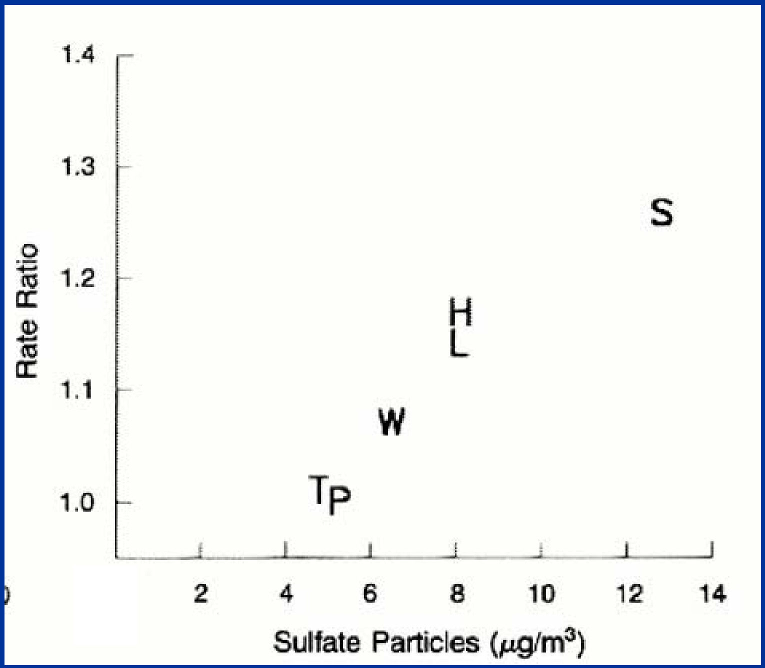
Enhanced Air Pollution Epidemiology using a Source-Oriented Chemical Transport Model

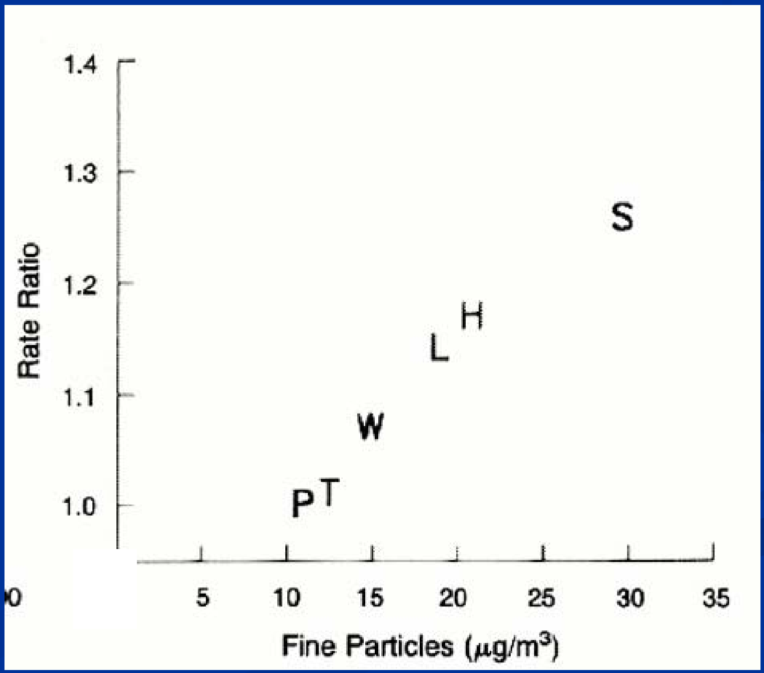
Michael J. Kleeman, Civil and Environmental Engineering, UC Davis
Shuhua Chen, Atmospheric Science, UC Davis
Qi Ying, Civil and Environmental Engineering, Texas A&M
Joel Kauffman, Environmental and Occupational Health Sciences, University of Washington
Paul Sampson, Statistics, University of Washington
Bart Ostro, California Office of Environmental Health Hazard Assessment
Peggy Reynolds, Northern California Cancer Center

June 10, 2009

Advancing Beyond the Basic Relationship Between PM2.5 and Health Has Proven Difficult

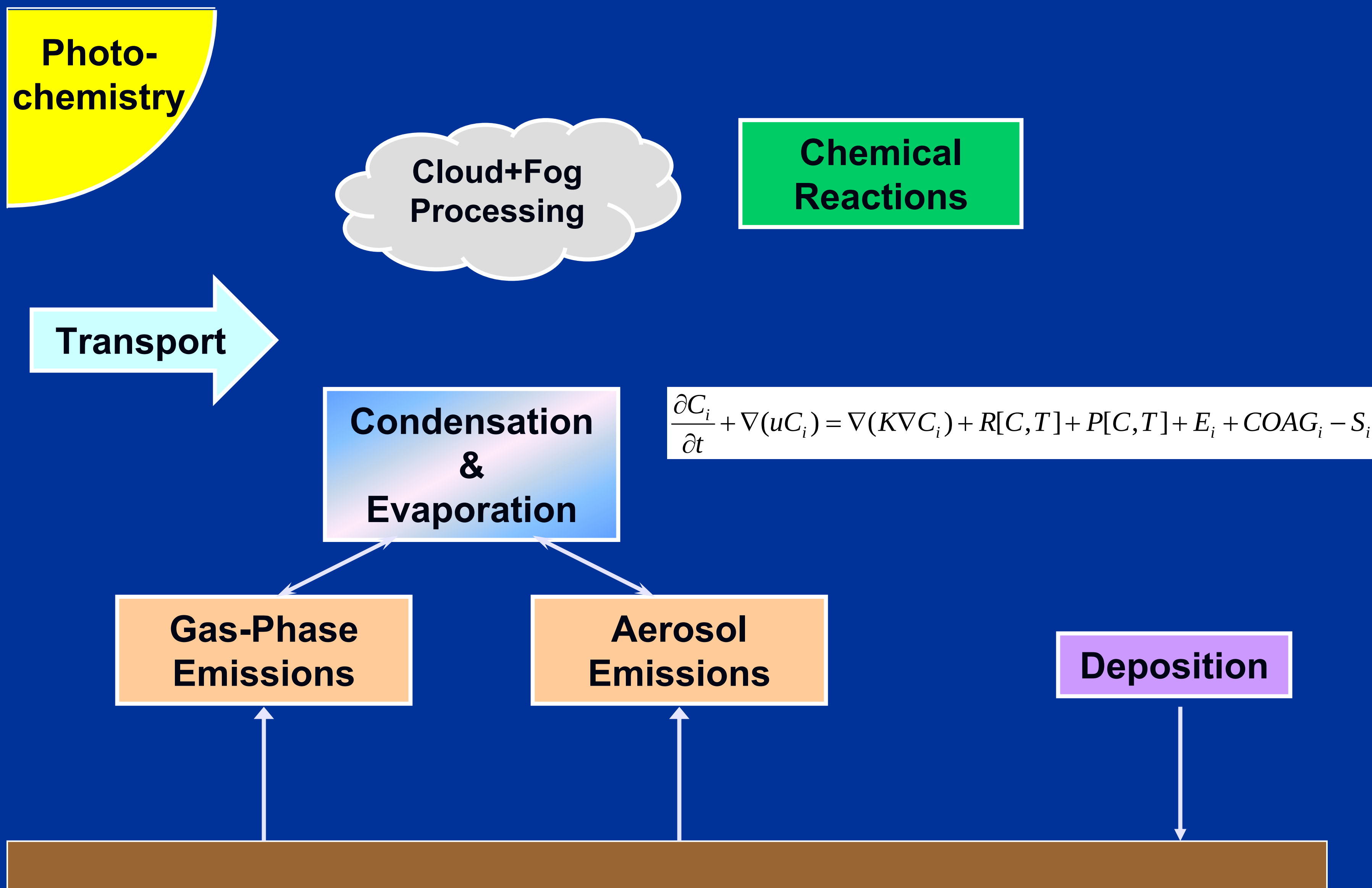
- Are central monitors providing poor exposure estimates that are masking the more detailed associations?
- "One in three" or "one in six" sampling schedules leave significant time gaps





Source: DOCKERY DW, POPE CA, XU XP, et al. "AN ASSOCIATION BETWEEN AIR-POLLUTION AND MORTALITY IN 6 UNITED-STATES CITIES", NEW ENGLAND JOURNAL OF MEDICINE 329 (24): 1753-1759 DEC 9 1993.

Air Quality Models



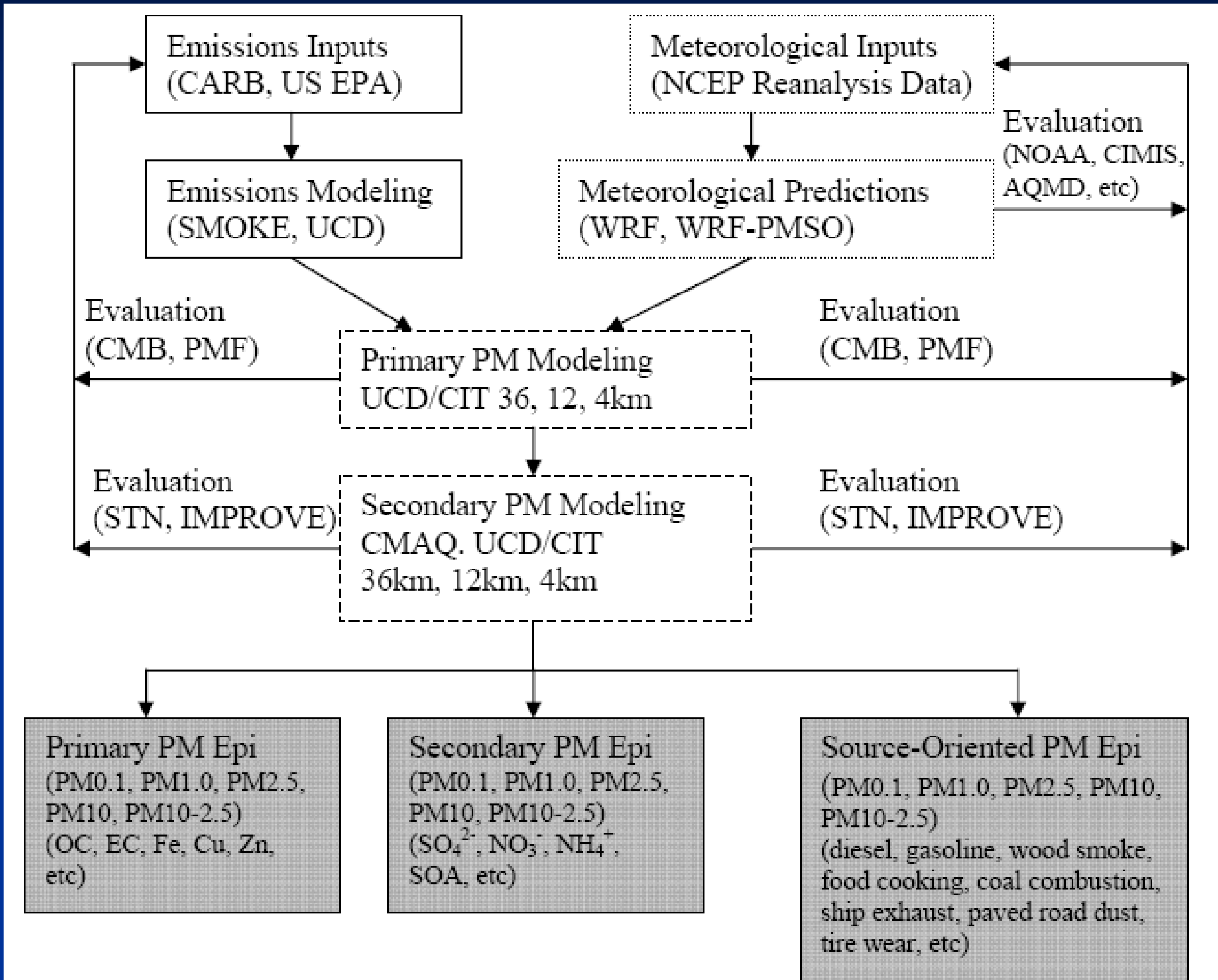
$$\frac{\partial C_i}{\partial t} + \nabla(uC_i) = \nabla(K\nabla C_i) + R[C,T] + P[C,T] + E_i + COAG_i - S_i$$

Figure courtesy of Prakash Bhave, US EPA.

Each Grid Cell in the Model Has:

- Gas phase species
  - O3, NO, NO2, NO3, N2O5, HNO3, HONO, HNO4, RNO3, PAN, PPN, NPHE, GPAN, PBZN, NH3, SO2, H2SO4, HCL, CO, CO2, MEK, HCHO, CCHO, RCHO, ACET, MGLY, PHEN, CRES, BALD, TOLU, C6H6, AAR1, AAR2, AAR3, AAR4, AAR5, AAR6, AAR7, OLE1, OLE2, OLE3, C7OL, C8OL, C9OL, ISOP, APIN, BPIN, HO2., RO2., OH, RCO3., etc
- Particle phase species
  - EC, OC, SO_3^{2-}, SO_4^{2-}, NO_3^-, Cl^-, NH_4^+, Na^+, Ca^{2+}, Fe, Cu, Mn, SOA, etc.
- Particle size distributions
- Source apportionment information
- Hourly time resolution

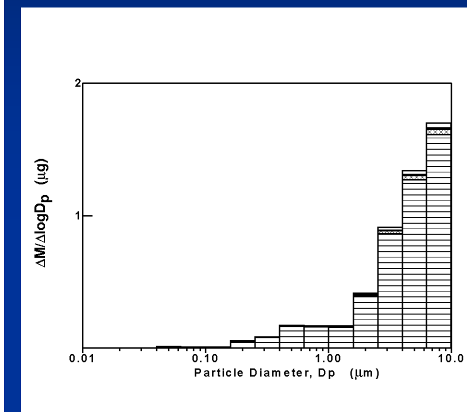
Our Project Design

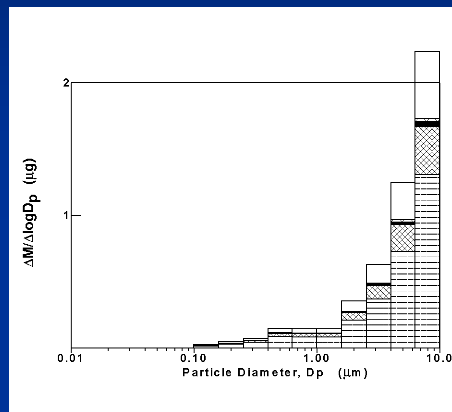


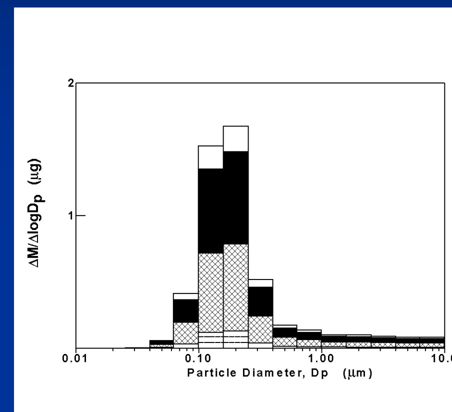
Hypothesis to Test

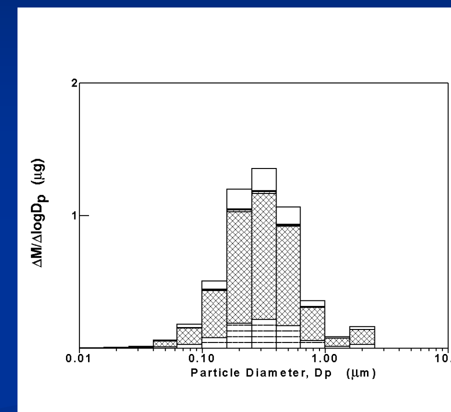
- ***Hypothesis 1:*** *Primary PM sources (diesel, gasoline, coal, etc) in the PM0.1, PM2.5, PM10, or PM10-2.5 size fractions are associated with acute and chronic human health effects.*
- ***Hypothesis 2:*** *Primary PM species (EC, OC, Fe, Zn, etc) in the PM0.1, PM2.5, PM10, or PM10-2.5 size fractions are associated with acute and chronic human health effects.*
- ***Hypothesis 3:*** *Exposure to PM generated by motor oil, diesel fuel, and/or gasoline fuel is associated with acute and chronic human health effects.*
- ***Hypothesis 4:*** *Simultaneous exposure to acidic particles and high concentrations of gas-phase oxidants is associated with acute and chronic human health effects.*
- ***Hypothesis 5:*** *Simultaneous exposure to particulate quinones and trace metals is associated with acute and chronic human health effects.*

Transforming the Regulatory Inventory Into a Source-Oriented Modeling Inventory







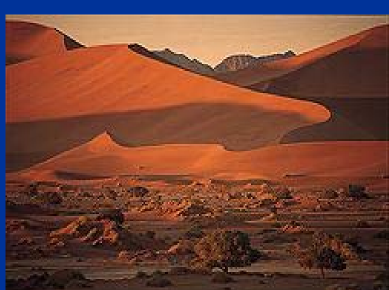


Crustal Material Other than Paved Road Dust

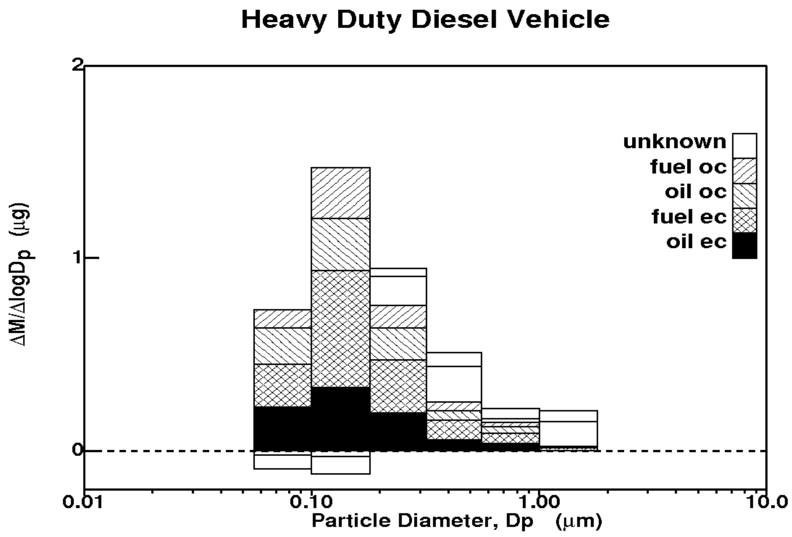
Paved Road Dust

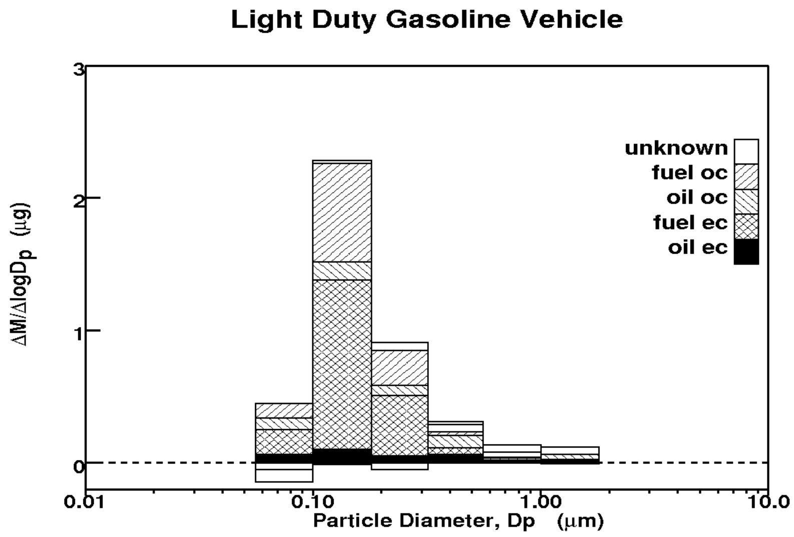
Diesel Engines

Meat Cooking



Source Profiles that Differentiate Motor Oil vs. Fuel Contributions to the Size Distribution of PM Emissions



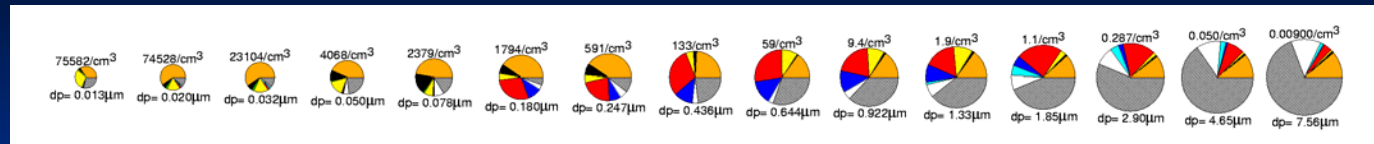


Quinone Emissions From Motor Vehicles

	Emission Rate [a,b] (µg L^{-1})											
	Light-duty Gasoline Vehicles by FTP						Heavy-duty Diesel Vehicles [c]					
	LEV (*9.3*)		TWC (*10*)		Smoker [d] (*8.8*)		1999 Idle-creep (*0.5*)		1999 HHDDT (*2.3*)		1985 HHDDT (*2.6*)	
Compound	gas phase	particle phase	gas phase	particle phase	gas phase	particle phase	gas phase	particle phase	gas phase	particle phase	gas phase	particle phase
BQN [f,j]	2-6 [a]	1.8 [k]	85	46	3200	1500	890 ± 600	180 [l]	510 ± 270	230 [l]	28000 ± 20000	1600
MBQN [f,i]					480	79 [l]	120 ± 40		35 ± 1		250 ± 30	
1,2-NQN [f,i]					340					10		44
1,4-NQN [f,i]					290		620 ± 160		120 ± 40	4.7	510 ± 50	27

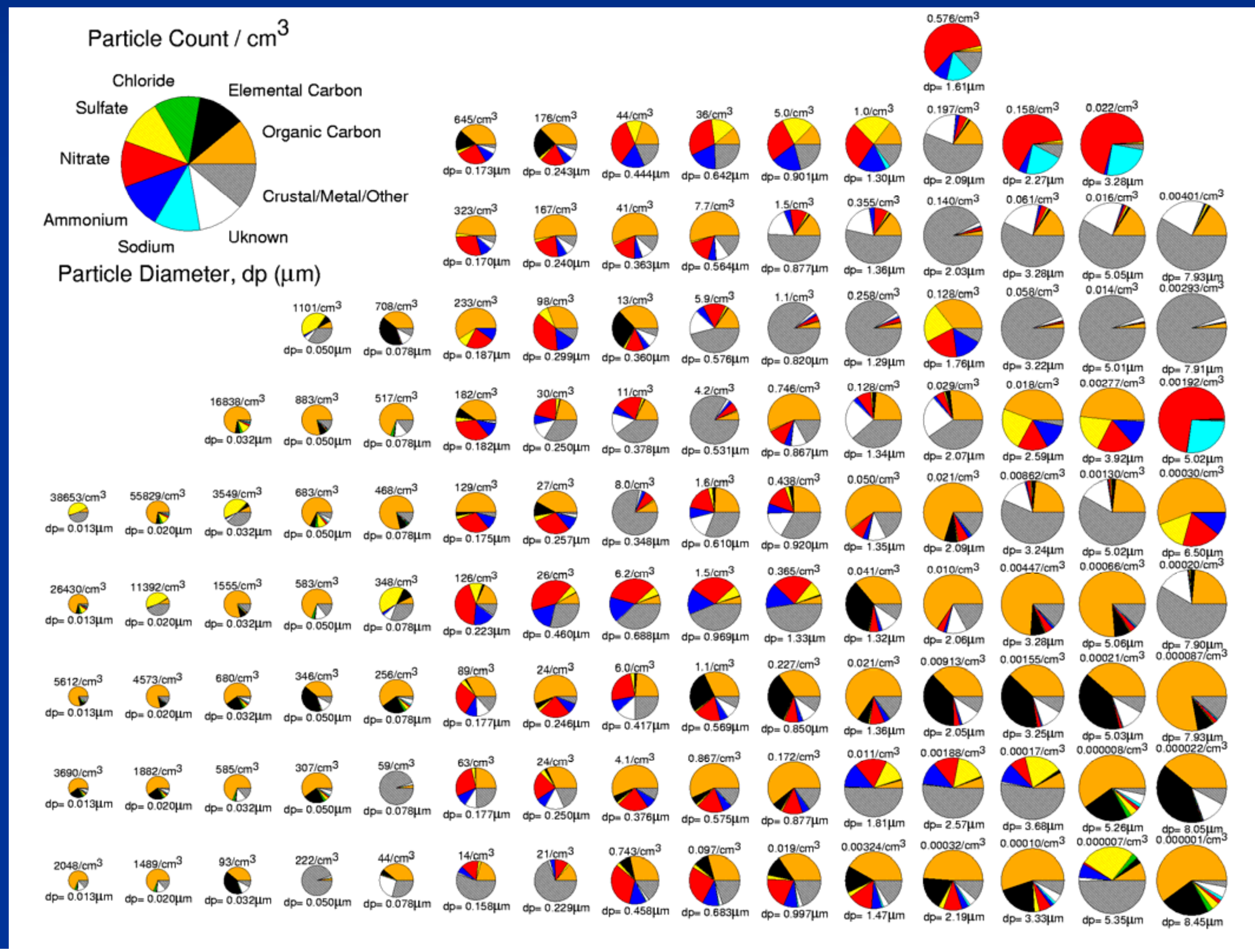
Source-Oriented External Mixture Representation

Internal Mixture



vs

Source-oriented external mixture



Model Evaluation CRPAQS PM2.5 Mass

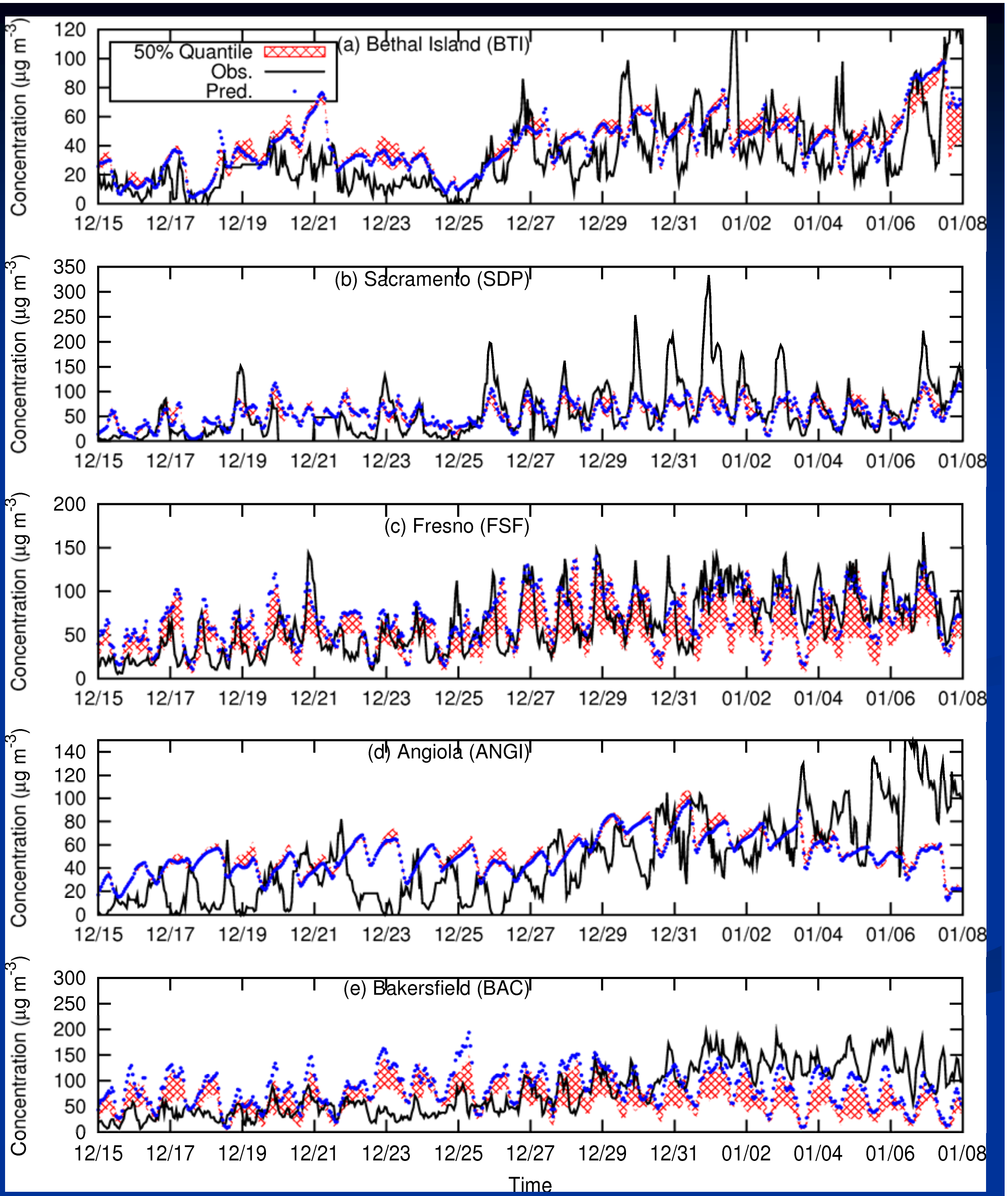
Black Line – measurements
Blue Line – predictions

Red Shading – Mid 50% Quantile within 10km of monitor

Major trends are captured at most stations

Under-prediction of mass at Angiola and Bakersfield near the end of the episode

Source: Q. Ying, J. Lu, P. Allen, P. Livingstone, A. Kaduwela, and M. Kleeman "Modeling Air Quality During the California Regional PM10/PM2.5 Air Quality Study (CRPAQS) Using the UCD/CIT Source-Oriented Air Quality Model – Part I. Base Case Model Results.", Atmos. Env., in press, 2008.

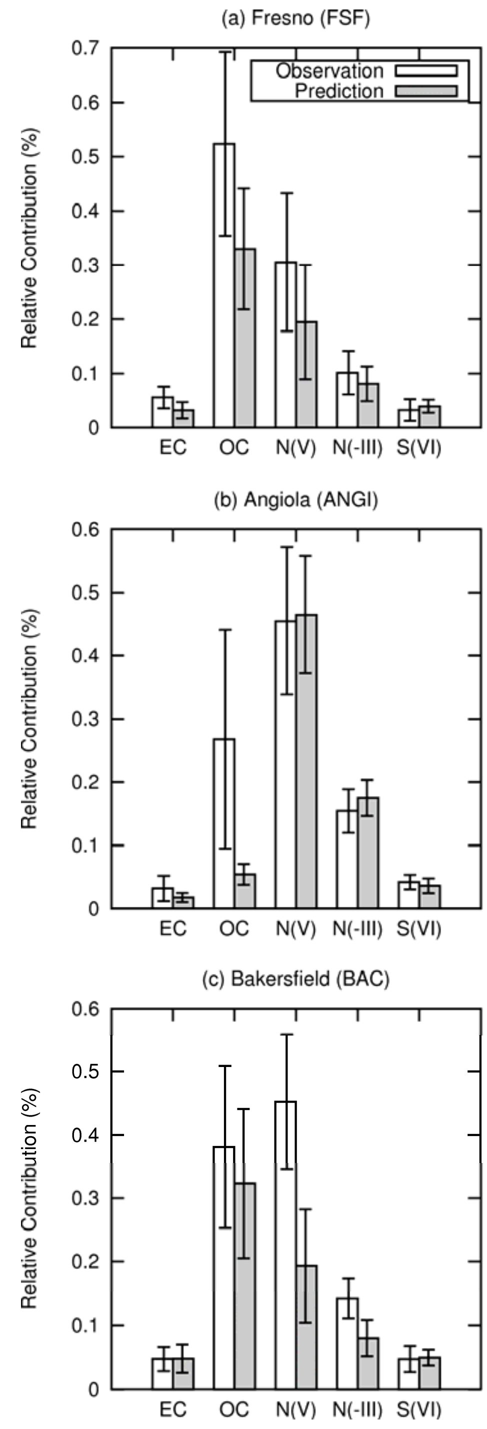


Model Evaluation Relative Component Contributions to PM

Average and standard deviation of predictions and observations is based on 55 samples

Urban locations (Fresno and Bakersfield)
Predictions and observations match except for nitrate under-prediction at Bakersfield (discussed previously)

Rural location (Angiola)
OC under-prediction. What primary sources are we missing? What SOA formation mechanisms are we missing?

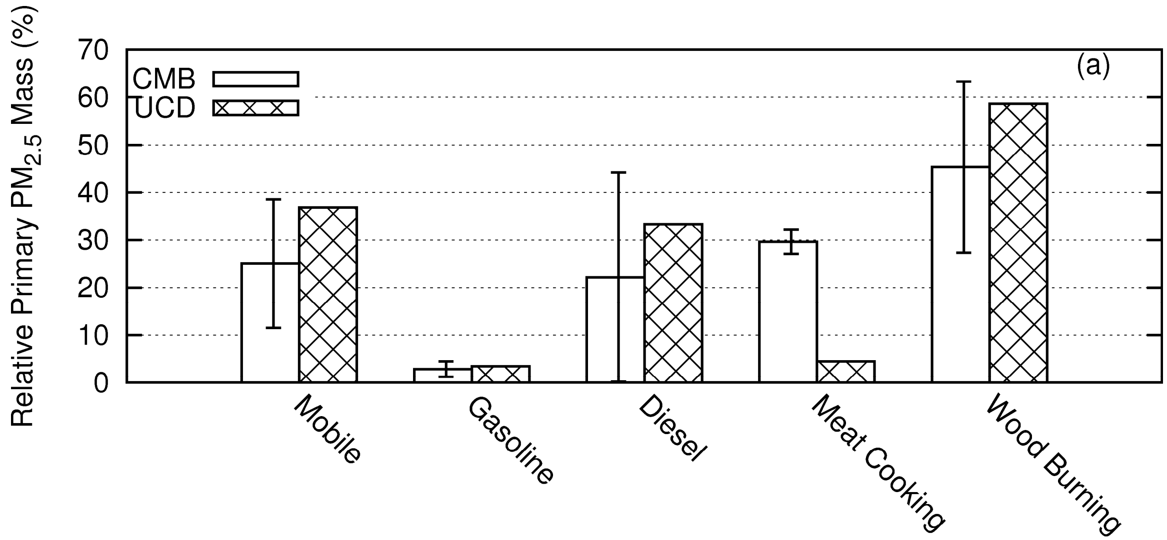


Source: Q. Ying, J. Lu, P. Allen, P. Livingstone, A. Kaduwela, and M. Kleeman "Modeling Air Quality During the California Regional PM10/PM2.5 Air Quality Study (CRPAQS) Using the UCD/CIT Source-Oriented Air Quality Model – Part I. Base Case Model Results.", Atmos. Env., in press, 2008.

Model Evaluation
Grid Model vs. CMB Source Apportionment

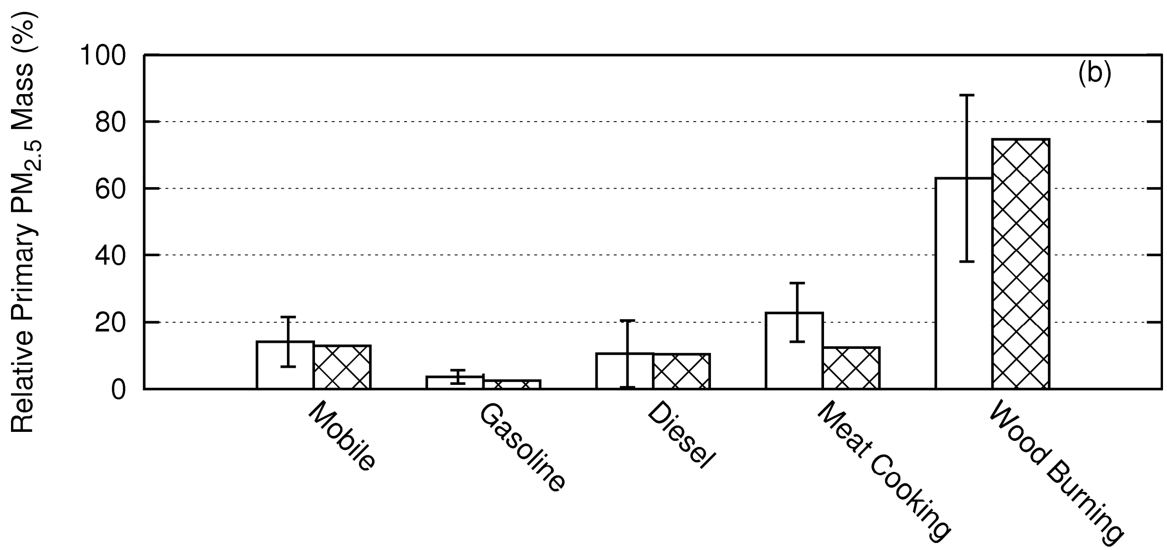
Angiola

**Dust sources removed from grid model



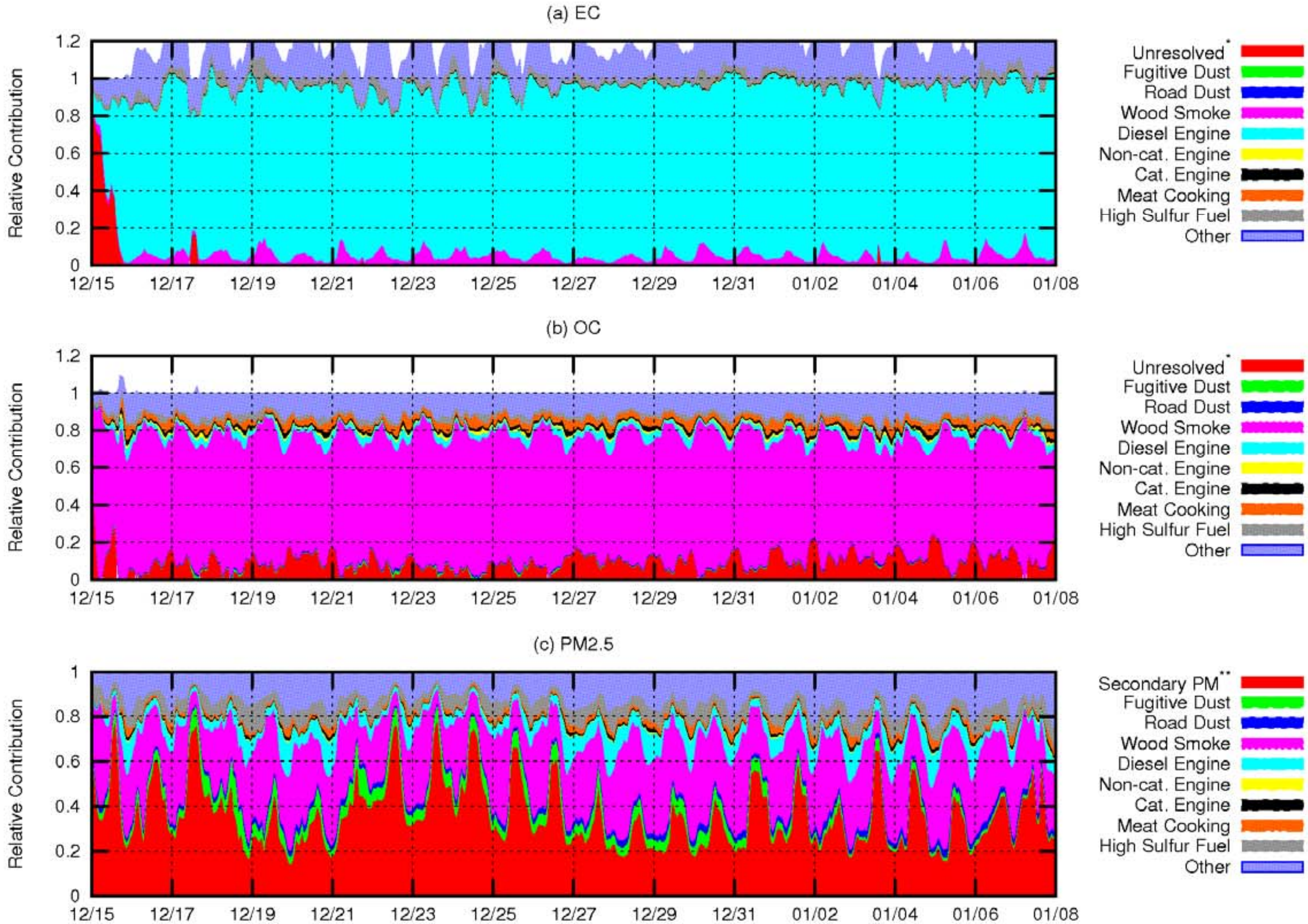
Fresno

**Dust sources removed from grid model



Source: Q. Ying, J. Lu, A. Kaduwela, and M. Kleeman “Modeling Air Quality During the California Regional PM10/PM2.5 Air Quality Study (CRPAQS) Using the UCD/CIT Source-Oriented Air Quality Model – Part II. Regional Source Apportionment of Primary Airborne Particulate Matter.”, Atmos. Env., in press, 2008.

Daily Variation of Predicted Source Contributions at Fresno Dec 2000-Jan 2001

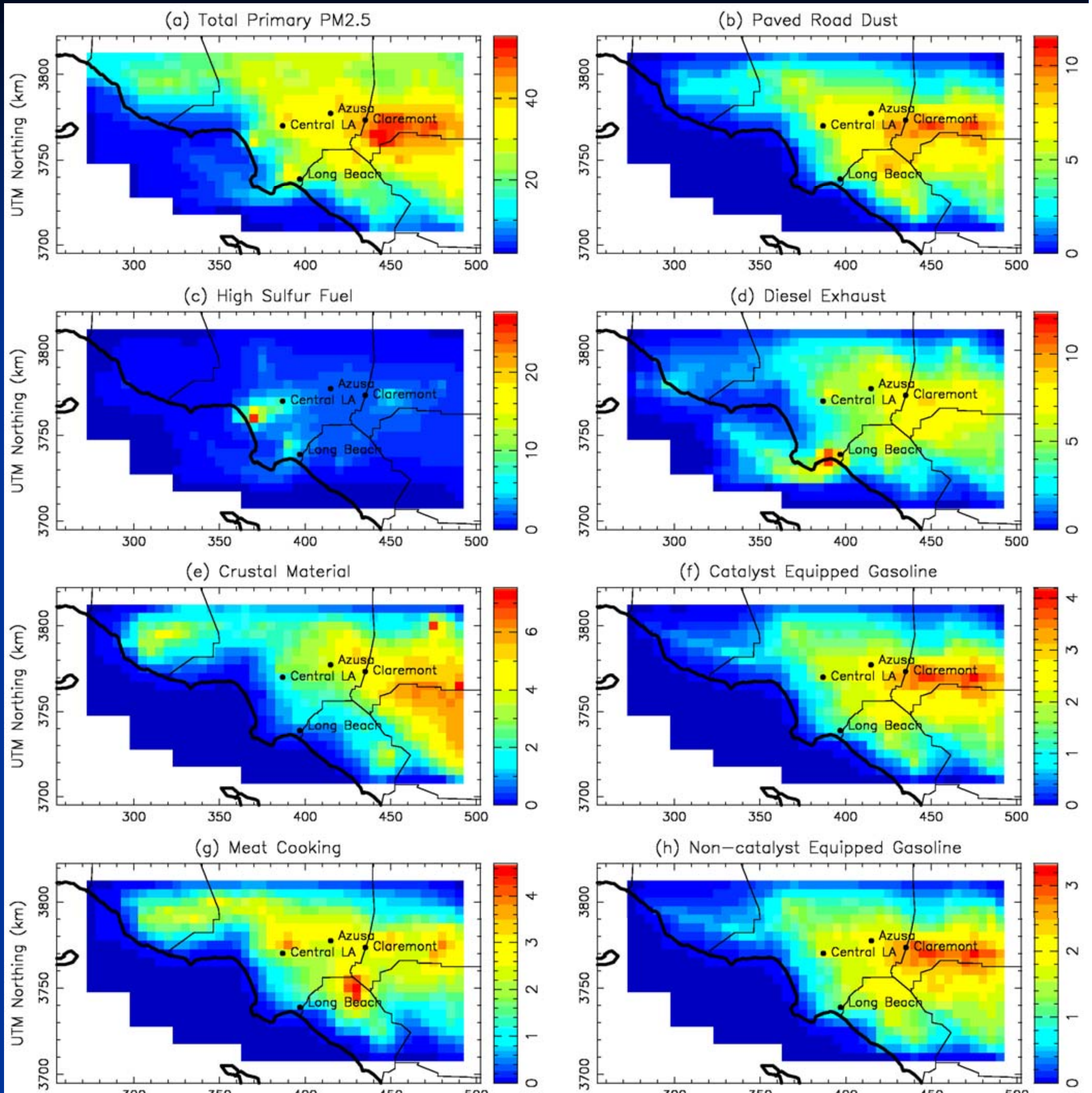


Source: 2008 Ying, Q., J. Lu, A. Kaduwela, and M.J. Kleeman. Modeling Air Qualitying During the California Regional PM10/PM2.5 Air Quality Study Using the UCD/CIT Source-Oriented Air Quality Model - Part II. Regional Source Apportionment of Primary Airborne Particulate Matter. Atmospheric Environment, accepted for publication.

Regional Source Apportionment Example:

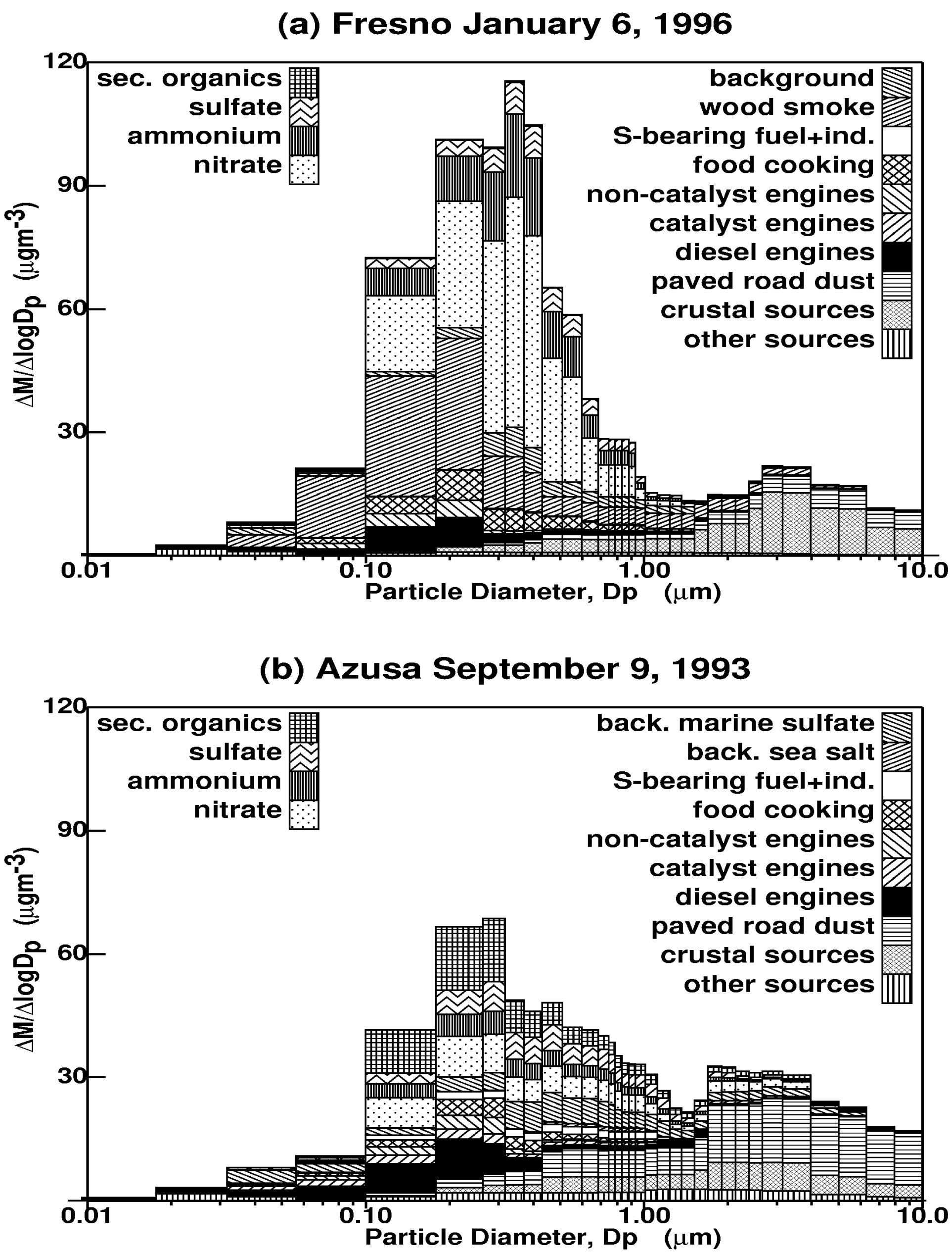
We can use the source-oriented model to predict the regional distribution of PM emitted from different sources.

Regional source contributions to PM in Los Angeles on September 25, 1996 .



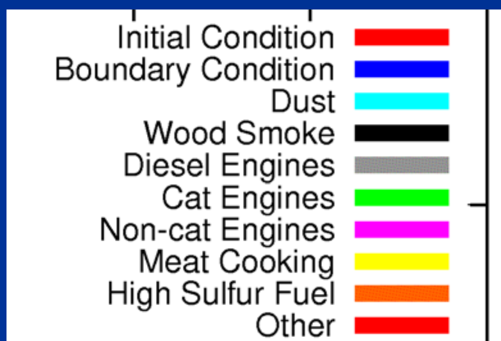
Source: 2005 Held T., Q. Ying, M.J. Kleeman, J.J. Schauer, M.P. Fraser. A comparison of the UCD/CIT air quality model and the CMB source-receptor model for primary airborne particulate matter. Atmospheric Environment. 39: 2281-2297.

Source-oriented models can predict source contributions to airborne particle size distributions

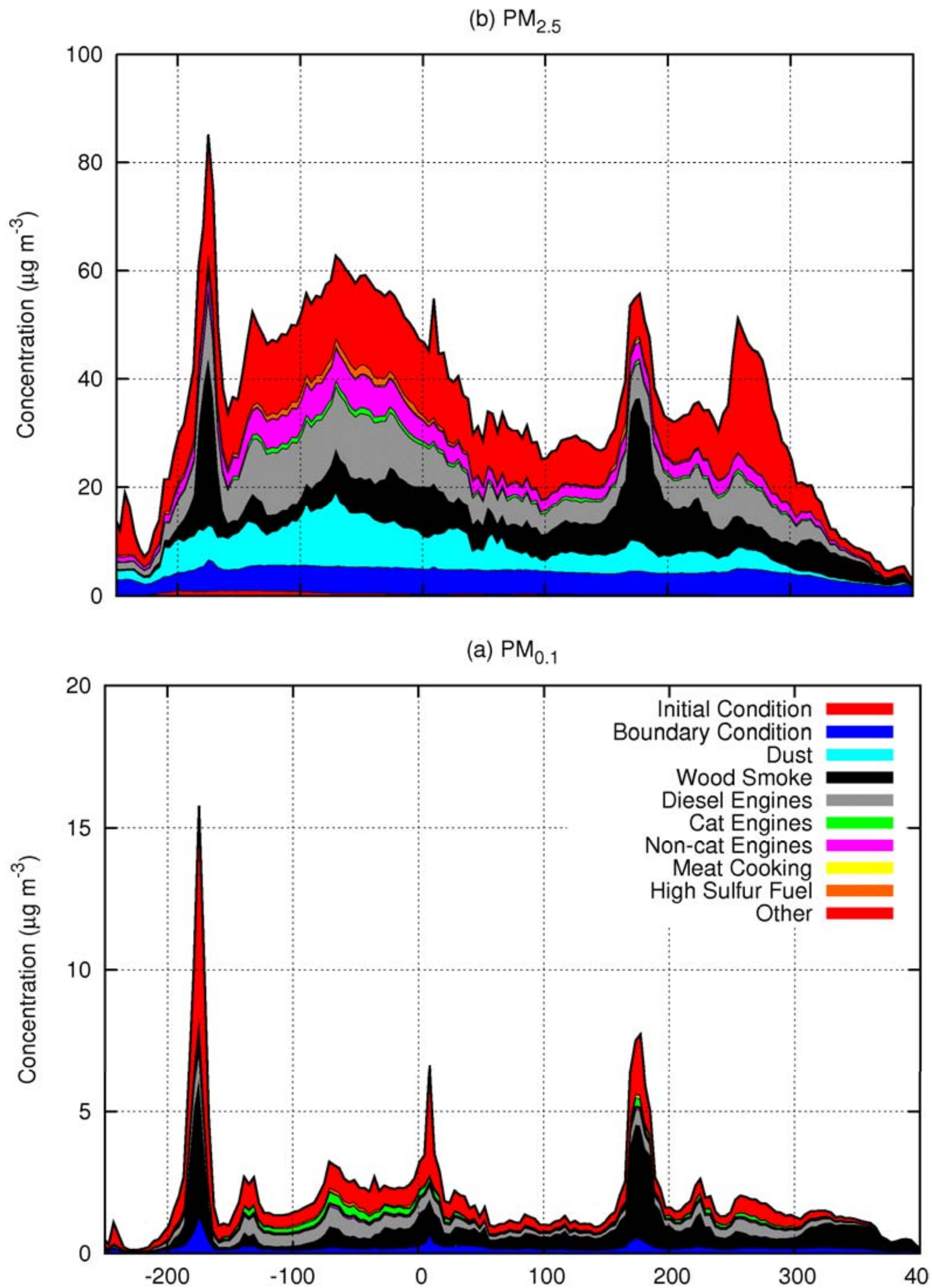


Source: 2005 Held T., Q. Ying, M.J. Kleeman, J.J. Schauer, M.P. Fraser. A comparison of the UCD/CIT air quality model and the CMB source-receptor model for primary airborne particulate matter. Atmospheric Environment. 39: 2281-2297.

Transect of PM Concentrations Between Sacramento and Bakersfield Dec 2000 – Jan 2001



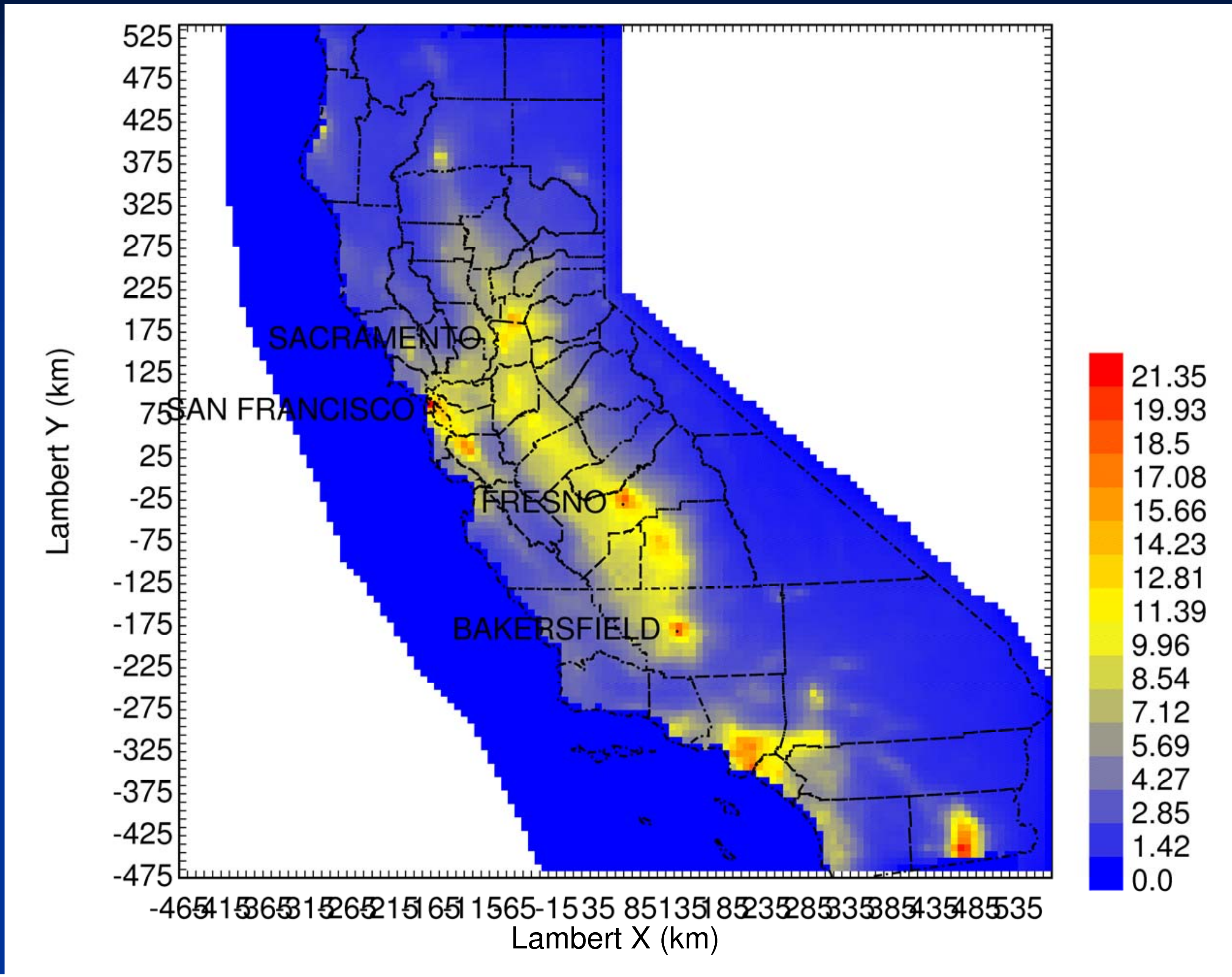
Source: 2008 Ying, Q. Lu J., Kaduwela, A. and Kleeman, M.J. Modeling Air Quality during the California Regional PM10/PM2.5 Air Quality Study (CPRAQS) using the UCD/CIT Source Oriented Air Quality Model - Part III. Regional Source Apportionment of Secondary and Total Airborne PM2.5 and PM0.1. Atmospheric Environment, accepted for publication.



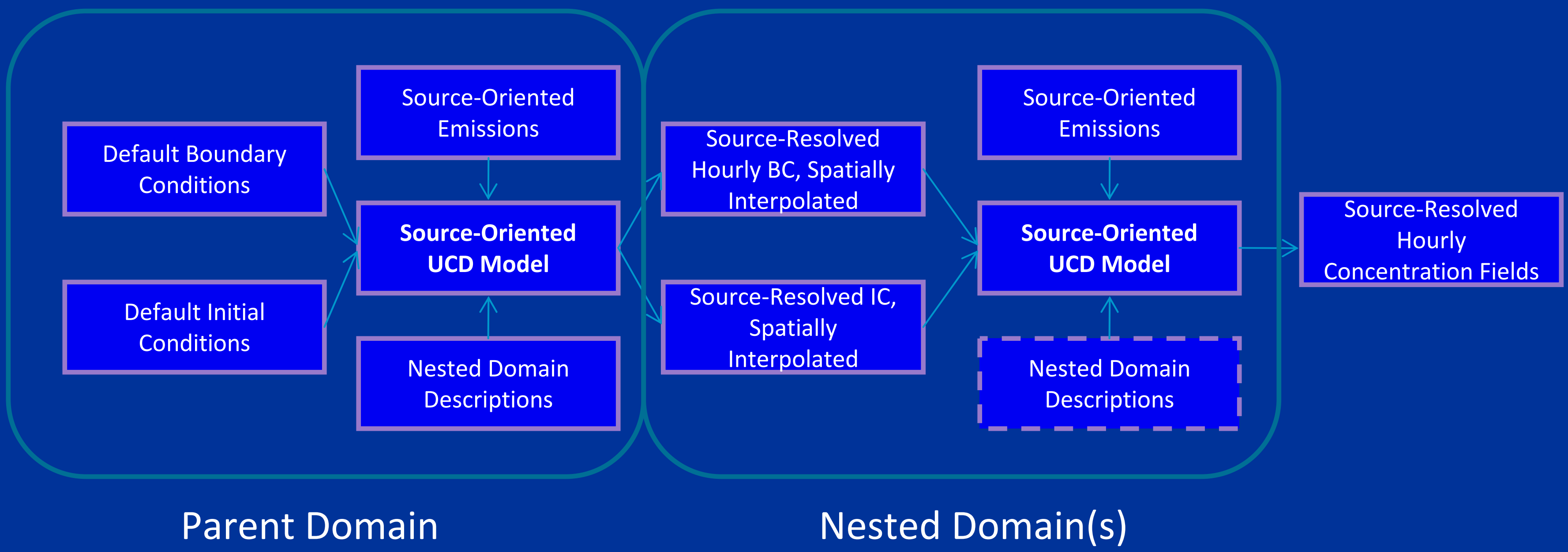
Computational Challenges Associated with Seven Years of Simulated Air Quality

- Meteorology simulations using WRF
  - 3 months of run time using 640 cores
  - 6 TB of output data
- Air Quality simulations using UCD+CMAQ
  - 5 months of run time using 1200 cores
  - 25 TB of output data
- All data will be available for download at conclusion of the project

Example: PM2.5 Averaged Between 2000-2006

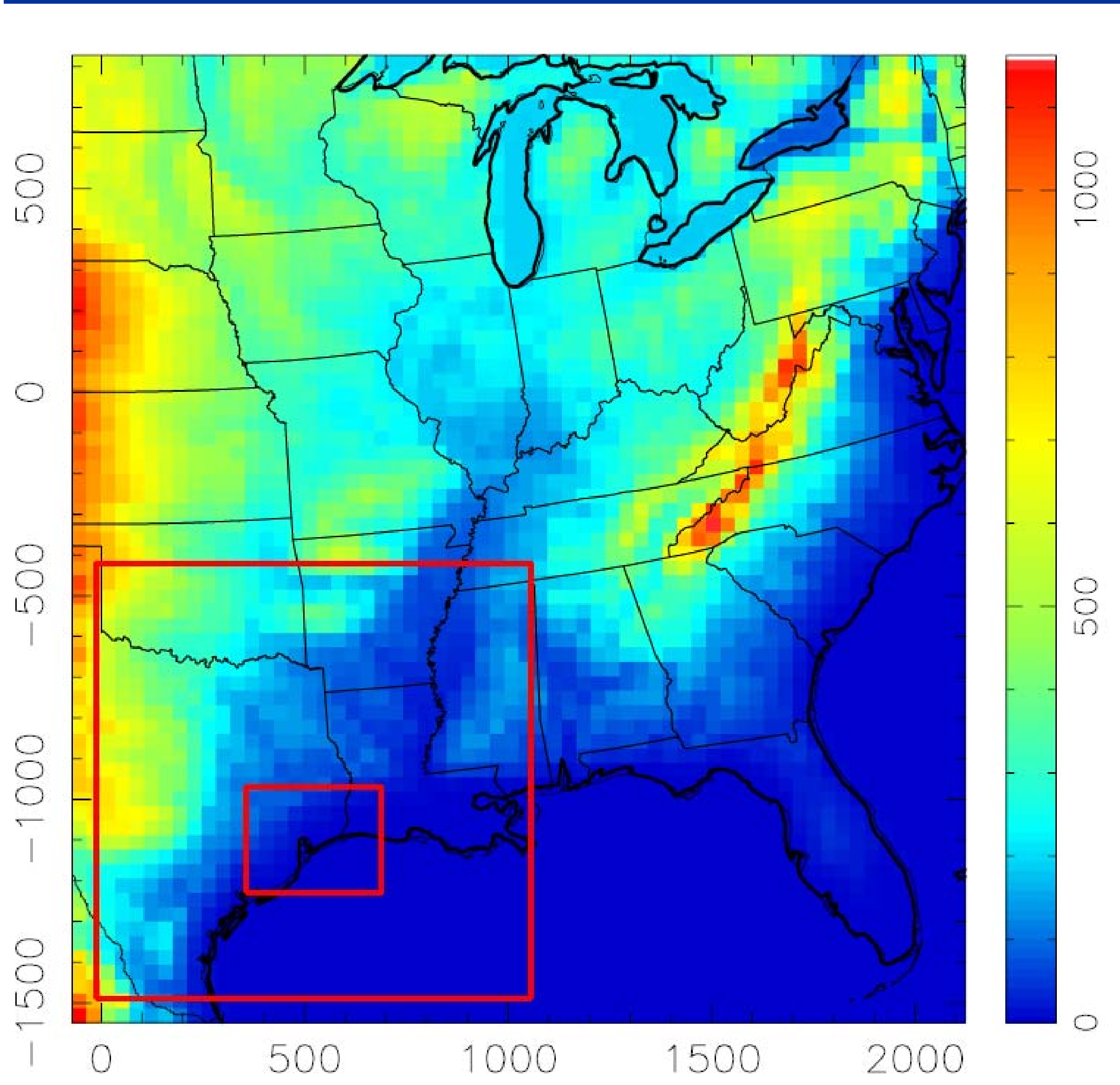


Texas A&M: One-Way Nesting in UCD Source-Oriented Air Quality Model



-Allow unlimited number of nested domains within a parent domain
-Allow multiple layers of nested domains

Preliminary Testing – TexAQS 2000



- 36 km East US
- 12 km East Texas
- 4 km Southeast Texas

- August 16, 2000 to September 7, 2000

Summary of Epidemiological Studies

- How do you find 50,000 deaths in a population of 300,000,000?
  - 50,000/300,000,000 = 1/6,000 (doesn’t consider sensitive populations)
- MESA – Cohort Study
  - 6,500 participants in Los Angeles CA, St. Paul MN, Chicago IL, New York City NY, Baltimore MD, and Winston-Salem NC
  - CIMT baseline evaluation in 2000-02
- CTS – Cohort Study
  - 133,000 current and former female public school employees in California
  - subjects enrolled in 1995, with mortality and hospital discharge data updated annually
- WHI – Cohort Study
  - 90,000 women from 45 cities in the continental U.S
  - initial evaluation between 1994-1998
  - annual updates for cardiovascular incidents and altered risk factors
- CALFINE – time series study of deaths in 9 California counties
  - Address for deaths 1999-2001
  - Zip code for deaths 2002-2005

Results from CALFINE Time Series Study: Respiratory Hospitalization and Components of Fine Particles

- Using time series analysis of acute exposures, we examined:
  - Hospital Admits for children age < 18 and < 5 for various respiratory diseases in six California counties from 2000 through 2003
  - Ambient concentrations of PM2.5 and several constituents, including EC, OC, NO3, SO4, SI, K and Zn

- Results:
  - Associations were observed between several components of PM2.5 and hospitalization for all of the respiratory outcomes examined.
  - For example, for total respiratory admissions for children < 5, exposure to the interquartile range of EC, OC and NO3 had an excess risk of:
    - EC: 4.7% (95% CI = 0.3, 9.3)
    - OC: 3.0% (95% CI = 0.4, 5.8)
    - Nitrates: 3.2% (95% CI = 0.5, 6.0)

- Conclusion: Components of PM2.5 were associated with hospitalization for several childhood respiratory diseases including pneumonia, bronchitis and asthma. (source: Ostro et al., EHP, 2009)

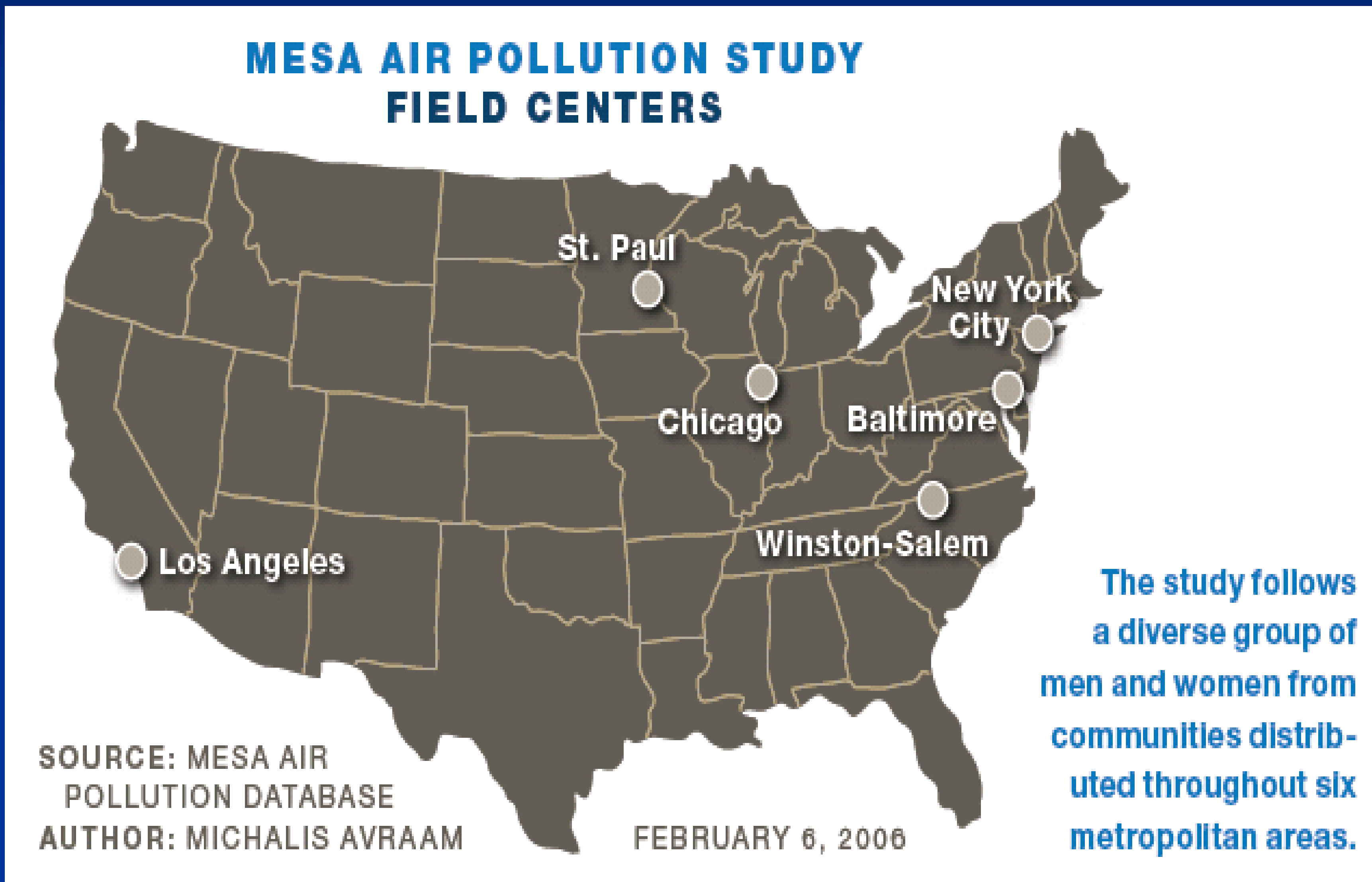
Results from California Teachers Cohort Study: Hazard ratios per 10 µg/m³ increment of PM2.5 and PM10

Model	# in analysis/ # events	PM2.5 1999-2002		# in analysis/ # events	PM10 1995-2002	
		HR	95% CI		HR	95% CI
All-cause mortality	89,962/3,056	1.19	(1.11, 1.29)	68,957/3,525	0.99	(0.95, 1.02)
Cardiopulmonary mortality	89,962/1,526	1.28	(1.15, 1.42)	68,957/1,739	1.00	(0.95, 1.05)
MI incidence[&]	88,916/1,224	1.28	(1.14, 1.45)	68,477/1,460	1.02	(0.97, 1.07)
Stroke incidence[&]	89,314/865	1.33	(1.15, 1.53)	68,671/1,040	1.02	(0.96, 1.08)

* All hazard ratios adjusted for smoking status, total pack years, BMI, marital status, alcohol consumption, second-hand smoke exposure at home, dietary fat, dietary fiber, dietary calories, physical activity, menopausal status, hormone replacement therapy use; and contextual variables (income, income inequality, education, population size, racial composition, unemployment).

[&]Includes both mortality and hospitalization

MESA Cohort Study



MESA Air Quality Monitoring for $PM_{2.5}$

- AQS/EPA fixed monitors
  - hourly, daily or every third day observations
- MESA Air fixed monitors
  - 2-week averages
- Home outdoor monitoring
  - rotating sets of 4 sites, each with two 2-week averages over 50 2-wk periods
  - total of at least 50 sites each monitored in two different seasons
- Speciated $PM_{2.5}$ at MESA Air fixed and home sites supplementing AQS STN sites

Monitoring Data Structure for $PM_{2.5}$

(2-week time scale)

Time (T=50)

Spatial Locations ($N \approx 125$)

Spatial Locations	Site	1	2	3	4	5	6	7	...	24	25	26	27	28	29	30	31	32	33	...	45	46	47	48	49	50
Fixed (EPA) (number varies by location)	1	X	X	X	X	X	X	X	...	X	X	X	X	X	X	X	X	X	X	...	X	X	X	X	X	X
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	20	X	X	X	X	X	X	X	...	X	X	X	X	X	X	X	X	X	X	...	X	X	X	X	X	X
Fixed (MESA) (5 sites)	1	X	X	X	X	X	X	X	...	X	X	X	X	X	X	X	X	X	X	...	X	X	X	X	X	X
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	5	X	X	X	X	X	X	X	...	X	X	X	X	X	X	X	X	X	X	...	X	X	X	X	X	X
Home Outdoor (100 sites)	1	X										X														
	2	X										X														
	3	X										X														
	4	X										X														
	5		X										X													
	6		X										X													
	7		X										X													
	8		X										X													
	...								...											...						
	97										X															X
	98										X															X
	99										X															X
	100										X															X

University of Washington: Sub 4 km Spatio-Temporal Model of Ambient Concentration

- Johan Lindström, Adam Szpiro, Lianne Sheppard
- Goal of sub-grid model
  - Predict relevant functions of outdoor concentration throughout areas where participants live (and work, etc)
  - Incorporate information from multiple time scales and spatial locations
- Inputs to sub-grid model
  - Geographic Information System predictors and coords
    - Spatial location
    - Road network & traffic calculations
    - Population density
    - Other point source and/or land use information
  - Monitoring data
    - Air monitoring from existing EPA/AQS network
    - Air monitoring from MESA Air data collection
    - Meteorological information
  - UCD/CIT 4 km grid model predictions

Our model

$$Y(s,t) = \mu(s,t) + \nu(s,t)$$
$$\mu(s,t) = \sum_{i=0}^{m} \beta_i(s) f_i(t)$$
$$\beta_i(s) = \sum_{j=1}^{m_i} X_{i,j}(s)\alpha_{i,j} + \varepsilon_i(s)$$

where

$$\nu(s) \in N(0, \Sigma_\nu)$$
$$\nu(s_1,t_1) \perp \nu(s_2,t_2) \quad t_1 \neq t_2$$
$$\varepsilon_i(s) \in N(0, \Sigma_i)$$

Adding model predictions as covariates

Option 1:

$$\mu(s,t) = \alpha_M M(s,t) + \sum_{i=0}^{m} \beta_i(s) f_i(t)$$

- One (few) additional parameters.
- Assumes a simple multiplicative bias and spatio-temporally varying additive bias.

Option 2:

$$\mu(s,t) = \beta_M(s) M(s,t) + \sum_{i=0}^{m} \beta_i(s) f_i(t)$$

$$\beta_M(s) = \sum_{j} X_{M,j}(s) \alpha_{M,j} + \varepsilon_M(s)$$

- Assumes that the multiplicative bias depends on location.
- Several additional parameters.
- Identifiability ???

Summary: Goals For Grid Models Applied in Epidemiology Studies

- Fill in spatial data between measurement stations using all known information about emissions, meteorology, and chemical reactions
- Fill in time data for 1-in-3 or 1-in-6 sampling days
- Provide a full description of gas species
- Provide a full description of PM species
- Provide a full description of particle size distributions
- Provide a full description of particle sources

Summary: Limitations to Overcome for Grid Models Applied in Epidemiological Studies

- 4km spatial resolution
- Results agree better with measurements at longer averaging times of ~1 week or more
- Computationally expensive to run for long cohort studies
- Requires help from atmospheric scientists to generate and evaluate the predictions