

US EPA ARCHIVE DOCUMENT

# Integrated Source/Receptor- Based Methods for Source Apportionment and Area of Influence Analysis

U.S. EPA STAR PM Source Apportionment  
Progress Review Workshop

June 21, 2007

Jaemeen Park, Talat Odman, Yongtao Hu, Florian  
Habermacher, and Ted Russell  
*Georgia Institute of Technology*

# Overview

- Introduction
- Source apportionment of PM<sub>2.5</sub>
  - Primary PM<sub>2.5</sub> using CMAQ
  - Regional/Secondary PM<sub>2.5</sub> using CMAQ-DDM
  - Improving emission inventories
    - Using metal tracer species
    - Inverse modeling using CMAQ-DDM
- Area of Influence Analysis

# Objectives

- Extend a recently developed source apportionment (SA) method for ozone to PM2.5/coarse
- Inter-compare results from source-apportionment methods (receptor and source-oriented approaches).
- Identify strengths and limitations of the approaches, focusing on the reasons for disagreement.
- Quantify uncertainties involved in the application of the various source apportionment methods.
- Refine and apply inverse modeling to improve emissions and source apportionment determinations.
- Develop the Area-of-Influence analysis technique.
- Assess the relative strengths of using Supersite data vs. routine monitoring data for SA applications.
- Provide source apportionment results to air quality managers and epidemiologic researchers.

# Approach

- Apply various modeling tools to conduct source apportionments
  - CMAQ-DDM3D
  - CMB
    - Regular, Molecular Marker, LGO w/gases
  - PMF
  - Inverse modeling (CMAQ-DDM-FDDA)
- Use the extensive data from the Supersites, SEARCH, ASACA and STN
  - Focus on SE, particularly Atlanta:
    - Atlanta Supersite: Extensive PM and gaseous data in summer 1999
    - SEARCH: SE, detailed PM and gaseous data since 1998
    - ASACA: Atlanta, daily PM composition since 1999
  - Larger scale focus using ESP data (July-August, 2001; January, 2002)
- Conduct uncertainty assessments

# Study Area and Periods

## Modeling periods:

August 1999

July 2001

January 2002

July 2005

January 2006

## Base inventories

EPA NEI

## Point sources in Georgia

EPA NEI 2002 (draft),

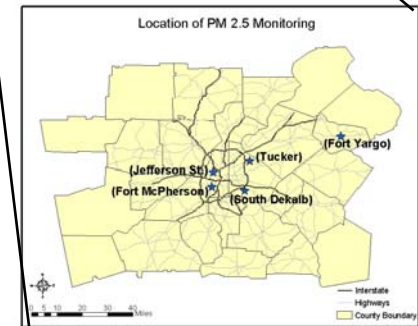
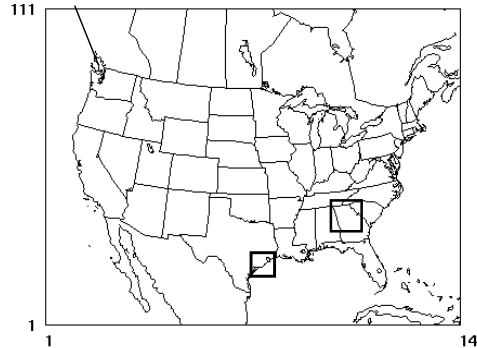
CEM data

## Forest fire, land clearing debris in 2002

VISTAS, 2005

## Residential meat cooking

New emissions were added



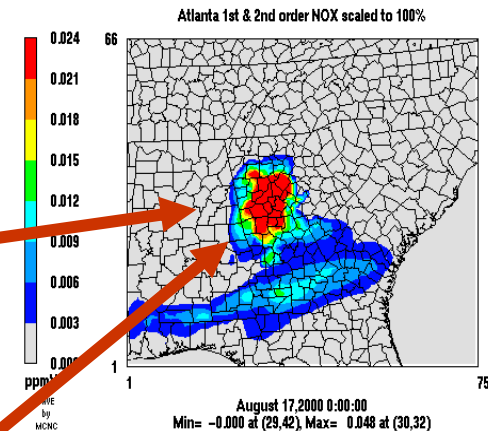
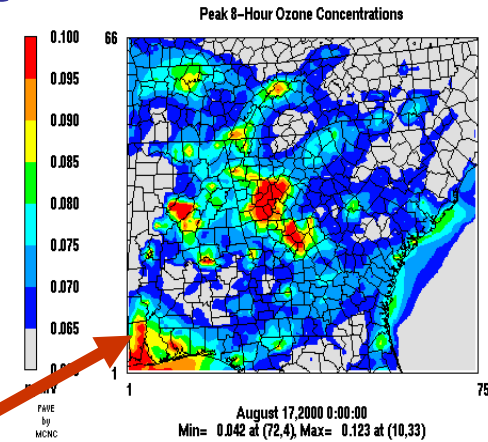
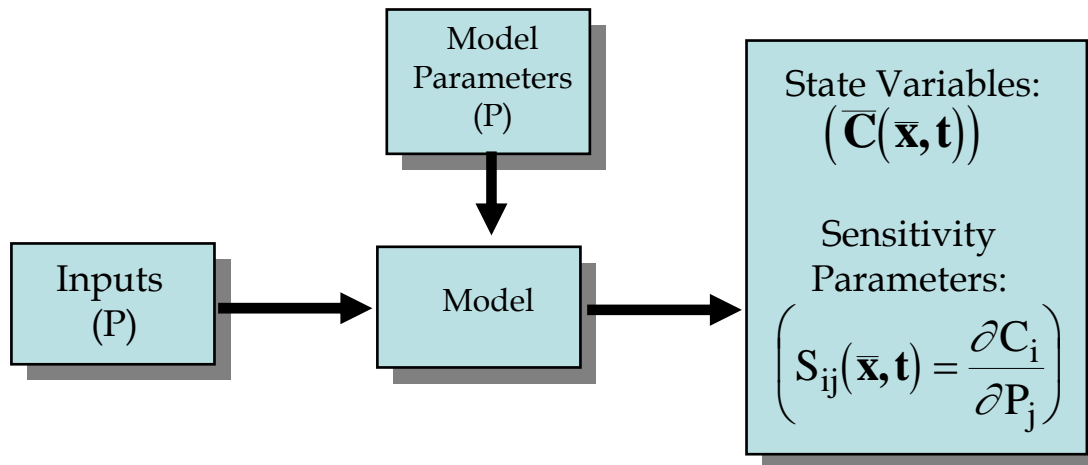
## SEARCH monitoring sites

- Urban sites : Atlanta, Jefferson St. (JST) Birmingham (BHM), Gulf port (GFP), Pensacola (PNS)
- Suburban sites: Pensacola (OLF)
- Rural sites: Oak Grove (OAK), Centreville (CTR), Yorkshire (YRK)

## ASACA

# DDM-Sensitivity analysis based Source Apportionment

- Given a system, find how the state (concentrations) responds to incremental changes in the input and model parameters:



If  $P_j$  are emissions,  $S_{ij}$  are the sensitivities/responses to emission changes, e.g., the sensitivity of ozone to Atlanta NO<sub>x</sub> emissions

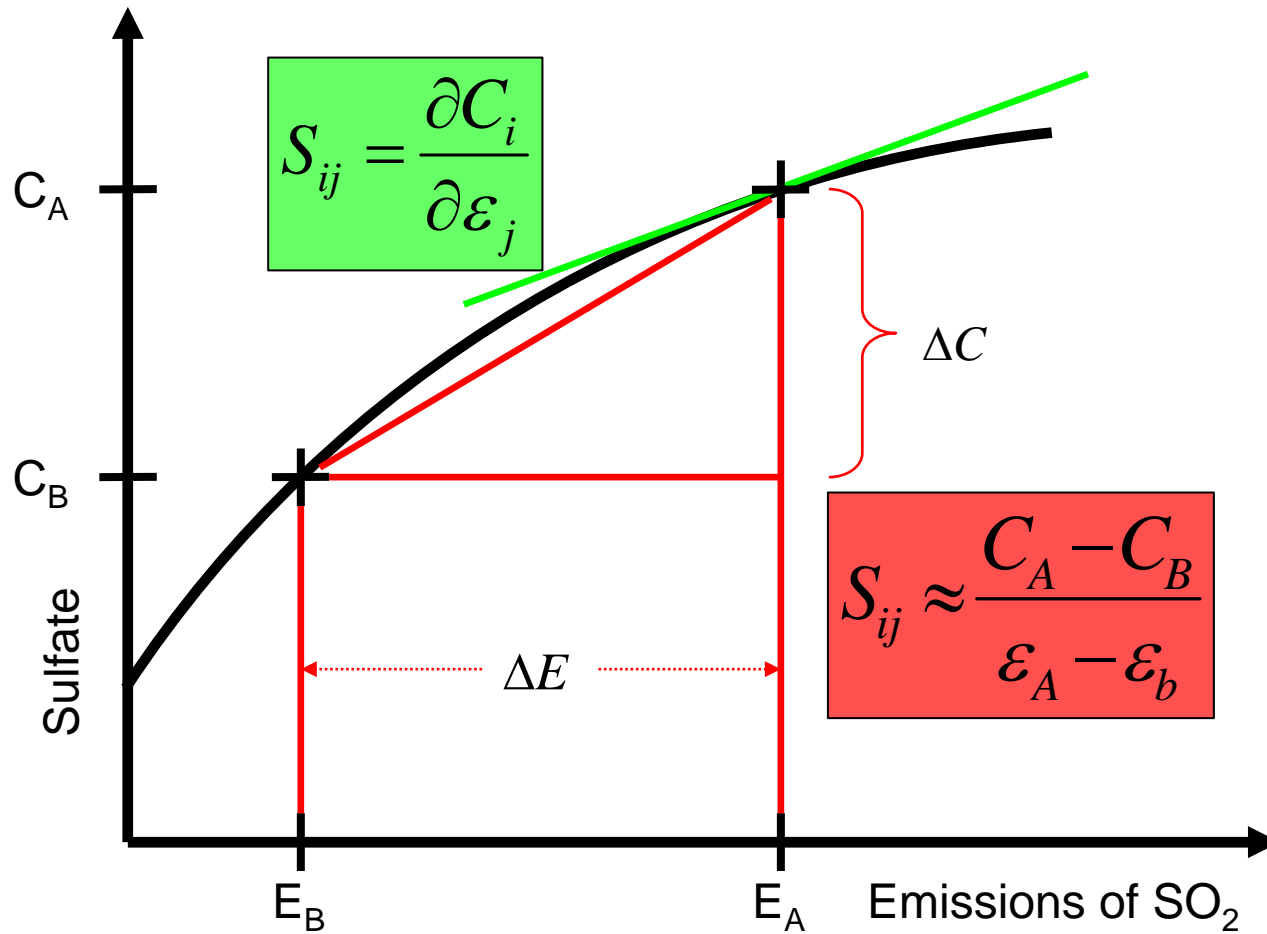
# Sensitivity Analysis with Decoupled Direct Method (DDM): The Power of the Derivative

- Define first order sensitivities as  $S_{ij}^{(1)} = \partial C_i / \partial E_j$
- Take derivatives of  $\frac{\partial C_i}{\partial t} = -\nabla(\mathbf{u}C_i) + \nabla(\mathbf{K}\nabla C_i) + R_i + E_i$
- Solve sensitivity equations simultaneously

	Advection	Diffusion	Chemistry	Emissions
$\frac{\partial C_i}{\partial t} =$	$-\nabla(\mathbf{u}C_i)$	$+ \nabla(\mathbf{K}\nabla C_i)$	$+ R_i$	$+ E_i$
$\frac{\partial S_{ij}}{\partial t} =$	$-\nabla(\mathbf{u}S_{ij})$	$+ \nabla(\mathbf{K}\nabla S_{ij})$	$+ \mathbf{J}S_j$	$+ \delta_{ij}E_i$



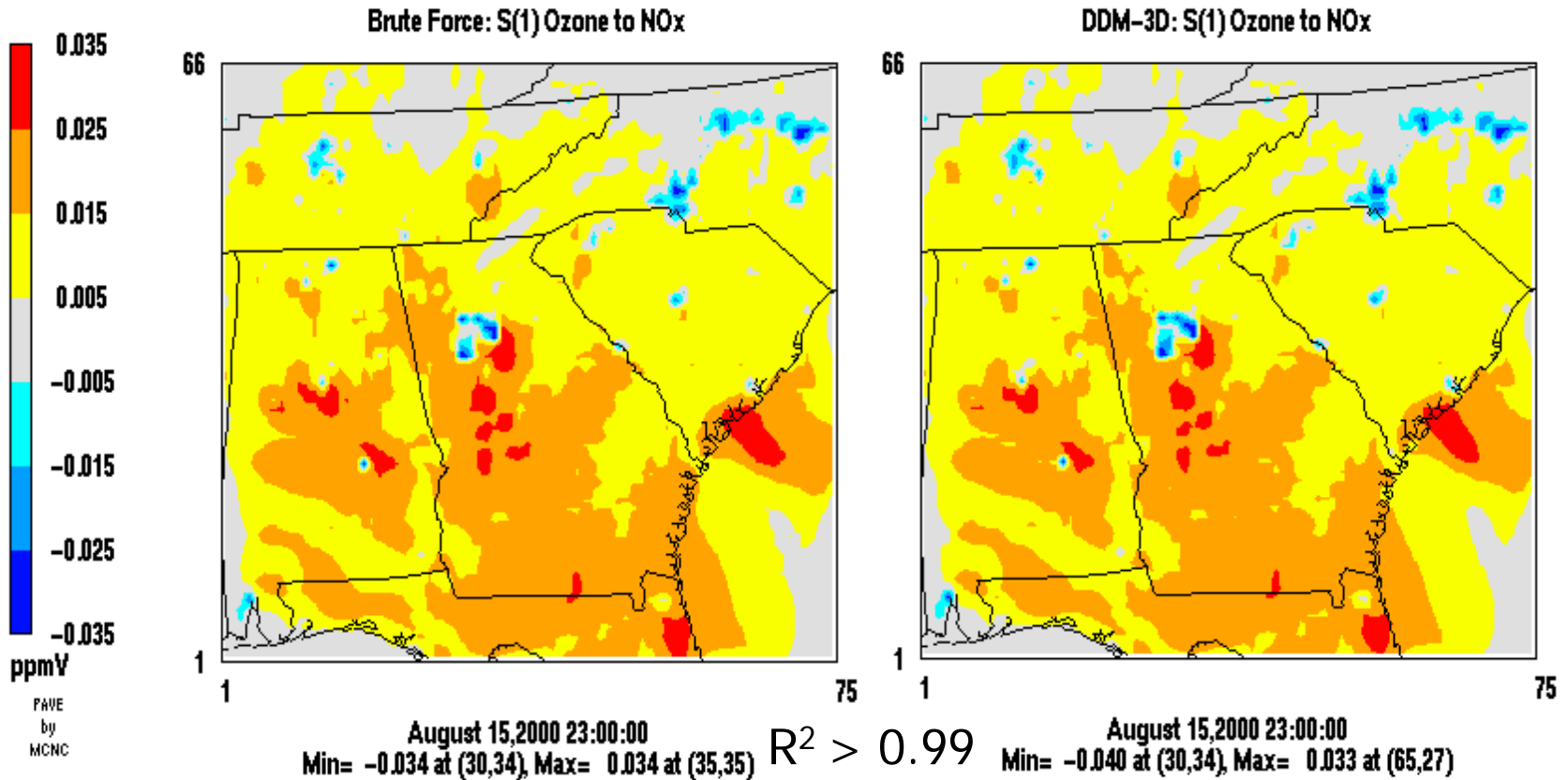
# DDM compared to Brute Force



# Consistency of first-order sensitivities

Brute Force (20% change)

DDM-3D



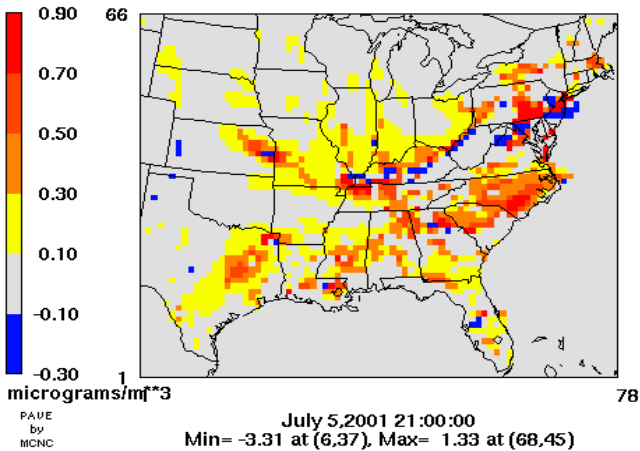
# Advantages of DDM-3D

- Computes sensitivities of all modeled species to many different parameters in one simulation
  - Can "tell" model to give sensitivities to 10s of parameters in the same run
- Captures small perturbations in input parameters
  - Strangely wonderful
- Avoids numerical errors sometimes present in sensitivities calculated with Brute Force
- Lowers the requirement for computational resources

# Evidence of Numerical Errors in BF

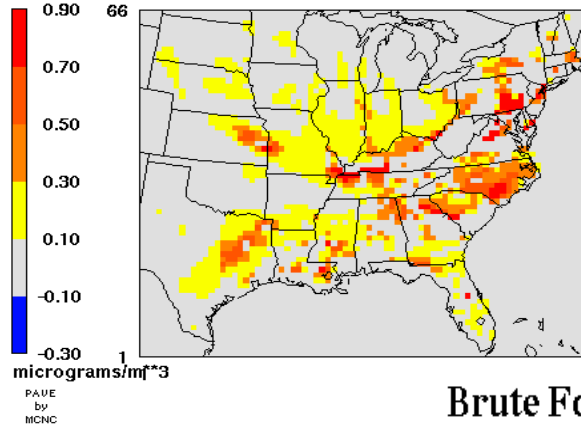
## Brute Force

Sensitivity of NH<sub>4</sub> to domain-wide emissions of SO<sub>2</sub>



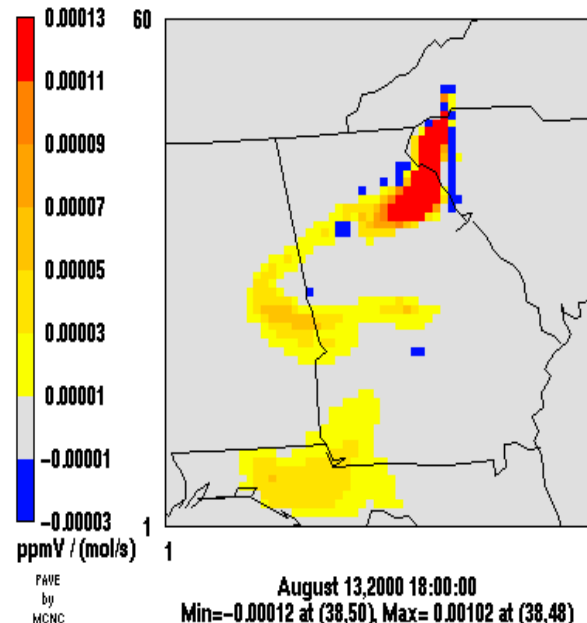
## DDM

Sensitivity of NH<sub>4</sub> to domain-wide emissions of SO<sub>2</sub>



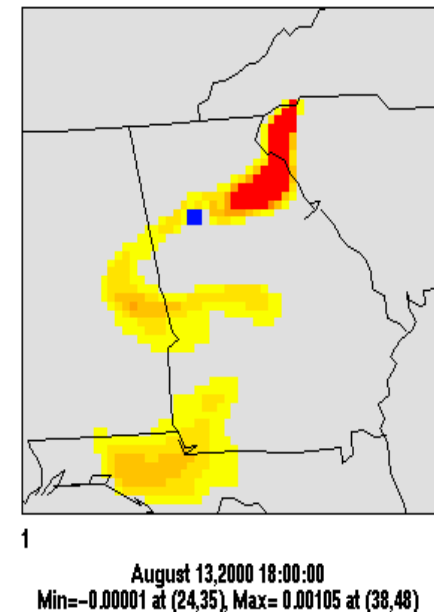
## Brute Force

Sensitivity of surface ozone to a 1 mol/s NO source emitted into (38,49), Layer 6



## DDM

Sensitivity of surface ozone to a 1 mol/s NO source emitted into (38,49), Layer 6

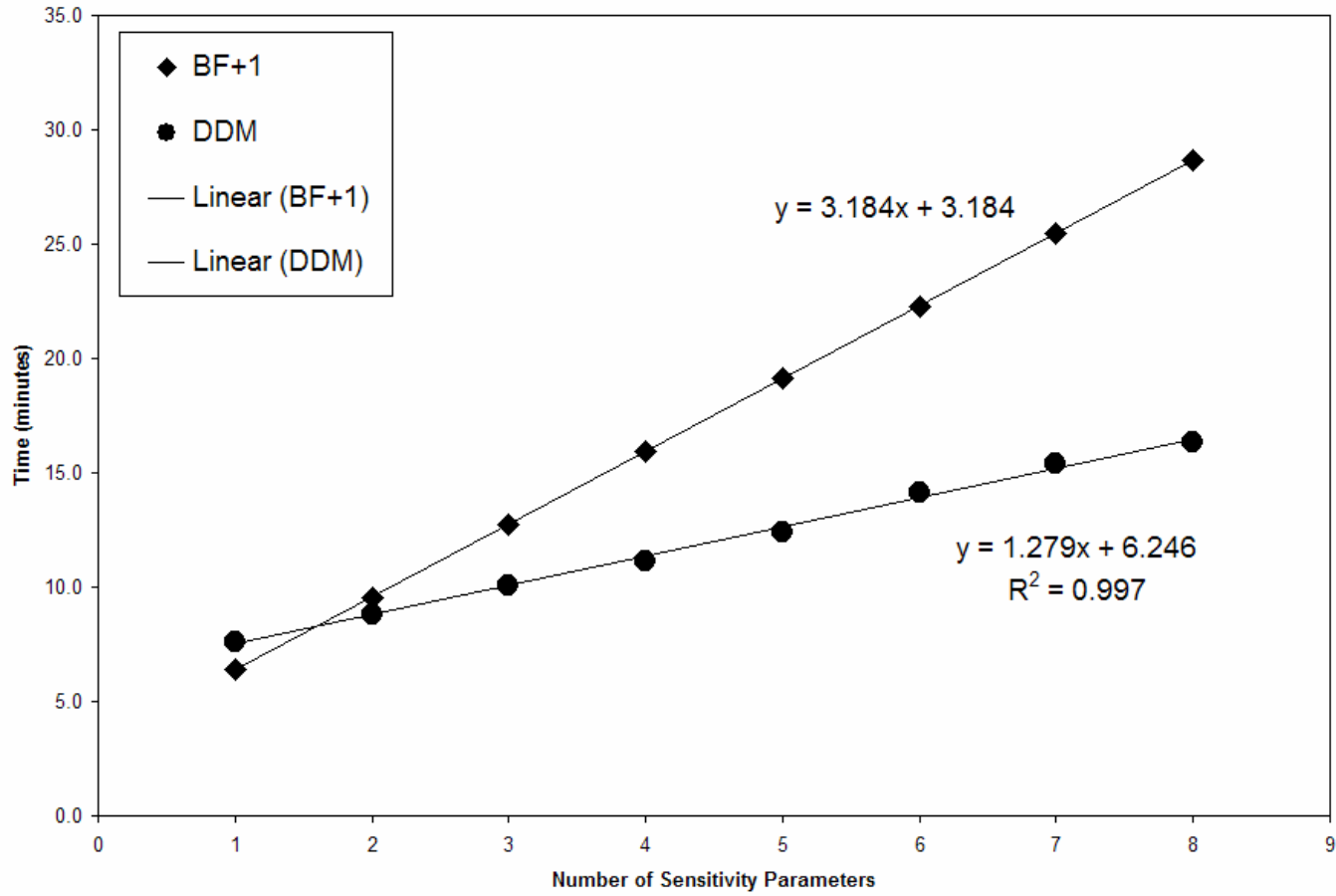


NH<sub>4</sub> sensitivity to domain-wide SO<sub>2</sub> reductions

NO<sub>x</sub> reductions at a point

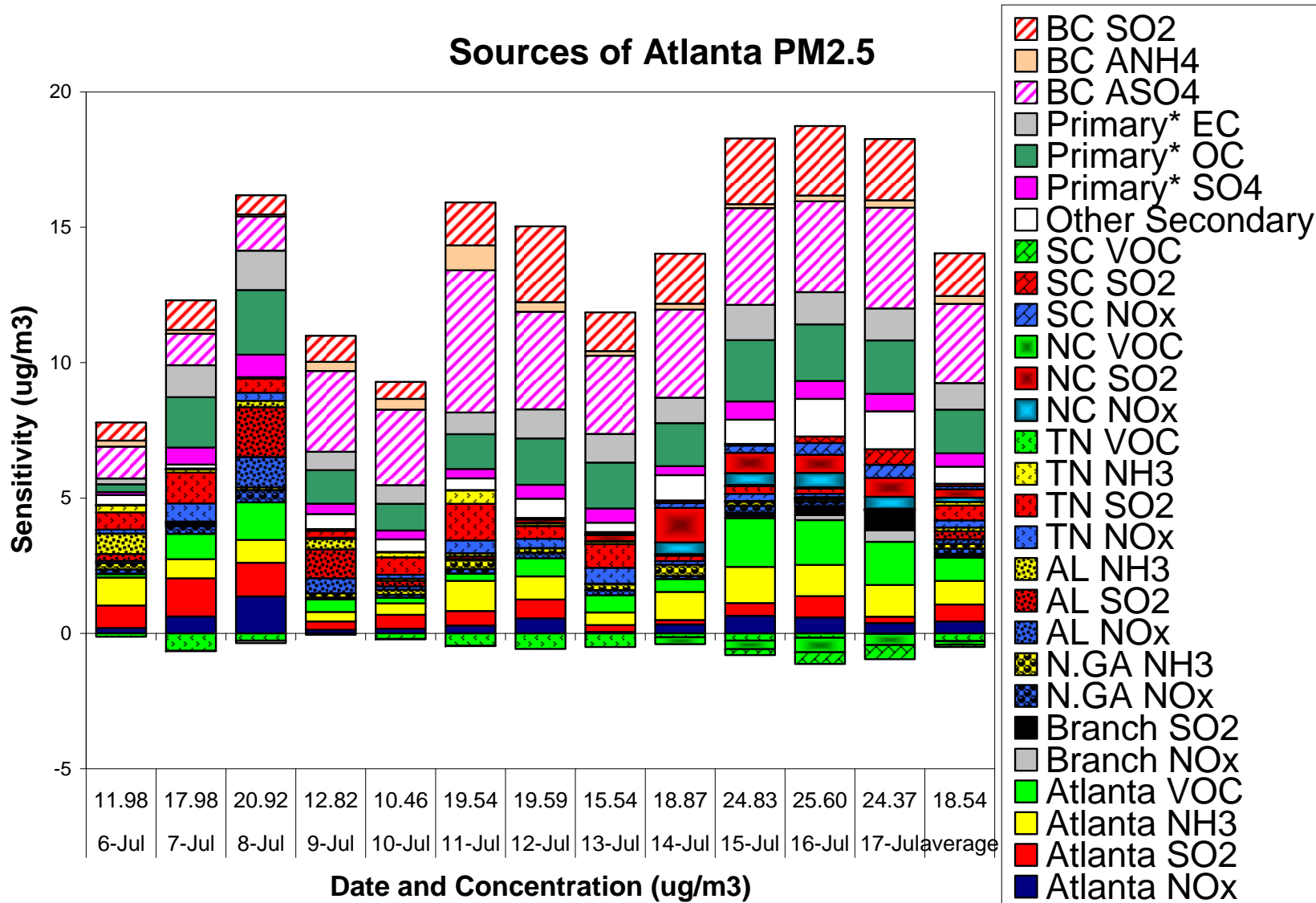
In recent study, brute force led to multiple maxima and minima being due to noise.

# Efficiency of DDM-3D



# Regional Source Apportionment of PM 2.5

## Using Direct Sensitivity: Application to Georgia



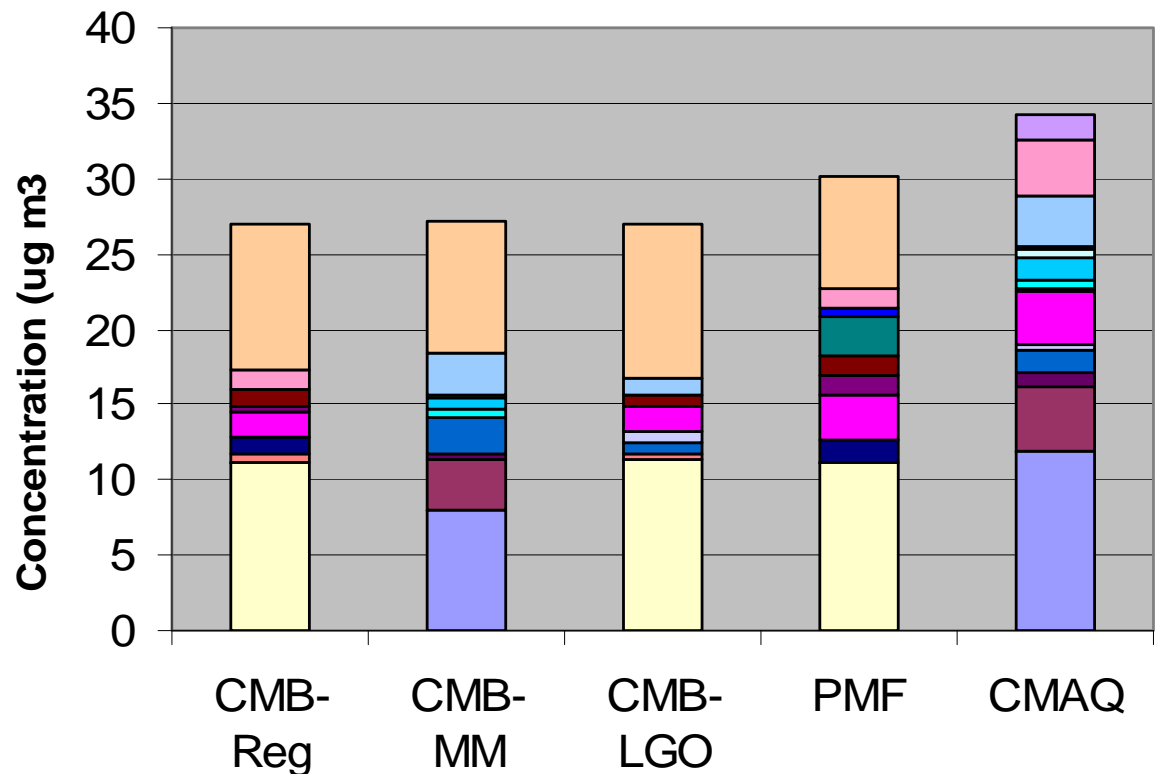
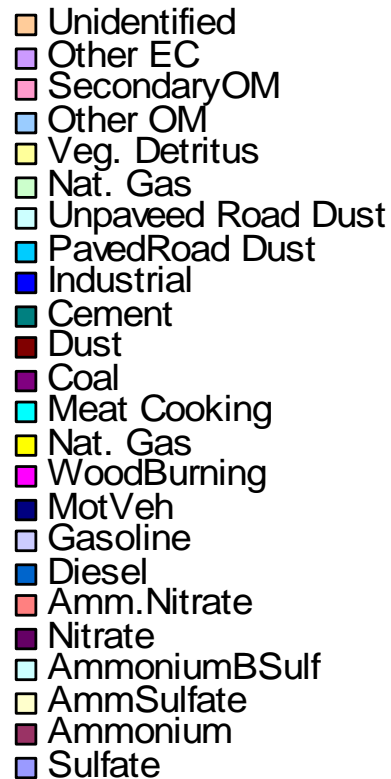
# Intercomparison of Source Apportionment Methods

- Apply a variety of methods to relatively rich data base of PM in the SE
  - Supersite, SEARCH, ASACA, STN
- Methods
  - CMAQ-DDM
  - PMF
  - CMB-Regular (typical analysis using STN-type data)
  - CMB-Molecular Marker (using organic molecular speciation)
  - CMB-LGO (optimized, using gas phase species, w/wo re-optimization of source profiles)
    - Adding gaseous species really helps: Don't stop monitoring CO, SO<sub>2</sub> and NO<sub>x</sub>!
    - Re-optimization of profiles made smaller difference

# Source apportionment of PM2.5 OC from different models

- Air quality model : CMAQ/DDM3D-PM

- Receptor models : CMB, PMF

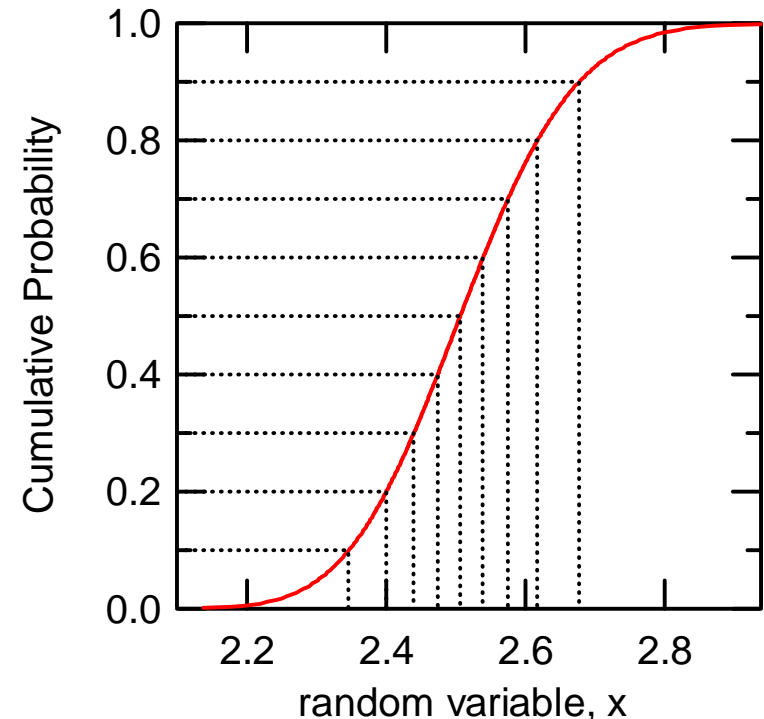


*N.b. Not all sources are common to all methods*

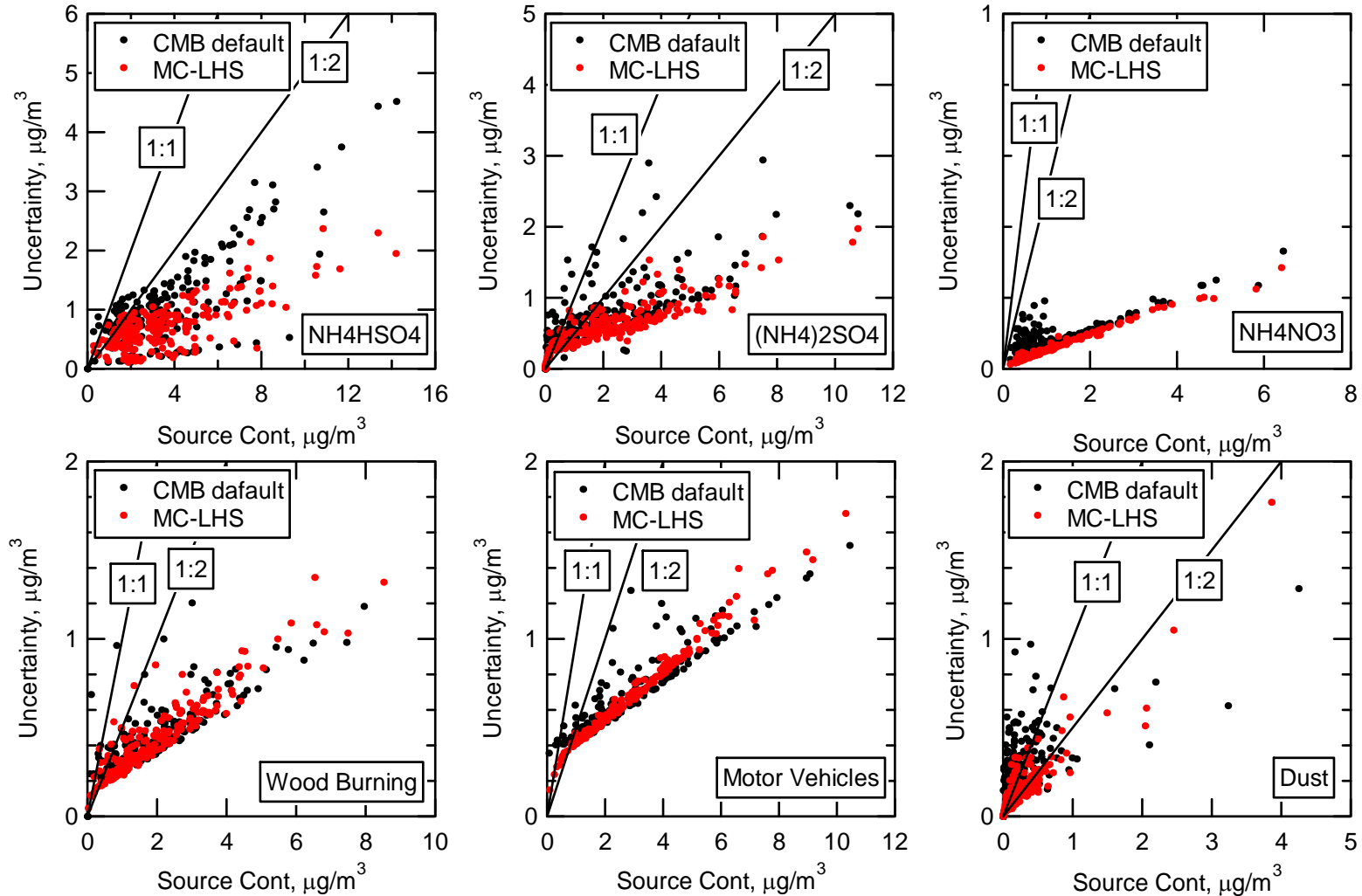


# Looking at Uncertainties: Monte Carlo Analysis of CMB with Latin Hypercube Sampling (LHS)

- Assume log-normally distributed variables in source profiles and ambient data
  - PM<sub>2.5</sub> data from Atlanta, GA (EPA STN): Jan 02 ~ Nov 03 (# of data points: 212)
- Construct CDF for each variable using uncertainties
  - Divide into 500 equal probable intervals
- Sample from each variable PDF 500 times
  - Constrain source profiles ( $\sum_{i=1}^n f_i \leq 1$ )
- 500 simulations using CMB

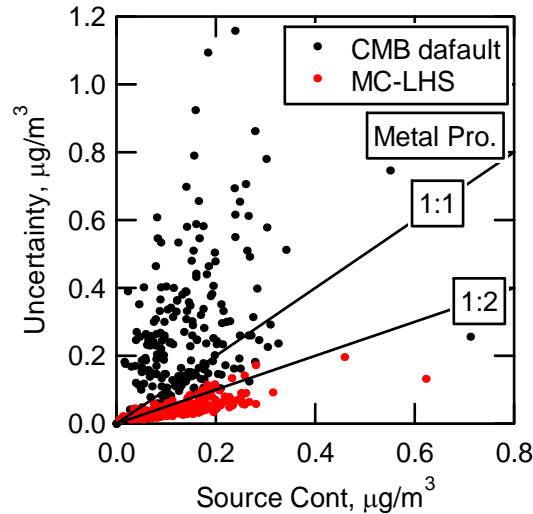
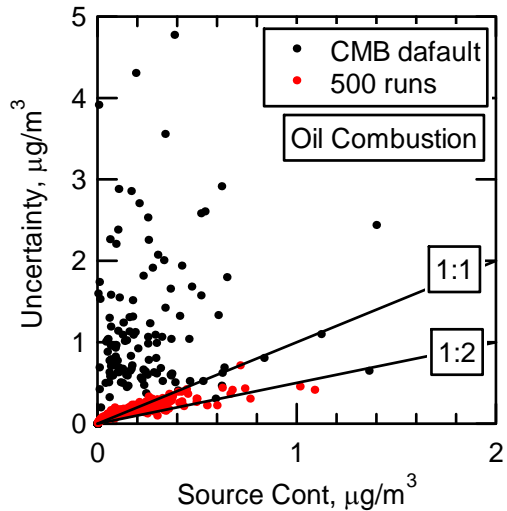
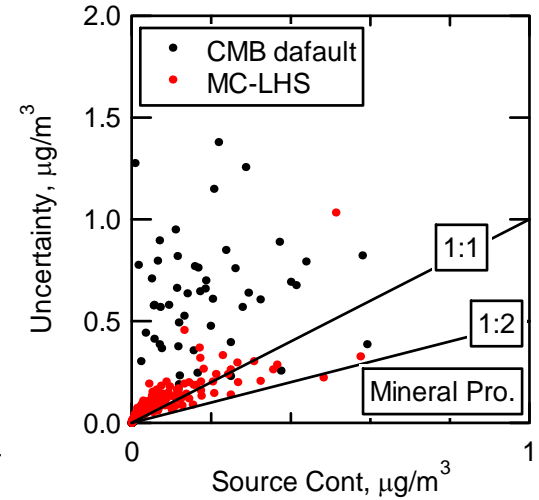
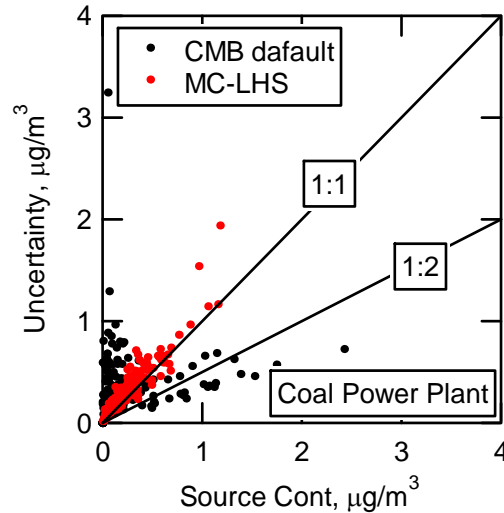
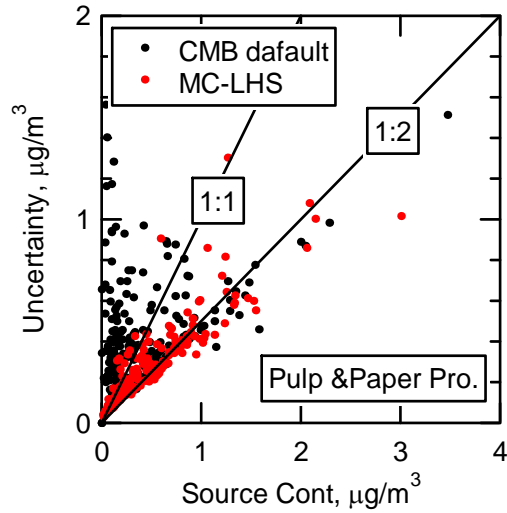


# Uncertainty vs. Source Contribution



If  $S_j < \sigma_{s_j}$  **non-detectable source**

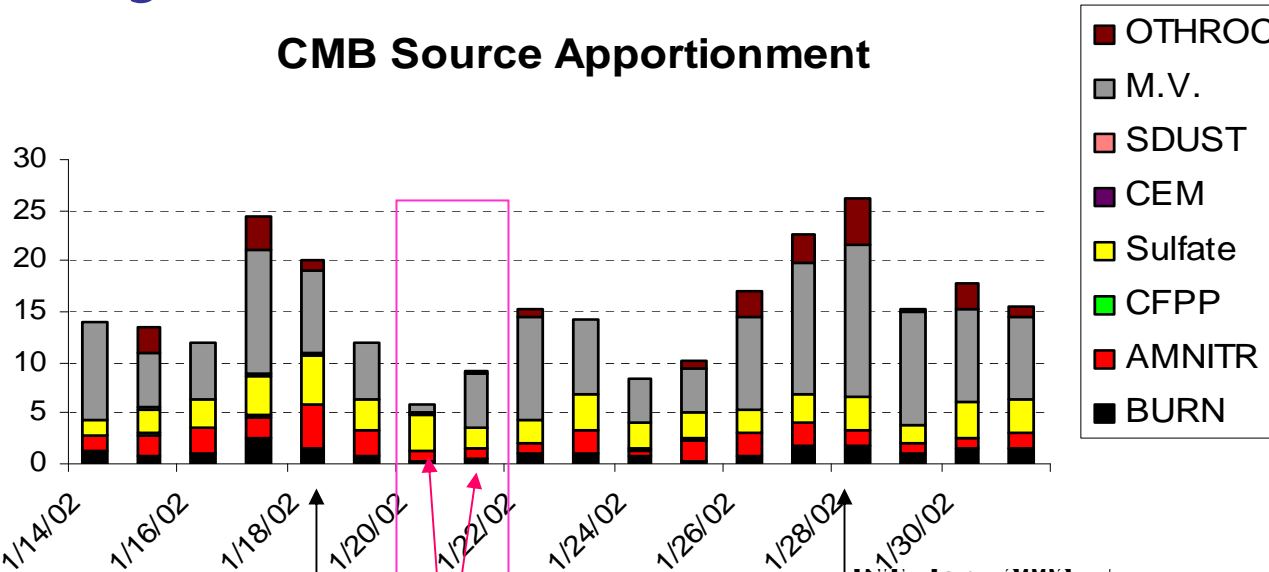
# Uncertainty vs. Source Contribution



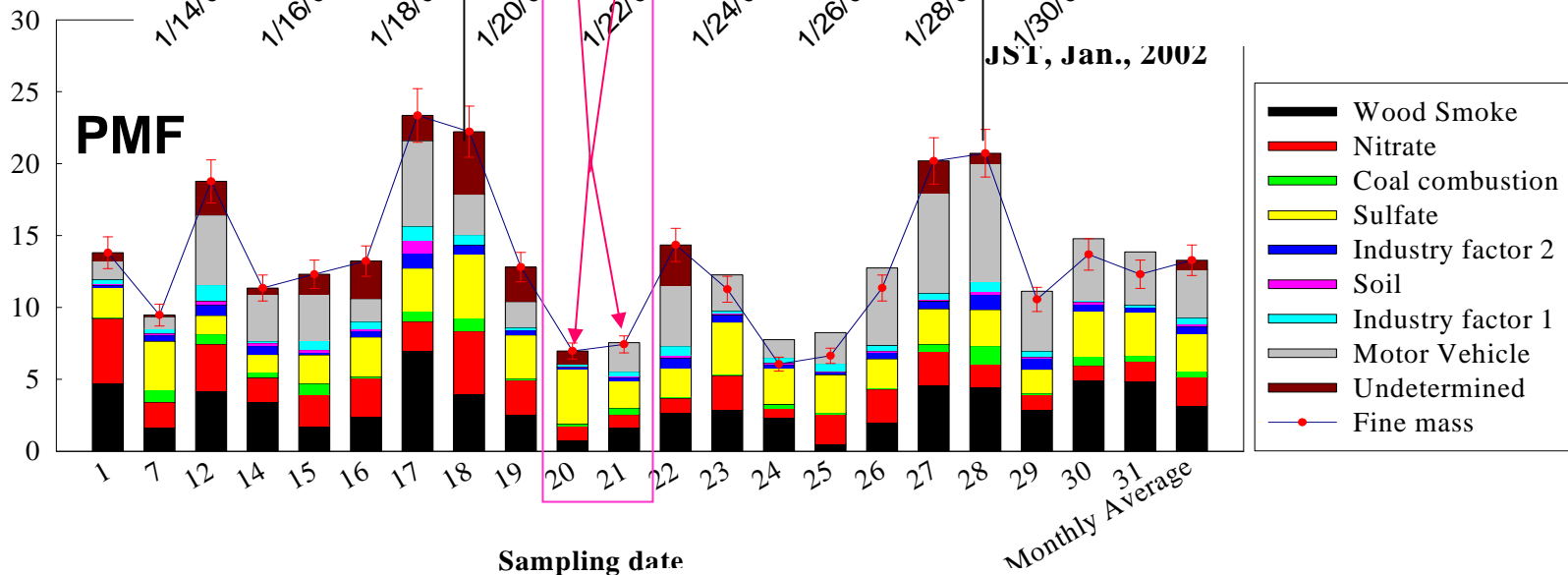
If  $S_j < \sigma_{s_j}$  **non-detectable source**

# Daily Variation: PMF vs. CMB-LGO

## CMB Source Apportionment



## PMF

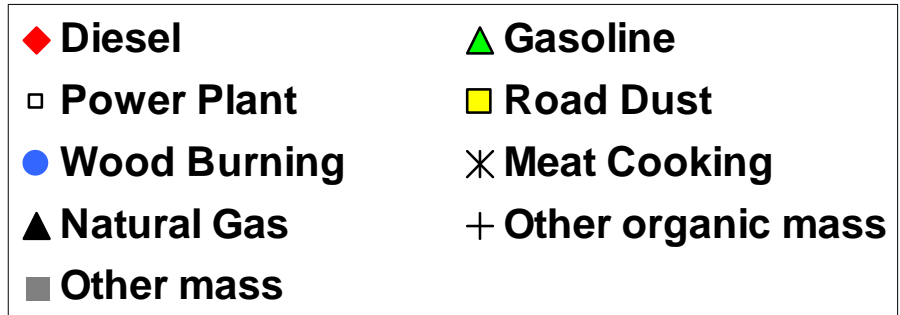
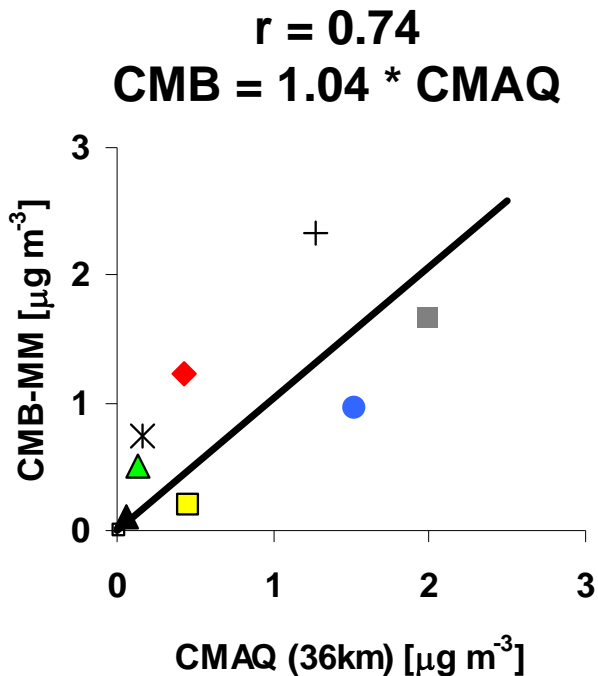


Note daily variability in relative source contributions

# Mass contributions to PM2.5: Comparison of CMB-MM and CMAQ

Averaged contribution  
over the eight SEARCH stations  
for July 2001 and January 2002

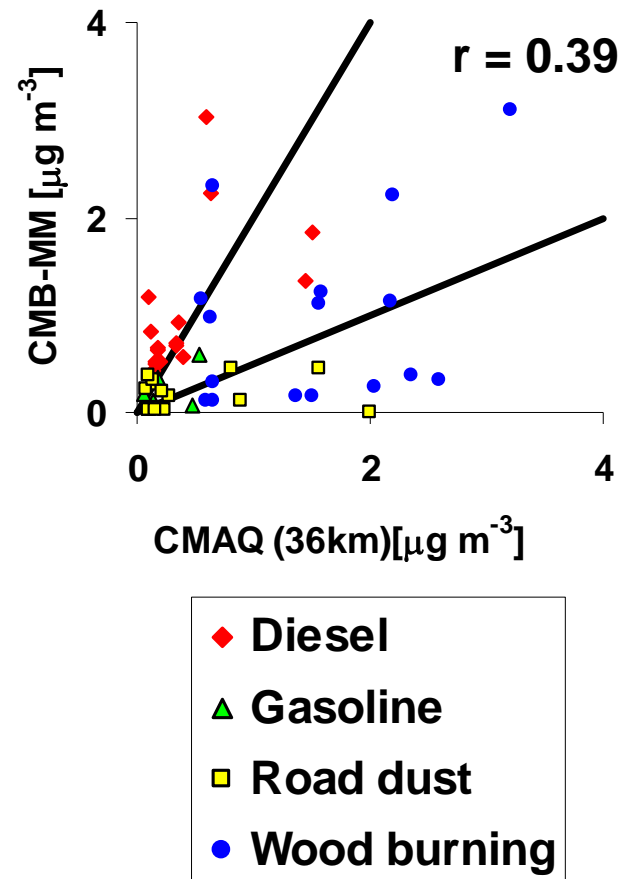
- Average of source contributions looks pretty good, particularly just looking at source impacts, but...



# Disaggregated some: not so good

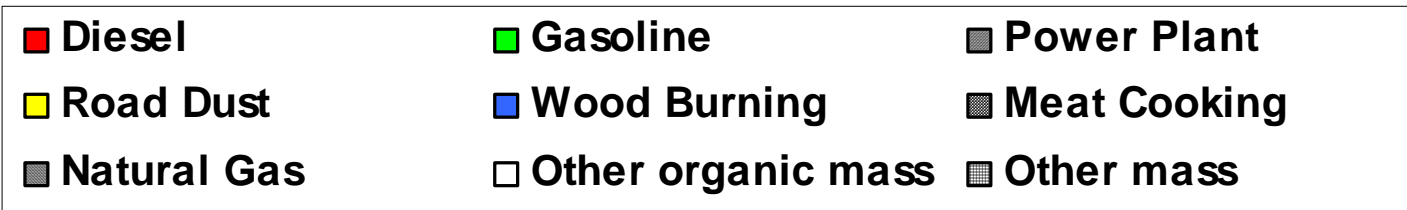
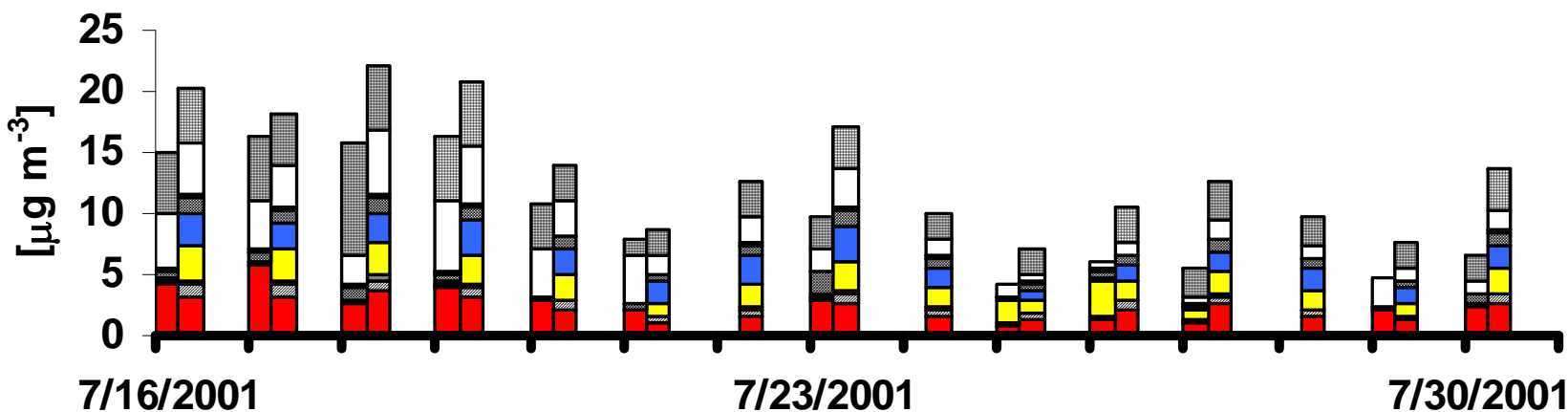
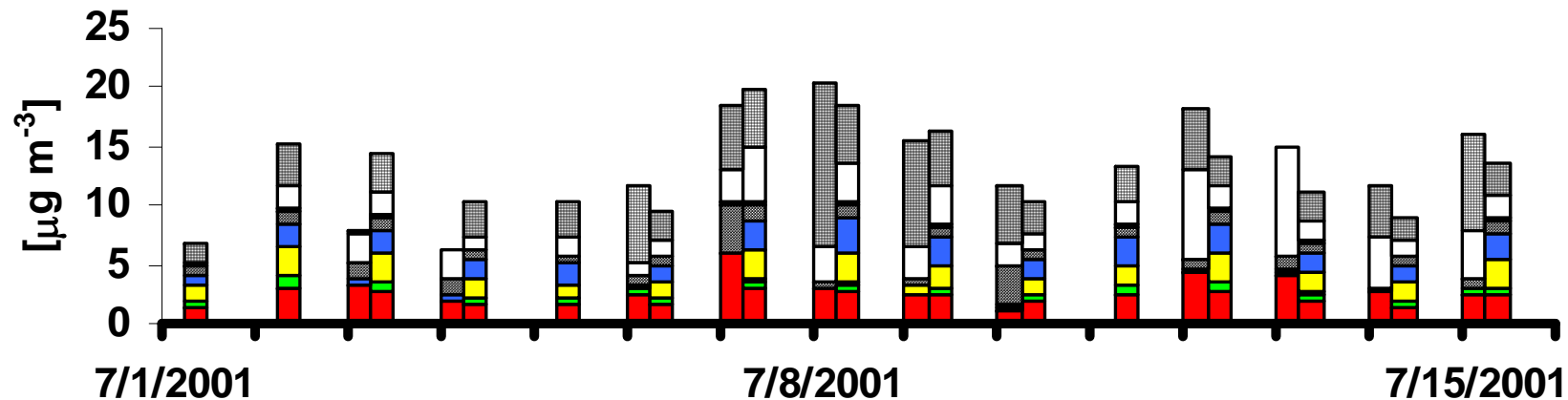
- If we look at the results by specific source at individual stations, not quite so good, and further, look at daily agreement...

Monthly contributions  
in SEARCH stations  
for July 2001 and January 2002



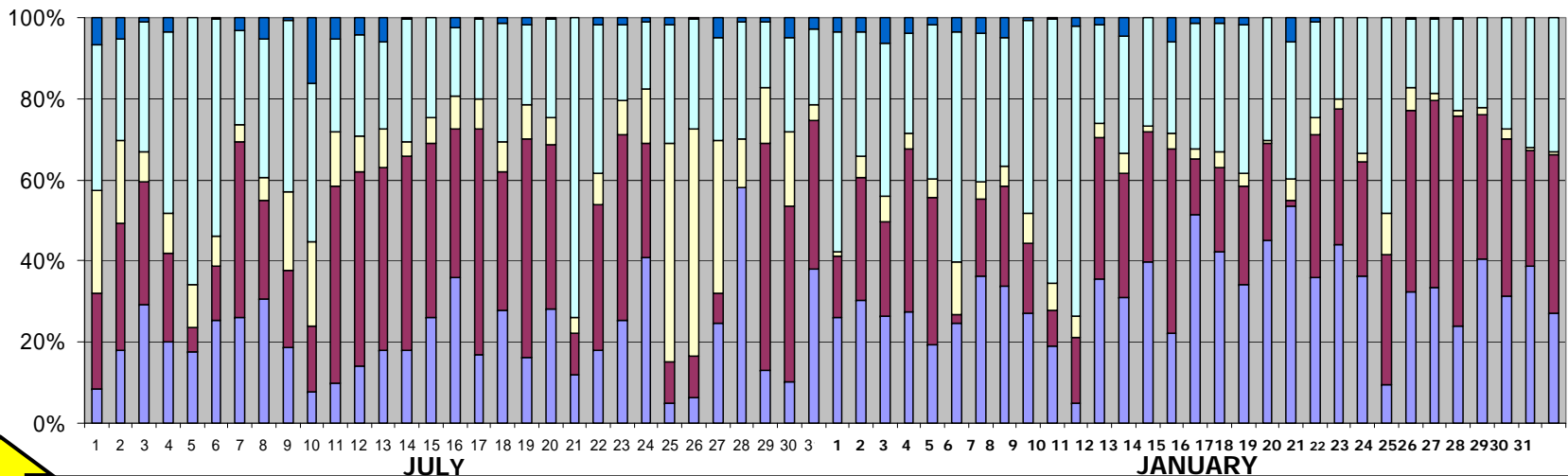
# Daily average mass contributions to PM2.5 in July 2001

CMB-MM and CMAQ (left to right)



# Fraction of primary PM<sub>2.5</sub> - JST

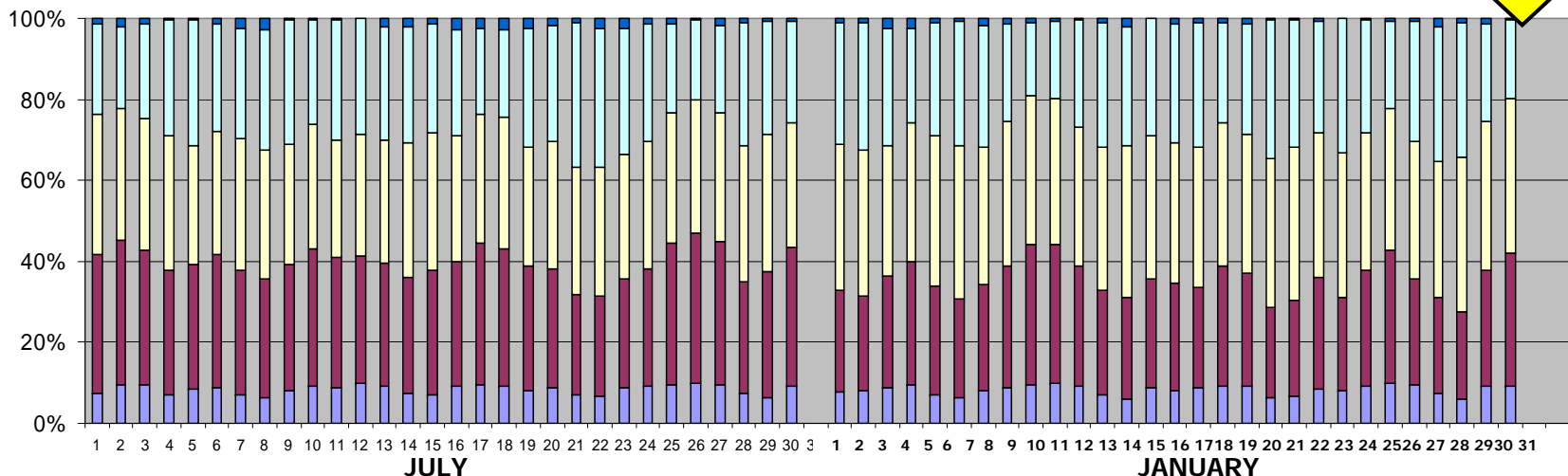
CMB-LGO



Significant daily variation

Little daily variation

CMAQ

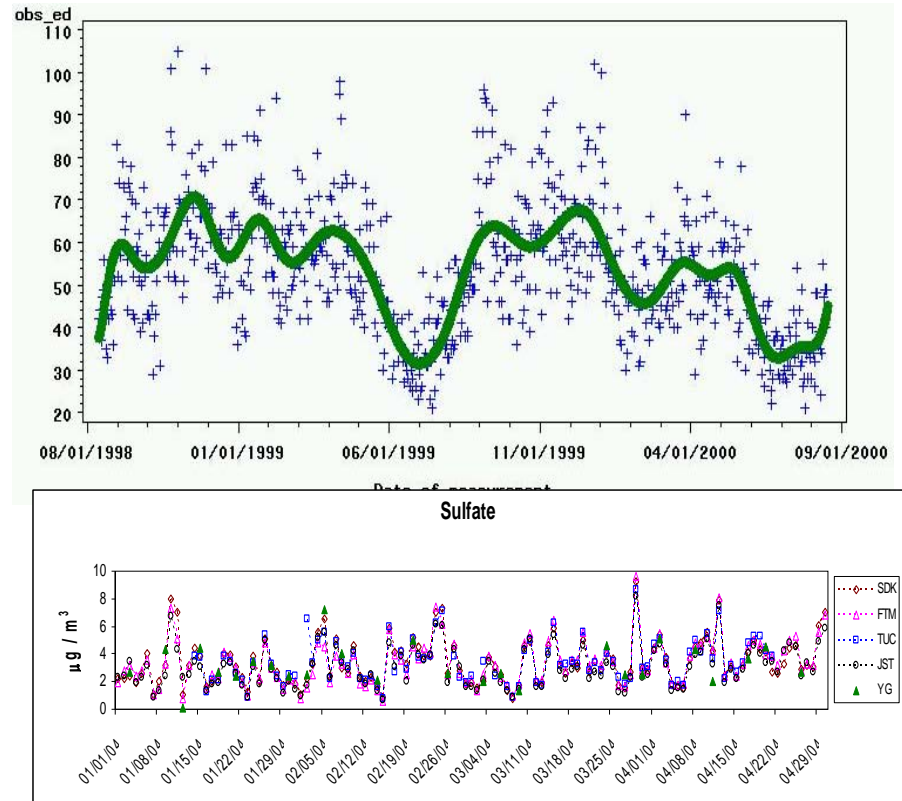


LDGV HDDV SDUST BURN CFPP



# Daily Variation is Important!

- Health associations are derived from how concentrations/outcomes deviate from the norm on a daily basis
  - Too little or too much (or wrong) will inhibit identification of outcomes and exposure-response relationship
    - Bias to the null and loss of power



# Cautionary notes on receptor and emissions-based air quality models

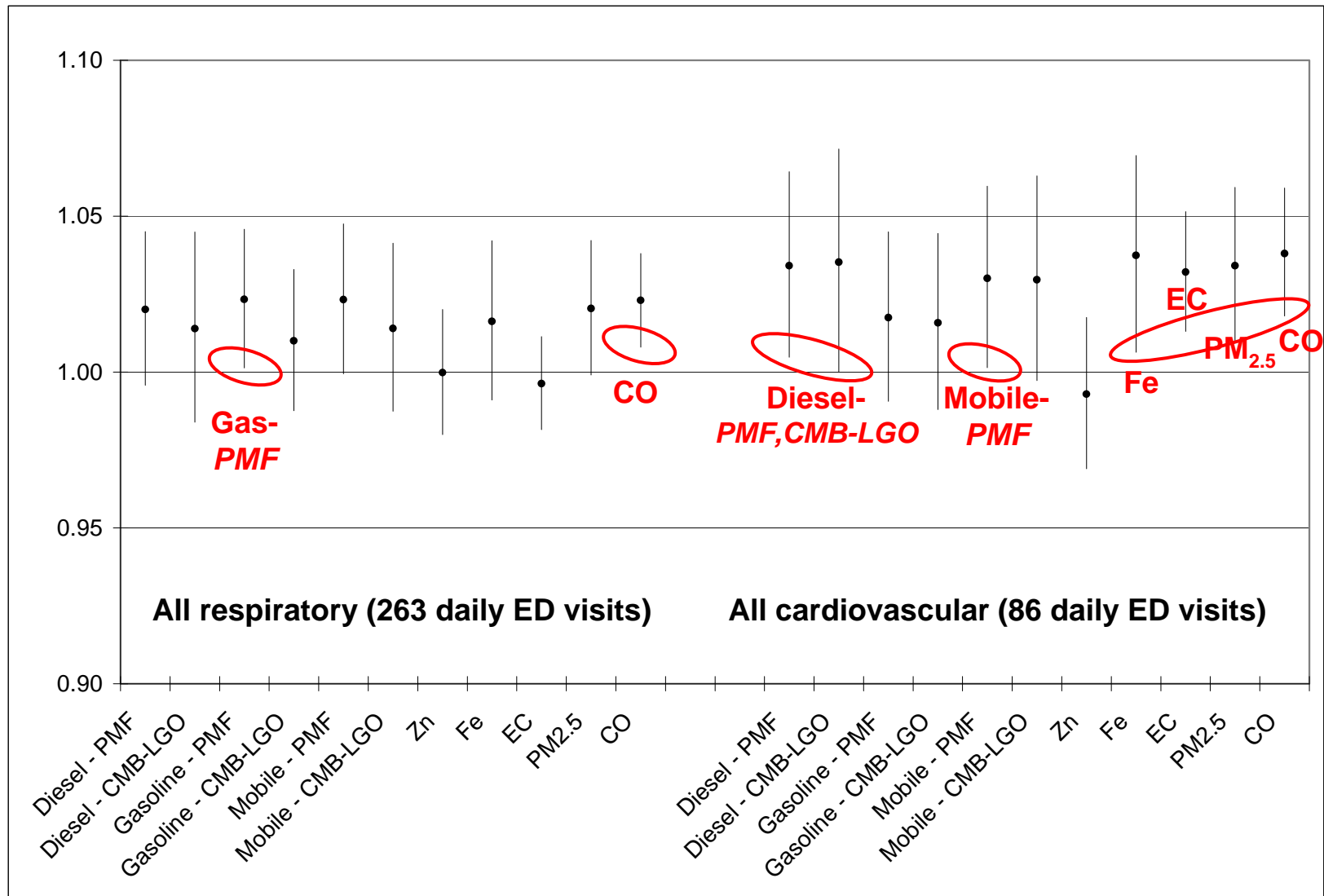
- Both approaches
  - Tend to agree relatively well on average
  - (usually) identify large vs. small sources
    - Sometimes by their absence
- Receptor models
  - Methods, based on largely the same data, give different results
    - Significant uncertainties
  - Gives more temporal variation in source impacts at a specific receptor site
    - Too much? (reasons to think so)
  - Not apparent how to conduct thorough evaluation and uncertainty analysis for all methods
- Emissions-based models
  - Propagate uncertainties in variety of inputs and process descriptions
  - Have less day-to-day variability (probably too little)
    - Meteorological models and inventories do not capture temporal variability well
    - May be more spatially representative
  - Can have obvious disagreements with the data
    - At least we know there is a problem!

# Application to Health Effects Associations

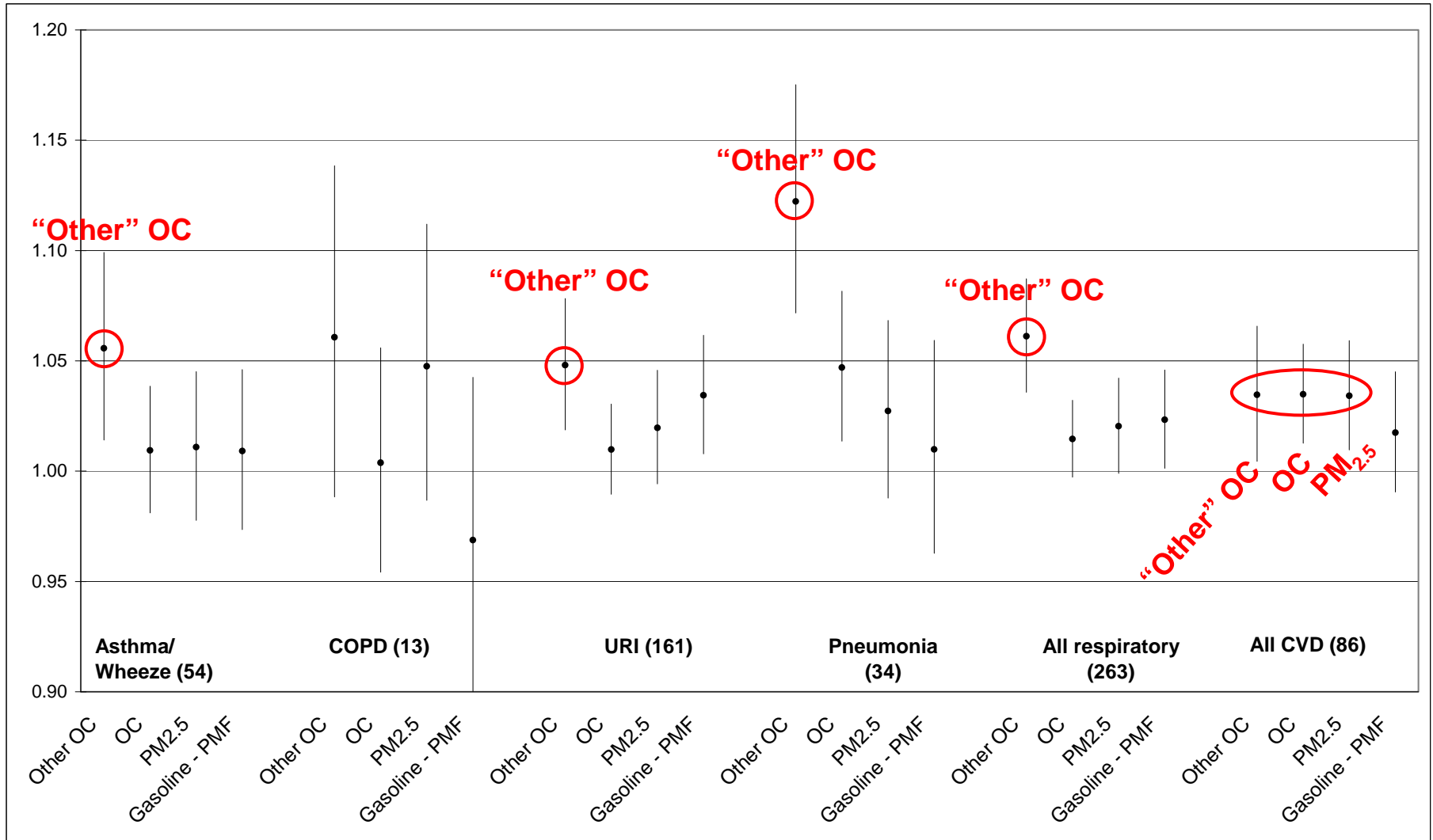
- Used CMB-LGO and PMF\*
- Applied in an emergency department time-series study (Rollins School of Public-Health, Emory University)
- Relative Risks (RRs) associated with change in inter-quartile-range of 3-day moving averages of PM<sub>2.5</sub> levels were estimated using Poisson generalized linear models.

\* - Kim et al., *Atm Env* 38, 3349-3362, 2004

# Source-specific RRs: Mobile sources



# Source-specific RRs: "Other" OC



*Preliminary results... further analysis appears to reduce other OC association*

# Analysis of Area of Influence (AOI)

- To identify which sources or regions might impact a specific receptor
- Uses source based sensitivity and AOI to determine the spatial distribution of emission influences
  - Evaluate the impact of specific existing sources
  - Predict the impact of future sources
- Uses source based sensitivity fields to generate receptor-based sensitivity fields
- Method is based on the DDM-3D functionality in CMAQ
- Computationally less intensive than adjoint modeling for multiple receptors

# AOI Development – Reverse Fields

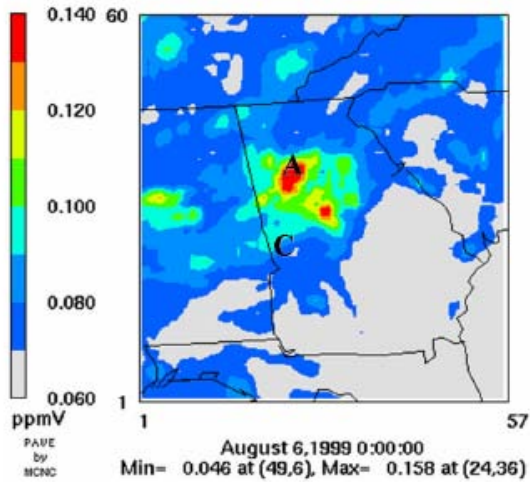
- Using the complete set of forward sensitivities (at each point in the domain), receptor oriented fields can be computed at any point using an inverse transformation:

$s_{ij,k}^* (\bar{x}_l, t)$  Forward sensitivity field for a source at  $k$

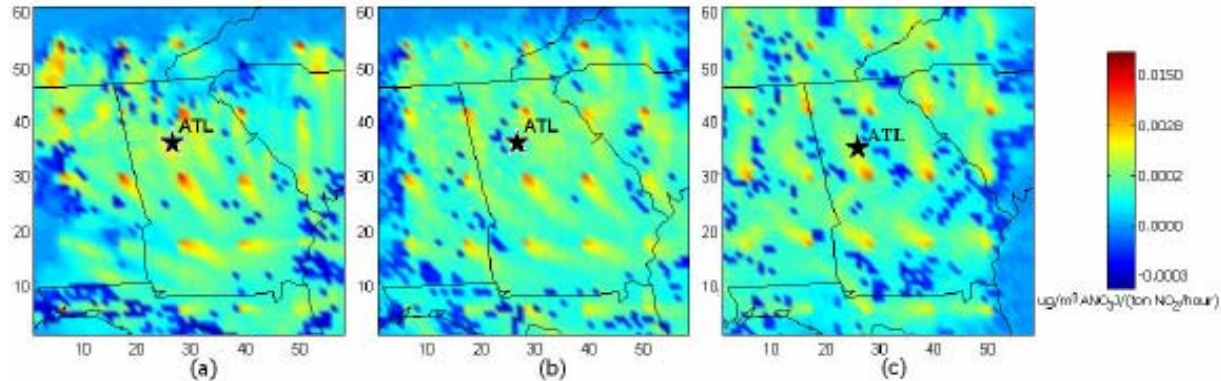
$z_{ij}^* (\bar{x}, \bar{x}_r, t)$  Reverse sensitivity for receptor located at  $\bar{x}_r$

$$z_{ij}^* (\bar{x}, \bar{x}_r, t) = \sum_{k=1}^N w_k (\bar{x}) * s_{ij,k}^* (\bar{x}_l, t)$$

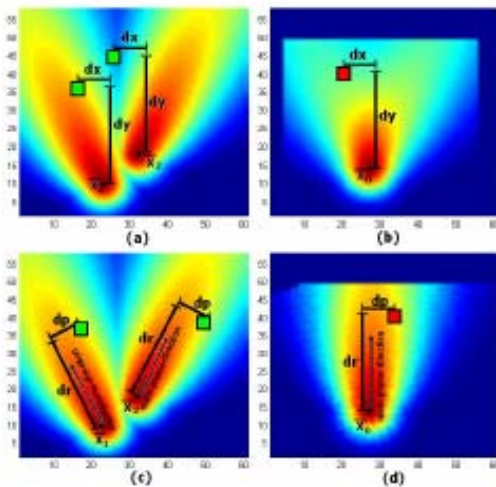
# AOI Development - Forward Fields/Back Inversion



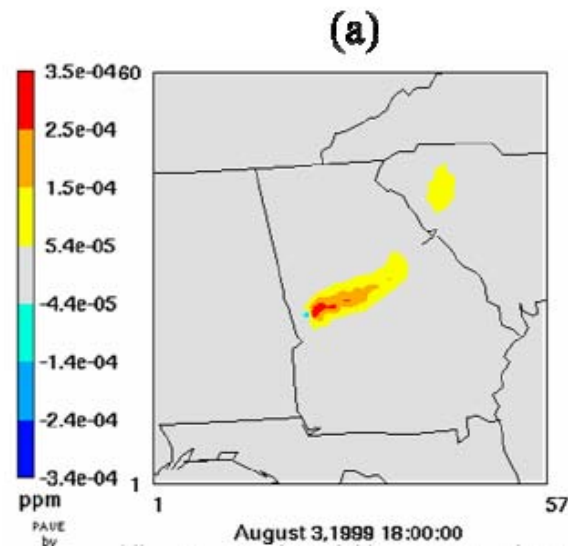
1. Choose a receptor



2. Calculate forward sensitivities of pollutants to emissions



3. Estimate backward sensitivities



4. Final AOI

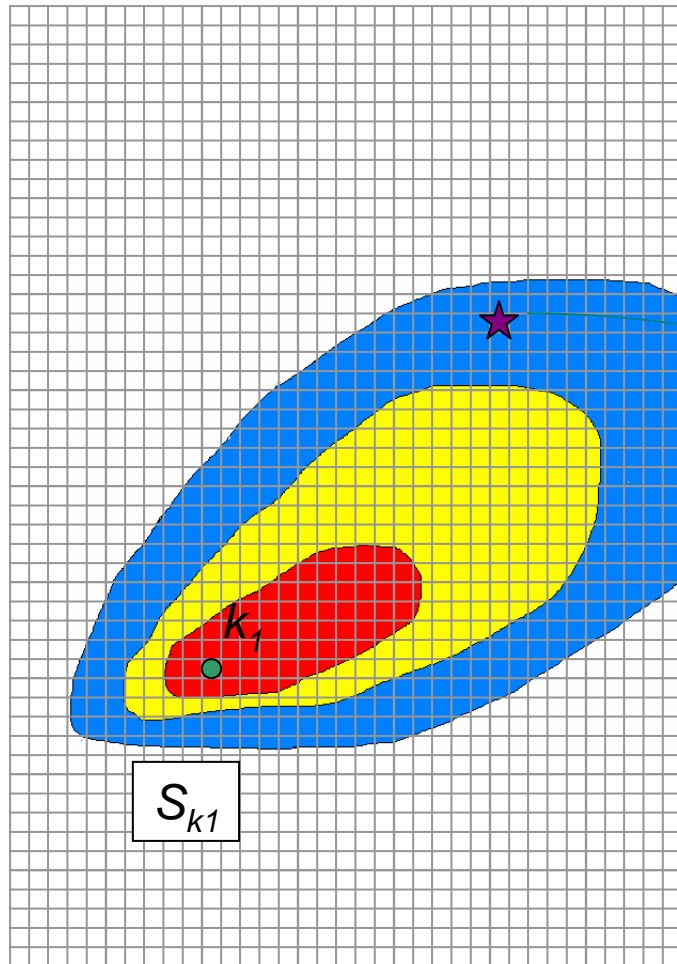


# Inversion

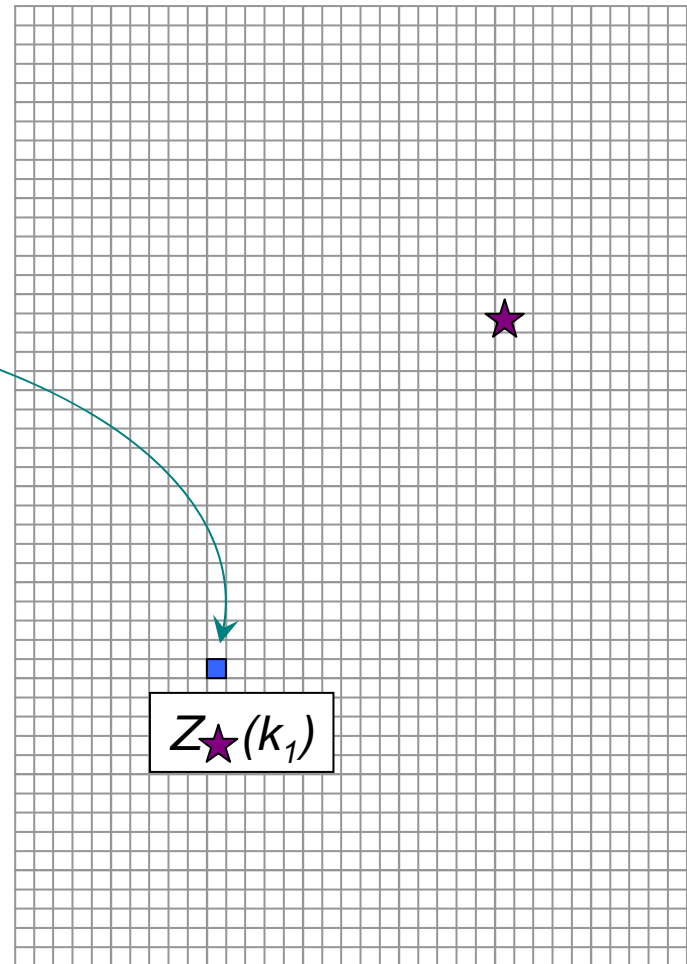
- The receptor based sensitivity field is known automatically after the interpolation

$$Z_{ij,r}(\bar{x}_k, t) = S_{ij,k}(\bar{x}_r, t)$$

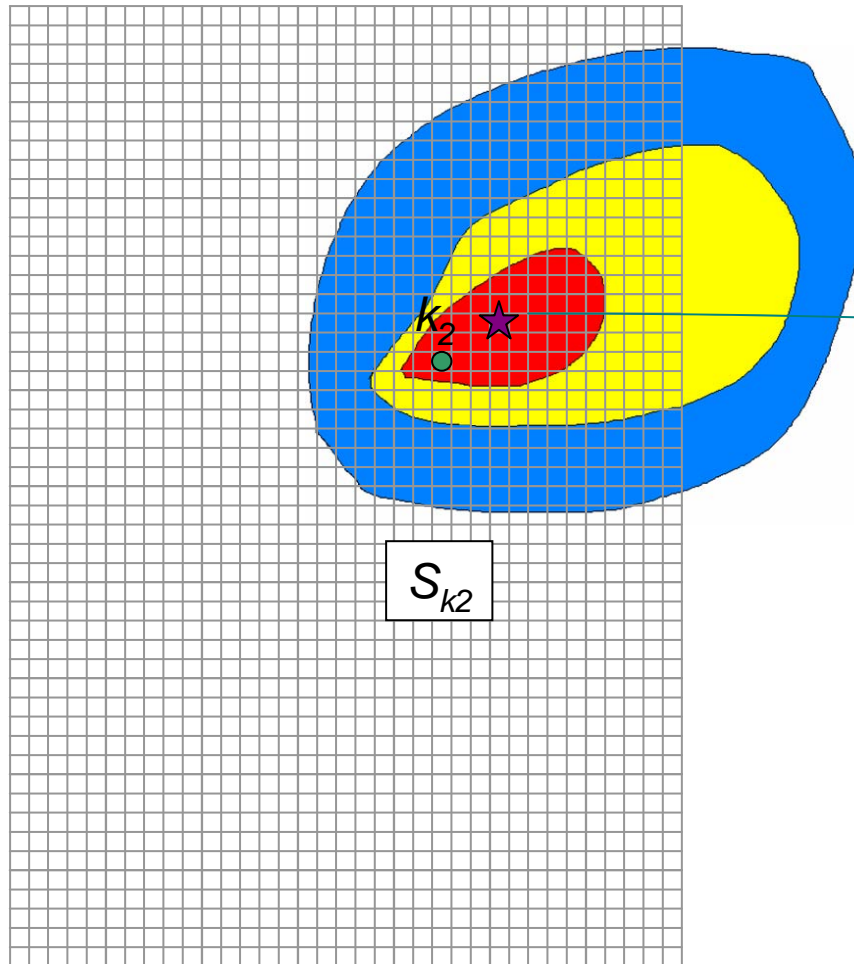
# Forward Field $k_1$



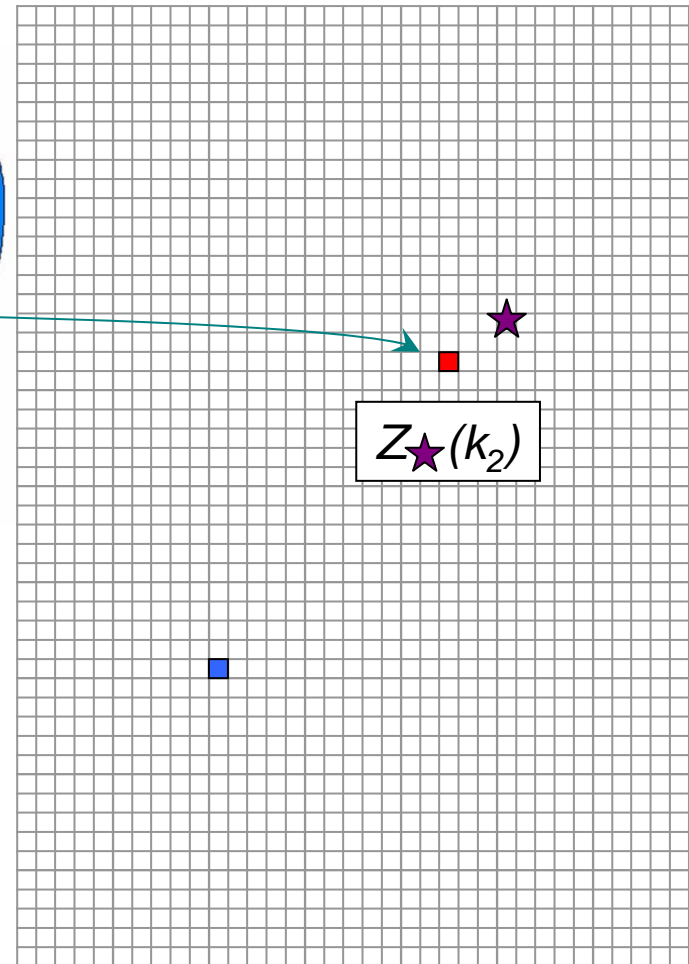
# Influence of $k_1$ on $\star$



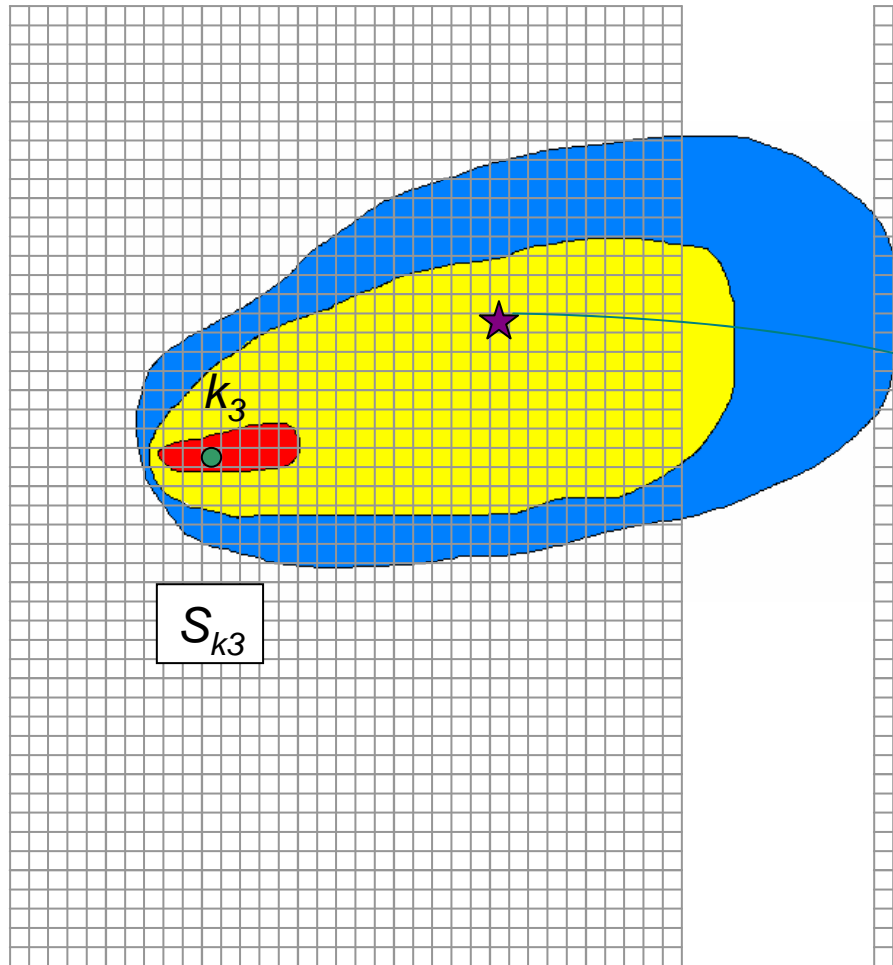
# Forward Field $k_2$



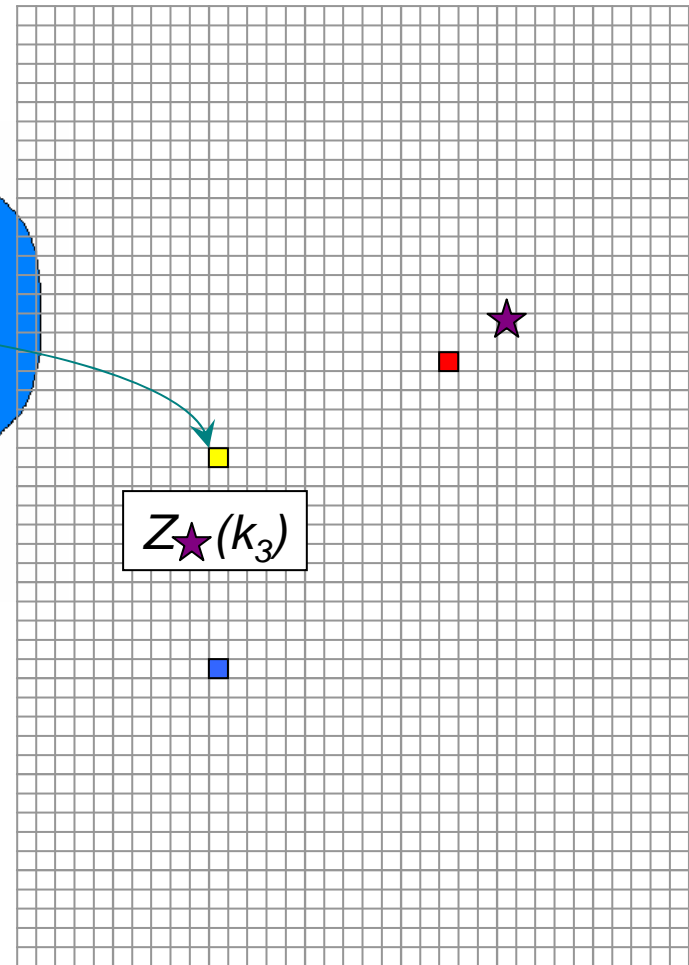
# Influence of $k_2$ on $\star$



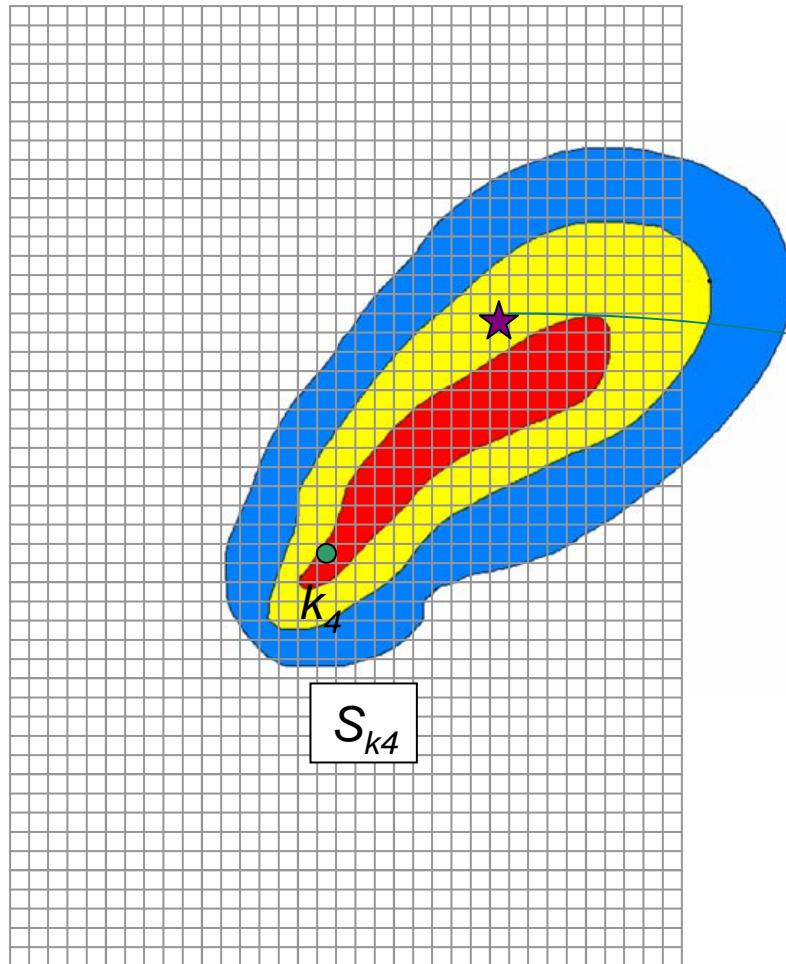
# Forward Field $k_3$



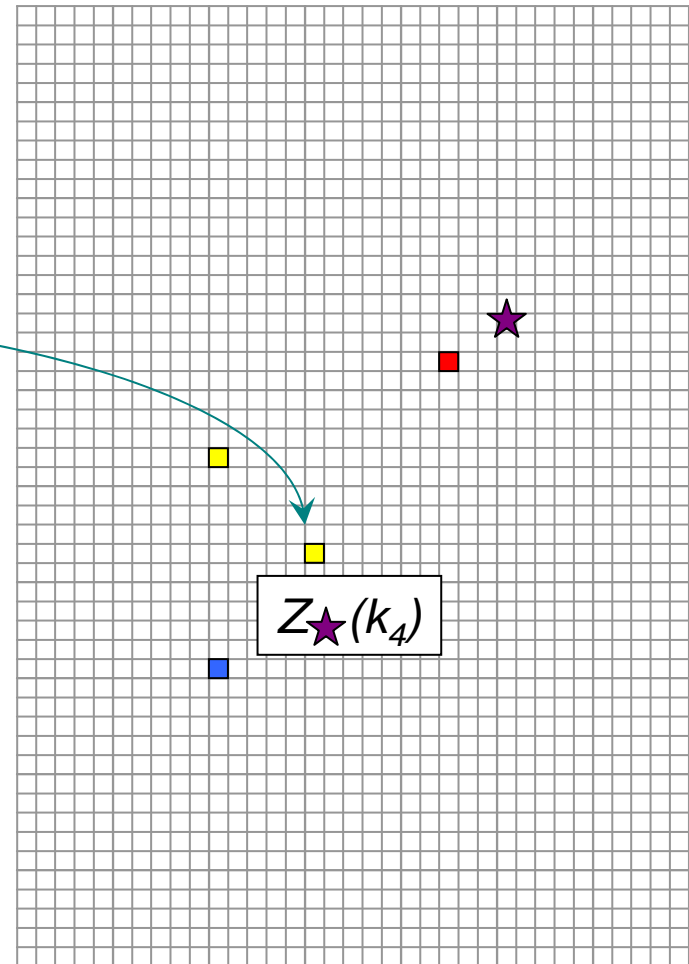
# Influence of $k_3$ on $\star$



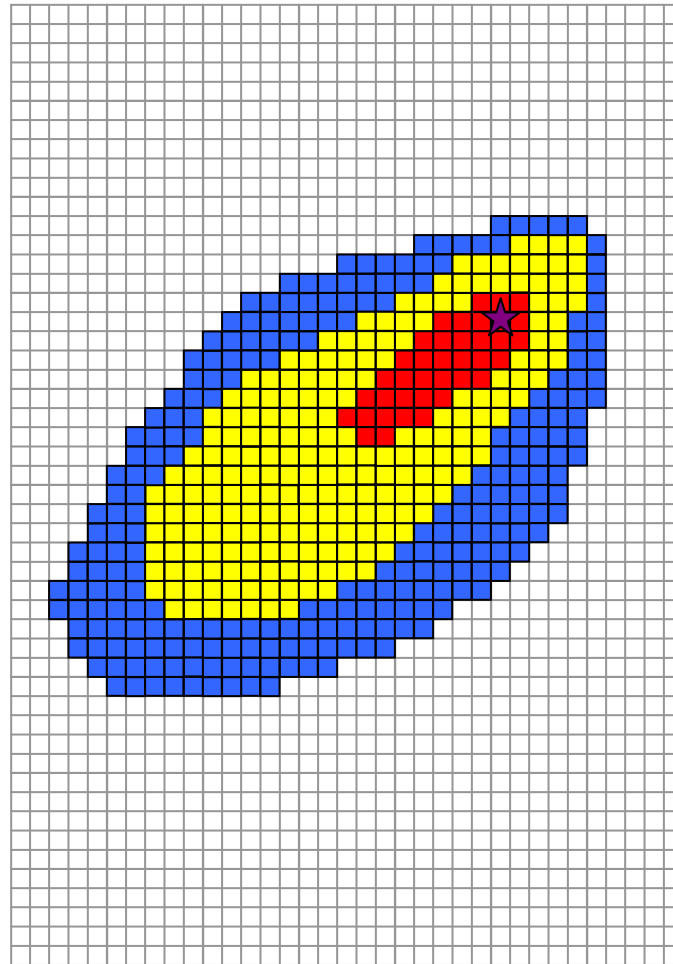
# Forward Field $k_4$



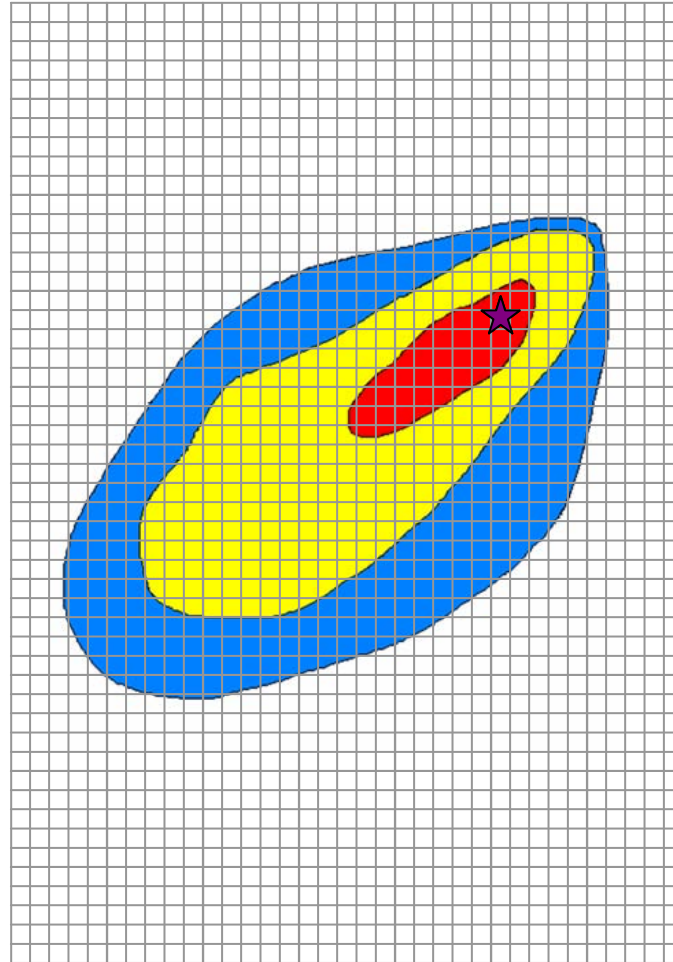
# Influence of $k_4$ on $\star$



AOI at ★

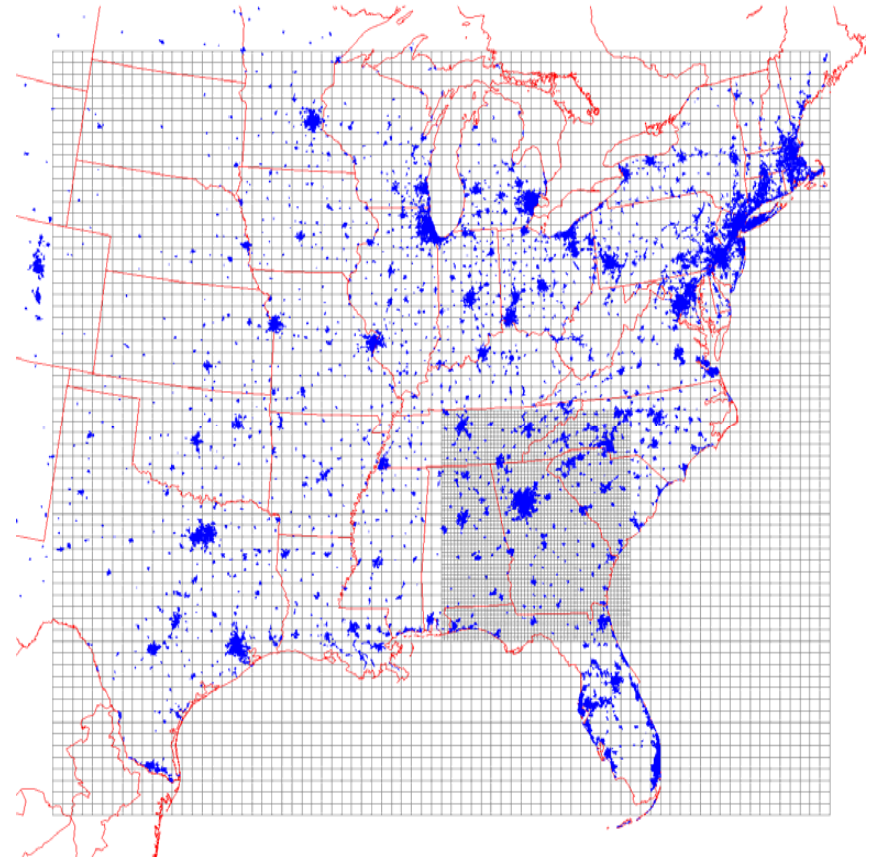


AOI at ★



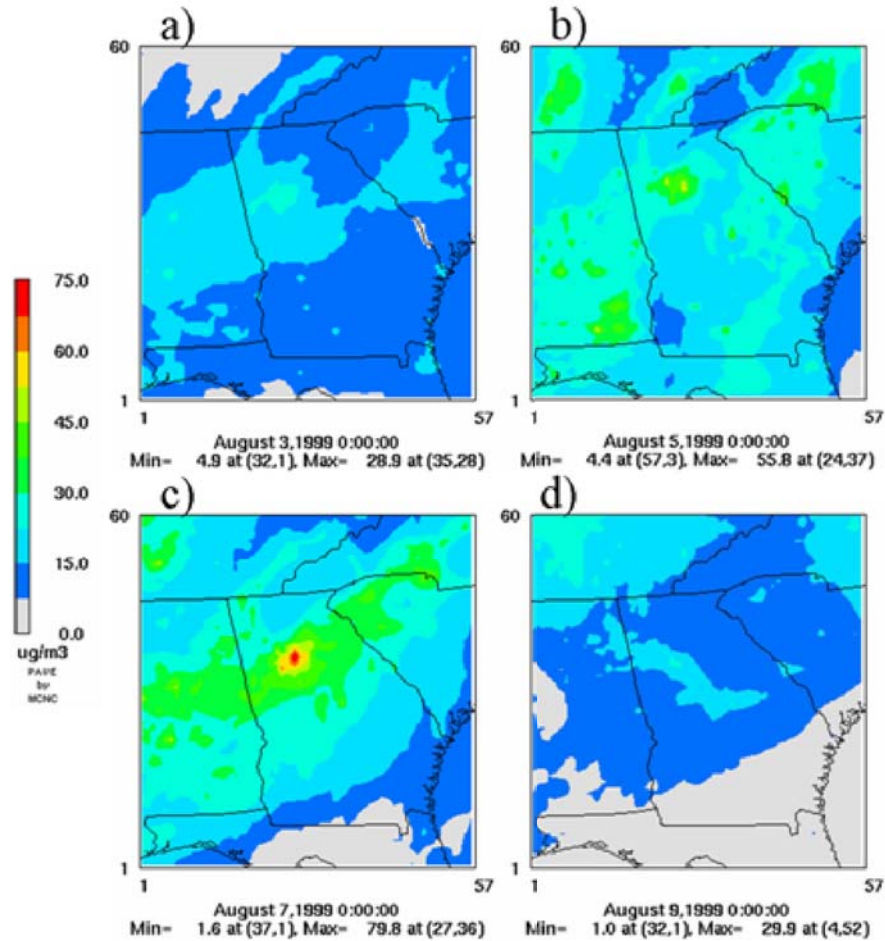
# Application – Atlanta, GA

- Episode: August 1-10, 1999
  - High PM<sub>2.5</sub> and ozone
  - Stagnant air trapped by a high pressure system directly over the southeast
  - Low wind speeds, high temperatures
- Domain: 12km resolution
- Nested in a larger 36km grid
- Meteorology: MM5
- Emission: SMOKE
- AQM: CMAQ w/DDM-3D





# Modeled PM<sub>2.5</sub>\* Levels



\*Aitken and Accumulation Modes of Sulfate, Nitrate, Ammonium, EC, OC, and “unspecified”

# Calculated Sensitivities

- Emissions

- $\text{SO}_2$
- $\text{NO}_x$
- $\text{NH}_3$
- anthropogenic VOC
- elemental carbon

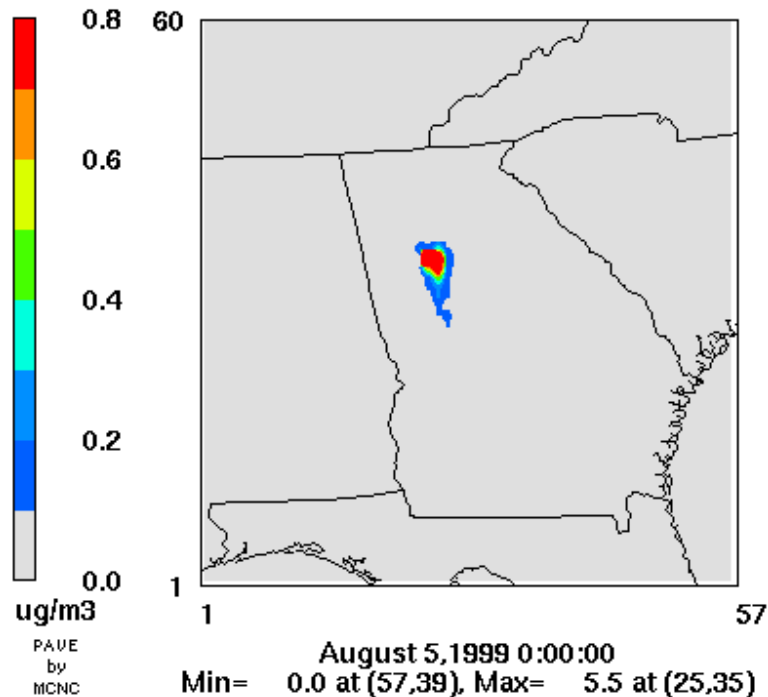
- Endpoint Pollutants

- Total  $\text{PM}_{2.5}$
- Sulfate
- Nitrate
- Ammonium
- EC
- Anthropogenic SOA
- Ozone

# AOI – EC from primary EC emissions

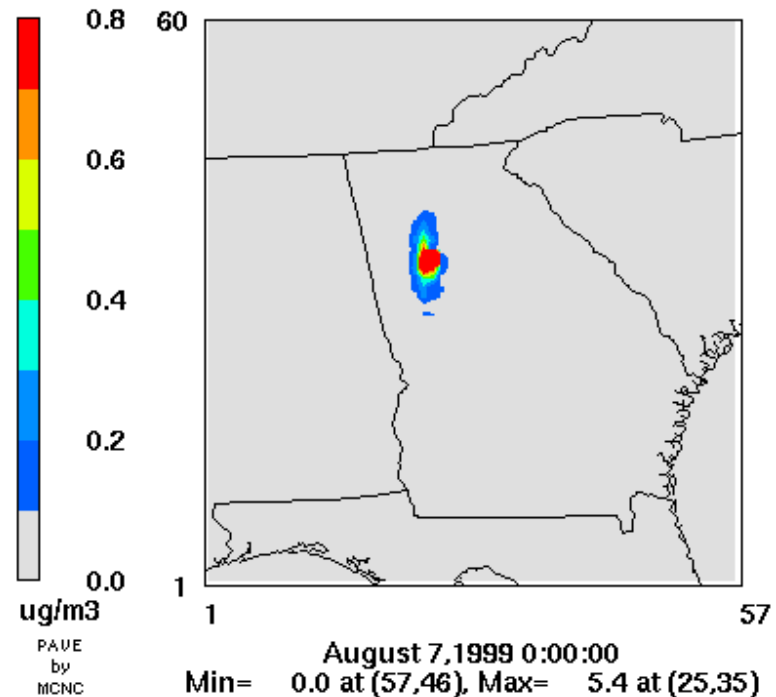
## EC Area of Influence

( $\mu\text{g}/\text{m}^3 \cdot \text{sec}$  EC emissions)  
24-hour average



## EC Area of Influence

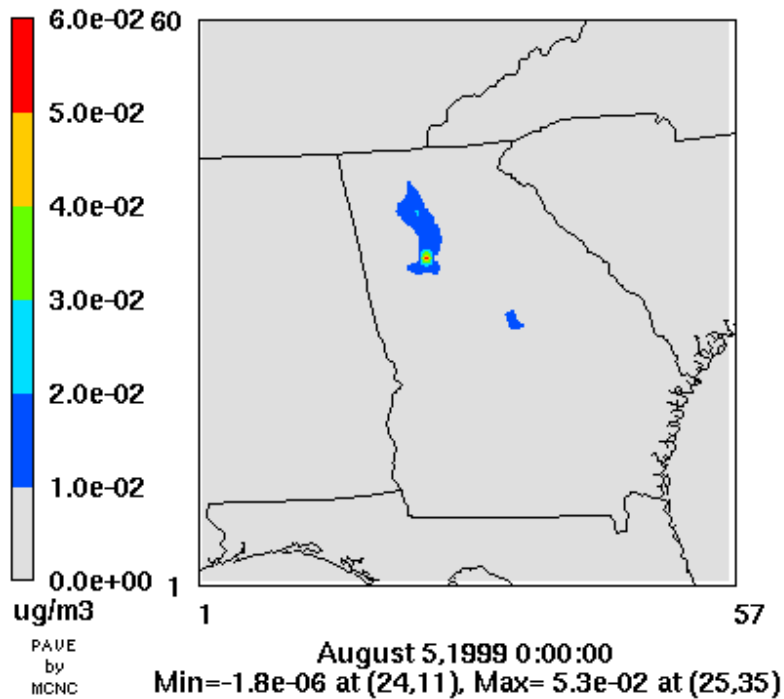
( $\mu\text{g}/\text{m}^3 \cdot \text{sec}$  EC emissions)  
24-hour average



# AOI – Sulfate from SO<sub>2</sub> Emissions

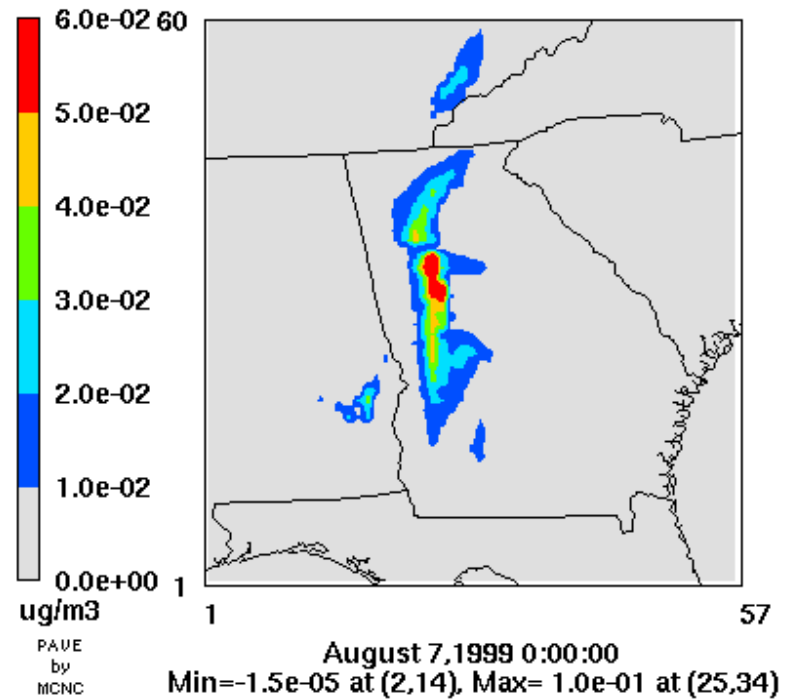
## Sulfate Area of Influence

(1 mole/sec SO<sub>2</sub> emissions)  
24-hour average

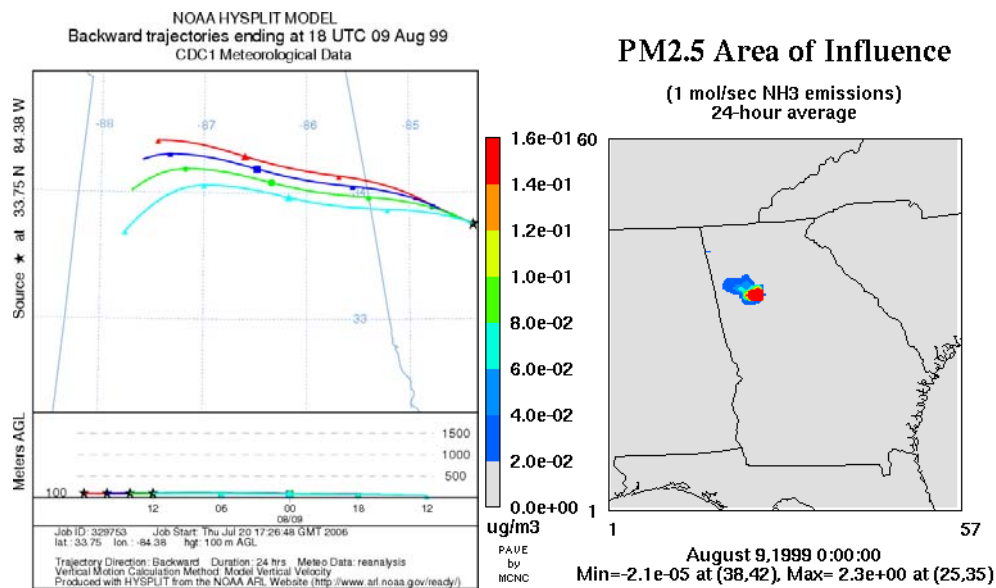
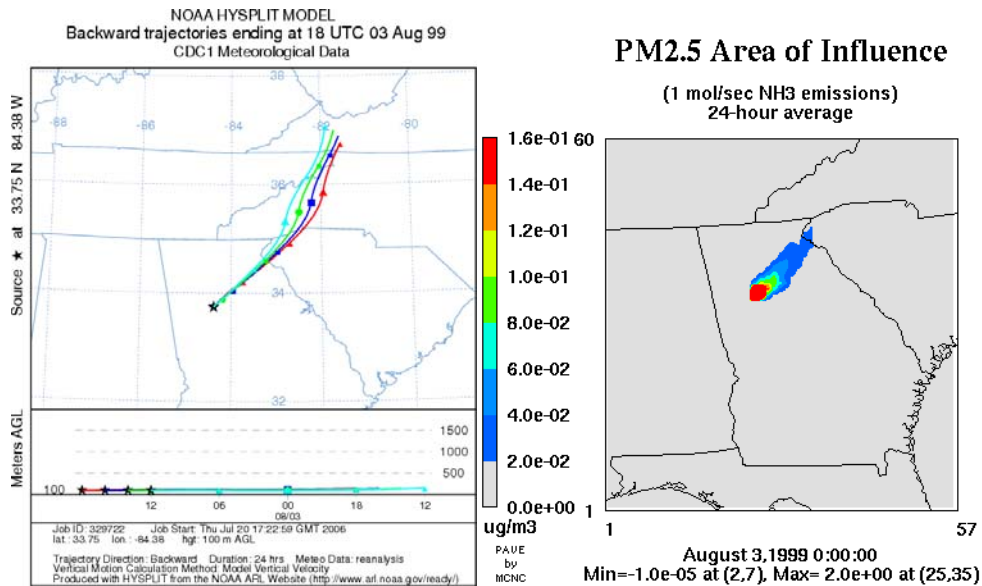


## Sulfate Area of Influence

(1 mole/sec SO<sub>2</sub> emissions)  
24-hour average



# HYSPLIT Trajectories



# Research Papers

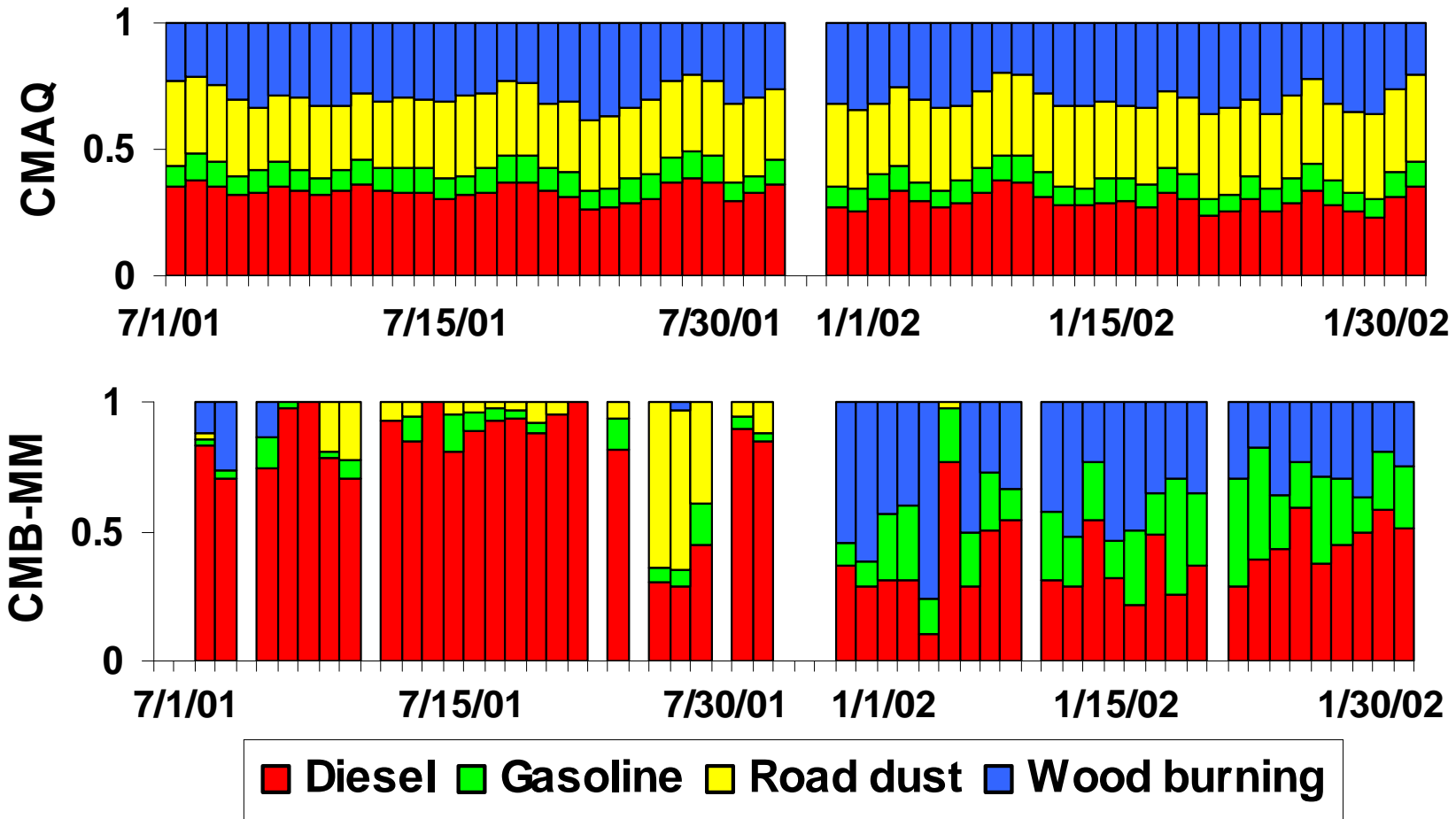
- Source apportionment of PM<sub>2.5</sub> using different models (A. Marmur, 2006; S. Lee, 2007; J. Baek, 2007)
  - CMAQ, CMB-MM, CMB-RG, CMB-LGO, PMF
- Improving emission inventories using tracer species (J. Baek, 2007)
- Regional source apportionment (S. Napelenok, 2006)
- Improving emission inventories using inverse modeling (S. Park, 2006; J. Baek, 2007; Y. Hu, 2007)
- Area of influence (F. Habermacher, 2007; S. Napelenok, 2006, S. Kwon, 2007)
- Use of SA results in Epidemiologic Studies (A. Marmur, 2006; J. Sarnet, 2006, 2007)

# Summary

- Sensitivity analysis based source apportionment fast
  - Reduces numerical noise issues
- No one source apportionment technique is a winner
  - Too many reasons to list
- Application of SA to epidemiologic studies has a number of model-dependent issues
  - Capturing diurnal, day-to-day and spatial variability/representativeness
- Area of Influence (AOI) approach is a computationally effective method to get complete fields of both reverse and forward sensitivities
  - Extensible to other models and planning (prescribed fires)
- Inverse modeling using, metals, ions and EC/OC suggests major biases in inventories
  - Need to be investigated... don't take as truth

# One issue:

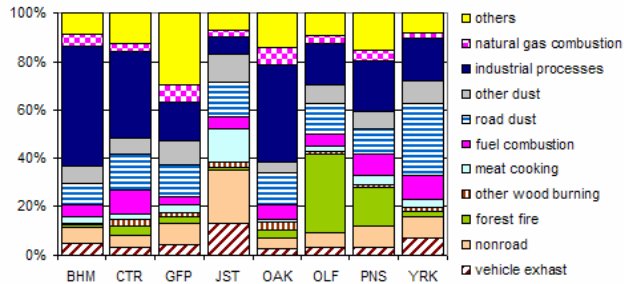
## Daily variation of fraction of major PM<sub>2.5</sub> sources at JST





# Inverse Modeling Using STN Tracer Species

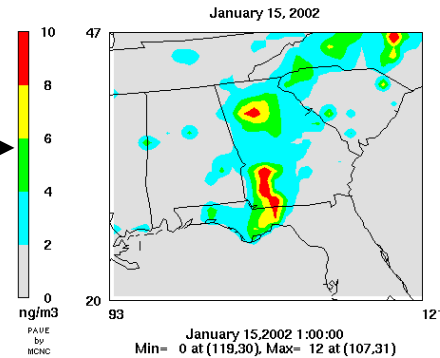
PM2.5 source contributions (Jul., 2001)



Source apportionment  
using CMAQ

Multiplication

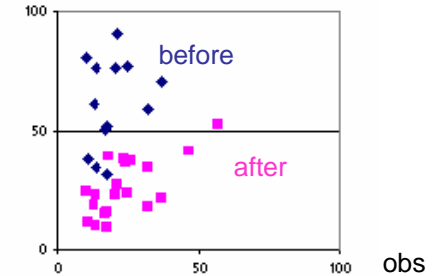
Potassium(K) concentration



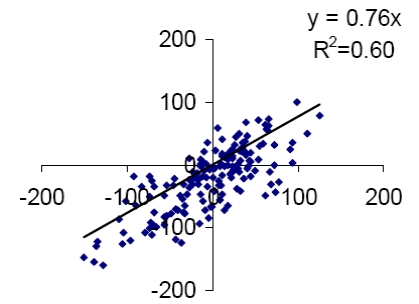
Tracer species  
concentration

Regression analysis

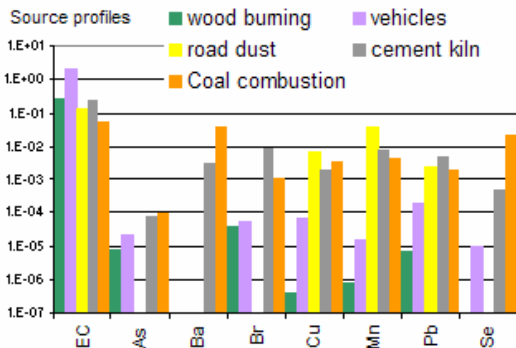
CMAQ



Scaling factors



Improved  
CMAQ simulations



Source profiles  
Used in CMB



Observations

# Quantitative Analysis: Regression analysis using tracer species

- Assumptions
  - Tracer species such as trace metals are non-reactive and conservative in the atmosphere
- Advantages
  - Require less resources
    - Combined with CMAQ Tracer & DDM methods
  - Site specific information
  - Source specific information
    - Mobile sources: EC, OC and Zn
    - Wood combustion: K, EC
    - Road/soil dust: Al, Si, Ca

# Regression analysis using tracer species – each species

- Representative tracers such as
  - EC
  - Silicon
  - Potassium
  - Zinc
  - Aluminum
- Can be used as a guideline to scaling factors of each source categories

