PECHAN

MEMORANDUM

Date: May 30, 2005

To: Lee Beck, U.S. Environmental Protection Agency, Office of Research and Development

From: Y. Hsu and S. Roe, E.H. Pechan & Associates, Inc.

Subject: Protocol for Expansion of the SPECIATE Database
EPA Contract No. 68-D-00-265, WA No. 4-46

This memorandum is intended to guide profile data collectors on how to collect and present source profile data to maximize their utility to SPECIATE users, to assist future SPECIATE managers in assessing whether the data should be incorporated, and to facilitate the mechanics of the actual inclusion.

Background

In order to ensure that future profile development meets the needs of the SPECIATE user community, Pechan has prepared several recommendations for speciation profile developers based on recent SPECIATE database updates and previous guidance from EPA (EPA, 2002) and other scientists (Watson and Chow, 2002). For this discussion, SPECIATE users are defined as individuals who: (1) conduct regional haze, PM2.5, and ozone modeling; (2) prepare speciated emissions inventories; (3) use the Chemical Mass Balance or other receptor models; (4) and/or verify profiles derived from ambient monitoring measurements by multivariate receptor models such as UNMIX.

Speciation Data Collection

Profiles are defined as the mass fractions of chemical species that make up a source-specific emission stream. Volatile organic compound (VOC) profiles should include the mass fractions of each of the species present. When all organic gas species are present (e.g. methane), these profiles are referred to as total organic gas (TOG) profiles. At a minimum, these profiles should include the 56 Photochemical Assessment Monitoring Station (PAMS) species, as well as any other species that are available.

Particulate matter (PM) profiles should include mass fractions for each of the species present. Minimum data requirements are for the major elements reported by the IMPROVE and PM2.5 Speciation Trends networks, water-soluble ions (sulfates and nitrates at a minimum, plus ammonium, potassium, sodium, chloride, fluoride, phosphate, calcium, and magnesium, if available), and carbon fractions (Total Carbon (TC), Organic Carbon (OC), and Elemental Carbon (EC)), preferably with other fractions that are defined by the method, such as the eight IMPROVE carbon fractions and carbonate carbon). Organic fractions, isotopic abundances, organic compounds, and single particle properties should be included, where they are reported

and well-defined. Test results from dilution sampling trains are recommended for use in SPECIATE, since these results come closest to representing the composition of emissions in the ambient air.

Profile data must contain information on the chemical abundance of each species noted above. These data can be defined as the fraction of mass emissions of PM/VOC/TOG or the mass emission rate of each species (e.g. lb/ton, g/VMT, etc.). In addition to the estimate of central tendency for each species (e.g. mean, median), an estimate of the variability of each species should also be provided (e.g. standard deviation). Priority should be given to profiles that express the mean and standard deviation of individual test profiles for representative samples. If statistics other than the mean and standard deviation are provided, the method used to estimate central tendency and variability should be described.

Available information on the analytical uncertainty for individual test profiles should be identified and described separately. For example, if the analytical method for a certain species is known to have a precision of +/- 20%, then this information should be listed for each applicable species.

Documentation

The primary reference for the profile should be cited as the source of documentation, not secondary references that might have compiled profile data from one or more primary references. Secondary references should be cited only when original profiles have been modified (i.e. by aerosol aging, different sample compositing, different normalization methods, etc.). The notes column in the SPECIATE database will be used to store this information, as well as additional descriptive information on the profile, such as vehicle model year, engine size, VIN, and other descriptors that might be used to document a mobile source profile.

Profile developers must provide extensive documentation of their results. This should include documentation of the entire experimental program. Where appropriate, this should include fuel type, operating parameters, type of facility, location, and date of test. Non-detects or incomplete analyses should be documented, so that the reader fully understands the analytical results.

Data Format

Profile developers should transmit data in a form that can be easily added to the SPECIATE database. The template in Table 1 is recommended. The SPECIATE data structure is completely documented in SPECIATE Database Update Documentation available at the SPECIATE project web site (http://projects.pechan.com/speciate). Information should be filled in as completely as possible, including references, test methods, analytical methods, Chemical Abstract System (CAS) numbers, data quality ratings, normalization basis, etc.

Data Normalization

Methods for profile normalization should be clearly documented, and the rationale for selecting the normalization basis should be stated. Normalization of organic gas data should be mass specific (i.e. mass species/mass TOG; emission rate species/emission rate TOG). Volume carbon basis is not recommended because it is more objective (assumptions are needed regarding the composition of unresolved species). Whenever possible, the total gas chromatograph (GC)-elutable organic gases normalization basis should be used and documented.

Normalization of PM data should be size-specific. Ideally, the profile will be normalized on total PM (with a specified upper size limit), PM10 and PM2.5. However, normalization based on other size fractions can also be accommodated in SPECIATE. The normalized mass can be measured or be the weighted sum of major chemical components (sulfate, nitrate, ammonia, soil elements with assumed or measured oxides, organic carbon, elemental carbon, and sea salt). Profiles normalized on total gravimetric mass are preferred; however if the sum of measured species basis is used, this should be noted and the reasoning for selecting this method stated.

Speciation Data Quality

Recommendations for or against inclusion of profiles in the SPECIATE will be based on the perceived overall quality of the profiles. There are no simple criteria that can be set to scrutinize speciation data for inclusion in the SPECIATE database. The supporting information housed within SPECIATE is therefore critically important. The SPECIATE database provides structure sufficient to thoroughly document profiles and their underlying analysis, and should be completed as thoroughly as possible when preparing profiles for potential inclusion in the database.

Each profile has a quality rating that is assigned by the profile developer. The quality rating protocol is completely documented in SPECIATE Database Update Documentation available at the SPECIATE project web site (http://projects.pechan.com/speciate). Speciation profiles developed from the following methods should be given a lower data quality rating:

1. Samples from combustion sources not collected by dilution sampling;
2. Low total speciated fractions (less than 80%);
3. PM profiles normalized by the “sum of species” mass, which assumes profiles of this type are fully speciated; and
4. Any noticeable outliers or other unreasonable test results (see examples provided below).

Additional profile quality considerations include:

- **Appropriate Method** – Reviewers experienced in analytical methods and application of speciation profiles will need to determine if characteristic compounds are present and properly measured. Sampling and analytical procedures need to be specific to the

source and documented as thoroughly as possible. For example, a recent study conducted for CARB used EPA Method TO-14 in a study of dairy farm emissions. Since this method was developed to test industrial sources, fatty acids and other important organic species were not included in the target species list.

- **Measurement Precision** – Low precision is expected for certain species; the data quality ratings should reflect this issue. In cases where the sampling or analytical methods are found to be wholly inappropriate for a given species, these data should not be included in SPECIATE. For example, it was recently found that the wet chemistry DNPH sampling procedure is not appropriate for acrolein measurement due to its poor recovery. .

- **Overall Test Program Confidence** – Results obtained from the test program should be consistent with expectations for that source, and if not, the differences should be sufficiently accounted for. For example, in an U.S. Air Force sponsored study (AFIERA/RSEQ, 1998) measuring aircraft exhaust compositions, a brief discussion in the measurement section showed that the contractor measured essentially the same concentrations of target compounds in the background air as in the samples collected from aircraft exhaust. As a result, toxic species were reported at relatively low emission rates in this study. In cases where there are significant unexplainable results, the data should not be included in the SPECIATE database.

- **Source Category-specific Considerations** – For certain source categories such as the pulp and paper industry, oxygenated compounds contribute significantly to organic gas emissions. The generic total hydrocarbon (THC) method using flame ionization detectors (FID) calibrated with hydrocarbon standards (e.g. hexane) does not properly characterize the total TOG or VOC emissions. For processes whose emissions are dominated by methanol, this compound (and other oxygenated species) should be sampled and quantified separately using gas chromatography calibrated with a methanol standard (see Someshwar 2003). Due to poor detector performance, the emission rates measured for THC were observed to be less than those measured specifically for methanol using an appropriate standard. Consequently, for this case, the THC is not suitable to serve as the normalization basis for this gas profile. The solution is to collect fully speciated data using appropriate methods and to consolidate all organic gases into a total organic gas profile for normalization.

Table 1. Descriptive Data Dictionary

	Field	Type[1]	Length[2]	Decimals	Description
					PM_PROFILE Table
Primary key	NUMBER	C	5		PM Profile Number (Primary Key)
	NAME	C	255		PM Profile Name
	QUALITY	C	3		Quality rating (A-E) of the profile
	CONTROLS	C	100		Emission Controls Description
	DATE	D			Date profile added
	NOTES	M			Notes
	TOTAL	N	6	2	Sum of species percentages for a given profile, excluding organic species in PM and inorganic gases in PM profiles.
	MASTER_POL	C	4		Indicates the pollutant to be used in calculation. Allowed value: 'PM' In the future, other values may be allowed (e.g. PM_PRI, PM_FIL, PM_CON)
	T_METHOD	M			Description of sampling method
	NORM_BASIS	C	25		Description of how profile was normalized
	ORIG_COMPO	C	1		Specifies whether the profile is original or composite. Allowed values: 'C','O'
	STANDARD	L	1		Indicates whether the profile is provided by the EPA SPECIATE (standard) or user-added. The database is constructed to allow users to add profiles.
	ORGANICS	L	1		Indicates whether or not the profile provides speciated organics in PM
	INCL_GAS	L	1		Indicates whether or not the profile includes inorganic gas species (e.g. SO_2, H_2S, NOx, etc.)
	TEST_YEAR	N	4		Indicates year testing was conducted
	J_RATING	N	4	2	Objective expert judgment rating based on general merit
	V_RATING	N	4	2	Vintage based on TEST_YEAR field (
	D_RATING	N	4	2	Data quality rating based on number of observations, robustness
	REGION	C	50		Geographic region of testing
	LOWER_SIZE	N	5	2	Identifies low end of aerodynamic diameter particle size, micrometers
	UPPER_SIZE	N	5	2	Identifies high end of aerodynamic diameter particle size, micrometers
Foreign key	SIBLING	C	5		GAS Profile number; samples taken from the same source and study, if exists.
					PM_SPECIES Table
Primary key	ID	N	9	0	Unique Identifier (Primary Key)

	Field	Type[1]	Length[2]	Decimals	Description
Foreign key	SPECIE_ID	N	9	0	Specie Identifier (The same as ID in SPECIE_PROPERTIES)
Foreign key	P_NUMBER	C	5		PM Profile number (Link to PM_Profile Table)
	PERCENT	N	7	3	Weight percent of pollutant (%)
	UNCERTAINT	N	7	3	Uncertainty percent of pollutant
	UNC_METHOD	C	25		Description of method used to calculate uncertainty
	ANALMETHOD	C	50		Description of Analytical method (e.g. XRF, IC, etc.)
				REFERENCE Table	
Primary key	ID	N	9	0	Unique Identifier (Primary Key)
Foreign key	P_TYPE	C	1		Indicates PM or GAS. Allowed values: P (PM), G (Gas)
Foreign key	P_NUMBER	C	5		Profile number (Link to PM_PROFILE and GAS_PROFILE tables)
	DATA_ORIGN	C	50		Source of data (e.g. EPA APPCD, Schauer, CARB, DRI, Literature, etc.)
	PRIMARY	L			Designates a reference as primary. When a profile is based on multiple references, this field allows one reference to be tagged as the primary reference.
	DESCRIPTIO	M			Stores the descriptive information about the profile.
	DOCUMENT	Object			Complete reference citation.
				GAS_PROFILE Table	
Primary key	NUMBER	C	5		GAS Profile Number (Primary Key)
	NAME	C	255		GAS Profile Name
	QUALITY	C	3		Quality rating (A-E) of the profile
	CONTROLS	C	50		Emission Controls Description
	DATE	D			Date profile added
	NOTES	M			Notes
	TOTAL	N	6	2	Sum of organic species percentages for a given profile
	MASTER_POL	C	4		Indicates the pollutant to be used in calculation. Allowed values: 'VOC', 'TOG', 'NMOG'.
	T_METHOD	M			Description of sampling method
	NORM_BASIS	C	25		Description of how profile was normalized
	ORIG_COMPO	C	1		Specifies whether the profile is original or composite. Allowed values: 'C','O'
	STANDARD	L	1		Indicates whether the profile is provided by the EPA SPECIATE (standard) or

	Field	Type[1]	Length[2]	Decimals	Description
					user-added. The database is constructed to allow users to add profiles.
	TEST_YEAR	N	4		Indicates year testing was conducted
	J_RATING	N	4	2	Objective expert judgment rating based on general merit
	V_RATING	N	4	2	Vintage based on TEST_YEAR field
	D_RATING	N	4	2	Data quality rating based on number of observations, robustness
	REGION	C	50		Geographic region of testing
Foreign key	SIBLING	C	5		PM Profile number; samples taken from the same source and study, if exists.
				GAS_SPECIES Table	
Primary	ID	N	9	0	Unique Identifier (Primary Key)
Foreign	SPECIE_ID	N	9	0	Species Identifier (Must be the same as ID in SPECIE_PROPERTIES)
Foreign	P_NUMBER	C	5		GAS Profile Number (Link to GAS_PROFILE table)
	PERCENT	N	6	2	Weight percent of pollutant (%)
	UNCERTAINT	N	7	3	Uncertainty percent of pollutant
	UNC_METHOD	C	25		Description of method used to calculate uncertainty
	ANALMETHOD	C	50		Description of Analytical method (e.g. GC/FID, GC/MS, HPLC/UV, etc.)
				KEYWORD Table	
Primary key	ID	N	9	0	Unique Identifier (Primary Key)
Logical key	P_TYPE	C	1		Indicates PM or GAS. Allowed values: P, G
Logical key	NUMBER	C	5		Profile Number (Link to PM_PROFILE and GAS_PROFILE Tables)
	KEYWORD	C	255		Keyword describing profile
				SPECIES_PROPERTIES Table	
Primary key	ID	N	9	0	Unique Identifier (Primary Key) (Link to PM_SPECIES and GAS_SPECIES tables)
	CAS	C	50		Chemical Abstract Service number assigned to pollutant (with hyphens) (blank if no CAS)
	EPA_ID	C	50		EPA Chemical Identifier; to be provided by EPA Substance Registry System for species without CAS.
	SAROAD	C	5		SAROAD code. There are many conflicts in EPA, CARB, DRI, and TCEQ versions; suggest to exclude SAROAD codes from the database.
	PAMS	L	1		Is PAMS pollutant? (Yes or No)
	HAPS	L	1		Is Hazardous Air Pollutant? (Yes or No)
	NAME	C	254		Pollutant name

	Field	Type[1]	Length[2]	Decimals	Description
	SYMBOL	C	9		Standard chemical abbreviation (provided by Eric Fujita, DRI)
	SPEC_MW	N	6	2	Species molecular weight
					MNEMONIC Table
Primary key	ID	N	9	0	Unique Identifier (Primary Key)
Foreign key	P_TYPE	C	1		Indicates PM or GAS. Allowed values: P (PM), G (Gas)
Foreign key	P_NUMBER	C	5		Profile number (Link to PM_PROFILE and GAS_PROFILE tables)
	DRI_PNUMBR	C	6		DRI profile number (Original DRI profile numbers)
	MNEMONIC	C	60		Alphanumeric Code unique to each profile. Used in CMB input files.

[1] Field types. C: Character; D: Date; L: Logical; M: Memo; N: Numeric; Object.
[2] Length – length allowed.

References:

AFIERA/RSEQ, 1998. *Aircraft Engine and Auxiliary Power Unit Emissions Testing for the US Air Force*, Environmental Quality Management Inc, and Roy F. Weston Inc., December 1998.

Someshwar, 2003. Arun Someshwar, *Compilation of 'Air Toxic' and Total Hydrocarbon Emissions Data for Sources at Kraft, Sulfite and Non-Chemical Pulp Mills – an Update*, Technical Bulletin No. 858, National Council for Air and Stream Improvement, February, 2003.

EPA, 2002. *Draft Guidelines for the Development of Total Organic Compound and Particulate Matter Chemical Profiles*, developed by Emission Factors and Inventory Group, U.S. EPA, September 25, 2002.

Watson and Chow, 2002. Watson, J. and J. Chow, *Considerations in Identifying and Compiling PM and VOC Source Profiles for the SPECIATE Database*, Desert Research Institute, August, 2002.