Speciation Data Validation & Analysis Tool (SDVAT)

USEPA National Air Monitoring Conference

November 6, 2006

Max R. Peterson

RTI International

Research Triangle Park, North Carolina

Background

- **EPA's PM2.5 Speciation Trends Network (STN)**
  - **Start date - February 2000**
  - **Laboratory support contractor - RTI International**
- **PM2.5 Sampling and Analysis Data Reported**
  - **Field data (15-20 parameters, varies by sampler type)**
  - **Laboratory analysis data**
    - **4 laboratories (Grav., Ions, XRF, OC/EC)**
    - **6 analyses (mass, sulfate, nitrate, anions, elements, OC/EC)**
    - **63 analytes (mass, 6 ions, 48 elements, 8 carbon fractions)**

Tools for Validation of $PM_{2.5}$ Chemical Speciation Data

- **Data Sources**
- **Data Validation**
- **Speciation Data Validation Analysis Tool (SDVAT)**
  - **Developed by RTI under contract to EPA in 2002**
  - **Updated in 2006**

DATA SOURCES

- **Field Sampling Data**
- **RTI Monthly Analytical Reports**

DATA SOURCES: Field Sampling Data Sent to RTI

- **Custody and Field Data Form (C&FDF or COC)**
  - **different C&FDF used for each supported sampler type**
- **Field Sampling Null Value And Validity Coding Form (FSNV&VCF)**
  - **same FSNV&VCF Form Used for All Samplers**
- **C&FDF and FSNV&VCF are:**
  - **supplied with sampling modules,**
  - **completed by field operator, and**
  - **returned to lab with sample.**
- **Field data are:**
  - **double-entered into RTI's database,**
  - **subjected to several verification and validation checks, and**
  - **reported back to agencies in monthly analytical reports.**

DATA SOURCES: Field Sampling Data Kept By Monitoring Agency

- **Operator's Notes and Site-Specific Information**
- **Electronic Data Transmittals from Sampler**
- **Systems Audit Reports**
- **Performance Audit Results**
- **Collocated Sample Results**
- **Shipping Records**

DATA SOURCES:
Completed Field Data Forms

- **Examples of completed Field Data forms represent a grab sample of forms arriving in RTI's Sample Handling & Archiving Laboratory (SHAL) on a single day.**
- **The forms in this set were selected based on a single criterion:**

  Field sampling personnel wrote at least one comment on the selected forms.

DATA SOURCES:
RTI Monthly Analytical Reports

- **Report.rtf**
- **Spreadsheet.xls**
- **MassSummary.xls**
- **ConcSummary.xls**

DATA SOURCES:
RTI Monthly Analytical Reports

- **Field Data**
  - **Transcribed by field operator from sampler screen to C&FDF**
  - **Double-entered into RTI's database**
  - **Included in monthly analytical reports.**
- **Laboratory Analysis Data**
  - **<u>Gravimetry</u>: mass**
  - **<u>X-Ray Fluorescence</u>: Al, Sb, As, Ba, Br, Cd, Ca, Ce, Cs, Cl, Cr, Co, Cu, Eu, Ga, Au, Hf, In, Ir, Fe, La, Pb, Mg, Mn, Hg, Mo, Ni, Nb, P, K, Rb, Sm, Sc, Se, Si, Ag, Na, Sr, S, Ta, Tb, Sn, Ti, V, W, Y, Zn, Zr**
  - **<u>Ion Chromatography</u>: Na^+, K^+, NH_4^+, NO_3^-, SO_4^{2-}**
  - **<u>Thermal-Optical Carbon Analysis</u>: OC, EC, TC, and OC Peaks (Pk1C, Pk2C, Pk3C, Pk4C, and PyrolC)**

DATA CODES & FLAGS

- **Data codes and flags in RTI monthly reports include:**
  - **Flags generated by the samplers**
  - **Informational flags generated by RTI (3 characters)**
  - **Null Value Codes (2 letters or 4 numeric digits) for ARS *(Null Value Codes are reported to ARS instead of the data values.)***
  - **Validity Status Codes (1 character) for ARS *(Validity Status Codes are reported to ARS along with the data values.)***
- **A complete list of the most recent codes and flags appears at the end of each RTF monthly report.**

DATA SOURCES:
RTI Monthly Analytical Reports, cont.

Filename	Content	Possible Use
Report.rtf	Rich text format report -- all reported data and flags (~5 pages/sample plus a list of all current flags with descriptions)	Print human-readable hard copy report for detailed review of data
Spreadsheet.xls	Spreadsheet report -- all reported data and flags (1 row/data item)	Import into a database
MassSummary.xls	Spreadsheet summary of valid analyte mass data (1 row/sample)	Overview and comparison (routine samples, FBs, and TBs)
ConcSummary.xls	Spreadsheet summary of valid analyte concentration data (1 row/sample)	Overview and comparison (routine samples)

DATA VALIDATION

- **Validation Levels (0, 1, 2, and 3)**
- **RTI Partial Validation**
- **State/Local Validation**

DATA VALIDATION: Validation Levels

- **Level 0 --** Basic review of data with respect to their provenance.
- **Level 1** -- Process of evaluating the correctness and acceptability of individual items or groups of items within the data set using statistical methods and other screening techniques.
- **Levels 2 & 3** -- May include correlations between sites, time-series analyses, collocated bias and precision, correlations between analytes and/or methods at the sampling site, modeling, etc.

NOTE: Activities for the verification/validation levels described above overlap.

DATA VALIDATION:
RTI Partial Validation

Level 0 (RTI)	Level 1 (RTI)
Sample Identification	Conservation of Mass:
Operator Observations	∑Species Conc ≈ Mass Conc
Sampler Flags	Conservation of Charge:
Shipping & Disassembly	Cation/Anion Ratio
Laboratory Checks (per SOPs)	Analyte Correlations:
Range Checking:	Sulfate/Sulfur Ratio
Flow Rate	Sodium(IC) ≈ Sodium(XRF)
Exposure Duration	Other Correlations
Elapsed Time before Retrieval	
Holding Times	

DATA VALIDATION:
RTI Partial Validation, cont.

For a complete description of RTI's Level 0 and Level 1 data validation procedures see:

Data Validation Process for the PM2.5 Chemical Speciation Network, Updated February 2005

This document can be downloaded from:

http://www.epa.gov/ttnamti1/files/ambient/pm25/spec/05datval.pdf

DATA VALIDATION: State/Local Validation

- Confirm RTI Level 0 and Level 1 Validation Checks
- Perform Level 2 and Level 3 Validation Checks
- For a discussion of monitoring agency validation of chemical speciation data see: *Quality Assurance Project Plan: $PM_{2.5}$ Speciation Trends Network Field Sampling*, 2000.

  This QAPP can be downloaded from:

  http://www.epa.gov/ttn/amtic/files/ambient/pm25/spec/1025sqap.pdf
- Other helpful documents can be downloaded from:

  http://www.epa.gov/ttn/amtic/specguid.html

SPECIATION DATA VALIDATION ANALALYSIS TOOL (SDVAT)

- **Overview**
- **Data Selection Features**
- **Data Analysis Features**
- **Data Manipulation Features**
- **Hands-On Practice**

SDVAT: Overview

- **The Speciation Data Validation Analysis Tool (SDVAT) was developed in 2002 to assist monitoring agencies in reviewing and validating data from RTI's Spreadsheet.xls monthly analytical reports.**
- **The SDVAT is useful for:**
  - **Combining data from multiple reports into a single database,**
  - **Creating charts and graphs for Level 2 and Level 3 checks,**
  - **Looking at temporal and spatial data trends.**
- **The *SDVAT User's Guide* (UG) explains features and uses in detail.**
- **Minimum System Requirements for the 2006 Version: Windows 2000 or Windows XP; 256 MB RAM (minimum); MS Excel 2000 (MS Office XP preferred); sufficient hard drive space.**

SDVAT: Data Manipulation Features

- **Import Data (UG 3 = User's Guide, Section 3)**

  Includes a "Migrate Data" option for moving data from a previous version of the SDVAT into the new version.

- **Export Data (UG 6.3)**
- **Delete Data (UG 6.2)**
- **Edit/Add Data (UG 6.1)**
- **Omit Data from Analysis (UG 4.8.2)**
- **Add Comments (UG 4.8.1)**

SDVAT: Data Selection Features

- **Select Site(s) (UG 4.2)**
- **Select Analysis or Analyses (UG 4.3)**
- **Select Analyte(s) (UG 4.4)**
- **Select Date Range (UG 4.5)**
- **Select Sample Type(s) (UG 4.6)**
- **Select Data Validity Type(s) (UG 4.7)**
- **Select Status for Omitted Data (UG 4.8)**

SDVAT: Data Analysis Features

Output:

- **Output from the SDVAT is always in the form of an Excel workbook.**
- **Each data analysis performed using the SDVAT creates a single Excel workbook with one or more charts each linked to its own separate worksheet containing the data used to generate the associated chart. In the workbook, Chart-1 is linked to data in spreadsheet Data-1, Chart-2 to Data-2, etc.**
- **The Excel workbooks can be viewed, printed, formatted, manipulated, etc. using Excel.**

SDVAT: Data Analysis Features

- **Data Completeness (UG 5.1)**
- **Time Series (UG 5.2)**
- **Mass Reconstruction (UG 5.3)**
- **Species Distributions (UG 5.4)**

SDVAT: Data Analysis Features

1. **<u>Data Completeness</u> – the ratio of valid data to total data and expressed as a percentage:**

$$\%\text{Valid} = \frac{\text{Number Valid Results}}{\text{Total Number Results}}(100\%)$$

NOTE: "Valid" data has its [OVERALL_INVALID] field equal to FALSE or NULL.

NOTE: Data completeness is used as an indicator of data capture and is calculated separately for each analyte/parameter.

SDVAT: Data Analysis Features

1. Data Completeness – Spreadsheet Output

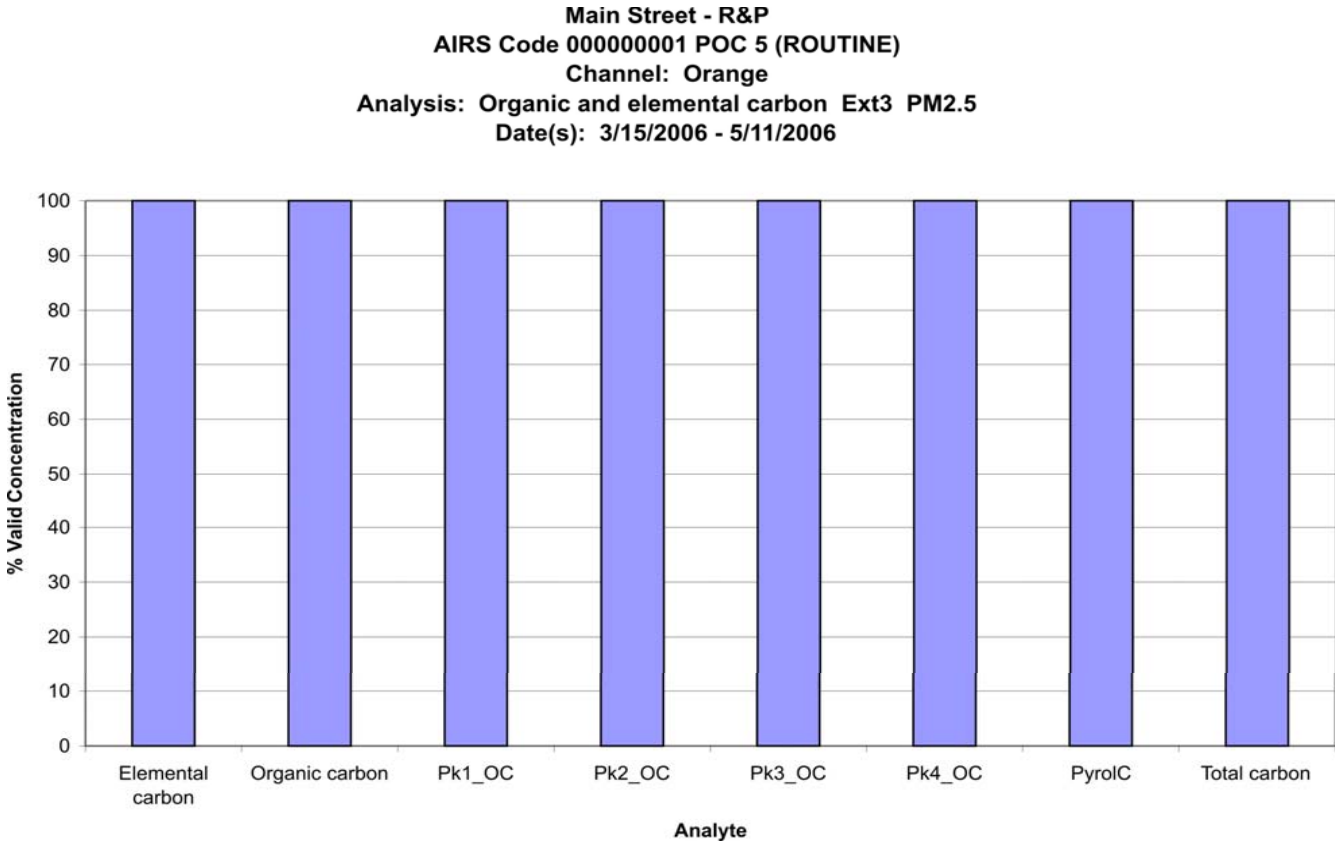
LOCATION_NAME	AIRS_CODE	POC	SAMPLE_TYPE	CHANNEL_NAME	ANALYSIS
Main Street - R&P	000000001	5	ROUTINE	Orange	Organic and elemental carbon Ext3 PM2.5

ANALYTE	TOTAL_COUNT	VALID_COUNT	PERCENT_VALID
Elemental carbon	9	9	100
Organic carbon	9	9	100
Pk1_OC	9	9	100
Pk2_OC	9	9	100
Pk3_OC	9	9	100
Pk4_OC	9	9	100
PyrolC	9	9	100
Total carbon	9	9	100

Note: Data output from the SDVAT is in the form of a spreadsheet with data fields linked to a chart.

SDVAT: Data Analysis Features

1. Data Completeness – Graphic Output



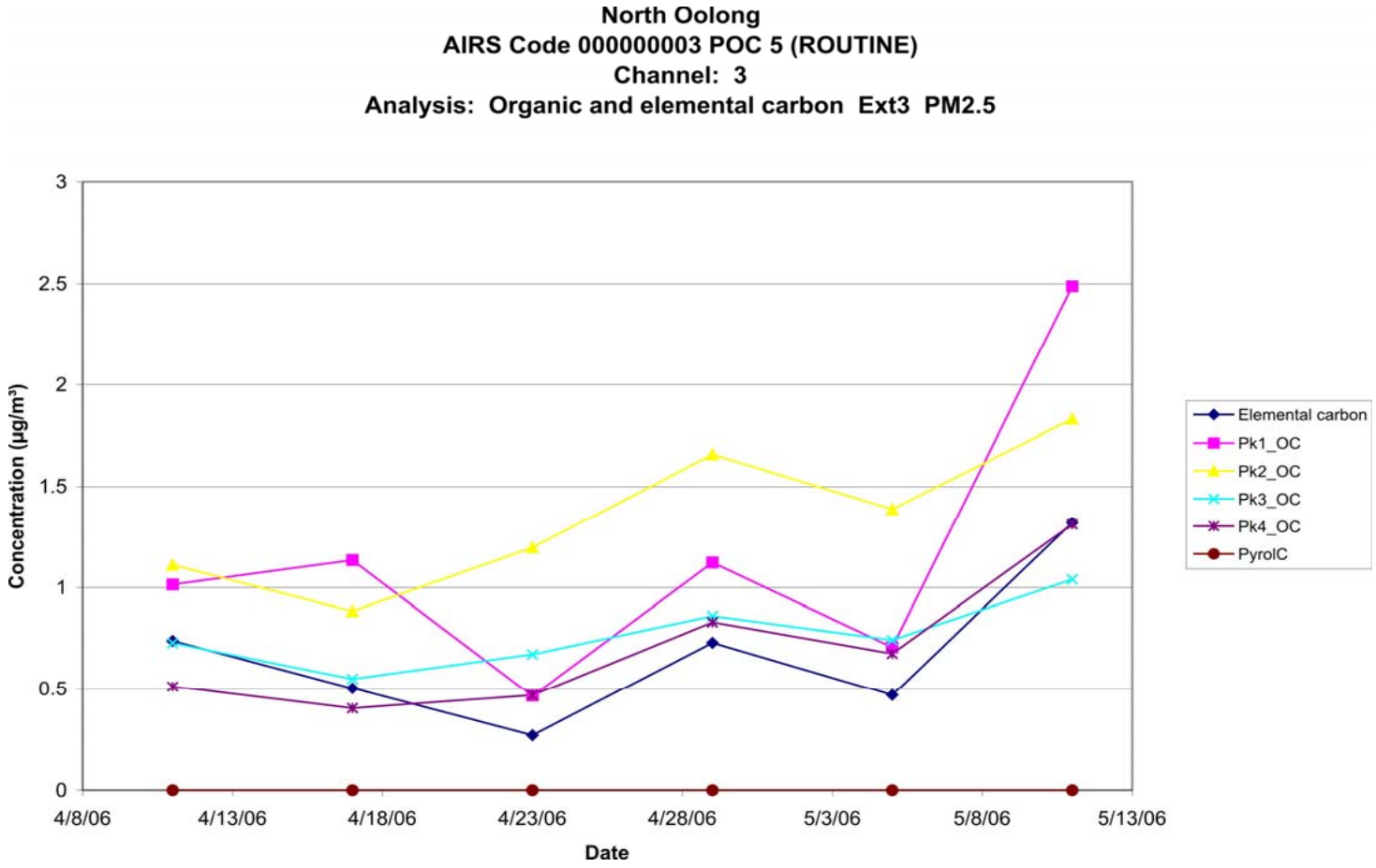
SDVAT: Data Analysis Features

2. **<u>Time Series</u> – A plot of the mass or concentration of an analyte or some field-reported value versus time.**

   NOTE: These are simple plots of values in the database versus sampling date and require no further calculations.

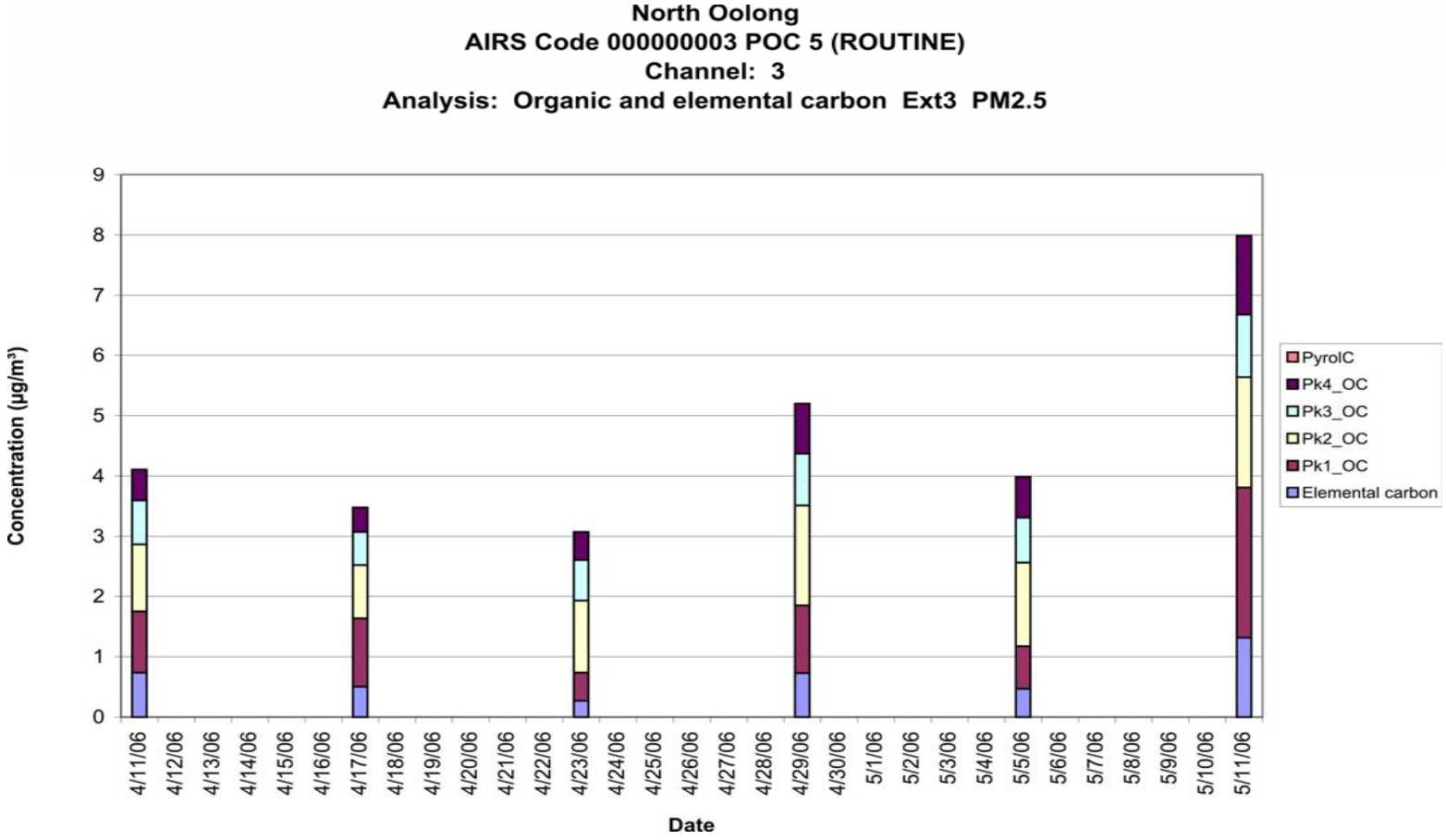
SDVAT: Data Analysis Features

2. Time Series –XY-Line Plot



SDVAT: Data Analysis Features

2. Time Series –Stacked Column Plot



SDVAT: Data Analysis Features

3. **<u>Mass Concentration Reconstruction Analysis</u> -- A plot of gravimetric mass concentration vs. reconstructed mass concentration.**

$$\left[\begin{array}{c}\text{Reconstructed} \\ \text{Mass} \\ \text{Concentration}\end{array}\right] = \left[\begin{array}{l}\sum\text{Anions(IC)} + \sum\text{Cations(IC)} + \text{Total Carbon} \\ + \sum\text{Trace Elements(XRF), excluding S, Na, K}\end{array}\right]$$

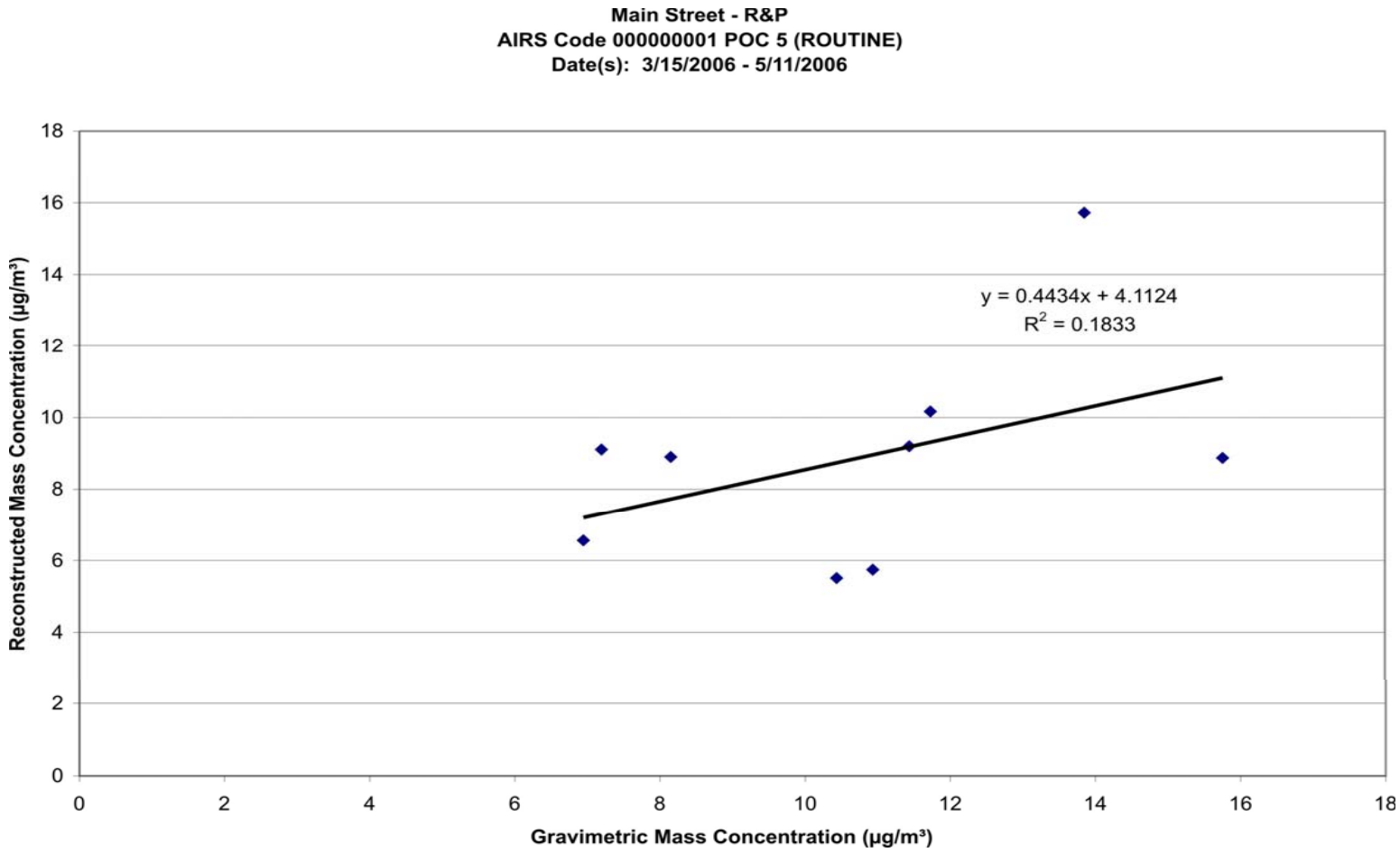
NOTE: Concentrations must be used in the equation. (Adding masses of analytes on filters does not work because different channels in the sampler have different flow rates and sample volumes.)

NOTE: S, Na, and K are excluded because they are measured as ions by IC.

NOTE: Performed on Routine Samples only.

SDVAT: Data Analysis Features

3. Mass Concentration Reconstruction Analysis – XY-Line Plot



SDVAT: Data Analysis Features

4. **<u>Species Distribution Analysis</u> -- Graphic representations of the composition of $PM_{2.5}$ by major components:**

- **Nitrate (total)**
- **Sulfate**
- **Ammonium**
- **Organic Carbon**
- **Elemental Carbon**
- **Crustal Component (calculated below)**
- **Other (calculated below)**

$$\left[\begin{array}{c}\text{Crustal}\\ \text{Component}\end{array}\right] = 2.2[Al] + 2.49[Si] + 1.63[Ca] + 2.42[Fe] + 1.94[Ti]$$

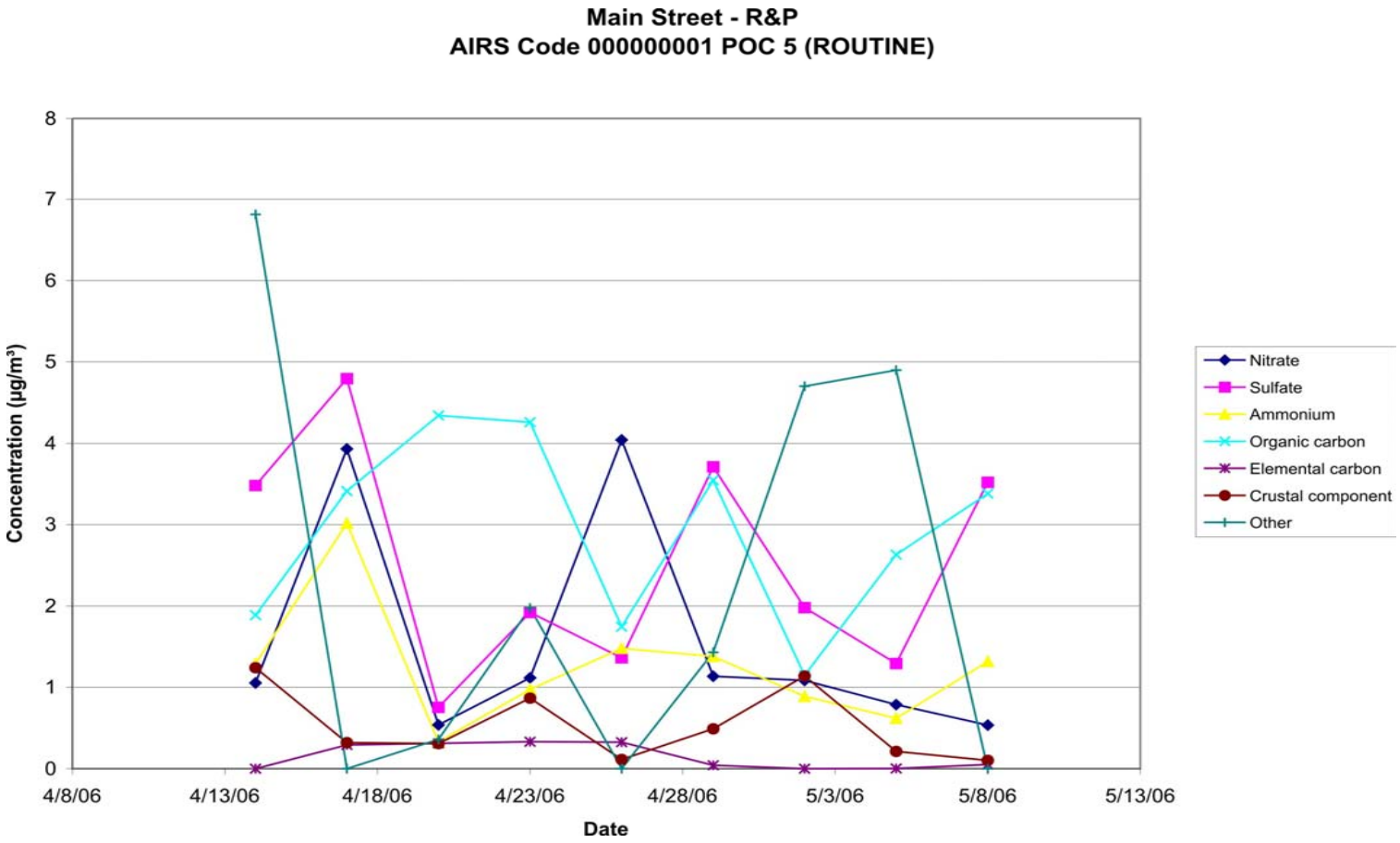
$$\text{Other} = \left[\begin{array}{c}\text{Gravimetric}\\ \text{Mass}\\ \text{Concentration}\end{array}\right] - \left[NO_3^- + SO_4^{2-} + NH_4^+ + OC + EC + \left[\begin{array}{c}\text{Crustal}\\ \text{Component}\end{array}\right]\right]$$

NOTE: Concentrations (not masses) are used in the calculations.

NOTE: If Other < 0, then Other = 0.

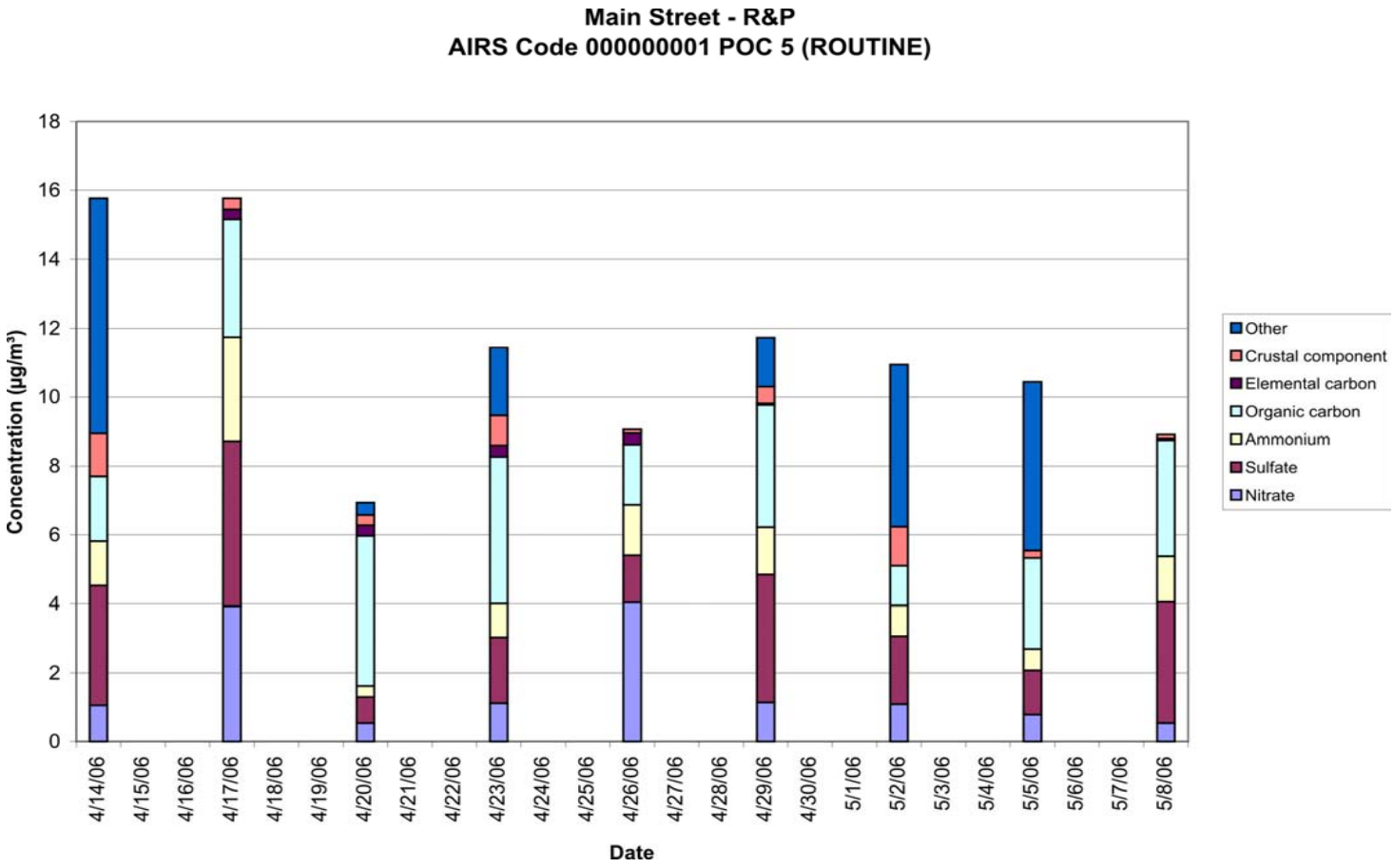
SDVAT: Data Analysis Features

4. Species Distribution Analysis –XY-Line Plot



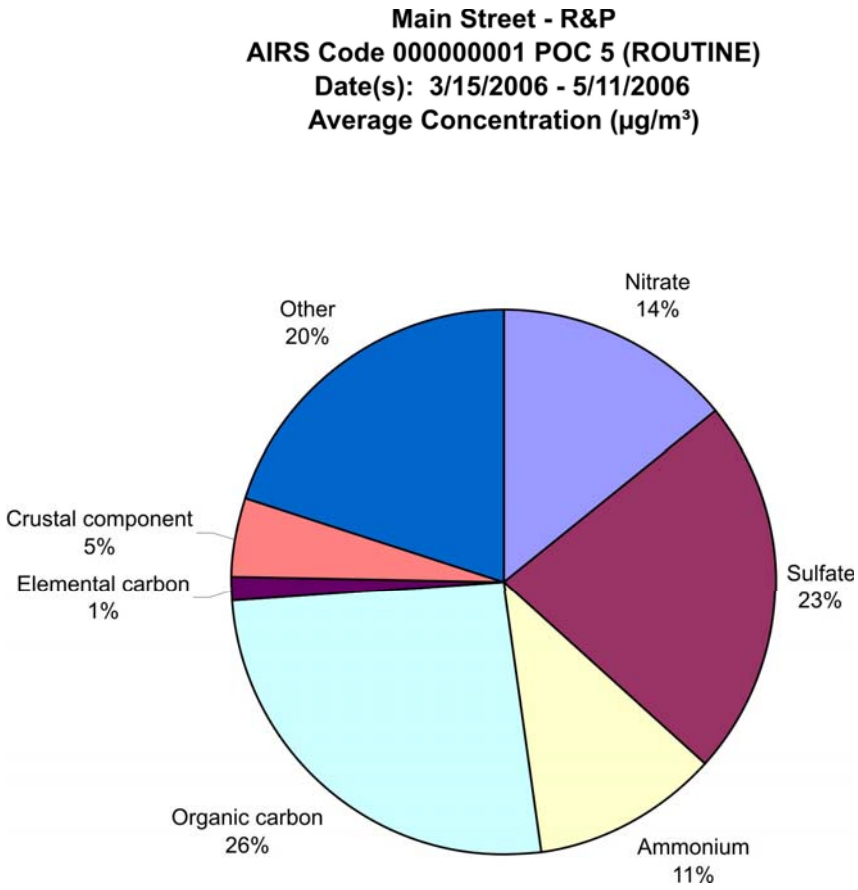
SDVAT: Data Analysis Features

4. Species Distribution Analysis –Stacked Plot



SDVAT: Data Analysis Features

4. Species Distribution Analysis –Pie Chart



SDVAT: Applicability of Selection Criteria

Selection Criterion	Data Completeness	Time Series	Mass Conc. Reconstruction	Species Distribution
Site(s)	✓	✓	✓	✓
Analysis(es)	✓	✓	N/A	N/A
Analyte(s)	✓	✓	N/A	N/A
Date(s)	✓	✓	✓	✓
Sample Type	✓	✓	Routine Only	Routine Only
Validity Type	N/A	✓	✓	✓
With Comment	✓	✓	✓	✓
Analysis Omit	✓	✓	✓	✓

Hands-On Practice

- Real data (names changed to protect the guilty)
  - 8 exercises using the SDVAT
  - 1 optional exercise using Excel to scan RTI data reports for flags
- Exercises include:
  - Running SDVAT data analyses
  - Excluding data
  - Editing data
  - Adding data
  - Assigning custom flags

Conclusions

- The SDVAT is a simple but powerful tool for $PM_{2.5}$ data analysis and validation.
- The new version of the SDVAT eliminates some of the annoying import problems experienced by users when RTI spreadsheet reports were changed to accommodate new ARS data fields.