Associating Health Impacts with Sources of Air Pollution: An Air Quality Engineer’s Thoughts

Ted Russell, et al.
Air Resource Engineering Center
Georgia Institute of Technology

April, 2005
Traditional Epidemiologic Approach

• Identify associations between air quality metrics and health endpoints:

Health endpoints

Statistical Analysis (e.g. time series)

Association
Coming Trend: 
Associate Sources with Impacts

Data

Air Quality Model

Source Impacts $S(x,t)$

Health Endpoints

Statistical Analysis

Association between Source Impact and Health Endpoints (e.g. Laden et al., 2000)
Why and Why Not?

• May not be measuring the species primarily impacting health
  – Observations limited to subset of compounds present
• Many species are correlated
  – Inhibits correctly isolating impacts of a species/primary actors
    • Inhibits identifying the important source(s)
• Multiple pollutants may combine to impact health
  – Statistical models can have trouble identifying such phenomena
• Identify stronger associations (?)
• Better (?) way to deal with spatial variations
• Ultimately want how a source impacts health
  – **We control sources**

Why Not?

• Added work
• Introduce increased uncertainties
  • Source apportionment, even based on observed data, introduces uncertainties
  • FA techniques may have much of the mass from another source
• Subtle issues
  • Using a time series technique to both identify “sources” as well as associate the “sources” with health outcomes may introduce/hide associations
PMF “Source” Profiles & Results

CMB vs. PMF

Georgia Institute of Technology
Uncertainties:
Sensitivity of CMB-MM to source profiles

**Diesel Exhaust**

Average difference = 55.7%

**Gasoline Exhaust**

Average relative STD = 90.6%

**Legend**
- S: SEARCH
- C/N/Y/J: Location
- S/W: Summer/Winter

**Graph Details**
- OC Concentration (ug/m3)
- DIESEL1
- DIESEL2
- C*CAR1
- C*CAR2
- CATCAR1
- CATCAR2
- NONCAT1
- NONCAT2

**Locations**
- SCS
- SCW
- SNS
- SNW
- SYS
- SYW
- SJS
- SJW

**Results**
- Diesel Exhaust:
- Gasoline Exhaust:
What can (should) air quality scientists do?

• Raise flags of concern
• Make more/better/more targeted observations
• Continue to improve our understanding of air pollutant dynamics across scales and pollutants
• Interpret observations and model results in a fashion that can be readily used by health scientists
• Provide source impacts (in space and time) with meaningful uncertainties
• Develop better source apportionment techniques
• Help identify emerging and future concerns
• “Action items”
  – Target 3+ cities for extensive monitoring and analysis (incl. spatial and temporal variability/uncertainty and modeling) to detail air quality and source impacts working with health effects researchers
Acknowledgements

• Students staff and faculty in the Air Resources Engineering Center of Georgia Tech
• SEARCH, Emory teams.
• GA DNR
• Georgia Power
• US EPA
• NIEHS
• Georgia Tech
CMB vs. PMF

Summer

Wood Combustion

Diesel Exhaust

Gasoline Exhaust

Road Dust

Secondary Nitrate

Secondary Sulfate

Winter

Wood Combustion

Diesel Exhaust

Gasoline Exhaust

Road Dust

Secondary Nitrate

Secondary Sulfate

Georgia Institute of Technology
Why Not?

• Added work
• Introduce increased uncertainties
  – Source apportionment, even based on observed data, introduces uncertainties
  – FA techniques may have much of the mass from another source
• Subtle issues
  – Using a time series technique to both identify “sources” as well as associate the “sources” with health outcomes may introduce/hide associations
Emissions, Air quality and Health

• (How) Can we use “air quality models” to help identify associations between PM sources and health impacts?
  – Species vs. sources
    • E.g., Laden et al., 2000
  – Need to know source strengths
    • Emissions inventory verification
Emissions Estimation

• Two Main Approaches
  – Traditional: Bottom up
    • Identify sources and estimate emissions
      – Emissions factors X Process rates
    • Typically very uncertain, but emissions linked to specific sources and locations
  – Alternative: Top down
    • Measure pollutants in the atmosphere
      – Calculate the level of emissions that would lead to observations
    • Weak link with actual sources and source locations
      – e.g., does not distinguish between NOx from automobiles or power plants or biomass burning
      – Information limited to specific receptors (lacks spatial link, requires many measurement locations for regional coverage)
      – Does not directly provide source strengths
  – Both have weaknesses and strengths
    • Need to bridge the two: Inverse modeling
Inverse Modeling and Sensitivity Analysis

- Inverse modeling involves using observations along with a physical model (e.g., traditional air quality) model to estimate model parameters and inputs, e.g., emissions.
Sensitivity analysis

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

If \( P_j \) are emission, \( S_{ij} \) are the sensitivities/responses to emission changes:

This is done automatically using DDM-3D
Emissions Inventory Assessment using Inverse Modeling/Four Dimensional Data Assimilation (FDDA)

**INPUTS**

- **Emissions inventory** (Mobile, area, biogenic, point sources)
- Other inputs that remain as defined in the base case scenario
- New emissions distribution by source that minimize the difference between observations and simulations

**Main assumption in the formulation:**
*A driving source for the discrepancy between predictions and observations is the emission estimates*

**Air Quality Model + DDM-3D**

- Pollutant distribution (spatial & temporal) (e.g. Ozone, NO\textsubscript{x}, NO\textsubscript{y}, SO\textsubscript{2}, CO, VOCs); and sensitivity fields

**Ridge regression Module**

- Observations taken from routine measurement networks or special field studies

---

Georgia Institute of Technology
Monitoring sites used for the 14-18 August 1999 episode

PM

Ozone

SO₂

NOₓ

CO

Georgia Institute of Technology
# Estimated emission adjustments for domain-wide emissions from FDDA

<table>
<thead>
<tr>
<th>Source</th>
<th>August 1999</th>
<th>SAMI July 1995</th>
<th>SAMI May 1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total CO</td>
<td>1.01</td>
<td>1.08</td>
<td>1.26</td>
</tr>
<tr>
<td>Total SO₂</td>
<td>0.92</td>
<td>1.13</td>
<td>1.08</td>
</tr>
<tr>
<td>Area source NOₓ *</td>
<td>1.62</td>
<td>1.77</td>
<td>1.50</td>
</tr>
<tr>
<td>Elevated point source NOₓ</td>
<td>1.48</td>
<td>1.31</td>
<td>1.24</td>
</tr>
<tr>
<td>Anthropogenic VOC *</td>
<td>2.47</td>
<td>2.21</td>
<td>2.84</td>
</tr>
<tr>
<td>Biogenic VOC</td>
<td>1.11</td>
<td>1.24</td>
<td>1.17</td>
</tr>
<tr>
<td>Total NH₃</td>
<td>0.56</td>
<td>0.52</td>
<td>0.59</td>
</tr>
<tr>
<td>Total fine OC PM</td>
<td>1.10 (0.60)</td>
<td>0.49</td>
<td>0.62</td>
</tr>
<tr>
<td>Total fine EC PM</td>
<td>0.56</td>
<td>N/C</td>
<td>N/C</td>
</tr>
</tbody>
</table>

* Includes mobile and area sources

Using only IMPROVE measurements

---

*Georgia Institute of Technology*
Ozone: Simulated vs. Observed

Peak Ozone

![Graph showing simulated vs. observed peak ozone levels.](image)
PM$_{2.5}$ at Atlanta sites

**Observed**

Average PM$_{2.5}$ concentration
28.4 µg/m$^3$

**Simulated**

Average PM$_{2.5}$ concentration
31.6 µg/m$^3$

Contribution normalized without considering unidentified measured mass or non-simulated species

*Georgia Institute of Technology*
**Advances**

- Inverse modeling identified biases found in emissions
  - Ammonia: 44% high bias (EPA recently found 40% bias)
    - Seasonal biases found
  - Black carbon: Forest fires too high
  - VOCs
  - NOx: diesels too low
- Follow on work (bottom-up assessments and use of satellite observations) have verified some of the biases, others under investigation
  - Laborious process
- Inventories corrected! (NH$_3$, Fires, Diesels)
  - Improved model performance, field study results make more sense
- Improved source impact assessment (source apportionments)
- Lots of good data better than some great data
PM (Source Apportionment) Models
(those capable of providing some type of information as to how specific sources impact air quality)

PM Models

Emissions-Based

Lag. Eulerian (grid)

Source Specific* "Mixed PM"

Hybrid

Receptor

CMB Norm.

SVD UNMIX PMF

Georgia Institute of Technology
Source-based Models

\[ \frac{\partial c_i}{\partial t} = -\nabla (u c_i) + \nabla (K \nabla c_i) + R_i + S_i \]
Source-based Models

• Strengths
  – Direct link between sources and air quality
  – Provides spatial, temporal and chemical coverage

• Weaknesses
  – Result accuracy limited by input data accuracy (meteorology, emissions…)
  – Resource intensive
SO$_4$ & its Change on July 15, 1995 for a 10% Reduction of 2010-OTW SO$_2$ Emissions from Different Regions

- **2010-OTW**: South East, Midwest, Northeast, Central, South East, FL + MS

<table>
<thead>
<tr>
<th>Region</th>
<th>Change in µg/m$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL + MS</td>
<td>-0.9</td>
</tr>
<tr>
<td>Central</td>
<td>-0.5</td>
</tr>
<tr>
<td>Midwest</td>
<td>-0.7</td>
</tr>
<tr>
<td>Northeast</td>
<td>-1.1</td>
</tr>
<tr>
<td>South East</td>
<td>-0.1</td>
</tr>
<tr>
<td>2010-OTW</td>
<td>+0.3</td>
</tr>
</tbody>
</table>
Receptor Models

Observed Air Quality $C_i(t)$

$C_i = \sum_{j=1}^{n} f_{i,j} S_j$

Source Impacts $S_j(t)$

$C_i$ - ambient concentration of specie $i$ ($\mu g/m^3$)
$f_{i,j}$ - fraction of specie $i$ in emissions from source $j$
$S_j$ - contribution (source-strength) of source $j$ ($\mu g/m^3$)
Receptor Models

• Strengths
  – Results tied to observed air quality
  – Less resource intensive (provided data is available)

• Weaknesses
  – Data dependent (accuracy, availability, quantity, etc.)
    • Monitor
    • Source characteristics
  – Not apparent how to calculate uncertainties
  – Do not add “coverage” directly
Traditional Source Apportionment

Results based on atmospheric measurements:
Not linked to source location or accounting for source strength
(e.g., a close source, that is weak, can have a huge impact at one receptor)
Source Apportionment: What Works?

• How well do the source apportionment approaches work?
  – Do they give consistent results?
• What are the various weaknesses and strengths?
• Approach
  – Compare method results using detailed data sets
    • SEARCH, ASACA, STN, EPA Supersites for July 2001 and January 2002
  – Identify problems
    • In progress
  – Assess consistency of results from
    • Receptor models: PMF, CMB (regular), CMB-Molecular Marker (MM), CMB-Optimized (LGO)
    • Source model: CMAQ
Source apportionment of Organic mass
CMB-MM (left), CMAQ (right)

July 2001

<table>
<thead>
<tr>
<th></th>
<th>BHM</th>
<th>CTR</th>
<th>GFP</th>
<th>JST</th>
<th>OAK</th>
<th>OLF</th>
<th>PNS</th>
<th>YRK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

January 2002

<table>
<thead>
<tr>
<th></th>
<th>BHM</th>
<th>CTR</th>
<th>GFP</th>
<th>JST</th>
<th>OAK</th>
<th>OLF</th>
<th>PNS</th>
<th>YRK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Other organic matter]
[Sec bio. OM.]
[Sec anthrop. OM]
[Other pri. OM]
[Meat cooking]
[Natural gas]
[Wood burning]
[Road dust]
[Gasoline]
[Diesel]
Fraction of primary PM - JST
What Works? Open Question!

- All methods still have issues
- CMB-MM
  - Sensitive to source profiles and ambient measurements
  - Some tracers may not be conserved (Robinson, 2005)
  - Relatively few source profiles
- CMB-Regular
  - Sensitive to profiles and elemental observations
    - Near XRF limits on many observations
- PMF
  - Sensitive to data, interpretation and observations
  - Not apparent how to assess uncertainties
- Source Model (CMAQ)
  - Sensitive to inventories
  - Does not have as large of daily variations as observed at point measurement sites
Bottom Line…

• Each of the approaches has issues, and there are reasons to question each of the results, so…
  – All models are wrong, but some might be useful

• Use of air quality models in health effects analysis raises issues beyond those atmospheric scientists have tended to address
  – Representativeness
  – Quantitative uncertainty analysis of reconstructed fields,
  – Etc.
  – Concern: A lot of good measurements is usually better than a few great measurements
    • Criticism of “short term” Supersites
Summary

- Effective air quality management requires knowing emissions accurately by source
  - Health effects and emissions trading programs are even more dependent upon accurate emissions estimates
- Inverse modeling has been very effective in our efforts to improve inventories
  - More powerful as satellite data becomes available
- Application of PM Source apportionment models in health studies more demanding than traditional “attainment-type” modeling
  - New (and relatively unexplored) set of issues
- Receptor models do not, yet, give same results
  - Nor do they agree with emissions-based model results (that’s o.k. for now)
  - Need a way to better quantify uncertainty
  - Probably lead to excess variability for application in health studies
  - Representativeness error
  - Not yet clear if model application, itself, increases representativeness error over directly using observations
- Emissions-based models
  - Likely underestimate variability (too tied to minimally varying inventory)
  - Performance is spotty
- Groups actively trying to reconcile differences
  - Hybrid approaches proving useful (extension of inverse modeling)
- Lots of good data often better than limited great data
  - Health, source apportionment and emissions analysis
Conclusions

• Effective air quality management requires knowing emissions accurately by source
  – Emissions trading programs are even more dependent upon accurate emissions estimates

• Traditional approach (bottom up) leads to uncertain emissions
  – Resource intensive to improve estimates (man-years) and can totally miss major sources (e.g., Houston area)

• Top-down approach does not provide source-based emissions estimates
  – Limited spatial information

• Inverse modeling can bridge the gap between methods
  – Maintains the strengths of both methods, fewer weaknesses

• After initial application, continued use will lead to further improved inventories
  – Relatively little effort to use after first applications
Di: do not translate any slides beyond here!
Sensitivity Analysis

• Calculate sensitivity of gas and aerosol phase concentrations and wet deposition fluxes to input and system parameters
  \[ s_{ij}(t) = \frac{\partial c_i(t)}{\partial p_j} \]

• Brute-Force method
  – Must run the model a number of different times
  – Inaccurate sensitivities may result due to numerical noise propagating in the model

• DDM - Decoupled Direct Method
  – Use direct derivatives of governing equations
  – Initial and boundary conditions, horizontal transport, vertical advection and diffusion, emissions, chemical transformation, aerosol formation, and scavenging processes
DDM-3D

3-D Air Quality Model

$NO^o$
$NO_2^o$
$VOC_i^o$
...
$T$
$K$
$u, v, w$
$E_i$
$k_i$
$BC_i$
...

$J$
$\frac{\partial R_i}{\partial k_j}$

$O_3(t,x,y,z)$
$NO(t,x,y,z)$
$NO_2(t,x,y,z)$
$VOC_i(t,x,y,z)$
...

$DDM-3D$ Sensitivity Analysis

$decoupled$

$s_{ij}(t) = \frac{\partial c_i(t)}{\partial p_j}$

Georgia Institute of Technology
Applications of Direct Sensitivity Analysis

- Regional response to controls
- Source receptor relationship
  - State-by-state emissions
  - Aggregated to find contributions to air quality in National Parks and cities
- Area of influence (AOI)
- Emissions reconciliation
  - Bottom-up/top-down inventory assessment
Sulfate Sensitivity to Domain-wide Reductions

Sulfate

*on-the-way* emissions
July 15, 2010  12:00

Sulfate Sensitivity
10% SO2 Reduction (domainwide)
July 15, 2010 12:00
Sensitivity Analysis: Regions

X: GSM
SO$_4$ & its Change on July 15, 1995 for a 10% Reduction of 2010-OTW SO$_2$ Emissions from SAMI States
Use AQ Models to Address Issues: Assess Errors, Provide Increased Coverage

Data → Air Quality Model → Air Quality C(x,t)

Health Endpoints

Association between Concentrations and Health Endpoints

Site Representative?

Monitored Air Quality C_i(x,t)
Effective Air Quality Management and Health Assessment: Necessary Components

• Emissions
  – Need to know who is emitting how much to know what controls are most effective
  – If using emissions trading (which is usually very cost effective), need to be able to set regional emissions caps
    • Requires knowing baseline levels
    • Emissions trading without knowing the emissions accurately is like trying to make a budget without knowing how much money you have or how much things cost: it does not work!

• Link between emissions, air quality and effects
  – Relationships are often non-linear and pollutant responses are linked (e.g., ozone and PM)
  – Requires comprehensive air quality model
  – Knowledge of associated health effects

• Ability to affect policy
Air Quality Modeling

Air Quality Goals

Controls

Emissions

Chemistry

Pollutant Distributions

Air Quality/Health Impacts

Meteorology

Air Quality Model

Georgia Institute of Technology
Association between CVD Visits and Air Quality

Georgia Institute of Technology

Metzger et al., 2004
Traditional

- Identify associations between air quality metrics and health endpoints:

Health endpoints

Statistical Analysis (e.g. time series)

Association