Multi-Pollutant Response Surface Modeling Using CMAQ: Development of an Innovative Policy Support Tool

Carey Jang, Sharon Phillips, Bryan Hubbell, Pat Dolwick, Norm Possiel, Tyler Fox

Air Quality Modeling Group

USEPA/OAQPS/EMAD

Outline

- Why do we need the response surface model?
- What is a response surface model (RSM)?
- Development of multi-pollutant RSM using CMAQ
- Steps in designing CMAQ RSM applications
  - Experimental design
  - CMAQ – SMOKE interface development
  - Evaluation & Validation
- Upcoming activities

Need for Response Surface Model

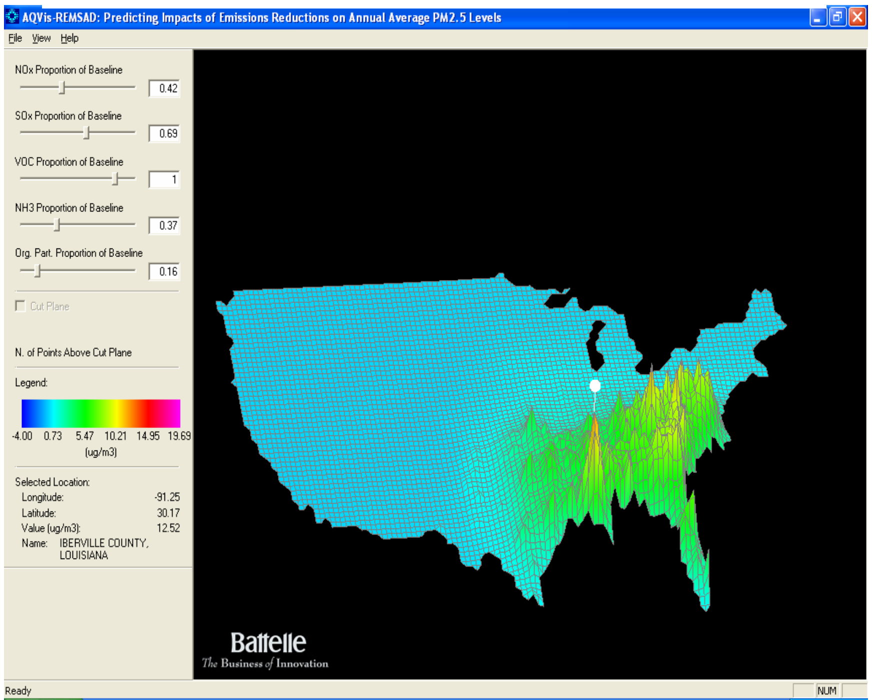
- Growing influential role of AQ models in guiding and supporting policy analysis and implementation for complex AQ issues such as PM, O_3, and air toxics
- Enormous computational costs (time & resource) of photochemical AQ modeling always present a challenge for time pressing need of policy analysis
- Current model operation in comparing the efficacy of various control strategies and policy scenarios is typically inefficient, if not ineffective
- An innovative policy support tool to address these issues in an economical manner is needed

What is a Response Surface Model?

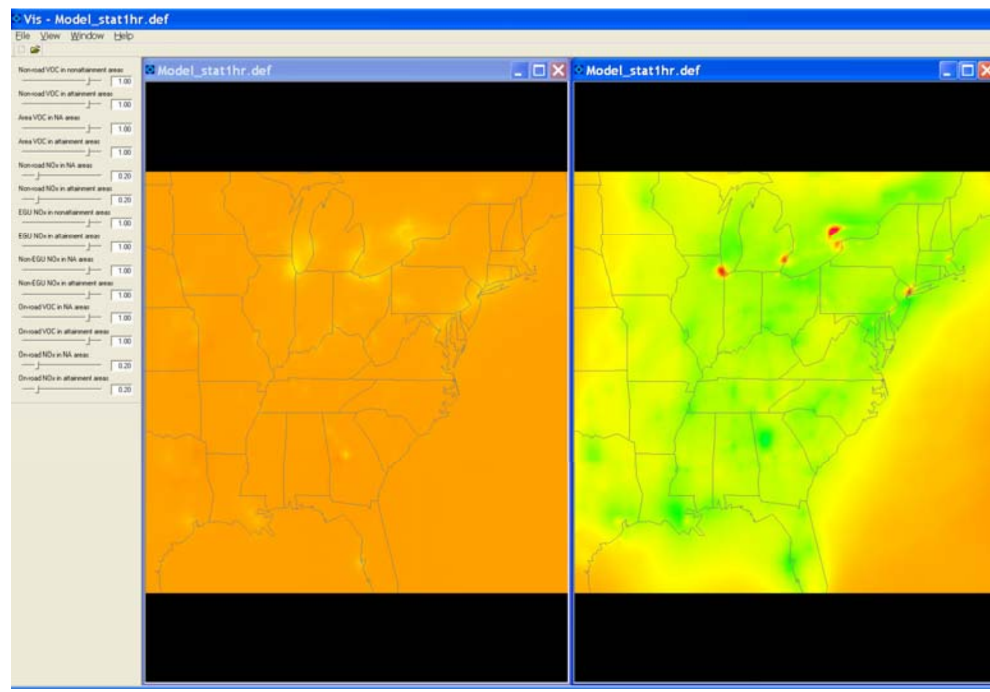
- Response Surface Model (RSM) is a “reduced form model” of a “model” (e.g. CMAQ)
- Based on a systematically selected set of model runs, statistical techniques can be used to represent the relationship between model inputs and outputs (e.g. emissions control and concentrations of PM & ozone)
- Once the “response surface” has been generated, it can be used to simulate the functions of computationally expensive photochemical air quality model
- Cross-validation can then be conducted to examine the validity of RSM to represent model responses

RSM Pilot Applications

RSM for $PM_{2.5}$ using REMSAD



RSM for O_3 using CAMx



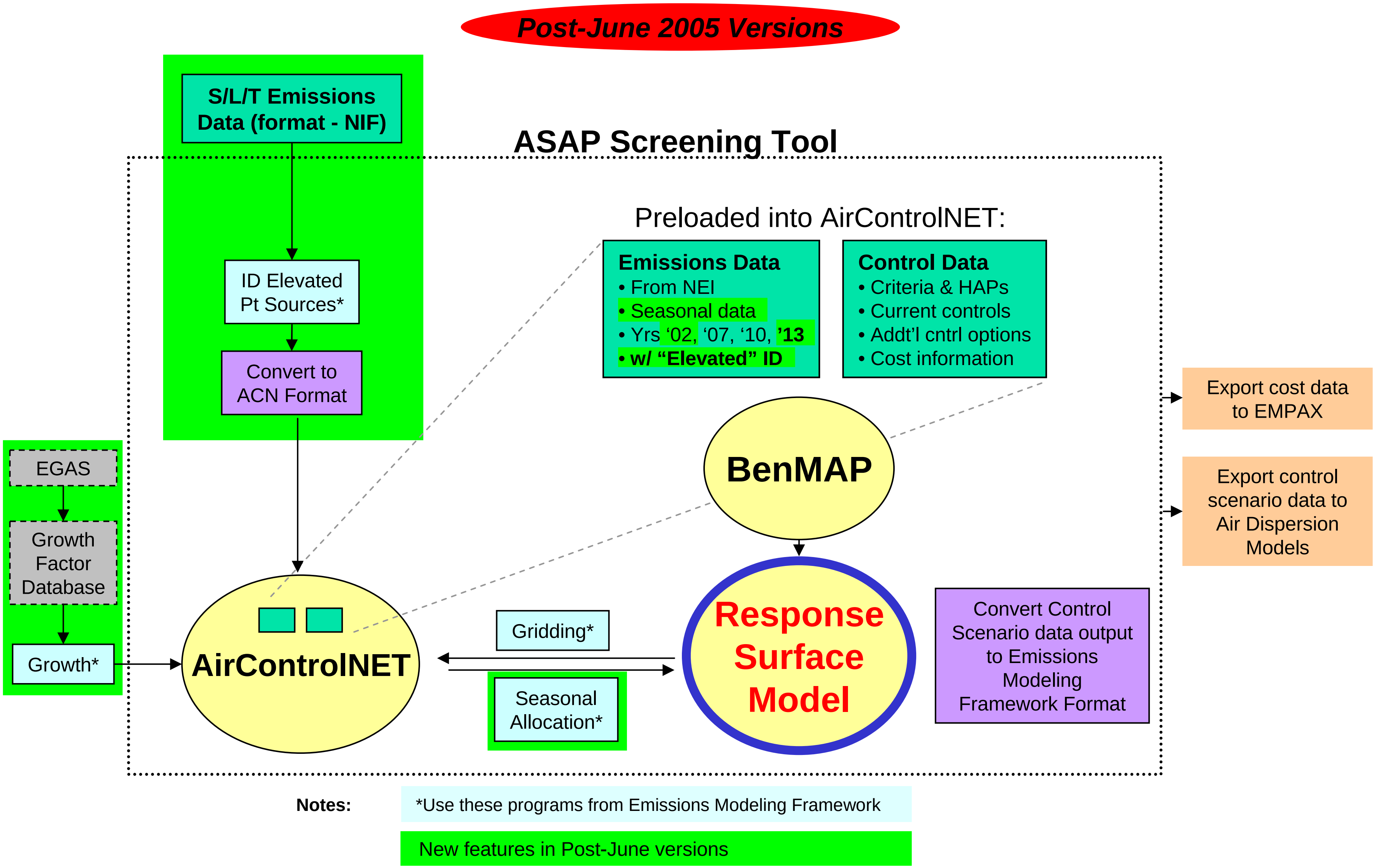
VOC controls vs. NOx controls

How the RSM can be used

- Strategy design and assessment screening tool
  - "What if " scenario analyses: provide real-time predictions of model responses to model inputs
  - Comparison of urban vs. regional controls
  - Comparison across source sectors
  - Comparison across pollutants
- Model sensitivity & uncertainty annalysis
  - Can be used to systematically evaluate the relative sensitivity of modeled ozone and PM levels to changes in emissions/met inputs
- Optimization
  - Can be used to develop optimal combinations of controls to attain standards at minimum cost

ASAP (Air Strategy Assessment Program)

- conceptual design for data flow -



Development of CMAQ RSM Applications

- Experimental Design
  - Selection of policy factors
    - Emission control factors
    - Regional vs. urban control
  - Selection of air quality model simulations
    - Continental U.S. modeling, 2010 CAIR Base, 36-km grid resolution
    - 240 runs (in 3 stages) for 4 months (Feb., April, July, Dec.)
- CMAQ – SMOKE Interface Development for RSM
  - Develop a module within CMAQ to read directly the pre-merged SMOKE sector files (e.g., 3-D point, 2-D mobile, etc.)
  - Allow RSM to directly control % changes of (1) emissions (2) specified areas
- Validation and Evaluation
  - Cross validation
  - Out-of-sample validation

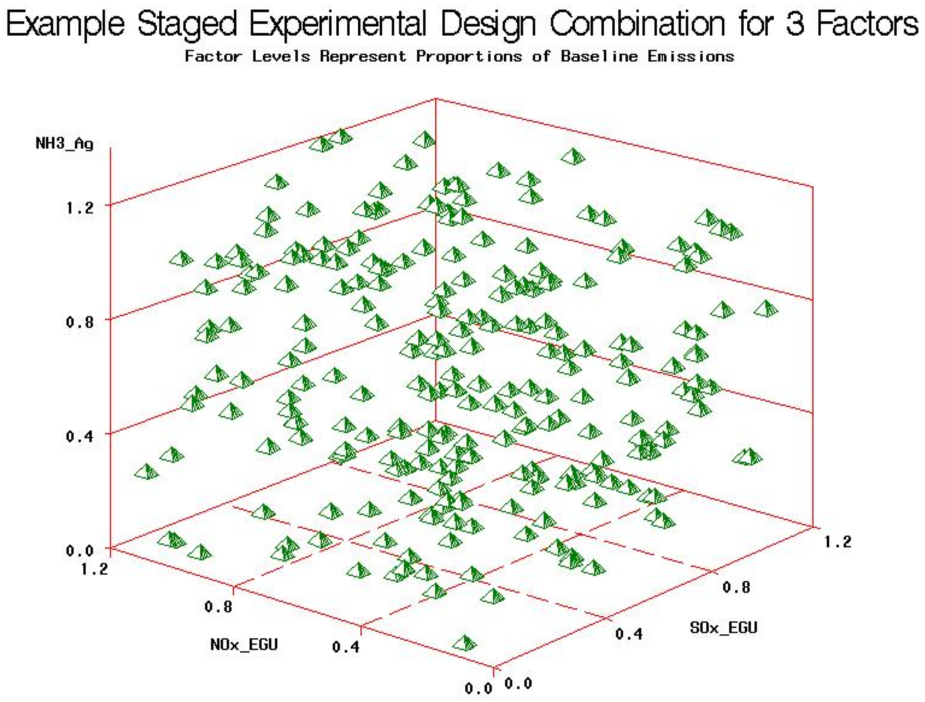
Experimental Design:
1. Selection of policy factors (1)

- 12 factors selected based on precursor emissions & source category relevant to policy analysis of interest

Run #	1) / NOx / EGU	2) NOx / NonEGU+ Area	3) NOx / Mobile	4) SOx / EGU	5) SOx / NonEGU_ Point	6) SOx / Area	7) VOC / All	8) NH3 / Area	9) NH3 / Mobile	10) POC& PEC / EGU+NonEGU	11) POC & PEC / Mobile	12) POC & PEC / Area
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	1.00000	1.000000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.000000	1.00000	1.00000
2	0.89177	1.020286	0.83054	0.96302	0.875918	0.90277	0.17062	0.17435	0.99388	0.394387	0.90896	0.87647
3	0.57572	0.266010	1.02190	0.55610	0.247603	0.80328	0.61750	1.00113	0.02458	0.501642	1.12849	0.52022
4	0.06134	0.462060	0.50556	0.39924	0.964653	1.13736	0.81932	0.66933	0.49475	0.688237	1.08155	0.49674
5	0.29622	0.607127	0.74570	0.06589	0.581318	0.43089	1.05862.	1.07534	0.63847	0.913331	0.51173	0.64017
. . . 240	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . , .	. . , .

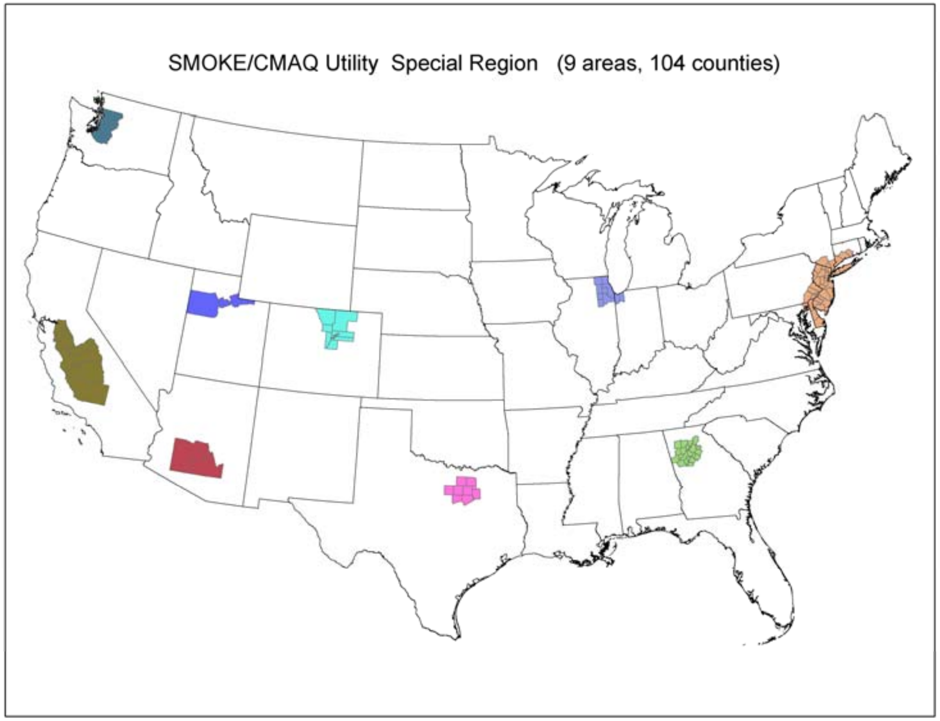
Experimental Design:
1. Selection of policy factors (2)

- Covers from zero to 120 percent of baseline emissions
- Staged Latin Hypercube (space filling design)
- 240 total runs, 120 runs in first stage, 60 runs each in stages two and three
  - Will allow testing of additional predictive power of additional model runs
- 10 additional model validation runs (out of sample)



Experimental Design:
1. Selection of policy factors (3)

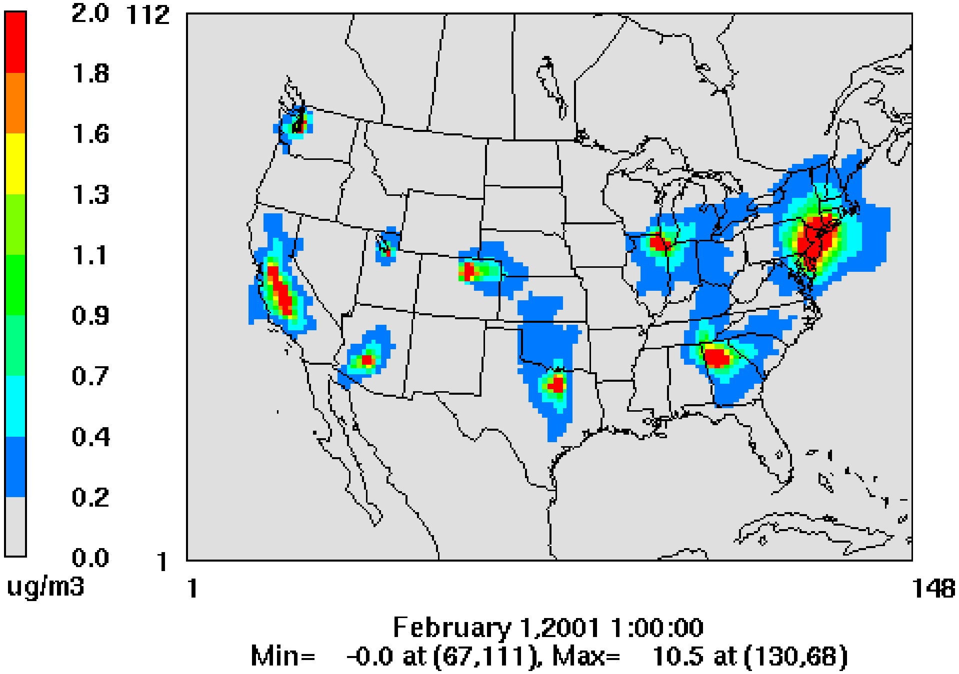
- Regional vs. Urban control: independent response surfaces for 9 urban areas, as well as a generalized response surface for the rest of model domain
  - Nine urban areas include: NY/Philadelphia, Chicago, Atlanta, Dallas, San Joaquin, Salt Lake City, Phoenix, Seattle, and Denver
  - Selected so that ambient PM2.5 in each urban area is largely independent of the precursor emissions in all other included urban areas



PM 2.5: Areas of Influence for All 9 Urban Locations

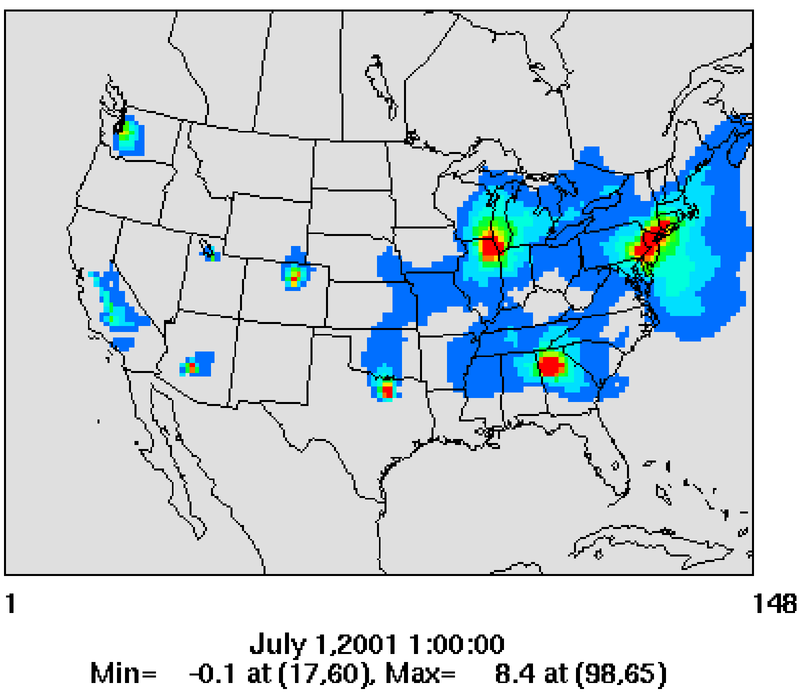
February 2001

(monthly avg.)



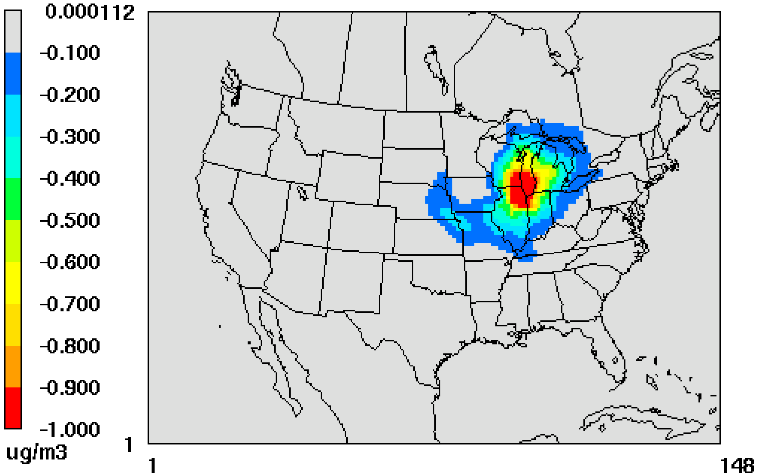
July 2001

(monthly avg.)

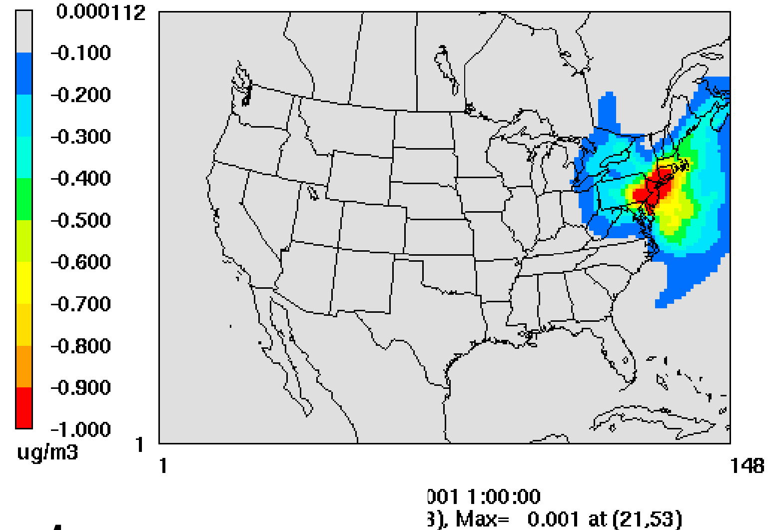


Areas of Influence for Selected Urban Locations

Chicago

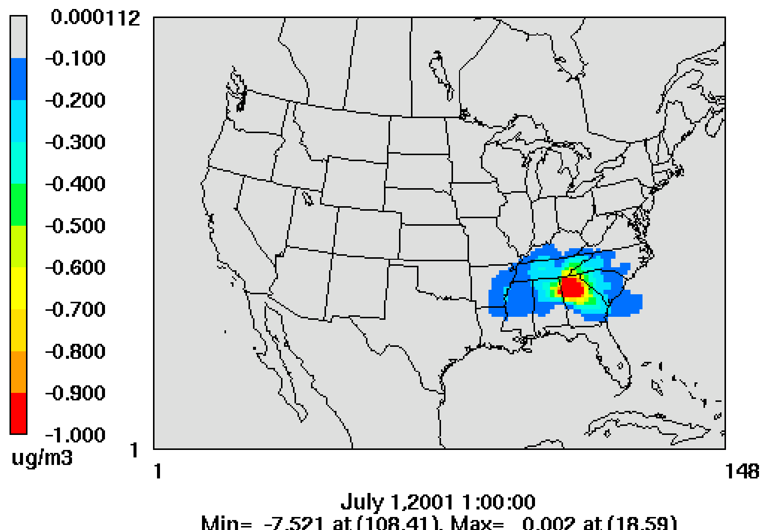


New York/Philadelphia



PM 2.5 July monthly avg.

Atlanta



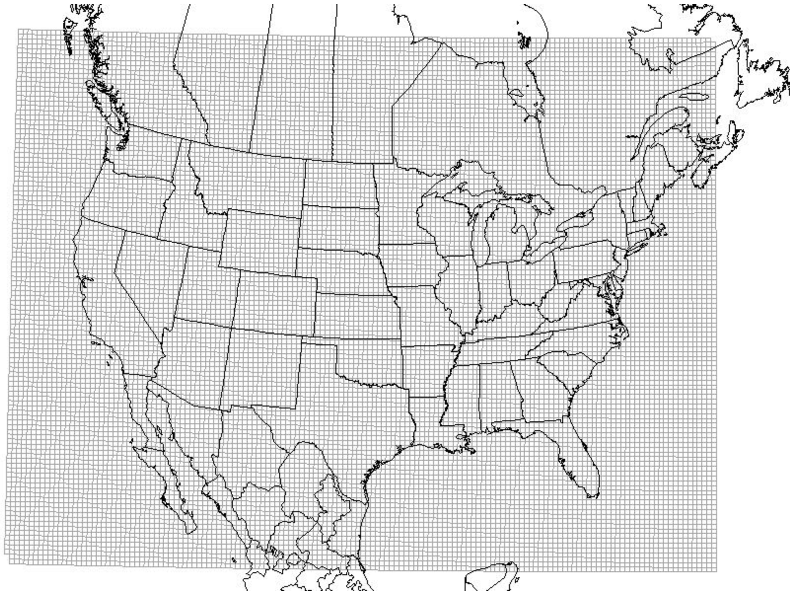
Small overlaps between Atlanta and Chicago influences in Western KY

Small overlaps between Chicago and NY influences in Ohio and Western NY. No overlap between Atlanta and NY

Experimental Design:
2. Selection of Model Simulations

- CMAQ model simulations
  - Domain = Continental U.S. 36-km CAIR modeling domain
  - 4 months, one from each season, February, April, July, October (months selected to provide best prediction of quarterly mean)
- Baseline Emissions Data
  - CAIR 2010 Base Case
  - Includes Tier 2, Heavy Duty Diesel Engines, and Nonroad Diesel standards, as well as the NOx SIP Call and MACT standards

CMAQ Modeling Domain



CMAQ – SMOKE Interface Development for RSM

- Hundreds of sector/pollutant control runs required for CMAQ RSM
- Current CMAQ emission inputs from SMOKE is a 3-D “combined” emissions file; it is cumbersome and inefficient (both processing & computer resources) to generate a large number of SMOKE emission files for RSM modeling
- Need to develop a module within CMAQ to read directly the pre-merged SMOKE sector source files (e.g., 3-D point, 2-D mobile, 2-D area, etc.)
- Allow RSM to directly control % changes of (1) emissions (2) specified areas in each model run

CMAQ – SMOKE Interface Development for RSM

Region A (9 urban areas)

Run #	1) / NOx / EGU	2) NOx / NonEGU+ Area	3) NOx / Mobile	4) SOx / EGU	5) SOx / NonEGU_ Point	6) SOx / Area	7) VOC / All	8) NH3 / Area	9) NH3 / Mobile	10) POC& PEC / EGU+NonEGU	11) POC& PEC / Mobile	12) POC& PEC / Area
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	1.00000	1.000000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.000000	1.00000	1.00000
2	0.89177	1.020286	0.83054	0.96302	0.875918	0.90277	0.17062	0.17435	0.99388	0.394387	0.90896	0.87647
3	0.57572	0.266010	1.02190	0.55610	0.247603	0.80328	0.61750	1.00113	0.02458	0.501642	1.12849	0.52022
4	0.06134	0.462060	0.50556	0.39924	0.964653	1.13736	0.81932	0.66933	0.49475	0.688237	1.08155	0.49674
5 . . . 240	0.29622	0.607127. . . .	0.74570	0.06589. . . .	0.581318. . . .	0.43089	1.05862.	1.07534	0.63847	0.913331. . . .	0.51173 . . , .	0.64017 . . , .

Region B (rest of domain)

Run #	1) / NOx / EGU	2) NOx / NonEGU+ Area	3) NOx / Mobile	4) SOx / EGU	5) SOx / NonEGU_ Point	6) SOx / Area	7) VOC / All	8) NH3 / Area	9) NH3 / Mobile	10) POC& PEC / EGU+NonEGU	11) POC& PEC / Mobile	12) POC& PEC / Area
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	1.00000	1.000000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.000000	1.00000	1.00000
2	0.89177	1.020286	0.83054	0.96302	0.875918	0.90277	0.17062	0.17435	0.99388	0.394387	0.90896	0.87647
3	0.57572	0.266010	1.02190	0.55610	0.247603	0.80328	0.61750	1.00113	0.02458	0.501642	1.12849	0.52022
4	0.06134	0.462060	0.50556	0.39924	0.964653	1.13736	0.81932	0.66933	0.49475	0.688237	1.08155	0.49674
5 . . . 240	0.29622	0.607127. . . .	0.74570	0.06589. . . .	0.581318. . . .	0.43089	1.05862.	1.07534	0.63847	0.913331. . . .	0.51173 . . , .	0.64017 . . , .

RSM Validation and Evaluation

- Cross validation
  - for each RSM iteration, one of the model runs is left out, the RSM is computed and used to predict the omitted run
  - RSM predicted changes in AQ are compared with CMAQ predictions and the mean square error (MSE) over all grid cells is computed for the run
- Out-of-sample validation
  - 10 additional CMAQ runs are conducted (not part of the experimental design and are not used in developing RSM)
  - RSM predictions for these model runs are compared with the CMAQ predictions and the MSE over all grid cells is computed for each run

Cross-Validation:

RSM Predictions vs. "True" CMAQ Values for July Total PM2.5 (120 rolling comparisons)

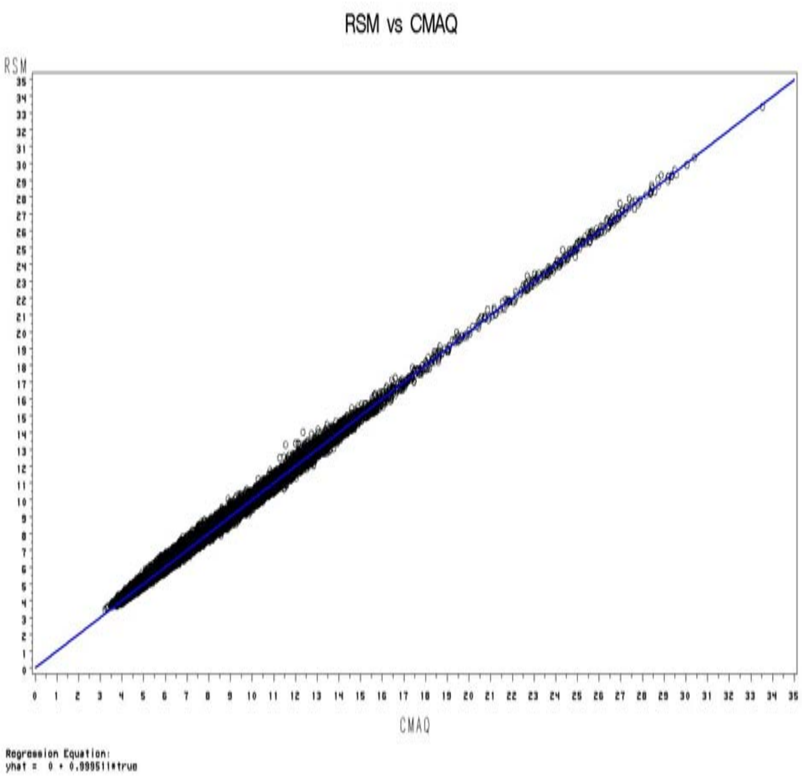


Table 1. Cross Validation Performance Metrics for Predicted July Total PM2.5 Mass (based on an evenly geographically distributed subsample of 700 grid cells, out of ~6,300 in the continental U.S.)

Performance Metric	Cross Validation (n=121)		
	Mean	Minimum	Maximum
Mean Bias ($\mu g/m^3$)	0.000	-0.063	0.130
Mean Error ($\mu g/m^3$)	0.027	0.006	0.130
Mean Normalized Bias (%)	0.02%	-1.58%	2.96%
Mean Normalized Error (%)	0.71%	0.21%	2.97%
Mean Fractional Bias (%)	0.01%	-1.61%	2.87%
Mean Fractional Error (%)	0.71%	0.22%	2.88%

Cross-Validation:

RSM Predictions vs. "True" CMAQ Values for October Total PM2.5 (120 rolling comparisons)

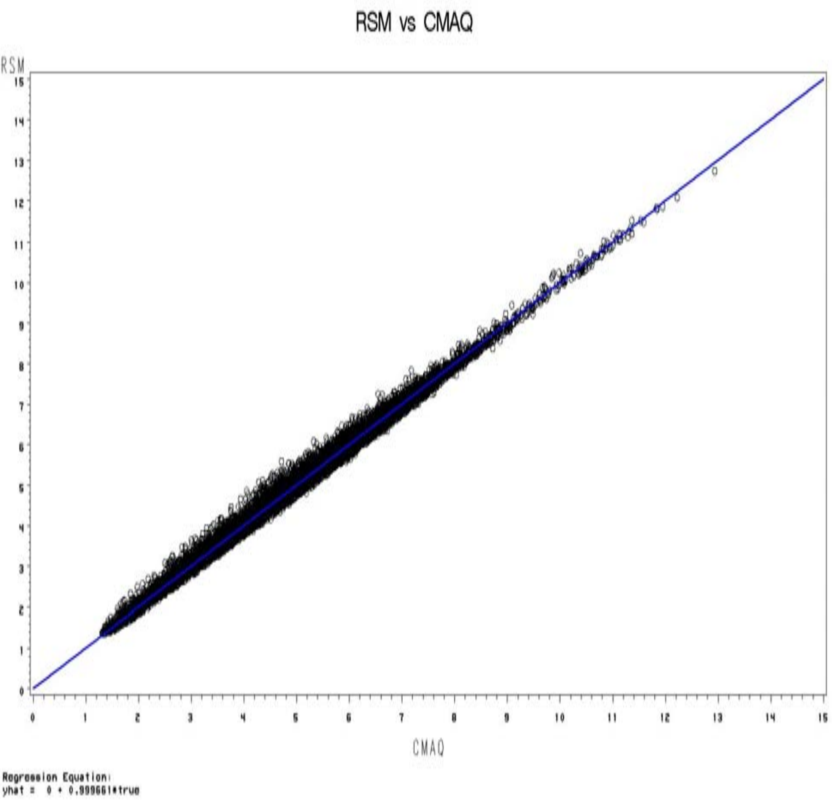
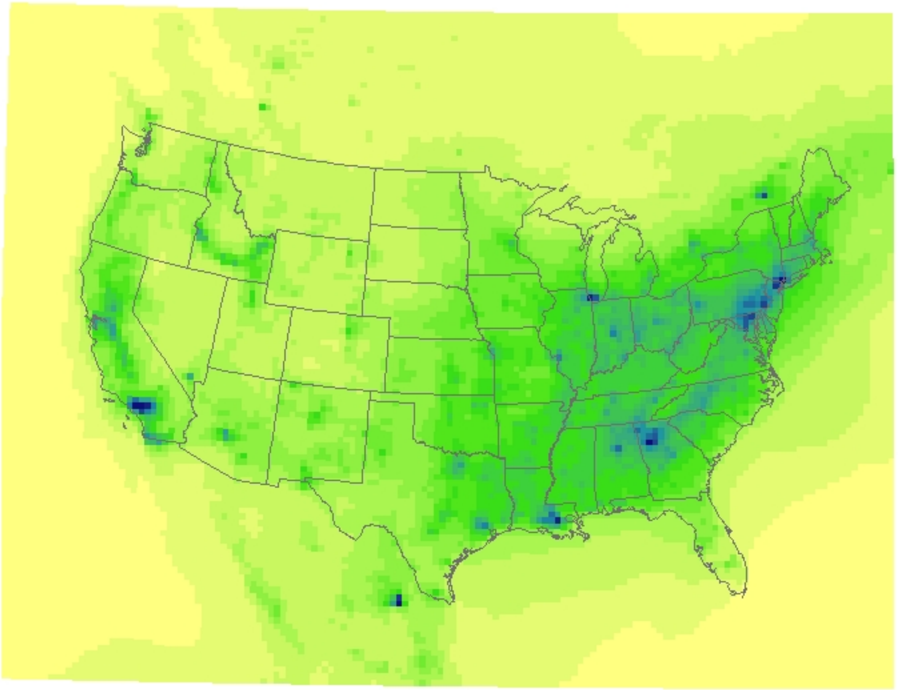


Table 2. Cross Validation Performance Metrics for Predicted October Total PM2.5 Mass (based on an evenly geographically distributed subsample of 700 grid cells, out of ~6,300 in the continental U.S.)

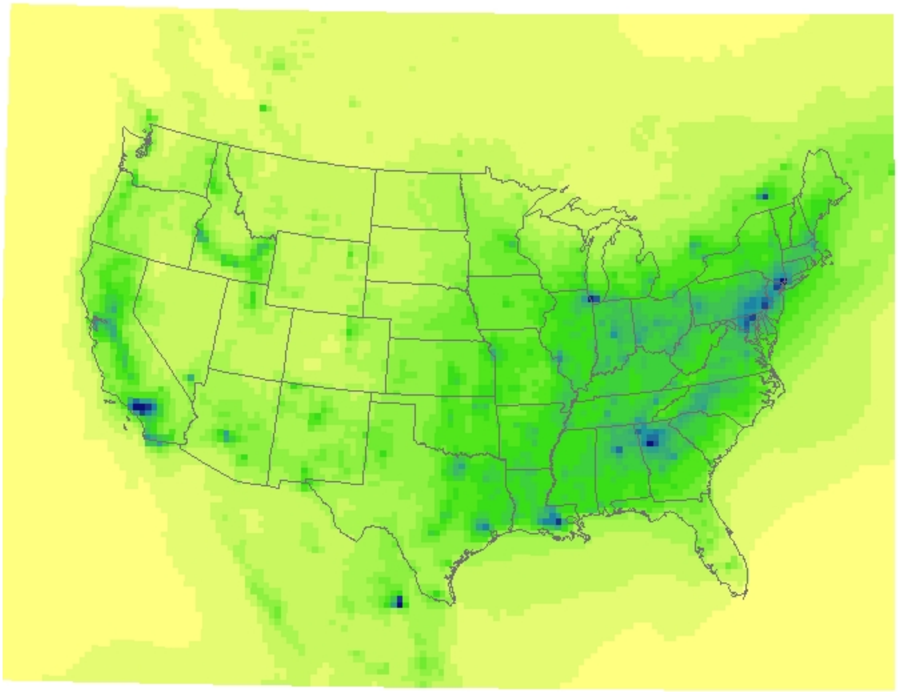
Performance Metric	Cross Validation (n=121)		
	Mean	Minimum	Maximum
Mean Bias ($\mu g/m^3$)	0.000	-0.100	0.221
Mean Error ($\mu g/m^3$)	0.047	0.007	0.221
Mean Normalized Bias (%)	0.03%	-2.70%	6.40%
Mean Normalized Error (%)	1.19%	0.18%	6.73%
Mean Fractional Bias (%)	0.01%	-1.61%	6.40%
Mean Fractional Error (%)	1.19%	0.18%	6.40%

Out-of-Sample Validation

RSM Predictions



True CMAQ Simulations



Total PM2.5 (ug/m3)

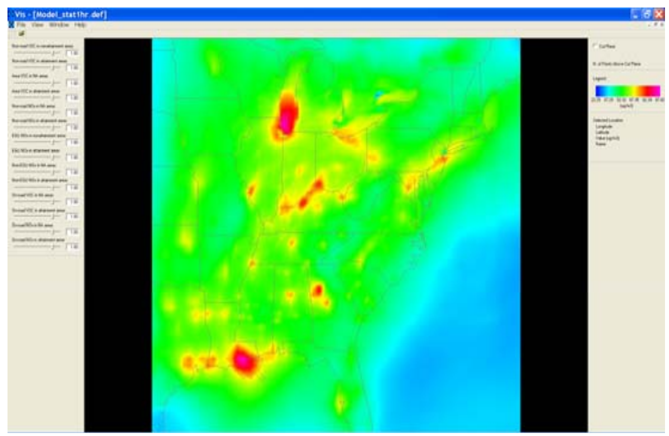
0.192111000 - 1.00000000	9.00000001 - 10.0000000	18.0000001 - 19.0000000
1.00000001 - 2.00000000	10.0000001 - 11.0000000	19.0000001 - 20.0000000
2.00000001 - 3.00000000	11.0000001 - 12.0000000	20.0000001 - 43.0000000
3.00000001 - 4.00000000	12.0000001 - 13.0000000	
4.00000001 - 5.00000000	13.0000001 - 14.0000000	
5.00000001 - 6.00000000	14.0000001 - 15.0000000	
6.00000001 - 7.00000000	15.0000001 - 16.0000000	
7.00000001 - 8.00000000	16.0000001 - 17.0000000	
8.00000001 - 9.00000000	17.0000001 - 18.0000000	

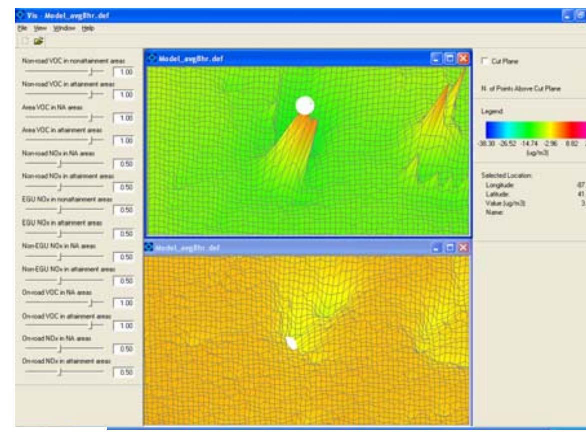
Total PM2.5 (October monthly avg.) :

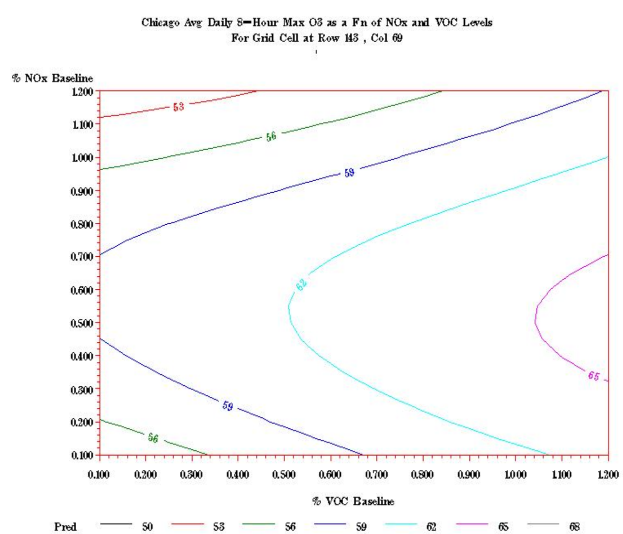
RSM vs. CMAQ simulation (Run 2 - Base)

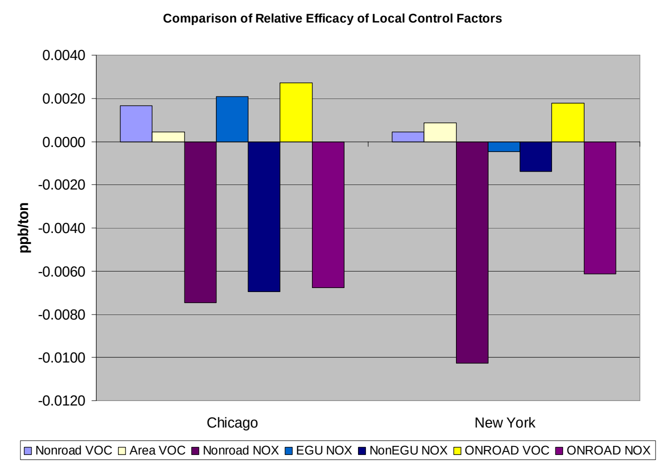
RSM Graphical Tool:
Visual Policy Analyzer

- Graphical analysis tool to allow for "real-time" RSM predictions of ozone, PM, visibility, and deposition
- Improvements will be implemented for CMAQ RSM

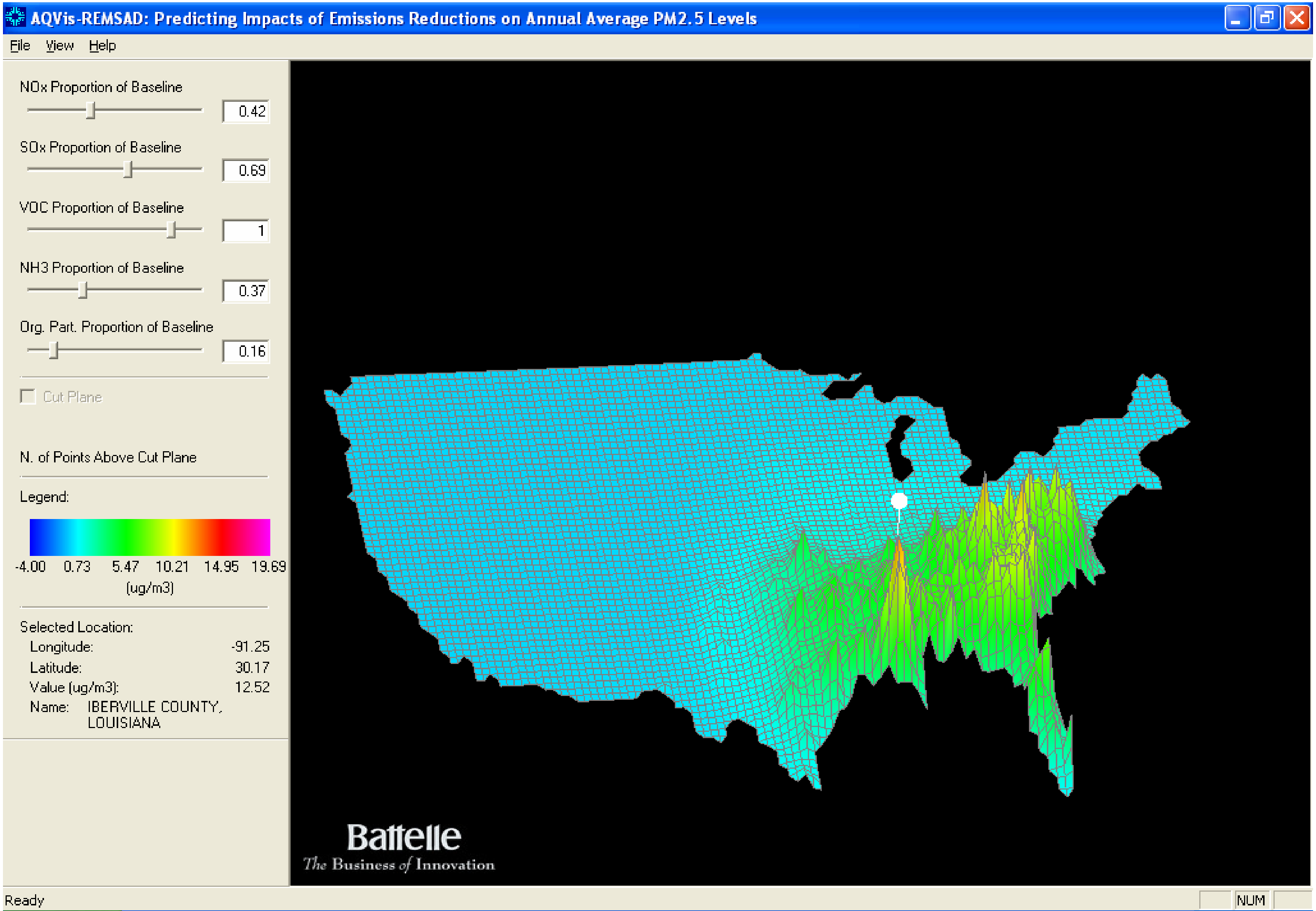








Demo of Prototype RSM Visualization Tool



Next Steps

- Planning for 12km "Local Scale" RSM for selected areas of concern
- Implementation of multi-pollutant ASAP version using CMAQ RSM
- Use RSM results to investigate/guide sector based O3/PM analyses
- Collaboration & outreach to AQ community (RPOs, academic, international, etc.) to facilitate transfer of methods and development of RSM tools