Employing Sub-Grid Variability (SGV) properties of Air Quality models towards improving exposure assessments

Presentation at EOSHI, NJ, June 20, 2007
American Chemistry Council Workshop on VOC inhalation exposure research

by
Jason Ching
ASMD/NOAA & NERL/EPA
RTP, NC
Ching.jason@epa.gov
Collaborators

• Mohammed Majeed
  – Delaware DNREC
• Vlad Isakov
  – NOAA/ARL & NERL/USEPA
• Winifred Hamilton
  – Baylor College of medicine
• Daewon Byun
  – IMAACs; University of Houston
• Wenyaw Chan
  – UT School of Public Health
• **Given:** Sub Grid Variability, SGV, is an inherent property of all grid models.

• **Introduce** concepts and a method for incorporating SGV with CMAQ simulations for use in Exposure Assessments (and other Applications)

• **Preliminary Results**
  – Philadelphia 12 km: 1995 and 2001
  – Delaware 12 km: Episodic (July 2001)
  – Houston 4 km study: Episodic (Texas 2000)
Statement of Problem

• Air Toxics HOT SPOTS: Characterization difficult with monitoring network. Coarse grid size for CMAQ operationally expedient but also unable to provide characterization.

• Modeled gridded fields provide average concentration values for a grid volume and does not represent the variabilities at smaller grid sizes.

• Exposures are from multiple pollutants over various exposure time scales and activity patterns, and most pronounced in vicinity of hot spots.
Neighborhood Scale (N-S) and SGV Modeling Prototype Paradigm

- CMAQ provides multiscale, grid resolved concentrations. Modeling at N-S is valuable when significant variability is present at that scale, but may still underestimate variability.
- SGV signals derived from (a) fine scale CMAQ, (b) from modeling of local sources and from (c) photochemistry in turbulent flows.
- SGV treated as concentration probability density functions (PDFs) are appropriate and essential information for improved human exposure assessments.
Schematic: Concentration variability within grid cells

Background
Simulation from air quality grid models yields a single concentration value for each grid.

However, within each grid there is an inherent amount of unresolved spatial variability as a result of the presence of variety of sources and turbulence-induced photochemistry.

Pointwise measurements will represent a value from the distribution of sub-grid concentration variability rather than the grid prediction value.

Thus, hypothetically, even if model and observation may be “Perfect” their values are NOT equivalent!!
Schematic: Concentration variability as function of grid size and location in modeling domain relative to observations at monitoring site.

Pollutant concentration, C8, is calculated by coarser grid air quality model (ex. 8km).

PDF (Probability Density Function) is derived from fine grid air quality model (ex. 1km).

SGV (Sub-Grid Variability) derived from CFD, LESchem or other techniques.

Pollutant concentration at point A = f (C8, PDF, SGV) (depending on time, chemical species, sources, transport).
Acetaldehyde relative histograms with Weibull probability density function fits (heavy lines) for 12 km grid cells derived from blocks of 81 1.33 km grid values at 15:00 LST from the CMAQ simulation of Philadelphia.

CMAQ 4 km gridded fields

Results of hybrid modeling method
Variability of benzene concentration at 4 km x 4 km grid cells for Philadelphia derived from ASPEN (Top) and ISCST3 (Bottom)

Note the degree of variability using local scale modeling (>1 order of magnitude)
Complementing SGV with CMAQ

**SAC** (SGV Adjusted CMAQ Concentration)

\[
\text{SAC} = C_g \times f_1 (C_{SGV}) \times f_2 \times f_3 \tag{1}
\]

- \(f_1\): Weighting function or SAC Factor based on model SGV
- \(f_2\): Function dependent on surrogate exposure parameters
- \(f_3\): Function dependent on photochemical-dynamical contributions

Functions (\(i=1\) is focus of current investigations) e.g.,

\[
f_1(C_{SGV}) = 1 + \frac{(C_{SGV})}{C_g} = 1 + \text{COV} \tag{2a}
\]

Other suggested \(C_{SGV}\) options include but not limited to:

- 95th or other percentile of the distribution \(\tag{2b}\)
- Peak (or range)/cell mean. \(\tag{2c}\)
Preliminary Results, Findings

- Concentration fields at different model grid resolutions
- SGV in coarse grid simulation
  - SGV statistics at 12 and 4 km grid cell resolution derived using outputs from 1.3 km grid simulations
- Examples of SACs
Distribution of CMAQ modeled pollutants for one 12 km grid cell from 1.3 km results in central Philadelphia, July 12, 1995)

Peak-to-Grid Mean
Range-to-Grid mean
Subgrid variation in 12 km grid cell (central Philadelphia) from 9x9 1.3 km grid cells (July 14, 1995)
Subgrid Variation in 12 km cell (central Philadelphia) from 9x9 1.3km cells (July 14, 1995)
PHILADELPHIA COUNTY

CMAQ Annual (2001) simulation of Benzene (ug/m3) using 4 km grids

SGVs & SAC factors are derived from ISCST3 model. Road links are shown as background.

COMMENTS:
CMAQ: Range of variability from 0-4 ug/m3

SAC factor:
(1) Not necessarily spatially correlated with gridded CMAQ results
(2) COV > 2X gridded values
(3) Peak/Mean > 10 X Mean
(4) 95th percentile ~5X Mean
Benzene (ppb) (July 2001) Wilmington, DE (A&B are urban and C is rural cell)

Black (A) Red (B) Green (C)
1+COV for small domain (Central Houston)

Benzene  Acetaldehyde  1-3 Butadiene

[Images of maps and scatter plots showing concentration levels for Benzene, Acetaldehyde, and 1-3 Butadiene over the dates 8/22 to 8/31.]
REVIEW

• SGV is an inherent property of all grid models. Its properties vary with grid size. We seek effective ways to utilize SGVs for applications.

• SAC, an ad hoc suggestion/method to incorporate SGV results with CMAQ.

• Results of SAC and SAC factors are based on SGVs from local scale and fine scale CMAQ model simulations for episodic and annual simulations.

• Characteristic of SAC factors varied for different:
  – pollutant species
  – emission densities and mixtures
  – grid sizes
  – averaging periods
What does this all mean?

- We haven’t exploited the SGV property of all grid models.

- Regarding personal exposure it seems likely that SGV in combination with grid modeled concentration would be considerably more realistic than using gridded values alone.

- This conjecture has not been tested, fully, but if so, it would have enormous implications and importance in performing exposure assessments.

- What are these $f_i$ functions?
Suggested Outline of SAC Boosting Study

“Use of Inherent Sub-Grid Variability to Amplify CMAQ-AT and Reduce Uncertainty in Human Exposure to VOC Assessments”

Elements for conducting such a study is present in a current study of hospital admissions in Houston TX.
The objective of this study is to analyze spatial relationships between hospital admissions by cardiovascular or respiratory (CVR) discharge diagnosis and air pollution in 337 4 x 4-km cells in Harris County, Texas, from July 1 to October 6, 2000.

Three month hospital admissions database (~35,000/month) ~35-40% CVR or approximately 30K-40K CVR patients.

Three months of 36/12/4 km CMAQ-AT simulations of air pollutants include 5 criteria pollutants and 16 HAPs. PCA with varimax rotation applied to reduce the number of variables representing the 16 air toxics.

Statistical analysis primarily uses a mixed-level generalized linear model (GLM) with Poisson response and log link to examine the relationship between age- and gender-adjusted hospitalization rate by diagnosis (the dependent variable) and CMAQ-AT simulated air pollution (exposure), adjusting for significant individual and domain-specific demographic factors.
Hypotheses for Proposed studies

• Boosting gridded exposure concentrations with within-grid data extrapolations measurably improves model performance by reducing uncertainty in the multivariate model as measured before and after application of these signal boosts.
  – Hot spots can be characterized using AQ modeling or combination of model and observations; but the appropriate scale resolution is yet to be determined
  – CMAQ at coarse scales need fine scale information for providing information on “hot spots” This is obtained using finer grid sizes and local scale modeling approaches
Approach using Inherent Sub-Grid Variability to Amplify CMAQ-AT and Reduce Uncertainty in Human Exposure Assessments

• (1) Derive and characterize the SGV concentration distributions for 4-km CMAQ-AT simulations using both nested 1 x 1-km runs and local-scale dispersion modeling applied to mobile (road link) and other source data;

• (2) Develop SGV weighting factors from these distributions and apply them to the 4x4-km CMAQ simulations for the health assessment;

• (3) Design and explore a set of optional amplifications to CMAQ-AT using additional “boosts” to these SGV weights or alternatively using GIS overlays of additional exposure factors such as proximity of patient residence to roadway, within-cell density of road links and other surrogates for exposure. These may give insights on $f_i$s

• (4) Explore methods to adjust CMAQ-AT concentrations with available observational data; and

• (5) Apply statistical methods to measure and compare the residual (random or structural) uncertainty under different signal boosting conditions.
3: Design and apply amplification factors to CMAQ using additional “boosts” to these SGV weights; incorporate GIS overlays of additional exposure factors including proximity of patient residence to roadway, within-cell density of road links and other surrogates for exposure.

- It is expected that the pollutant concentration gradients along the roadside pollutants vary greatly depending on the species of interest. We characterize pollutant gradients in three ways:
  - (a) Compare CMAQ-AT results at higher resolution (i.e., 1 km) with the 4-km resolution for the cells including distinctive patterns of road segments.
  - (b) Apply parametric formulas of air pollution gradients from roadways (using literature estimates of values from the on-going EPA highway study program); and
  - (c) Study an independent approach to characterize SGV in this particular hospital admissions study by utilizing actual geocoded patient addresses, i.e., point data, and creating a buffer zone around each residence. Characterize each of the geocoded patients in study based on the daily VMT and other road link parameters in patient’s zone.
A Research Opportunity

- Intriguing new methodology for improving exposure assessments
- Health endpoint is actual geocoded addresses of hospital admission patients with CVR release diagnosis
- Considerable effort expended to date to produce a high quality and large quantity of model and hospital data for Houston.
- Collaboration and synergisms between the Houston team, EOSHI and EPA towards advancing tools and techniques with available data sets is very promising.
- Envision collaborative investigations to have a very high benefits to cost bases. The outcomes would be highly rewarding achieved at a relatively low cost and risk, i.e. Leveraging onto initial costs already expended.
- Research plans could be readily developed. Implementation would be move quickly due to considerable experience and analyses conducted to date.
- Similar programs to test the fundamental hypotheses can build upon the success of the results of the Houston prototype.
Your interest and participation towards furthering developments on SAC methodology and to the propagating of their operational implementations and advanced applications is welcome.

Go SAC it to ‘em

Thank you

Disclaimer: The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce’s National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views.