PM Fine Database and
Associated Visualization Tool

Steven Boone
Director of Information Technology
E. H. Pechan and Associates, Inc.
3622 Lyckan Parkway, Suite 2002; Durham, North Carolina 27707; 919-493-3144 x117

Dr. Frank Divita, Jr.
and
Brian Eisemann
E. H. Pechan and Associates, Inc.
5528-B Hempstead Way; Springfield, Virginia 22151; 703-813-6700

Venkatesh Rao
United States Environmental Protection Agency
Office of Air Quality Planning and Standards, Air Quality Trends Analysis Group
c/o OTAQ/ASD
2000 Traverwood Drive; Ann Abor, MI 48105; 734-214-4016

ABSTRACT

A national network of monitors that analyze the concentration of particulate matter less than 2.5 microns in diameter ($PM_{2.5}$) in the atmosphere has been deployed as a result of a 1997 presidential directive establishing the $PM_{2.5}$ National Ambient Air Quality Standard (NAAQS). This network is intended to facilitate reporting of $PM_{2.5}$ data to the general public. Special purpose monitors are utilized to help identify potential $PM_{2.5}$ problem areas, to assist in identifying their boundaries, to better define diurnal patterns, to help determine the spatial scale of high concentration areas, to help characterize the chemical composition of $PM_{2.5}$ especially on high concentration days.

In order to facilitate the analysis of Environmental Protection Agency's (EPA) urban $PM_{2.5}$ Speciation Network data for mass and composition and to compare the $PM_{2.5}$ Speciation Network data to $PM_{2.5}$ speciation data collected from the IMPROVE rural network, the PM Fine database and associated visualization tool were developed. PM Fine-related data were extracted from the EPA's Air Quality Subsystem (AQS) and the IMPROVE web site and were placed into a database structure designed to compatibly house data from both systems. The database structure was derived from static fields in the AQS and IMPROVE databases and with fields with values derived from calculations performed on existing AQS and IMPROVE fields.

Microsoft Visual Basic was used to build a graphical user interface viewer for the composite database. The tool allows users to select a geographic area like state and county or metropolitan statistical area (MSA), select monitors, temporal ranges and specific monitoring parameters. The tool produces statistics such as minimum, maximum, average and number of data points of the selected monitoring parameters. The tool produces bar charts and pie graphs to facilitate the comparison of data from multiple parameters.

INTRODUCTION

A national network of monitors that analyze the concentration of particulate matter less than 2.5 microns in diameter ($PM_{2.5}$) in the atmosphere has been deployed as a result of a 1997 presidential directive establishing the $PM_{2.5}$ National Ambient Air Quality Standard (NAAQS). This network is intended to facilitate reporting of $PM_{2.5}$ data to the general public. Special purpose monitors are utilized to help identify potential $PM_{2.5}$ problem areas, to assist in identifying their boundaries, to better define diurnal patterns, to help determine the spatial scale of high concentration areas, to help characterize the chemical composition of $PM_{2.5}$ especially on high concentration days.

The EPA $PM_{2.5}$ Speciation Database is a product of new regulatory initiatives for $PM_{2.5}$, ozone, and regional haze that have significantly expanded the requirements for monitoring ambient air from those that previously supported Clean Air Act (CAA) programs. Previous PM monitoring guidelines required at State, local and National Air Monitoring Sites (SLAMS and NAMS, respectively) collected PM_{10} mass concentration with high-volume samplers. New EPA monitoring initiatives expand the PM monitoring program to include installation of PM Speciation sites, in addition to providing funding for additional SLAMS and NAMS.

Since 1999, more than 50 Routine PM Speciation sites have been funded by EPA and installed in urban areas. The sampling schedule for EPA PM Speciation sites usually consists of one 24-hour sample every three days, although some sites operate daily. These sites are designed to provide chemical analyses of PM samples. The results of chemical analyses will 1) provide an improved understanding of the emissions and dynamic atmospheric processes that influence particle formation and distribution; 2) determine the sources of pollutants that contribute to elevated PM concentrations; 3) develop methods useful to decision makers in formulating and comparing candidate control strategies; and 4) provide reliable means for estimating the impacts of control strategy options developed for PM.

The Interagency Monitoring of Protected Visual Environments (IMPROVE) monitoring network has been operating for many years and contains a vast amount of data that can be used to supplement the EPA Speciation database, especially in rural areas. The network collects measurements that provide PM speciation data during time intervals comparable to EPA Speciation sites. For example, most IMPROVE samples collect a 24-hour filter sample on a 3-day or 6-day cycle.

Data validation is of the most important functions of data processing. Data validation is performed on several levels within the PM Speciation network. Sample validation consists of procedures which identify deviations from measurement assumptions and procedures. Three levels of validation are applied which will result in the assignment of a rating to each measurement: 1) valid; 2) valid but suspect; or 3) invalid.

Level I sample validation is raw data with field data cross-check flags, and Level II sample validation evaluates the physical consistencies and identifies outliers. Level I sample validation takes place in the field or in the laboratory and consists of the following: 1) flagging samples when significant deviations from measurement assumptions have occurred; 2) verifying computer file entries against data sheets; 3) eliminating values for measurements which are

known to be invalid because of instrument malfunctions; 4) replacement of data from a backup data acquisition system in the event of failure of the primary system; and 5) adjustment of measurement values of quantifiable calibration or interference biases.

Level II sample validation takes place after data from various measurement methods have been assembled in the master data base. Level II applies consistency tests based on known physical relationships between variables to the assembled data. Examples of these tests include the following: 1) the sum of all chemical species in a particulate matter sample should be less than or equal to the gravimetric mass of that sample; 2) size-segregated particle concentrations should be less than total particle concentrations; 3) the sum of all major species (with oxide forms included) should exceed 75% of the measured mass; and 4) analyses of the same species by different methods should yield compatible results (e.g., sulfur by x-ray fluorescence [XRF] and sulfate by ion chromatography [IC]).

The EPA Speciation and IMPROVE databases are assumed to be at validation level II. The IMPROVE database includes by default blank, corrected ambient measurements along with associated analytical uncertainties.

SYSTEM REQUIREMENTS

In order to facilitate the analysis of Environmental Protection Agency's (EPA) urban $PM_{2.5}$ Speciation Network data for mass and composition and to compare the $PM_{2.5}$ Speciation Network data to $PM_{2.5}$ speciation data collected from the IMPROVE rural network, the PM Fine database and associated visualization tool were developed. PM Fine-related data were extracted from the EPA's Air Quality Subsystem (AQS) and the IMPROVE web site and were placed into a database structure designed to compatibly house data from both systems. The database structure was derived from static fields in the AQS and IMPROVE databases and with fields with values derived from calculations performed on existing AQS and IMPROVE fields.

The requirements for the system were derived by EPA and included the ability to select monitoring sites by geographic area, time period and specific monitoring species. Business rules for the system stem from clean air regulations, the regional haze rule and visibility in Class I areas. Functionality includes the analyses of speciated PM data over different periods and geographic areas using various techniques such as tabular data and visual graphs like bar and pie charts.

Some of the requirements identified for the tool consisted of a desktop tool environment capable of supporting the functional requirements such as Microsoft Visual BASIC, and run-time operating systems of Microsoft Windows 95 to Windows XP. The tool was to employ a point-and-click interface and produce graphs using predetermined colors for specific species.

The initial functional hierarchy consisted of five major categories that included:

Base of Data

The base of data to be used for the PM Speciation Database consists of the Speciation Trends Network (STN) data residing in the EPA's Air Quality System (AQS) and Interagency

Monitoring of Protected Visual Environments (IMPROVE) Network data available from the web at *http://vista.cira.colostate.edu/improve/Default.htm*. The functionality provided enables viewing information from each database independently as well as the ability to compare the difference between STN and IMPROVE values data. Data can be selected by particulate species category.

Temporal Selection

The visualization tool provides the capability of selecting data by beginning month and ending month. Readings can be viewed summarized by month or for the entire period. The time period selection is dynamic and can be changed at the user's discretion.

Geographic Selection

Information can be selected by state and county, metropolitan statistical area (MSA) and/or Health Effects Institute (HEI) region.

Data and Graphs

Information for the selected site(s), time period(s) and species can be viewed in raw data format, statistics for each site (min, max, average and number of data points) and/or on stacked bar graphs and pie charts.

Base Data Updates

The visualization tool's database is preloaded with one year's data. Additional year's information can be downloaded from IMPROVE and AQS, when available, and imported into the tool for additional analyses.

DESIGN PARAMETERS

Before the development began, Microsoft Excel, Microsoft Visual FoxPro and Micorsoft Visual BASIC were considered for the tool's front end development environment. The data, while downloaded in ASCII format, was converted to Microsoft Visual FoxPro database table format for ease of handling and updates. Graphics Server Technologies' Graphics Server was selected as the tool to generate the graphs and charts for the application. Subsequently Microsoft Visual BASIC was selected as the primary development environment because it provides EPA with flexibility in implementation, distribution and enhancement choices and can seamlessly integrate all of the needed tools and features into one application.

DATABASE DESIGN

The database design accounts for the increase in size of the locally contained IMPROVE and STN data. A Visual FoxPro database table can handle very large quantities of data. The database design utilized open database connectivity (ODBC) so that other back-end databases can be used in the future, if needed. The internal design structure of the database incorporates a single flat file that encompasses data fields from both the IMPROVE and STN databases. There

are numerous internal tables that relate the information to the user such as geographic information, site information, species and species colors. All fields were designed to allow for the future import of STN and IMPROVE data. IMPROVE and STN structures were used for field constraints (type, length and decimals). This will expedite the transfer of data from IMPROVE and STN and will minimize potential data overflow and data mismatch problems.

PROGRAM FLOW

The application presents two main tools to the user, a Site Information tool and a Site Comparison tool as shown in Figure 1. A Graph Color Selector Tool is also available to manage the colors for each of the particulate species.

Figure1
PM Visualization Tool Main Screen
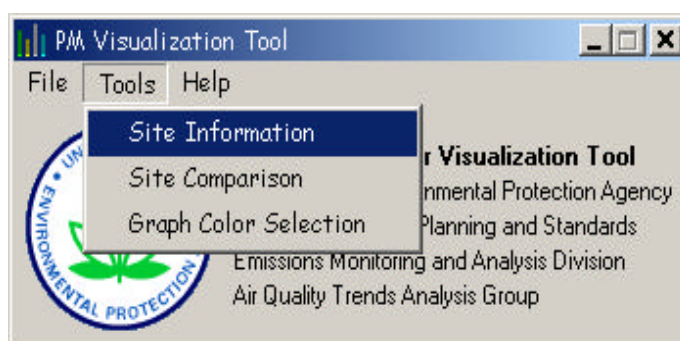
The first tool, Site Information, allows a user to view specific information about one or more sites of either STN or IMPROVE particulate data, as shown in Figure 2. The user works down the screen from left to right, top to bottom in making selections for viewing specific data. First, a user selects the desired date range of the data for viewing. Next a geographic range is selected, HEI region, state and county and/or MSA. A list of sites will appear for the selected geographic region, and the user may selected multiple sites to include in the analysis. The temporal period of the analysis is selected (monthly or entire period), and the desired species to be included in the analysis are checked. Finally the user has the option of viewing the raw data from STN or IMPROVE (Figure 3), statistics of the selected data (Figure 4) or a stacked bar graph (Figure 5). A pie chart (Figure 6) can be generated from the stacked bar graph by right clicking directly on the stacked bar chart. All graphs can be exported using the Export button on the lower left portion of the graph window. The graph export feature can create either Windows Bitmap (.BMP) or JPEG (.JPG) formatted graphics files for inclusion in reports, presentations and websites.

Figure 2
Site Information Tool Selection Screen

Figure 3
Site Information Tool Data Table View

Site 06073-0003-5 Data

Site ID	POC	Site Type	Sampler	Latitude	Longitude	Lab	Date	Al	Ca	Crustal	EC	Fe
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/05/13	0.069	1.129	0.260	0.105	1.010
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/05/19	0.022	0.272	0.290	0.036	1.370
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/05/25	0.024	0.380	0.580	0.053	3.360
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/05/31	0.028	0.566	0.820	0.087	5.060
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/06/06	0.061	0.637	0.610	0.082	4.620
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/06/12	0.055	0.440	0.410	0.042	1.120
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/06/18	0.063	0.700	0.590	0.098	1.010
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/06/24	0.049	0.389	0.320	0.045	1.030
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/06/30	0.043	0.428	0.560	0.064	2.020
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/07/06	0.034	0.310	0.460	0.034	0.980
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/07/12	0.025	0.370	0.310	0.054	0.720
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/07/24	0.058	0.707	0.474	0.093	1.670
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/07/30	0.065	0.584	0.460	0.068	1.480
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/08/05	0.066	0.949	0.462	0.127	2.200
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/08/11	0.061	0.540	0.440	0.078	2.700
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/08/17	0.099	1.167	0.657	0.153	3.450
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/08/23	0.071	1.152	0.529	0.128	1.770
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/08/29	0.069	0.826	0.355	0.092	2.840
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/09/04	0.063	0.986	0.519	0.125	1.160
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/09/10	0.002	0.047	0.308	0.003	1.940
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/09/16	0.080	0.921	0.510	0.091	2.610
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2001/09/22	0.091	1.129	1.120	0.175	2.770

Figure 4
Site Information Tool Statistics Table

Site 06073-0003-5 Stats

SiteID	Entire Period	Data Points	Crustal Min	Crustal Avg	Crustal Max	EC Min	EC Avg	EC Max
06073-0003-5	02/2000 - 02/2002	61	0.047	0.935	2.806	0.260	1.106	3.450

Figure 5
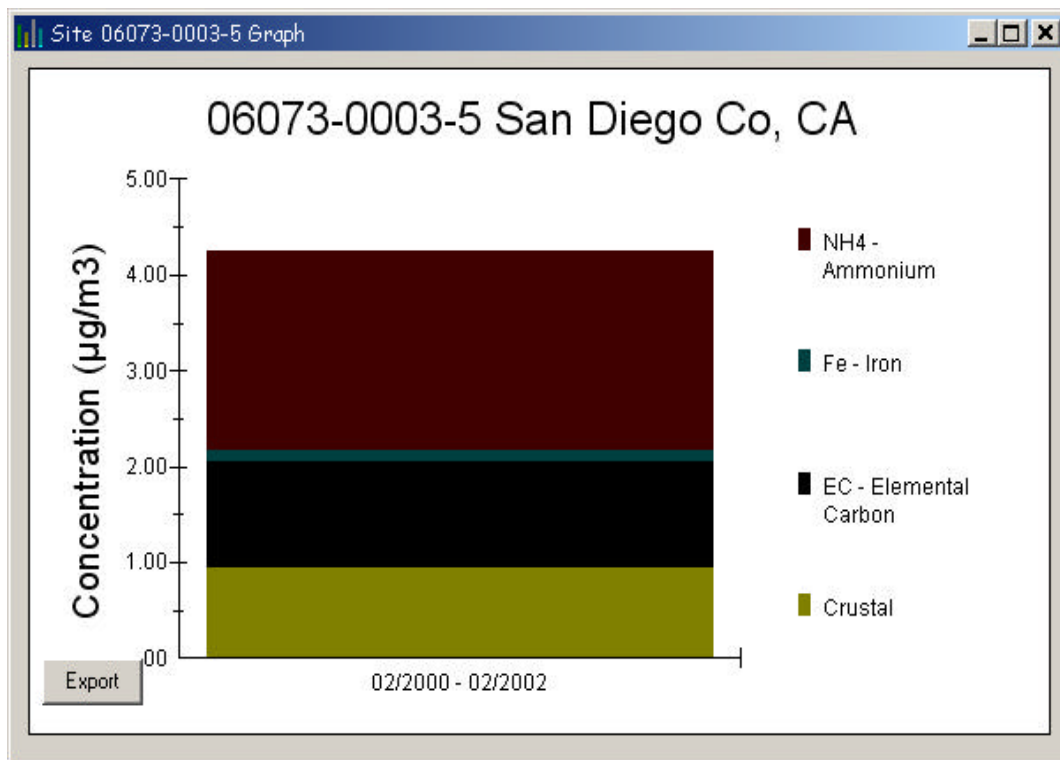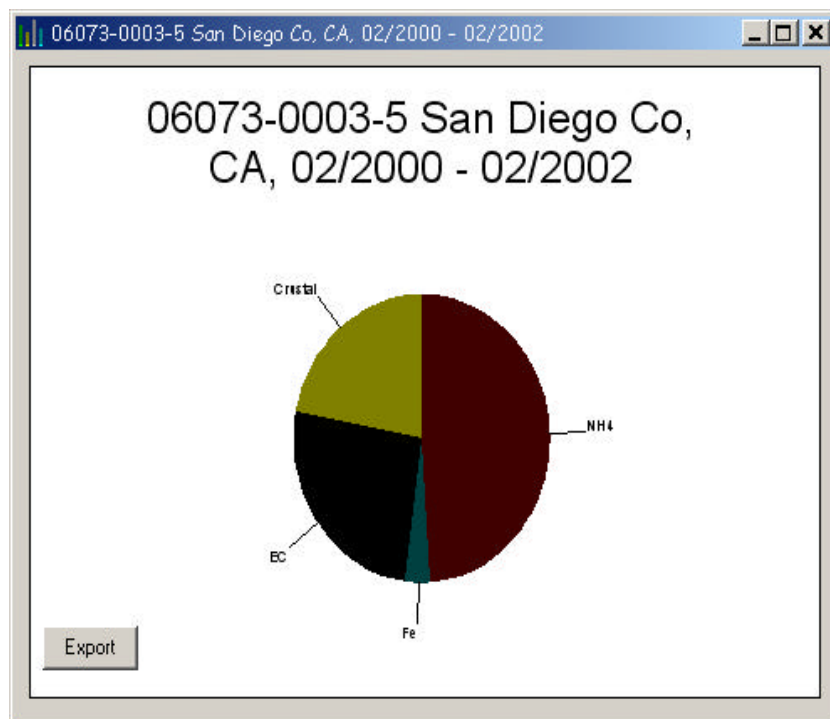**Site Information Tool Stacked Bar Graph**

Figure 6
Site Information Tool Pie Chart

The second tool, Site Comparison, allows a user to view specific information about an STN site and one or more IMPROVE sites Figure 7. As in the Site Information tool, the user works down the screen from left to right, top to bottom in making selections for viewing specific data. First, a user selects the desired date range of the data for viewing. Next an STN Site is selected. A filter appropriate for the analysis being performed is applied to the IMPROVE data and one or more IMPROVE sites (depending on availability) can then be selected for analysis. The temporal period is selected (monthly or entire period), and the desired species to be included in the analysis are checked. Finally the user has the option of viewing the raw data from STN and IMPROVE sites (Figure 8), statistics of the selected data (Figure 9) or a stacked bar graph (Figure 10). Note that the stacked bar chart shows the delta (from the statistics table) between the averages of the STN and IMPROVE sites.

Figure 7
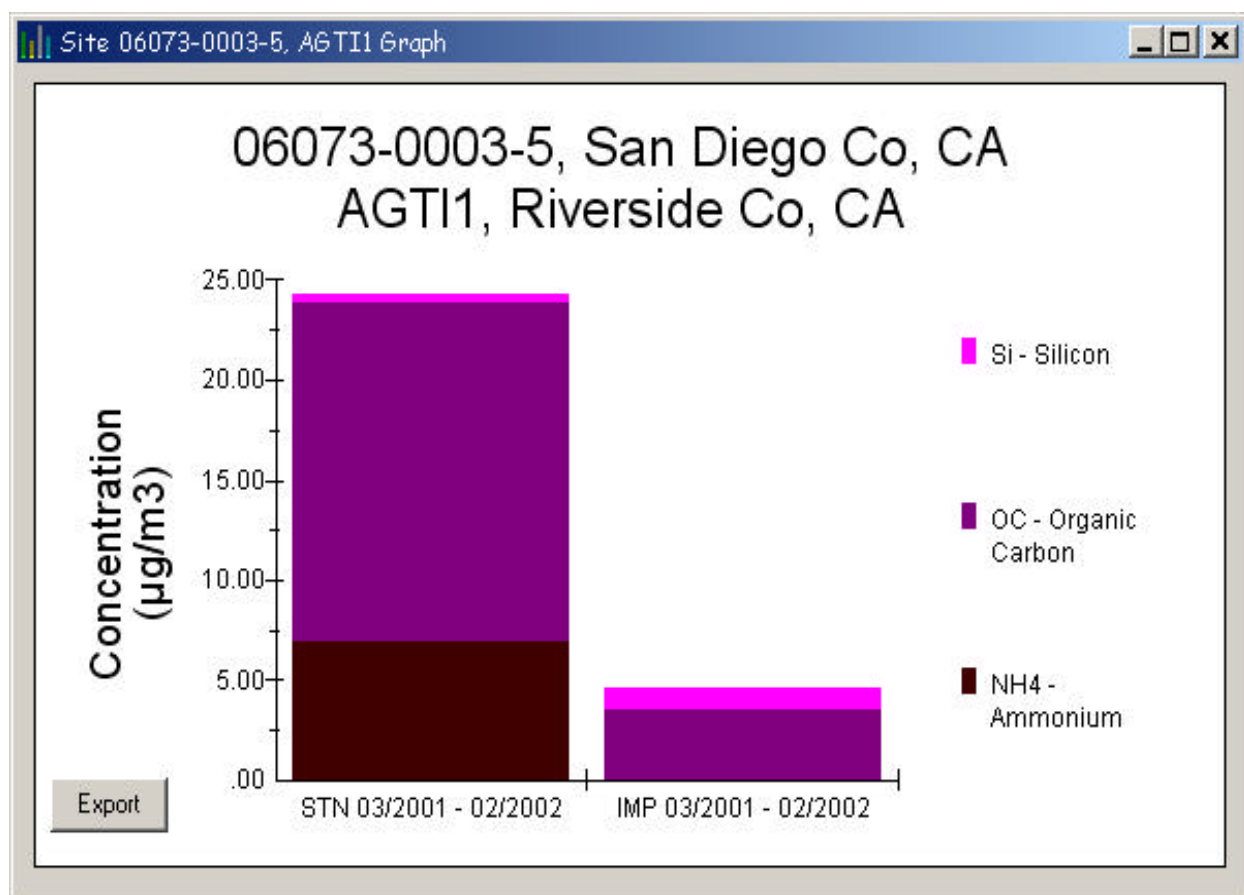**Site Comparison Tool Selection Screen**

Figure 8
Site Comparison Tool Data Table

Site 06073-0003-5, AGTI1 Comparison Data

Site ID	POC	Site Type	Sampler	Latitude	Longitude	Lab	Date	Al	Ca	Crustal	EC
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/01/20	0.097	1.115	1.890	0.133
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/01/23	0.078	1.440	0.950	0.145
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/01/26	0.083	1.491	2.040	0.226
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/01/29	0.009	0.182	0.880	0.035
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/01	0.067	1.132	2.760	0.203
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/07	0.090	1.183	1.500	0.210
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/10	0.118	1.722	0.910	0.177
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/13	0.113	1.817	1.150	0.245
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/16	0.037	0.613	0.570	0.085
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/22	0.109	1.774	1.170	0.290
06073-0003-5	5	TRENDS	SASS	32.7914	-116.942	RTI	2002/02/28	0.061	0.842	0.800	0.093
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/02	0.014	0.165	0.559	0.029
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/05	0.016	0.149	0.493	0.019
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/08	0.012	0.151	0.619	0.028
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/11	0.030	0.131	0.124	0.010
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/14	0.000	0.000	0.000	0.000
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/17	0.000	0.000	0.000	0.000
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/20	0.000	0.000	0.000	0.000
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/23	0.000	0.000	0.761	0.000
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/26	0.024	0.252	0.497	0.031
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/03/29	0.000	0.000	0.525	0.000
AGTI1	-99	IMPROVE	IMPR...	33.4636	-116.971	IM...	2001/04/01	0.017	0.265	0.431	0.029

Figure 9
Site Comparison Tool Statistics Table

Site 06073-0003-5, AGTI1 Stats

SiteID	Period	Data Points	NH4 Min	NH4 Avg	NH4 Max	OC Min	OC Avg
delta (06073-0003-5 - AGTI1)	01/03/2	...	0.040	2.074	6.900	2.360	5.679
06073-0003-5	03/2001 - 02/2002	61	0.040	2.074	6.900	2.360	6.860
AGTI1	03/2001 - 02/2002	122	0.000	0.000	0.000	0.000	1.180

Figure 10
Site Comparison Tool Stacked Bar Graph

NEXT STEPS

The PM Visualization Tool is to be put into production and available in the near future. The import of additional year's information is currently being implemented. The tool will be made available from EPA's and Pechan's web sites. There are several enhancements under consideration for the PM Visualization Tool which include (but are not limited to):

- Allowing more user control over graphs;
- Providing preset combinations of species;
- Incorporating additional graph types;
- Allowing for multiple temporal periods in one analysis;
- Adding a GIS front-end for selections and back-end for viewing results; and,
- Developing a web Version.

In summary, the PM Visualization Tool is an off-the-shelf application that facilitates the analysis of urban and rural air quality information regarding particulate matter species.