Appendix E

Speciated Modeled Attainment Test (SMAT) Documentation

Procedures for Estimating Future PM2.5 Values by Application of the Speciated Modeled Attainment Test (SMAT)

EPA has issued draft guidance (EPA, 2001) that describes a procedure for combining monitoring data with outputs from simulation models to estimate future concentrations of $PM_{2.5}$ mass. The guidance recommends that model predictions be used in a relative sense to estimate changes expected to occur in each major $PM_{2.5}$ species. These species are sulfates, nitrates, organic carbon, elemental carbon, crustal and un-attributed mass which is defined as the difference between measured $PM_{2.5}$ and the sum of the other five components. EPA is using the "SMAT" procedure to estimate the ambient impact of national rules and legislation, including the Clear Skies Act and the Transport Rule for reducing the impact of interstate pollutant transport.

The guidance describes a sequence of key steps that are recommended in processing the data. The following is a brief summary of those steps:

(1)   Derive current quarterly mean concentrations (averaged over three years) for each of the six major components of $PM_{2.5}$. This is done by multiplying the monitored quarterly mean concentration of Federal Reference Method (FRM) derived $PM_{2.5}$ by the monitored fractional composition of $PM_{2.5}$ species (at speciation monitor sites) for each quarter in three consecutive years. (e.g., 20% sulfate x 15 $\mu g/m^3$ $PM_{2.5}$ = 3 $\mu g/m^3$ sulfate).

(2)   For each quarter, apply an air quality model to estimate current and future concentrations for each of the six components of $PM_{2.5}$. Take the ratio of future to current predictions for each component. The result is a component-specific *relative reduction factor* (RRF). (e.g., given model predicted sulfate for base is 10 $\mu g/m^3$ and future is 8 $\mu g/m^3$ then RRF for sulfate is 0.8).

(3)   For each quarter, multiply the current quarterly mean component concentration (step 1) times the component-specific RRF obtained in step 2. This leads to an estimated future quarterly mean concentration for each component. (e.g., 3 $\mu g/m^3$ sulfate x 0.8 = future sulfate of 2.4 $\mu g/m^3$).

(4)   Average the four quarterly mean future concentrations to get an estimated future annual mean concentration for each component. Sum the annual mean concentrations of the six components to obtain an estimated future annual concentration for $PM_{2.5}$.

EPA will use the Federal Reference Monitor (FRM) data for nonattainment designations. Therefore it is critical that FRM data is used in the speciated modeled attainment test described above. As can be seen from the list of steps, the modeled attainment test is critically dependent on the availability of species component mass at FRM sites. Since roughly 80% of the FRM sites will not have collocated speciation monitors, a spatial interpolation methodology was developed to estimate component species mass at the FRM locations. This method was further

utilized to estimate PM2.5 and component species mass at every grid cell in the study domain. Additional ambient data handling procedures were also developed. Below we describe an example application of the procedures, for a study domain that extends over a large portion of eastern US. The study domain is defined for grids of dimension ½ degree longitude by 1/3 degree latitude (~36 km X 36 km) covering the area enclosed within -100 to -67 longitude and 25 to 49 latitude. Base case and future year model predictions are available for each grid cell (72 rows by 66 columns) that make up the study domain.

Ambient Data preparation

PM$_{2.5}$ quarterly averages at FRM sites for 1999-2001 were calculated using data from the Air Quality System (AQS). The resulting data set contained 325 sites that meet the completeness criteria needed to determine the PM$_{2.5}$ NAAQS attainment status. Each of the PM$_{2.5}$ sites was uniquely associated with one of the grid cells in the study domain.

Speciated PM2.5 data from both the Interagency Monitoring of Protected Visual Environments (IMPROVE) and EPA's speciation trends network[1] (STN) were used to derive mean concentrations of each of the six PM$_{2.5}$ components. No attempt was made to resolve differences in measurement and analysis methodology between the two networks[2]. Since three years of urban speciation data were not available, the latest full year of data was used. Quarterly average concentrations between April 2001 through March 2002[3] were retained for sites that had at least 15 monitored values (50% completeness for 1 in 3 day sampling). The quarters were defined as follows: Q2 = April - June 2001; Q3 = July 2001 - September 2001; Q4 = October - December 2001 and Q1 = January - March 2002. Figure 1 shows the spatial distribution of IMPROVE and STN stations that met this completeness criteria for first quarter of 2002.

---

[1]The network is referred to as the "STN", but all urban speciation sites were used, not just the trends sites.

[2]There are certain differences in sampling and analysis techniques which may affect the results of this application. The data from both networks were treated similarly whenever possible. Further comparison studies and analyses are needed to develop data sampling and handling procedures that may make the data from the two networks more similar.

[3] The 1st quarter of 2002 was the most recent quarter of data available from both the IMPROVE and STN networks at the time of the analysis. The ambient speciation data will be updated as newer data and more sites become available.
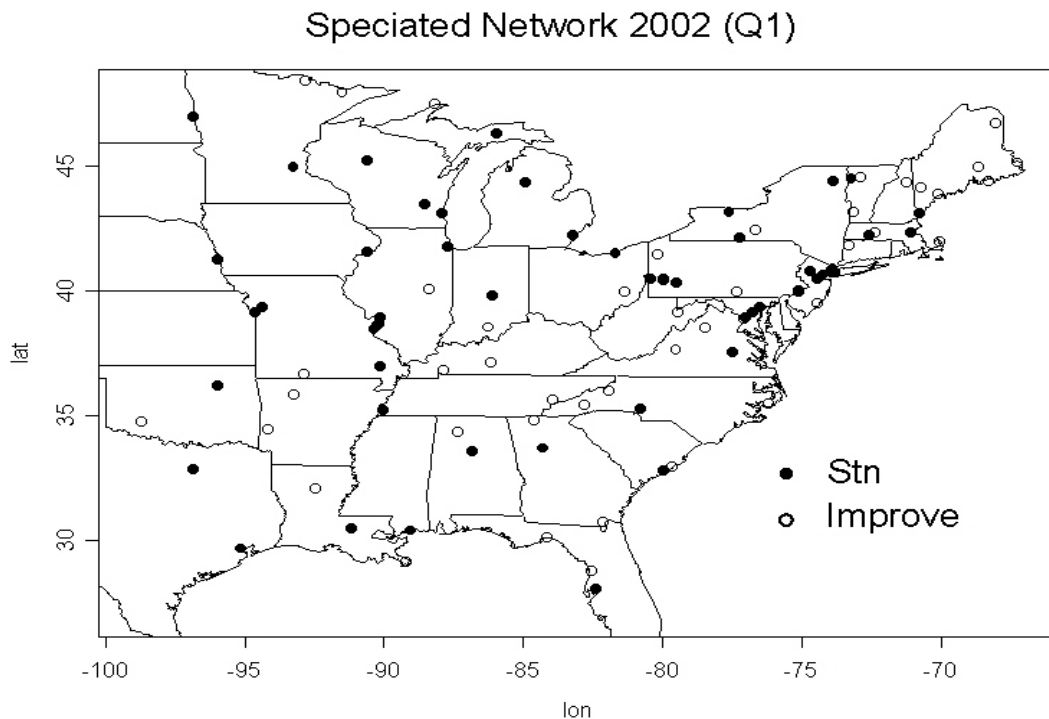
Speciated Network 2002 (Q1)

Figure 1. Speciated stations with at least 15 quarterly samples

Note: The number of stations meeting completeness criteria for the four quarters:

Quarter 2, 2001 – 82 sites,
Quarter 3, 2001 – 103 sites
Quarter 4, 2001 – 106 sites
Quarter 1, 2002 – 105 sites

As noted in the modeling guidance, the mass associated with each component must be estimated based on assumptions about chemical composition. Table 3.4 in the modeling guidance provides recommended default assumptions which were applied for each of the species except sulfate and carbon compounds[4]. Because ammonium is reported in the STN, it was possible to analyze the degree to which sulfate measured on the filter was actually neutralized. The analysis, concluded that, on average, sulfate was not completely neutralized, resulting in use of the factor 1.25 rather than the value (1.375) recommended in the guidance. The 1.25 factor

---

[4]As recommended in the modeling guidance, organic carbon was multiplied by 1.4 and particulate nitrate was multiplied by 1.29.

was derived through a mass balance of measured ammonium, sulfate, and nitrate at the STN sites. It was assumed that all particulate nitrate was in the form of ammonium nitrate. The measurements of nitrate ion and particulate ammonium are known to be uncertain. The calculation of the ammoniation of sulfate is subject to these uncertainties. Therfore, a single domainwide annual average value of 1.25 was used for all sites due to the uncertainties in the measurements of ammonium and nitrate. This value assumes that sulfate is, on average, partway between ammonium bisulfate and ammonium sulfate.

The elemental and organic carbon mass from the STN was adjusted downward based on measurements from field blanks which indicate a positive bias. The blank corrections were based on a draft report which examined the blank carbon data in the STN network (RTI, 2002). The carbon corrections are shown below in Table 1. The values were taken from Table 4.1 from the RTI report. The monitor dependent blank corrections were made to the quarterly average concentrations at each STN site. The IMPROVE carbon measurements are blank corrected by the IMPROVE program.

Sampler Type	Elemental Carbon (ugC/m3)	Organic Carbon (ugC/m3)
URG MASS	0.03	0.29
R and P 2300	0.22	0.90
Anderson RAAS	0.09	1.19
R&P 2025	0.07	0.77
MetOne SASS	0.11	1.42

Table 1. Carbon blank corrections

Finally, un-attributed mass was calculated for each of the STN monitors with a co-located FRM monitor. Un-attributed mass was not calculated for the IMPROVE sites since there were no collocated FRM PM2.5 data available. The results produced generally small positive estimates of un-attributed mass although for some sites, the estimate was negative. The un-attributed mass did not follow any clear spatial or temporal patterns. Due to the relatively random pattern of the un-attributed mass, a single quarterly value of un-attributed mass was used at each site. Table 2 summarizes the quarterly average un-attributed mass data. A quarterly average un-attributed mass value was calculated at each STN site by applying the un-attributed percentage to the quarterly average site specific FRM mass.

	Quarter 1 (Jan-Mar 02)	Quarter 2 (Apr-June 01)	Quarter 3 (July-Sept 01)	Quarter 4 (Oct - Dec 01)
Num of Monitoring sites	47	31	43	46
Avg FRM PM2.5 mass ($\mu g/m^3$)	12.17	13.51	14.43	11.97
Avg species mass sum ($\mu g/m^3$)	12.12	13.41	13.70	11.74
Un-attributed ($\mu g/m^3$)	0.05	0.10	0.73	0.23
Percent Un-attributed	0.4 %	0.7 %	5.0 %	1.9 %

Table 2. Average Un-attributed Mass of PM$_{2.5}$

Species Component Estimation

Only a small fraction of PM$_{2.5}$ sites have measured species information. For this reason, an objective procedure was developed for using the speciated component averages from the IMPROVE and STN networks to estimate concentrations of species mass at all FRM PM$_{2.5}$ monitoring sites. Kriging was adopted as the method for estimating PM$_{2.5}$ component mass at PM$_{2.5}$ sites since software is readily available and can produce estimates of prediction error. Kriging was performed using an S-PLUS software package known as FIELDS (NCAR, 2002) developed by scientists at NCAR to perform generalized kriging and efficient spatial analysis of large data sets.

The Krig function in FIELDS estimates the parameters of the spatial field using the Generalized Cross Validation (GCV) error as the criterion for parameter estimation. A simple exponential covariance function was used to describe the variogram. Outputs from Krig include the parameter estimates (range, nugget and sill) along with predicted values at each of the PM$_{2.5}$ monitor locations. Once the kriging equations were established for each species, prediction of quarterly average species concentrations were made for each of the FRM sites (325) and for each grid cell in the modeling domain. The latter predictions were made so that estimated PM$_{2.5}$ concentrations could be obtained for the entire modeling domain, allowing for a more complete spatial assessment of future PM$_{2.5}$ levels. Figures 2 and 3 illustrate the spatially interpolated concentration fields for nitrates (quarter 1) and sulfates (quarter 3).
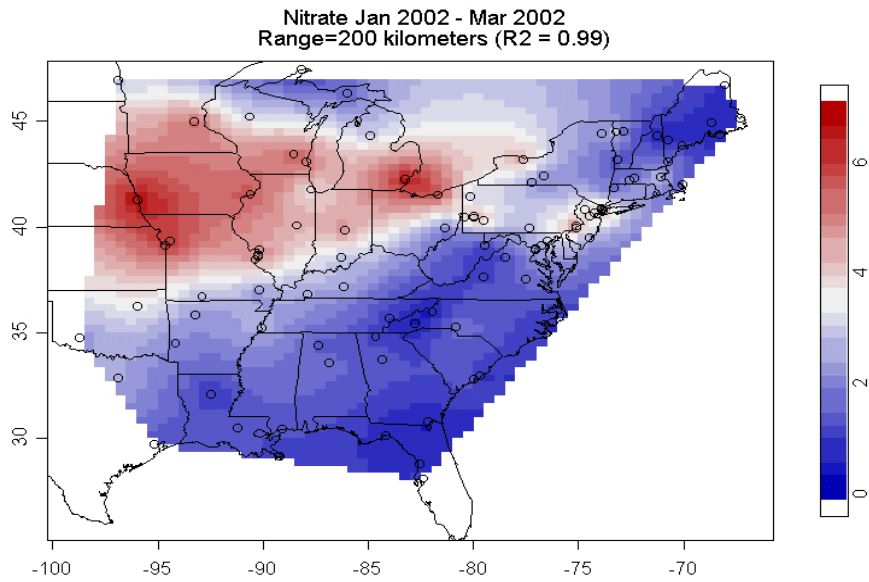
Figure 2   **Spatially Interpolated Nitrate Quarterly Average Concentrations (quarter 1)**
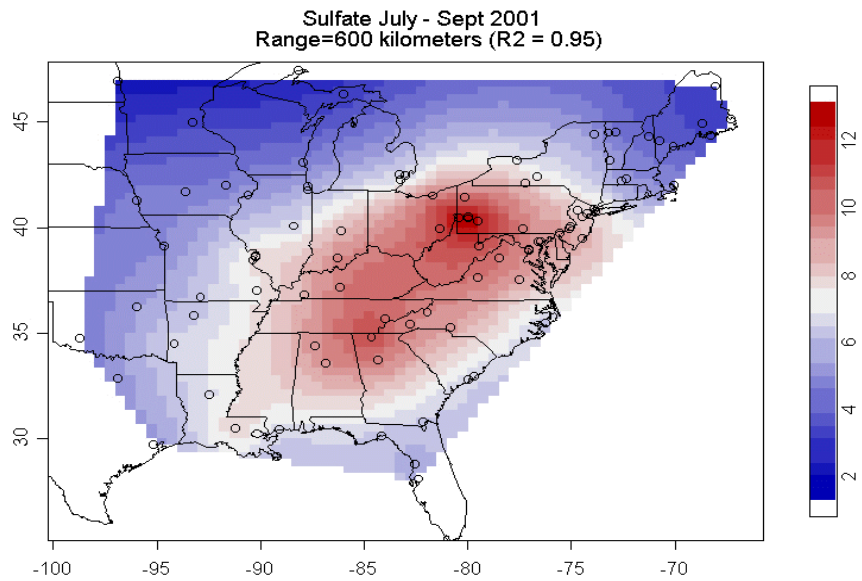

Figure 3   **Spatially Interpolated Sulfate Quarterly Average Concentrations (quarter 3)**

Kriging was not used for spatial interpolation of un-attributed mass since it only available for some of the STN sites and because there was no discernable spatial trend. Instead the quarterly average of the un-attributed mass from the STN sites was first expressed as a fraction of the average $PM_{2.5}$ mass. The estimated fractions for each quarter were previously shown in Table 2.

For each quarter, predicted concentrations for each of the six species are combined with quarterly $PM_{2.5}$ FRM averages to derive composition concentrations in the following manner. First, the un-attributed mass at each $PM_{2.5}$ site was estimated by multiplying the average fraction of un-attributed mass by the quarterly average $PM_{2.5}$ concentration for that site. For example, if a site in quarter 3 had an average $PM_{2.5}$ mass of 20 $\mu g/m^3$, then the un-attributed mass would be 20 $\mu g/m^3$ x 0.05 = 1 $\mu g/m^3$. The total $PM_{2.5}$ mass that is identifiable was calculated by subtracting the estimated un-attributed mass from each quarterly average $PM_{2.5}$ value. Next, the component mass of each of the five identifiable species was estimated by multiplying the fraction of each species by the identifiable portion of the quarterly $PM_{2.5}$ mass. This procedure is repeated for each $PM_{2.5}$ site and quarter to complete the calculation of current or baseline ambient concentrations used as the basis for future estimates of $PM_{2.5}$ mass and its components. Table 3a shows an example of the un-attributed mass calculation and the species fractions for an FRM site in quarter 2. The species fractions in table 3a are derived from the quarterly interpolated (Kriged) spatial fields for each of the five species. Multiplying the un-attributed mass fraction of 0.7% (from table 2) times 17.0 (FRM mass from table 3a) yields the identifiable mass of 16.88. The identifiable mass can then be split into individual species component mass estimates by using the fractions in table 3a.

FRM Mass ($\mu g/m^3$)	% Un-atributed mass	Identifiable Mass ($\mu g/m^3$)	% Sulfate	% Nitrate	% Organic aerosol	% Elemental Carbon	% Crustal
17.0	0 .7	16.88	32.1	11.4	38.9	9.9	7.7

Table 3a. Un-attributed mass and species fractions for an FRM site in quarter 2

Table 3b shows the resultant mass for each of the component species at the same FRM site. The species mass is calculated by multiplying the fraction of each component by the identifiable mass. **The sum of the components is the observed FRM PM2.5 mass concentration (17.0 $\mu g/m^3$)**.

FRM Mass ($\mu g/m^3$)	Un-atributed Mass ($\mu g/m^3$)	Sulfate Mass ($\mu g/m^3$)	Nitrate Mass ($\mu g/m^3$)	Organic aerosol Mass ($\mu g/m^3$)	Elemental Carbon Mass ($\mu g/m^3$)	Crustal Mass ($\mu g/m^3$)
17.0	0.12	5.42	1.92	6.57	1.67	1.30

Table 3b. Resultant species mass at an FRM site in quarter 2

Estimating Future Year PM₂.₅

Future concentrations of PM$_{2.5}$ component species are estimated by assuming that the quarterly average component concentration will change in exactly the same proportion as the model predicted change. Model predicted changes in species concentrations (from a current year to a future year) are used to calculate "relative reduction factors". Relative reduction factors are calculated for each grid cell and species as the ratio of the quarterly average future predictions to the current base predictions. The relative reduction factor for each species is then multiplied by the estimated current year ambient species mass for the site to estimate future species concentrations. These future species concentrations at each FRM site are then summed over the five species to estimate the identifiable portion of future quarterly average PM$_{2.5}$ concentration. The current year quarterly average estimate of un-attributed PM$_{2.5}$ mass is added to the future quarterly average identifiable PM$_{2.5}$ mass estimate. The four quarterly values are then averaged to obtain the estimated future annual average PM$_{2.5}$ for each FRM site.

FRM sites close to or co-located with an STN monitor will have the least "error" in the estimation of species fractions[5]. There is more uncertainty associated with FRM monitoring sites that are not located near a speciation site. It should be noted that the sole use of the interpolated speciation data is to calculate the mass fractions of each of the PM2.5 components. All of the future year design value calculations at FRM sites are "anchored" by the FRM data itself.

The results of the analysis at each of the FRM monitoring sites (with complete data) will be used in analyses such as Clear Skies and the Transport Rule. Application of SMAT with Kriged spatial fields allows us to take advantage of the design value information at each FRM site. In this way, a more complete attainment/nonattainment picture can be derived by not limiting the predictions of future year design values to only speciation monitoring sites.

Additional Spatial Information

PM$_{2.5}$ concentrations can also be estimated over the entire field of grid locations that define the study domain (i.e., 72 x 66 grid cells). This requires that the quarterly average PM$_{2.5}$ also be kriged to estimate PM$_{2.5}$ average concentrations for each grid cell. Because the majority of PM$_{2.5}$ measurement sites are urban oriented, the PM$_{2.5}$ mass reported for the IMPROVE sites are also included in the spatial interpolation process to help minimize potential urban bias in more rural locations. Figure 4 shows the spatially interpolated base year (1999-2001) PM$_{2.5}$ annual concentration field and figure 5 shows the projected future base case (2010) PM$_{2.5}$ concentration field.

---

[5]The species fractions at co-located FRM and speciation sites can be calculated without the use of spatial fields. However, for this application, the species fractions for **all** FRM sites were derived from the spatial fields. This allowed for consistent calculations at all sites.
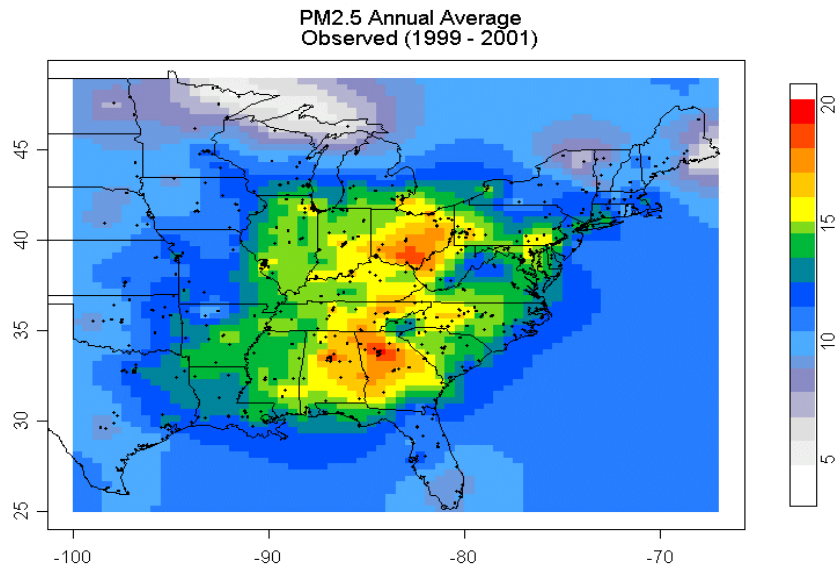
PM2.5 Annual Average
Observed (1999 - 2001)

Figure 4 **Example spatial fields of base year (1999-2001) annual average PM2.5 design values**
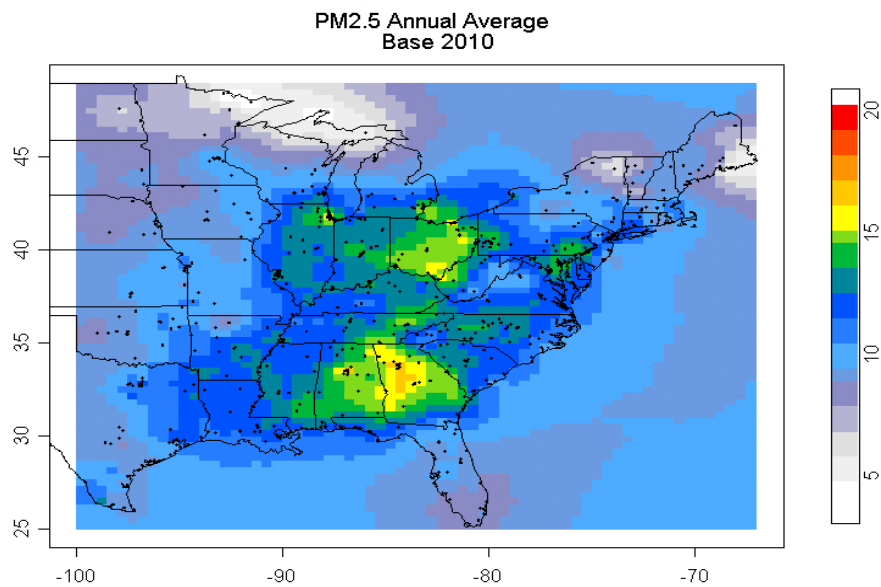

PM2.5 Annual Average
Base 2010

Figure 5 **Example spatial fields of future year (2010) annual average PM2.5 design values (calculated from relative reduction factors from the REMSAD model)**

Summary of Outputs

Future year design values can be calculated at monitoring sites which have co-located FRM and speciation monitors. Kriging of speciation data allows the calculation of future year design values at all FRM monitoring sites. Additional Kriging of all PM2.5 data (FRM and IMPROVE) allows the calculation of future year design values at all model grid cells. Table 4 shows the available outputs of the modeled attainment test with spatial fields.

	Ambient PM2.5 Data From:	Ambient Speciation Data From:
Case1- FRM monitoring sites	FRM monitor	Interpolated (Kriged) speciation data
Case 2-All grid cells	Interpolated (Kriged) PM2.5 data	Interpolated (Kriged) speciation data

Table 4. Sources of data for speciated modeled attainment test with spatial fields

There are uncertainties associated with many aspects of the analysis. There is uncertainty associated with collection and analysis of the ambient data (e.g. positive organic carbon artifacts and negative particulate nitrate artifacts associated with the ambient data collection and analysis), post-processing of the ambient data (e.g, assumptions regarding the 1.25 factor for sulfate or the 1.4 factor applied to organic carbon), interpolation of the data to the FRM sites and grids (e.g. Kriging error and replication of species gradients), use of the model predicted changes in species (e.g. errors and uncertainty in the model science and inventories), etc.

We have the most confidence in future estimates of PM2.5 at FRM monitoring locations (case 1). Therefore, the results of this analysis at each of the FRM monitoring sites (with complete data) will be used for regulatory purposes.

It is unclear at this point how the design value results from the grid cells with no monitoring data will be used (case 2). At a minimum, the results at all grid cells can be used qualitatively to evaluate model performance, to inform the "conceptual model", and to portray a visual picture of the estimated current year and future year design values. The full spatial fields may also be able to replace and/or enhance the hot spot modeling and screening test required in the PM and ozone modeling guidance.

Use of SMAT with Spatial Fields for SIPs

The details of this application of SMAT are specific to the short term use of the FRM and STN data in estimating future year PM2.5 concentrations. The use of a single year of speciation data interpolated to a modeling grid is necessary at this time, due to the relatively sparse ambient

data sets. The amount of available ambient data will increase significantly in the future. When ambient data is needed for SIP development, there will be at least 3 years of complete speciation data at hundreds of sites. In many areas, the coverage of speciation data may be adequate so that interpolation of the data through spatial fields is not necessary. This application should serve as an example that can be replicated in the short term, but the techniques and assumptions will likely evolve over the long term.

References

EPA, Guidance for Demonstrating Attainment of Air Quality Goals for $PM_{2.5}$ and Regional Haze (Draft). January 2, 2001. Available on EPA' TTN web page (www.epa.gov/scram)

RTI, Analysis of Speciation Network Carbon Blank Data. Draft Report submitted to EPA, August 30, 2002 under Work Assignment 4-05 for contract 68-D-032, Research Triangle Institute.

NCAR Fields Manual. Home page of the Geophysical Statistic Project. URL address located: http://www.cgd.ucar.edu/stats/Software/Fields/MAN.shtml.