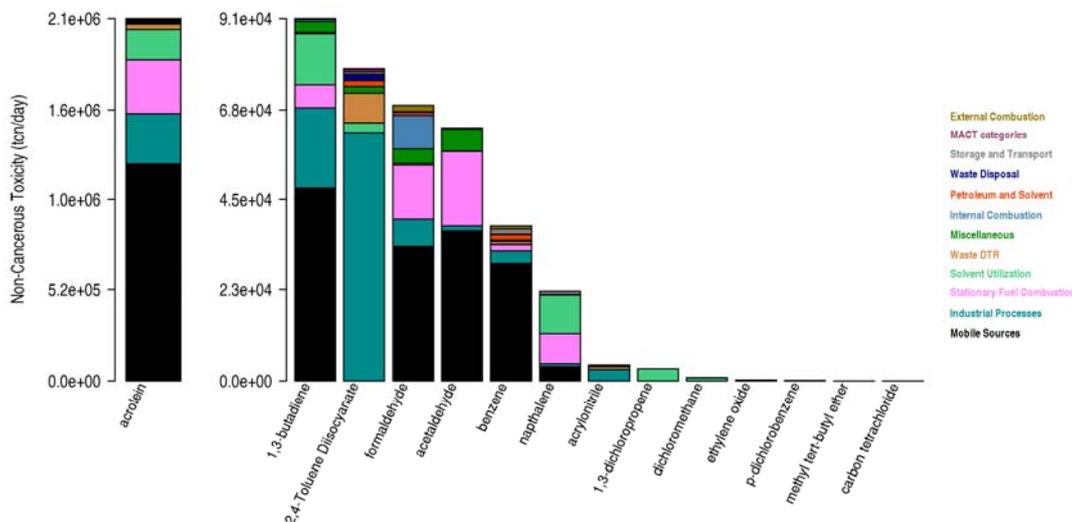


Figure 3. Relative toxicity weighting of 14 high risk HAPs with respect to their non-cancer effects, based on emissions in Figure 1 (metric tons day⁻¹) divided by the compound-specific RfC (mg m⁻³).

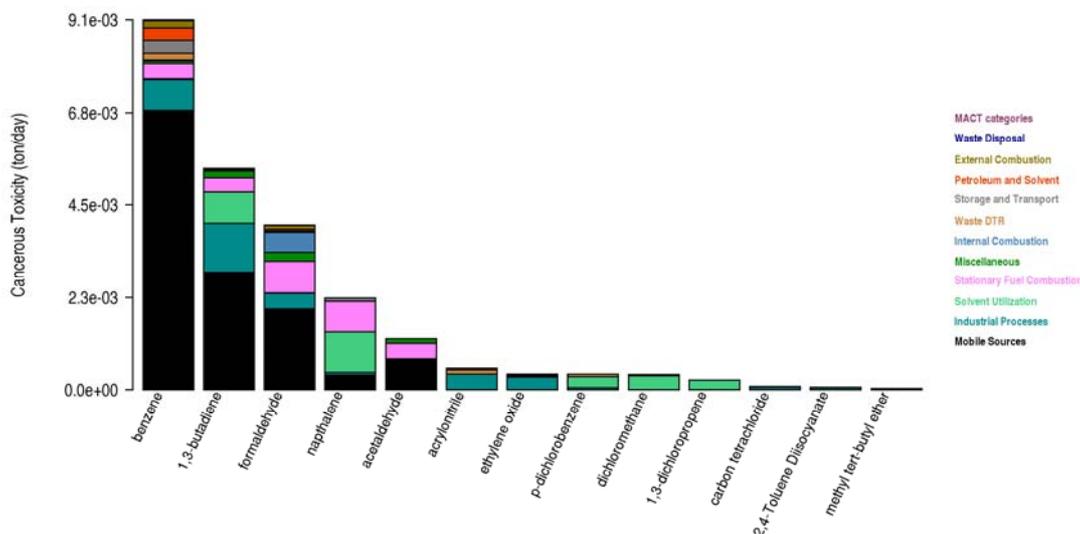


Figure 4. Relative toxicity weighting of 14 high risk HAPs with respect to their cancer effects, based on emissions in Figure 1 (metric tons day⁻¹) multiplied by the compound-specific URE ($\mu\text{g}^{-1} \text{ m}^3$). Acrolein has no URE for cancer and is not shown.

2.4. Comparison of SPECIATE results to NEI toxics estimates

In recent years, EPA has undertaken efforts to develop an inventory of HAPs, and include it as part of the NEI (Pope et al., 2002). To the extent that emission factors are used in the NEI, HAP emissions from the NEI and those obtained from Equation (2) could be identical if the same set of source data were used to create the emission factors and speciation profiles. However, data sets used to create speciation profiles must contain a complete characterization of the individual species making up the total whereas species-specific emissions factors or total VOC emission factors require only measurement of a single compound. Therefore, speciation data sets are much more expensive to conduct and, thus, are much more limited in number. As a result, perfect agreement between specific HAPs in the NEI and speciated VOCs from equation (2) is not expected. In addition, there has not previously been coordination between groups assembling the NEI and SPECIATE to use consistent datasets where they are available. In cases where data are derived from the same sources, we expect for the emission estimates from the two approaches to be reasonably consistent. In cases where they are significantly different, additional investigation is warranted for the pollutant and/or the source category.

In Figures 5 and 6, we show the same information as in Figures 3 and 4, but we have extracted the HAP emissions directly from the

NEI, rather than through speciation of the VOC emissions. While the general identification of the most significant pollutants is very similar, the magnitude of emissions and the contributions of major sources sometimes differ significantly. Figure 7 compares emissions for 96 different HAPs derived from the SPECIATE database versus those extracted directly from the NEI. Correlations for individual source categories vary from near zero (no correlation) for the “Internal Combustion” and “MACT” categories to 0.96 for the “Mobile Source” category. More investigation is needed for the categories that show especially low R^2 values (below 0.3) in order to estimate emissions of HAPs with more confidence. These categories include external combustion sources, industrial processes, MACT categories, waste disposal, treatment and recovery, and miscellaneous sources. Even for source categories in which correlations are high, the magnitudes of emission estimates are often not consistent between these two methods. For instance, mobile source and internal combustion HAP emissions estimates in the NEI are respectively 31% lower and 81% lower than those calculated using SPECIATE. When comparing emission estimates across all source categories, cancer-weighted emission estimates from the subset of 14 high-risk HAPs from the NEI inventory are 48% lower than from SPECIATE calculations, and non-cancer weighted emissions are 50% lower.

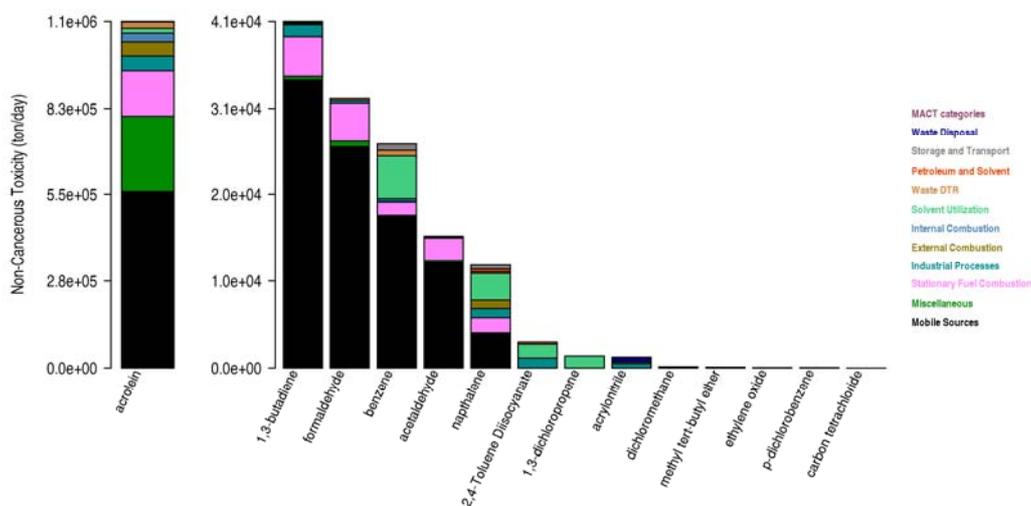


Figure 5. Relative toxicity weighting of 14 high risk HAPs with respect to their non-cancer effects, based on emissions extracted directly from the NEI (metric tons day⁻¹) divided by the compound-specific RfC (mg m⁻³).

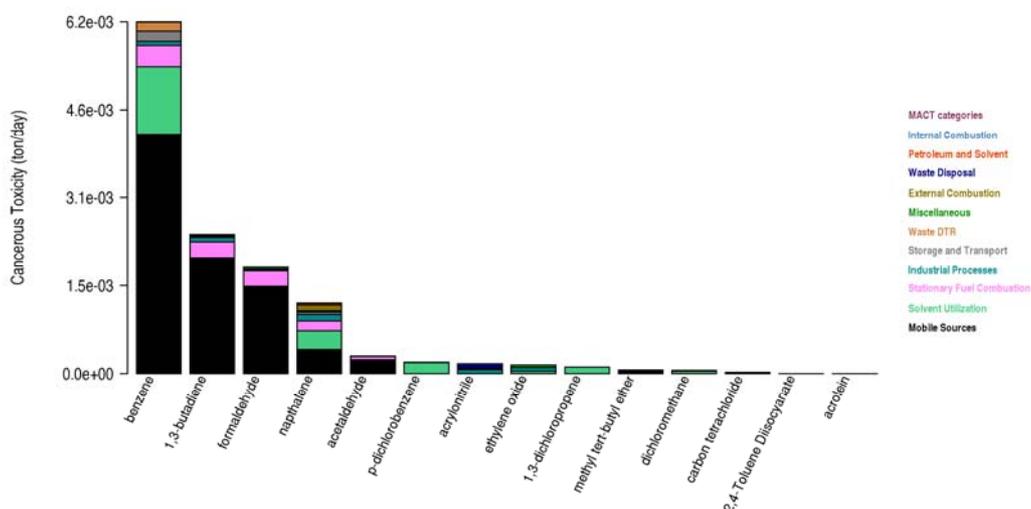


Figure 6. Relative toxicity weighting of 14 high risk HAPs with respect to their cancer effects, based on emissions extracted directly from the NEI (metric tons day⁻¹) multiplied by the compound-specific URE ($\mu\text{g}^{-1} \text{m}^3$) for each HAP. Acrolein has no URE for cancer and is not shown.

There are numerous reasons for these discrepancies, including differences between the speciation fractions and methods used to develop the HAP emission in the NEI. One example is with the mobile sources category. In the NEI, VOC and HAPs are consistent as they are generated using the same emission factor model and activity estimates. However, the EPA's mobile source emissions modeling methodology for benzene accounts for the percent of benzene in the fuel, whereas the speciation profile does not account for this fuel property. In addition, emission models used by the EPA account for changes in evaporative and exhaust emissions of specific compounds for different vehicle technologies, while SPECIATE profiles do not. In this case, it is obvious that the direct emission model estimate should be used rather than the speciation profile. However, an emission model is only applicable to the most important air toxics while SPECIATE has a complete list of all such compounds. The example of emissions from commercial charbroiling is less clear. Emission factors used in the NEI for HAPs were based on testing done to characterize emissions from cooking operations from street vendors in Mexicali, California (EPA, 1999), whereas the speciation profile is based on speciated organic gas emission measurements of hamburger meat charbroiling at a large institutional-scale natural gas fired charbroiler (Schauer et al., 1999). For this source category, it would be appropriate to conduct an evaluation to determine which sampled source is more representative of typical emissions in the United States.

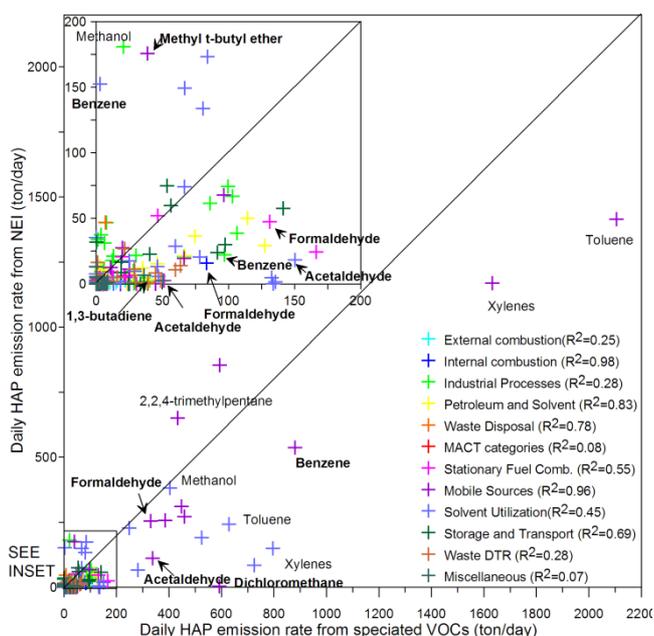


Figure 7. Emissions of 96 HAPs from 12 source categories via Equation (2) versus those extracted directly from the NEI. A few of the major HAPs and/or species with large differences are indicated.

In general, since the reporting of HAP emissions is not mandated with the same rigor as VOC reporting, the source of the emission data of the HAP and VOC emissions can be different for the same emission source and potentially inconsistent. For example, a particular emission source in the NEI may have its HAP emissions based on the toxics release inventory and its VOC (and other criteria air pollutants) reported by a state agency. The NEI does not report any benzene emissions from "Waste Disposal, Treatment, and Recovery; On-site Incineration", yet the VOC speciation profile for this source category includes benzene. EPA has been attempting to rectify the inconsistencies between VOC and HAP in the NEI that result from the inventory not being fully integrated. Strum et al. (2006) were the first to report on this in an investigation of the 2002 NEI.

Analyses using SPECIATE, as shown here, can help identify and prioritize sources of toxics that have a large impact on risk or are significant ozone contributors for which speciation profile or emission estimation methods can be improved. The SPECIATE database is especially valuable for this type of analysis because the speciation profiles have been recently evaluated and they will continue to be updated as fuels, industrial usage, and solvent compositions change over time. For example, refineries have produced many different gasoline fuels over the years to meet Federal and State regulations as well as for economic reasons. A relatively high-risk HAP, Methyl t-butyl ether (MTBE), was a common oxygenated fuel additive for many years but has now been phased out due largely to concerns about contamination of water supplies from leaking underground gasoline tanks. Ethanol has been used as an additive to replace MTBE. In addition to ethanol, refineries may add butane and other highly volatile species to boost ignition for winter season or to increase the Octane number. Because on-road vehicles are major sources of compounds with both a high ozone-formation potential and high toxicity, the SPECIATE database critically needs accurate profiles for liquid gasoline, tailpipe emissions, and headspace speciation from a large variety of different gasoline samples around the U.S. We plan to update SPECIATE on a periodic basis with refined profiles as new information becomes available. Updates will be posted at EPA's CHIEF website (<http://www.epa.gov/ttn/chief/>). Although SPECIATE is a valuable tool in this type of analysis, caution should be used to choose appropriate and high quality profiles.

3. Applications

3.1. Source apportionment of air pollution

SPECIATE can assist in providing a complete picture of pollutant sources. One of the major intended uses of SPECIATE is as a library of source profiles for use in source apportionment, particularly receptor modeling. At least two broad classes of receptor models currently exist: the chemical mass balance (CMB) approach, and models that employ some form of factor analysis. Both approaches seek to fit the following model to ambient species measurements:

$$x_{ij} = \sum_{k=1}^p a_{ik} \cdot s_{kj} + e_{ij} \quad (3)$$

where x_{ij} is the measurement in sample i of species j , p is the number of source profiles, a is the contribution of source k to sample i , s is the fraction of species j in profile k , and e_{ij} is the residual error.

In the CMB approach, predetermined source profiles (s_{kj}) are input to Equation (3) and values of a_{ik} are fitted to make the source profiles sum to measured ambient air samples. Ideally source profiles would be derived from measurements of species emissions in a given study's local airshed, but the logistics of such measurements can frequently be prohibitive. SPECIATE can provide the profiles for input to CMB when measurements from local sources are unavailable. This application has already seen usage by researchers such as Balakrishna and Pervez (2009) and Chen et al. (2001).

In the factor analytic approach, no species profiles are given to the model. Rather, various mathematical techniques are used to decompose a dataset of species measurements taken over various time periods into a product of matrices that have the form of Equation (3). Finding a solution in which there is confidence that s_{kj} values represent those of source profiles is a challenging undertaking, and numerous techniques such as principle components analysis (PCA) and positive matrix factorization (PMF) have evolved

over the years to solve this problem. SPECIATE can assist the factor analytic approach by serving a confirmatory role. Species profiles output by factor analysis models can be compared to data in SPECIATE to help determine their source identity. Reff et al. (2007) even posed the possibility of automated matching of PMF outputs to SPECIATE profiles, although we have not yet determined an acceptable metric for such profile matching.

3.2. Interpretation of ambient data

SPECIATE profiles can help to further understand ambient measurements, such as interpreting trace element measurements. Elements (quantified by XRF or PIXE) are often used to estimate total soil dust concentrations. This information is important both for identification of major PM source categories as well as characterization of physical aerosol properties such as light scattering and absorption. The Interagency Monitoring of PROtected Visual Environments (IMPROVE) monitoring network currently measures speciated PM concentrations at 13 urban and 178 remote sites (IMPROVE, 2010). Soil mass is calculated from the IMPROVE trace elements with Equation (4), which was developed based on a single soil profile (McDade, 2008):

$$SOIL = 3.48 Si + 1.63 Ca + 2.42 Fe + 1.94 Ti \quad (4)$$

El-Zanan et al. (2005) and Malm and Hand (2007) have suggested that this equation may not accurately represent total soil mass for all soil types. SPECIATE offers an extensive database against which this hypothesis can be tested. Here, we compile 51 soil profiles from SPECIATE v4.2 and from the published literature (Mori et al., 2002; Kim et al., 2003; Labban et al., 2004; Kandler et al., 2007; Cao et al., 2008; Jeong, 2008; Zhang et al., 2008; Abed et al., 2009). Using Equation (5), we then calculate the ratio of the actual soil mass to the mass obtained by applying the IMPROVE soil equation [Equation (4)]. The weight percentages of Si, Ca, Fe, and Ti in soil were calculated based only on the mineral portion of the soil mass (i.e. %Si = 100 Si/[soil mass – (soil sulfate + soil nitrate + soil ammonium + soil EC + soil OC)]).

$$\frac{M_{actual}}{M_{IMPROVE}} = \frac{100}{3.48 (\%Si) + 1.63 (\%Ca) + 2.42 (\%Fe) + 1.94 (\%Ti)} \quad (5)$$

Figure 8 shows the results of these calculations grouped by soil type. The boxes indicate the 25th and 75th percentile values and the whiskers extend to the maximum and minimum values. This analysis shows that the IMPROVE soil equation can either under- or overestimate the soil mass in ambient aerosols by as much as 50%. For example, actual agricultural soil mass is 13% lower than that estimated by the IMPROVE equation, but actual desert soil mass is 32% higher than the mass predicted by the IMPROVE equation.

In another example, we use SPECIATE to interpret lead in road dust over time as lead has been phased out of gasoline. Reff et al. (2009) have reported that unpaved road dust is the single largest source of Pb in the U.S. emission inventory. They attributed this to the sheer magnitude of total PM_{2.5} emissions from unpaved roads (700 000 tons year⁻¹) coupled with a relatively small mass fraction of Pb in the source samples (0.03%). This raises two policy-relevant questions:

1. What portion of the Pb emissions is due to the natural content of Pb in Earth's crust?
2. Is the anthropogenic contribution to these Pb emissions decreasing over time?

We can begin to address these questions by comparing profiles of unpaved road dust available in the SPECIATE database with published values of Pb measured in the Earth's crust. In Figure 9, Pb values from all 122 unpaved road dust profiles in SPECIATE are plotted against the year in which the source samples were

collected. The source samples appear to show a declining trend in Pb content with time, presumably associated with the phase out of leaded gasoline that began in 1974. In a few of the recent source samples, the Pb content in unpaved road dust approaches that of the Earth's crust (gray line in Figure 9). Using 17 ppm as the natural background (McLennan, 2001), one can deduce that only 6% of the unpaved road dust Pb inventory of Reff et al. (2009) is attributable to the natural crustal composition of Earth. This analysis illustrates a unique utility of the SPECIATE database. By serving as a historical repository of source profile data rather than a library of profiles representing a single receptor-modeling study period, it is possible to study temporal trends in the source contributions to atmospheric pollution.

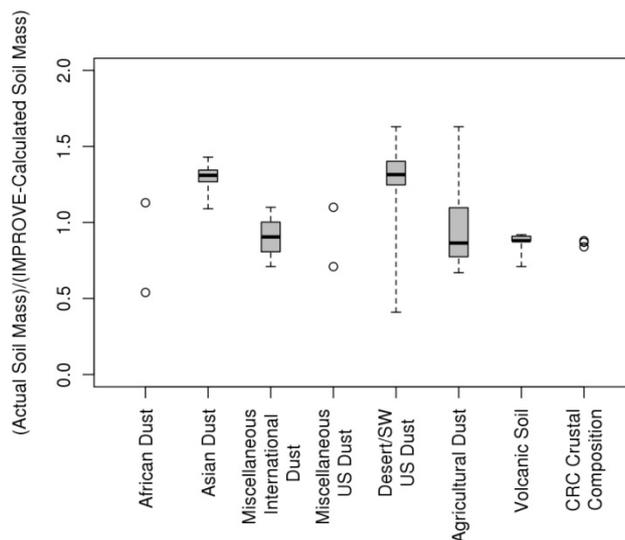


Figure 8. Bias in IMPROVE soil equation for different soil types.

3.3. Inputs for regulatory models

Photochemical air quality models are used to simulate air pollution fate and transport and are important tools in the regulatory process. Within these models, chemical reactions that affect concentrations of major pollutants (ozone, NO_x, VOC, PM) are represented using simplified chemical mechanisms. Common chemical mechanisms either group compounds based on reactivity with hydroxyl radicals or break compounds into functional groups which necessitates developing speciated emissions. EPA's NEI contains estimates of total anthropogenic emissions of NO_x, VOC, PM, and other pollutants in the US.

SPECIATE is routinely used to convert total emissions from specific sources into the speciated emissions needed for models. Here we compare the impact of using VOC and PM profiles from the new SPECIATE database relative to those from previous modeling platforms in the Community Multiscale Air Quality (CMAQ) modeling system (version 4.6) (Byun and Schere, 2006). In this example, CMAQ was applied to simulate air quality for the eastern United States for July 2001 at a horizontal grid resolution of 12-km. Meteorological data for the model were generated using the Pennsylvania State University/National Center for Atmospheric Research Mesoscale Modeling System Generation 5 (MM5) (version 3.5) (Grell et al., 1994). Anthropogenic emissions were derived from the 2001 National Emissions Inventory and the biogenic emissions were generated using the Biogenic Emissions Inventory System (version 3.13) (Schwede et al., 2005).

We examine the differences in using the two profiles by calculating the Mean Bias (MB):

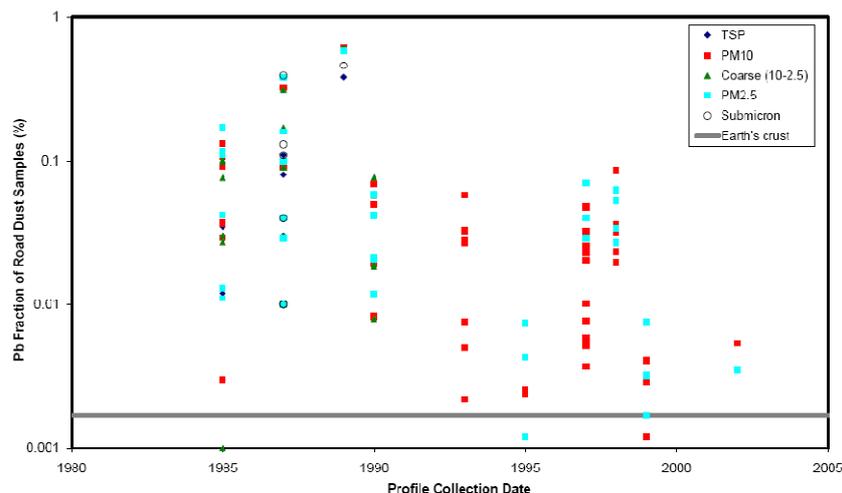


Figure 9. Lead content in thirty years of road dust samples included in the SPECIATE database. Note that 24 road dust profiles are taken from an earlier version of SPECIATE in which the test year was not recorded. For plotting purposes, those source samples are assumed to be collected in 1985.

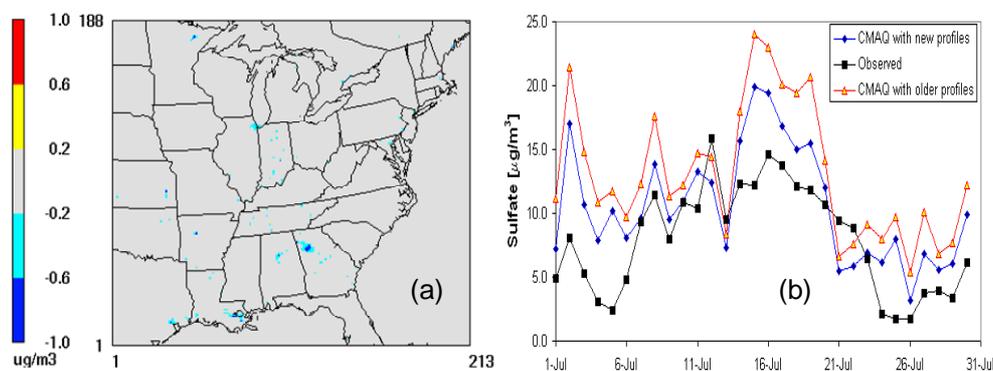


Figure 10. (a) Difference in monthly mean aerosol sulfate obtained with new and older emissions speciation profiles (new minus old), (b) comparison of predicted aerosol sulfate with observed data from Jefferson Street monitoring site in Atlanta, Georgia.

$$MB = \frac{1}{N} \sum_{1}^N (C_m - C_o) \quad (6)$$

where N is the number of observations, C_m is model predicted concentration and C_o is the observed concentration. MB for ozone is estimated by comparing model results with ambient monitoring data from EPA's Air Quality System (AQS). MB for aerosols is estimated by comparing model results with ambient monitoring data from the IMPROVE, Clean Air Status and Trends Network (CASTNet), Speciation Trends Network (STN), and Southeastern Aerosol Research and Characterization (SEARCH) networks. Differences in the monthly mean daily maximum 8-hr ozone between the two scenarios are less than 0.4 ppbv. Changes to MB for aerosol sulfate at the IMPROVE, STN, and CASTNet is less than $0.5 \mu\text{g m}^{-3}$. However, aerosol sulfate obtained with new profiles are different than those obtained with older emission profiles at SEARCH network sites. As shown in Figure 10a, mean aerosol sulfate between the two scenarios are similar except for some urban areas. The largest impact occur in Atlanta, Georgia where aerosol sulfate obtained with the new emissions speciation profiles decreased by almost $5.0 \mu\text{g m}^{-3}$ compared to those obtained with the older emission speciation profiles. Predicted and observed aerosol sulfate from Jefferson Street monitoring site at SEARCH network in Atlanta, Georgia are presented in Figure 10. Predictions obtained with both scenarios capture the day-to-day variability of observed aerosol sulfate. While predictions obtained with both scenarios are greater than observed data, predicted sulfate obtained with newer emissions speciation profiles are closer to and

in better agreement with observed data. Mean observed aerosol sulfate was $8 \mu\text{g m}^{-3}$ while mean predicted values with the new and older emissions speciation profiles are $10.5 \mu\text{g m}^{-3}$ and $13.1 \mu\text{g m}^{-3}$, respectively. MB for aerosol sulfate at this location improve from $5.1 \mu\text{g m}^{-3}$ with the old emissions speciation to $2.6 \mu\text{g m}^{-3}$ with the new emissions speciation profiles.

In the future, the new other gases profiles in SPECIATE may also help to improve model predictions. Current air quality models seriously under-predict nitrous acid (HONO) compared to observed data. Sarwar et al. (2008) compared CMAQ (CMAQv4.6) predicted HONO concentrations to observed data from the 2001 Northeast Oxidant and Particle Study and reported that predicted HONO concentrations were significantly lower than the observed data. The model did not capture the diurnal variability of the observed data either. CMAQv4.6 did not include HONO emissions and heterogeneous reactions that can produce HONO in the atmosphere. When HONO emissions and heterogeneous reactions were incorporated into the model, comparison of predicted HONO to observed data and the diurnal variability also improved. In that study, HONO emissions from on-road and off-road vehicles were estimated using a value of 8×10^{-3} for the HONO/NO_x emissions ratio following the study of Kurtenbach et al. (2001). The new SPECIATE database contains data for speciation of NO_x into NO, NO₂, and HONO for aircraft engines. Such data can be used to speciate NO_x emissions from aircraft into NO, NO₂, and HONO emissions for use in air quality models to improve model predictions of HONO concentrations.

3.4. Inventory and analysis of compounds affecting radiative forcing

Although SPECIATE was originally envisioned to help in analyzing sources of air pollution, it may prove valuable in quantifying the compounds important for global climate modeling. Methane (CH_4) is an important greenhouse gas, with a global warming potential about 21 times as large as carbon dioxide (CO_2) on a 100-year horizon (Ramaswamy et al., 2001). In combination with the TOG emissions inventory, speciation profiles in the SPECIATE database can be applied to calculate a CH_4 emission inventory and used for greenhouse gas emissions control policy, rule making, and greenhouse gas inventory reporting purposes. For example, as much as 25% of the TOG emissions from biomass burning (e.g., grassland and forest fires, fireplaces) are in the form of CH_4 , whose emissions can be quantified by multiplying source specific speciation profiles to TOG emissions. Similarly, CH_4 from spark-ignition vehicles, which emit about 12% of TOG as CH_4 , can be estimated in the same manner. However, since the fraction of TOG that is methane varies somewhat among vehicle technologies, use of an emissions model that will specifically calculate methane is preferred.

Black carbon (BC) is a PM species with climate warming potential but there is currently a large uncertainty about the magnitude of this forcing. PM speciation profiles in the SPECIATE database have been applied by several state agencies to calculate elemental carbon (EC) emissions and transform the mass emission estimates for EC into carbon dioxide equivalents (CO_{2e}) in order to present the emissions within a greenhouse gas context, with the assumption that EC is the same as BC. The BC mass emission estimates were derived by multiplying the PM emission estimates by the appropriate aerosol fraction for BC. The method is based on the modeling of Jacobson (2002) and his updates to this work (Jacobson, 2005).

4. Recommendations

This paper has introduced the SPECIATE v4.2 database and has shown its value in a number of applications. It provides a valuable repository of information of source samples collected over a 40 year period. SPECIATE v4.2 and its data browser enable the community to characterize emissions by species and source category, thus benefiting the air quality modeling and source-receptor modeling applications. Additional efforts are needed to capture newly collected source profiles based on data submitted via the protocol for database expansion. The user community can support SPECIATE development by supplying electronic data with full references.

While the database has been revised and many profiles have been added, the SPECIATE Workgroup has identified and prioritized needs for profiles in future versions of the SPECIATE database. The current priority is to update mobile sources profiles, coal-fired utility profiles especially for control processes, and profiles for sources of black carbon. This is an ongoing process in which newly identified data sources of speciation profiles, such as a newly discovered database or a published technical paper, are prioritized by the Workgroup and added to the database according to priority ranking and available resources.

In addition, previous reviews have outlined areas for future development. An emission inventory assessment (NARSTO, 2005) concluded that accurate speciation profiles are essential for making appropriate regulatory decisions about air quality policy. The assessment emphasized the need to update VOC speciation profiles for both anthropogenic and biogenic source categories and the continuing challenge of adding or assessing the contributions from specific compounds as new chemicals are identified as health risks. NARSTO's ten priorities included size-segregated, speciated emissions of fine particles and their precursors including black and

organic carbon emissions; toxic and hazardous air pollutants; emissions from on-road vehicles; emissions of ammonia from agricultural and other area sources; speciated, spatially and temporally resolved organic emissions from biogenic sources; emissions of VOCs and organic HAPs from petrochemical and other industrial facilities; emissions from off-road mobile sources including farm and construction equipment, aircraft and airport ground equipment, commercial marine facilities, and locomotives; emissions from open biomass burning, including agricultural and forest prescribed burning, wildfires, and residential backyard burning; residential wood combustion in woodstoves and fireplaces; and paved and unpaved road dust.

Reff et al. (2009) identified several key sources that lack specific $\text{PM}_{2.5}$ source profiles in the database: road construction dust, heavy construction dust, off-road diesel agricultural equipment, off-road diesel mining equipment, railroad locomotives, marine vessels, gasoline-powered boats, bagasse boilers, waste incinerators, solid waste boilers, and coal mining, cleaning and material handling. These sources are currently mapped to more general categories that may not accurately represent their emissions. Future studies and updates to SPECIATE should concentrate on obtaining speciation estimates for these source types.

Analysis presented in this paper has also identified some areas for improvement. Results from comparison of VOC emissions in the NEI combined with SPECIATE profiles versus direct estimates of toxics species in the NEI identifies internal combustion, MACT categories, solvent utilization, and mobile source as sources categories with significant differences that are candidates in need of additional investigation to identify, understand and correct the discrepancies. In the future, it is recommended that groups developing emissions factors and speciation profiles work collaboratively to make sure that the best data is used for both purposes.

Finally, if SPECIATE is to be used for CH_4 and BC emission estimates in the future, it will be critical to evaluate the profiles with radiative forcing metrics specifically in mind. Since CH_4 is not classified as a VOC by the EPA, previous evaluations of TOG speciation may not have focused particular attention on this compound. BC and EC are aerosol classifications that are often used interchangeably. However, in practice the definition for BC may encompass organic carbon compounds that absorb light but are not purely graphitic. More analysis is needed to determine if current EC speciation profiles are appropriate for estimating BC emissions.

The SPECIATE project is a work-in-progress; comments based on review of the database and documentation are welcome. Profiles characterizing the emissions from source categories are needed from all stakeholders in the air quality management process – industry, government, academia, contractors, and consultants. The protocol for supplying updated profiles is available in the SPECIATE v4.2 documentation, <http://www.epa.gov/nrmrl/pubs/600r09038/600r090>. Please submit your data for incorporation into the database so that analyses utilizing the SPECIATE database can be based on the best information available and air quality management decisions can be based on the latest scientific information.

Acknowledgements

The authors would like to thank Bill Lonneman, Doug Solomon, Joe Somers, and Tom Pierce from the US EPA for their helpful feedback on this work. The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency's administrative review and approved for publication.

References

- Abed, A.M., Al Kuisi, M., Khair, H.A., 2009. Characterization of the Khamaseen (spring) dust in Jordan. *Atmospheric Environment* 43, 2868-2876.
- Balakrishna, G., Pervez, S., 2009. Source apportionment of atmospheric dust fallout in an urban-industrial environment in India. *Aerosol and Air Quality Research* 9, 359-367.
- Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews* 59, 51-77.
- Cao, J.J., Chow, J.C., Watson, J.G., Wu, F., Han, Y.M., Jin, Z.D., Shen, Z.X., An, Z.S., 2008. Size-differentiated source profiles for fugitive dust in the Chinese Loess Plateau. *Atmospheric Environment* 42, 2261-2275.
- Carter, W.P.L., 2010. Development of the SAPRC-07 chemical mechanism. *Atmospheric Environment* (In Press, doi:10.1016/j.atmosenv.2010.01.026).
- Carter, W.P.L., 1994. Development of ozone reactivity scales for volatile organic compounds. *Journal of the Air and Waste Management Association* 44, 881-899.
- Chen, K.S., Lin, C.F., Chou, Y.M., 2001. Determination of source contributions to ambient PM_{2.5} in Kaohsiung, Taiwan, using a receptor model. *Journal of the Air and Waste Management Association* 51, 489-498.
- Clean Air Act Advisory Committee, "Recommendations to the Air Quality Management Working Group," December 2004.
- El-Zanan, H.S., Lowenthal, D.H., Zielinska, B., Chow, J.C., Kumar, N., 2005. Determination of the organic aerosol mass to organic carbon ratio in IMPROVE samples. *Chemosphere* 60, 485-496.
- Grell, G., Dudhia, J., Stauffer, D., 1994. A description of the fifth-generation Penn State/NCAR Mesoscale model (MM5), NCAR Tech. Note NCAR/TN-398+STR.
- Hales, J.M., 2007. The Reactivity Research Working Group-A Chronological Overview and Evaluation. In NARSTO, NARSTO: Pasco, WA, 2007.
- Interagency Monitoring of Protected Visual Environments (IMPROVE), 2010. <http://vista.cira.colostate.edu/improve/Overview/Overview.htm>.
- Jacobson, M.Z., 2005. Erratum: "Control of fossil-fuel particulate black carbon and organic matter, possibly the most effective method of slowing global warming" (Journal of Geophysical Research D: Atmospheres (2002) vol. 107 (D19) (4410) 10.1029/2001JD001376). *Journal of Geophysical Research D:Atmospheres* 110, 1-5.
- Jacobson, M.Z., 2002. Control of fossil-fuel particulate black carbon and organic matter, possibly the most effective method of slowing global warming. *Journal of Geophysical Research D:Atmospheres* 107, art. no. 4410.
- Jeong, G.Y., 2008. Bulk and single-particle mineralogy of Asian dust and a comparison with its source soils. *Journal of Geophysical Research D: Atmospheres* 113(D2), art. no. D02208.
- Kandler, K., Benker, N., Bunde, U., Cuevas, E., Ebert, M., Knippertz, P., Rodriguez, S., Schutz, L., Weinbruch, S., 2007. Chemical composition and complex refractive index of Saharan Mineral Dust at Izana, Tenerife (Spain) derived by electron microscopy. *Atmospheric Environment* 41, 8058-8074.
- Kim, K.H., Choi, G.H., Kang, C.H., Lee, J.H., Kim, J.Y., Youn, Y.H., Lee, S.R., 2003. The chemical composition of fine and coarse particles in relation with the Asian Dust events. *Atmospheric Environment* 37, 753-765.
- Kurtenbach, R., Becker, K.H., Gomes, J.A.G., Kleffmann, J., Lorzer, J.C., Spittler, M., Wiesen, P., Ackermann, R., Geyer, A., Platt, U., 2001. Investigations of emissions and heterogeneous formation of HONO in a road traffic tunnel. *Atmospheric Environment* 35, 3385-3394.
- Labban, R., Veranth, J.M., Chow, J.C., Engelbrecht, J.L.P., Watson, J.G., 2004. Size and geographic variation in PM₁, PM_{2.5}, and PM₁₀: source profiles from soils in the western United States. *Water Air and Soil Pollution* 157, 13-31.
- Luecken, D.J., Mebust, M.R., 2008. Technical challenges involved in implementation of VOC reactivity-based control of ozone. *Environmental Science and Technology* 42, 1615-1622.
- Malm, W.C., Hand, J.L., 2007. An examination of the physical and optical properties of aerosols collected in the IMPROVE program. *Atmospheric Environment* 41, 3407-3427.
- McDade, C.E., 2008. IMPROVE Standard Operating Procedure. Davis, CA, Crocker Nuclear Laboratory, University of California. Available at: http://vista.cira.colostate.edu/improve/Publications/SOPs/ucdavis_sop/sop351_V2.pdf.
- McLennan, S.M., 2001. Relationships between the trace element composition of sedimentary rocks and upper continental crust. *Geochemistry Geophysics Geosystems* 2, art. no. 2000GC000109.
- Mori, I., Nishikawa, M., Quan, H., Morita, M., 2002. Estimation of the concentration and chemical composition of kosa aerosols at their origin. *Atmospheric Environment* 36, 4569-4575.
- National Research Council, 2004. Research Priorities for Airborne Particulate Matter: IV. Continuing Research Progress, National Academies Press, Washington, DC, 2004.
- NARSTO, 2005. Improving Emission Inventories for Effective Air Quality Management Across North America—A NARSTO Assessment, NARSTO-05-001, September 2005.
- National Emissions Inventory (NEI), 2010. Available at: <http://www.epa.gov/ttn/chief/net/2005inventory.html#inventorydata>.
- Pope, A., Billings, R., Finn, S., Wilson, D., 2002. The development of the 1999 national inventory for HAPs. Presented at the 11th International Emissions Inventory Conference: Emissions Inventories- Partnering for the Future, Atlanta, GA.
- Ramaswamy, V., Boucher, O., Haigh, J., Hauglustaine, D., Haywood, J., Myhre, G., Nakajima, T., Shi, G.Y., Solomon, S., 2001. Radiative forcing of climate change, in Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, edited by J. T. Houghton et al., pp. 349–416, Cambridge Univ. Press, Cambridge, U. K.
- Reff, A., Bhawe, P.V., Simon, H., Pace, T.G., Pouliot, G.A., Mobley, J.D., Houyoux, M., 2009. Emissions inventory of PM_{2.5} trace elements across the United States. *Environmental Science and Technology* 43, 5790-5796.
- Reff, A., Eberly, S.I., Bhawe, P.V., 2007. Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. *Journal of the Air and Waste Management Association* 57, 146-154.
- Sarwar, G., Roselle, S.J., Mathur, R., Appel, W., Dennis, R. L., Vogel, B., 2008. A comparison of CMAQ HONO predictions with observations from the Northeast Oxidant and Particle Study. *Atmospheric Environment* 42, 5760-5770.
- Schauer, J.J., Kleeman, M.J., Cass, G.R., Simoneit, B.R.T., 1999. Measurement of emissions from air pollution sources, 2. C1 through C30 organic compounds from medium duty diesel trucks. *Environmental Science and Technology* 33, 1578-1587.
- Schwede, D., Pouliot, G., Pierce, T., 2005. Changes to the biogenic emissions inventory system version 3 (BEIS3), 4th Annual CMAS Models-3 Users' Conference, September 26-28, 2005, UNC-Chapel Hill, NC. Available at http://www.cmascenter.org/html/2005_conference/abstracts/2_7.pdf.
- Strum, M., Houyoux, M., Mason, R., Beidler, A., Stanley, C., Luecken, D., 2006. EPA's MultiPollutant Modeling with the 2002 NEI: Integrating Criteria and Toxics. In 15th International Emission Inventory Conference, New Orleans, LA, 2006.
- U.S. Environmental Protection Agency, 2010a. Available at: <http://www.epa.gov/ttn/chief/emch/index.html#2005>.
- U.S. Environmental Protection Agency, 2010b. Available at: <http://www.epa.gov/ttn/atw/orig189.html>, <http://www.epa.gov/ttn/atw/pollutants/atwsmod.html>.
- U.S. Environmental Protection Agency, 2010c. Available at: <http://www.epa.gov/ttn/atw/natamain/>.

- U.S. Environmental Protection Agency, "SPECIATE 4.2—Speciation Database Development Documentation, EPA/600-R-09/038, (<http://www.epa.gov/ttn/chieff/software/speciate/index.html>), Research Triangle Park, NC, June 2009.
- U.S. Environmental Protection Agency, 1999. Emissions from Street Vendor Cooking Devices (Charcoal Grilling), U.S. EPA, Washington, DC, June, 1999.
- Zhang, R., Han, Z.W., Shen, Z.X., Cao, J., 2008. Continuous measurement of number concentrations and elemental composition of aerosol particles for a dust storm event in Beijing. *Advances in Atmospheric Sciences* 25, 89-95.