Enhancing EPHT with Satellite-Driven $PM_{2.5}$ Exposure Modeling and Epidemiology – Year 1 Report

Yang Liu, Ph.D.
Rollins School of Public Health
Emory University

NASA Public Health Review Meeting
San Antonio, TX
September 27-29, 2010

Project Team

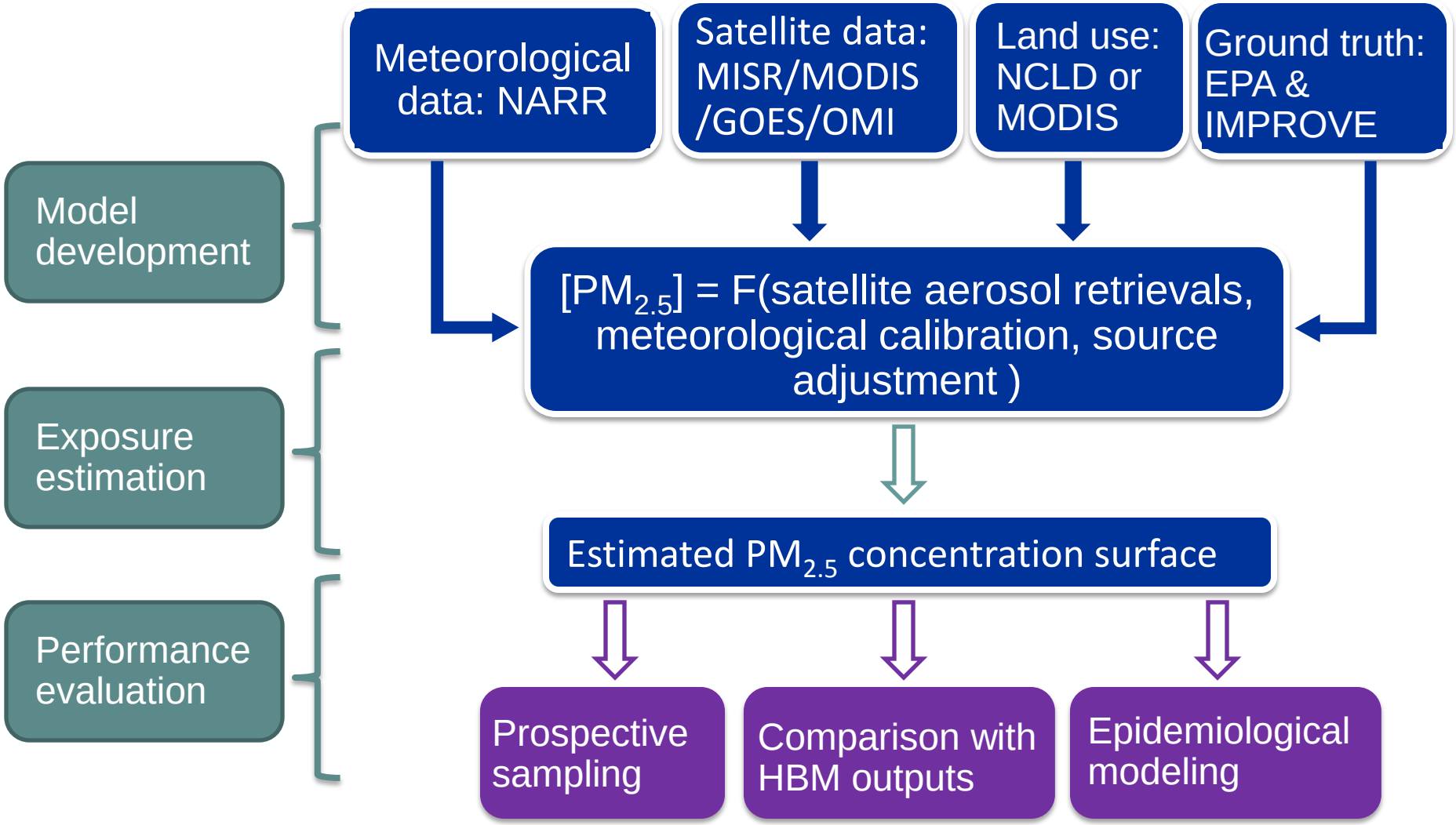
- Emory/RSPH: Yang Liu, Jeremy Sarnat, and Mitch Klein
- MSFC/USRA: Dale Quattrochi, Bill Crosson, Mohammad Al-Hamdan, Maury Estes, Sue Estes, Sarah Hemmings, Gina Wade
- CDC/NCEH: Judy Qualters, Paul Garbe, Helen Flowers, and Ambarish Vaidyanathan

Sponsored by NASA Applied Science Program grant NNX09AT52G under J. Haynes

Research Objectives

- Extend the spatial coverage of the $PM_{2.5}$ indicators in Tracking Network with satellite data
- Provide timely estimates of county average $PM_{2.5}$ health indicators
- Evaluate satellite $PM_{2.5}$ estimates as a alternative exposure data source in environmental epidemiologic studies and using independent ground sampling

Technical Approach

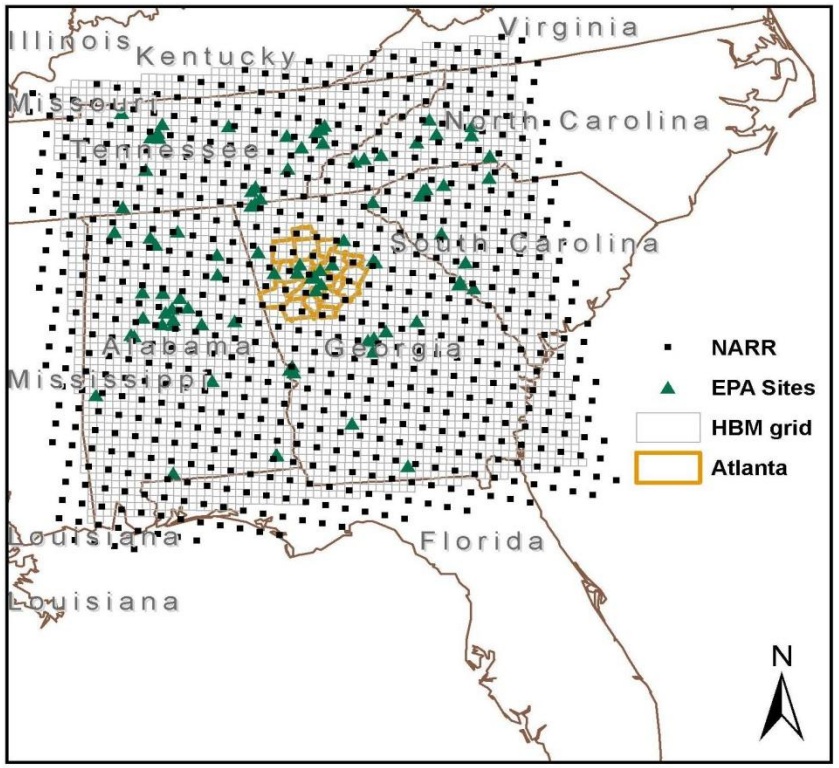


Year 1 Milestones

1. Design of modeling grid
2. Aerosol data collection and processing
3. MODIS cloud cover data processing
4. NLCD/NDVI land cover / land use data analysis
5. HBM data preparation
6. Meteorological data collection and processing
7. Initial data integration

Status summary: we have completed all proposed tasks on time, and done some extra work.

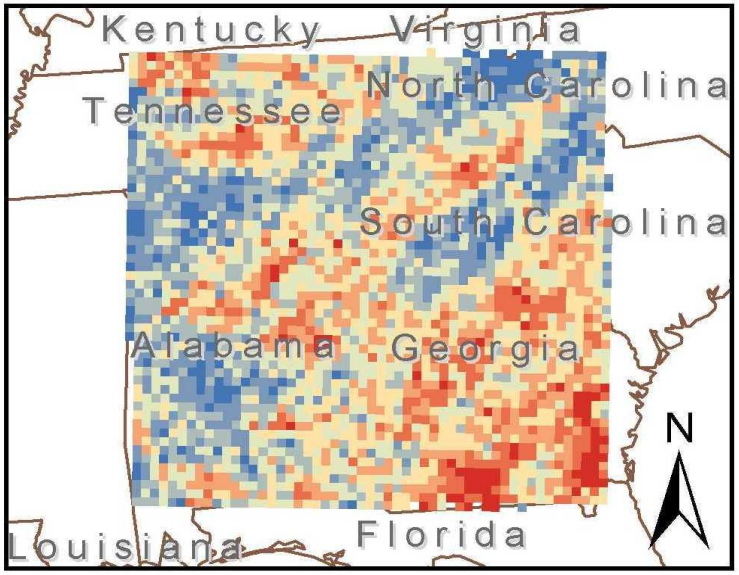
1. Study Domain Design



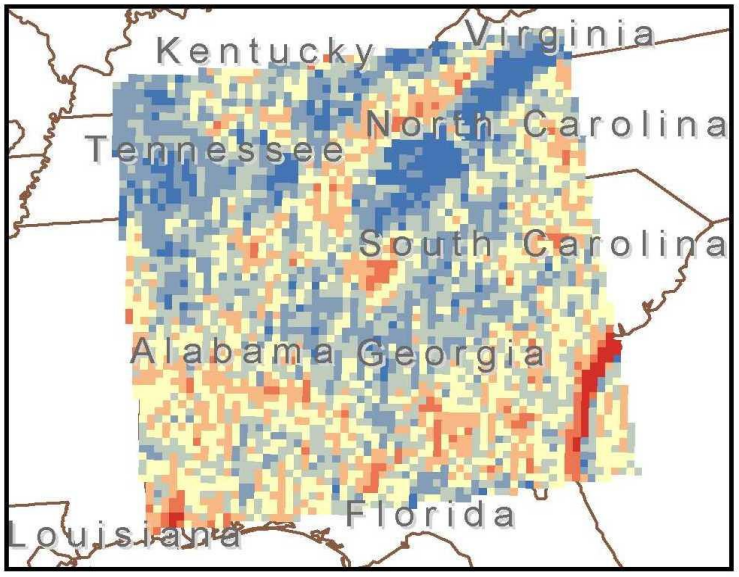
- Health study domain: 20-county Metro Atlanta area (150 x 150 km^2)
- Exposure modeling domain: 600 x 600 km^2
- Switched from 10 km grid compatible with MODIS data to standard 12 km CMAQ grid compatible with CDC HBM data

2. Aerosol Data Processing

- MOD04 and MYD04, Collection 5, 2001- 2007
- OMI level 2 aerosol index and type, 2004 - 2007
- MISR MIL2ASAE, v22, 2001 – 2007
- GOES GASP aerosol data, 2001 – 2007



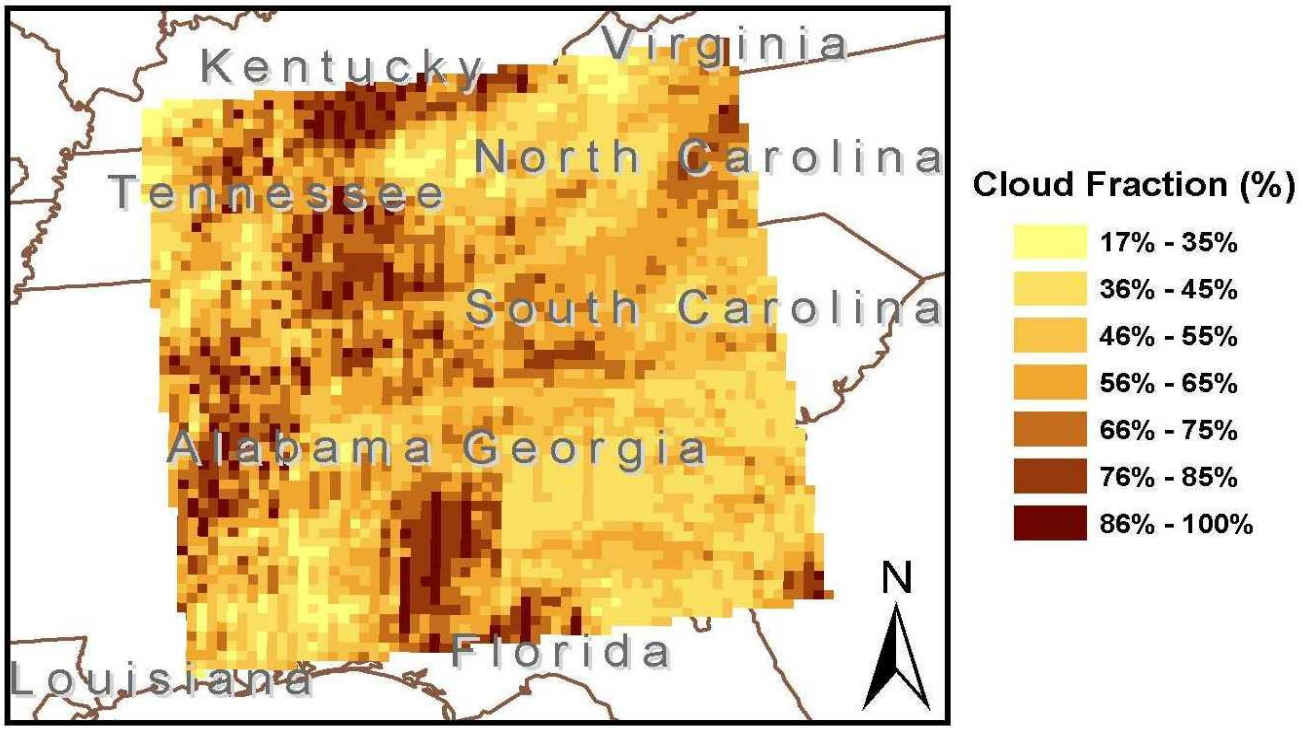
2001 Annual Mean GOES AOD



2001 Annual Mean MODIS AOD

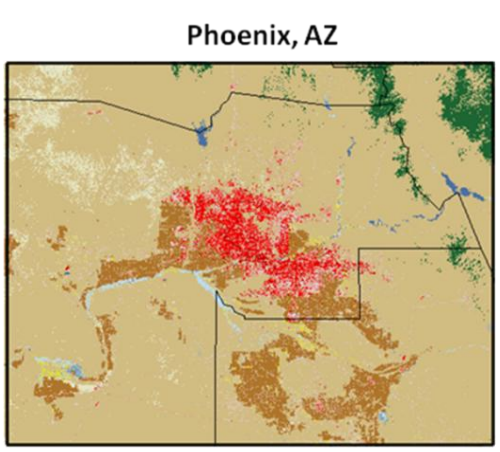
3. MODIS Cloud Data Processing

- MYD06 5 km resolution cloud fraction, 2002 – 2007
- Calculated mean cloud cover in each 12 km grid cell

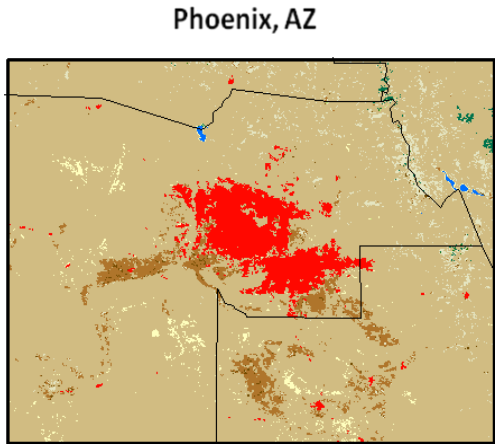


4. Land Use Data Analysis (1)

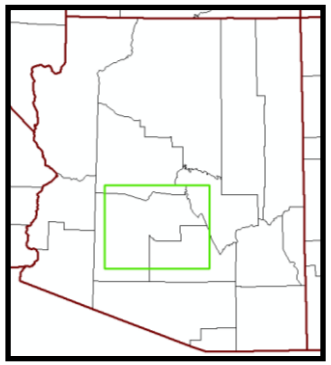
- NLCD 2001, 30 m resolution, static
- MODIS land data, MOD12Q1, dynamic



2001 NLCD (30 m)



2006 MODIS LCLU (500 m)

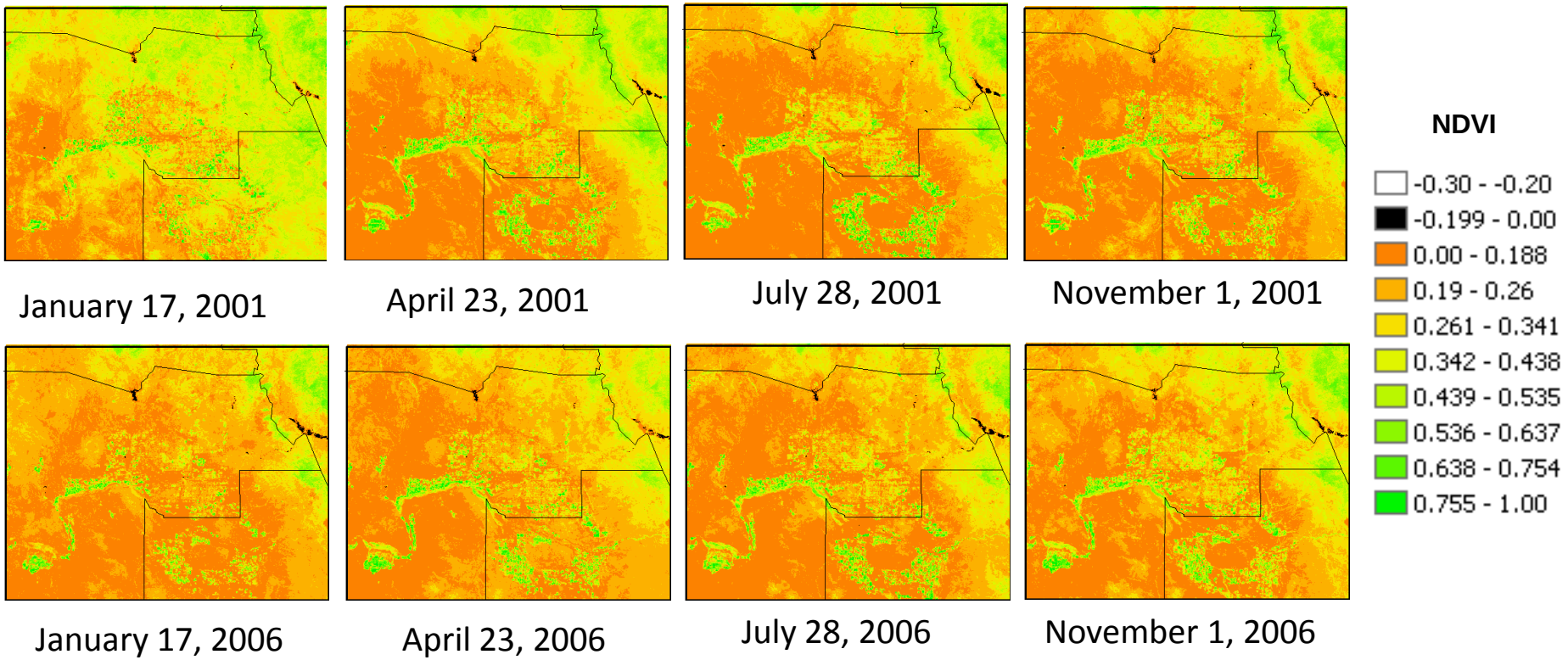


Differences in spatial patterns are visible but highly comparable.

Conclusion: use NLCD for its more detailed LU classes.

4. Land Use Data Analysis (2)

- MODIS NDVI, MOD13A1, 16 day, 500 m

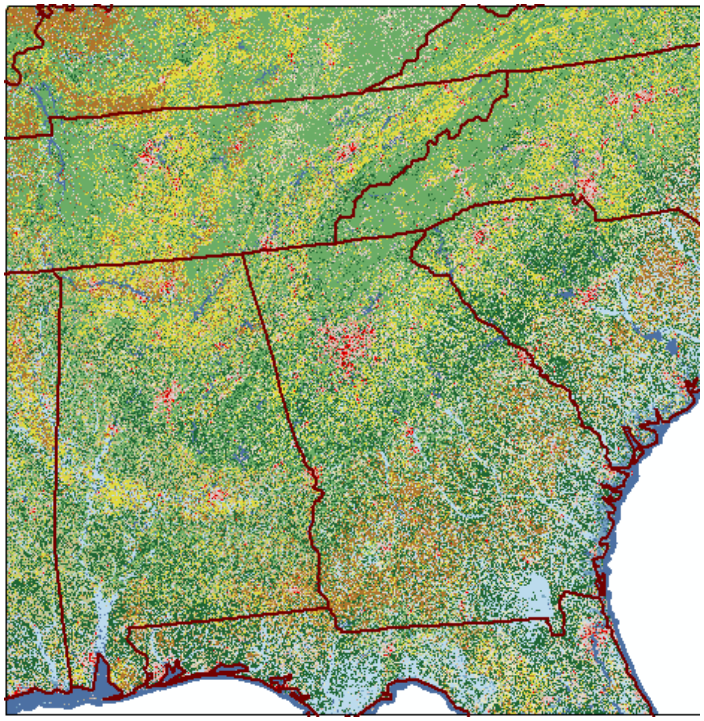


Both seasonal and inter-annual variations exist. Potential indicator of OC emissions. Withheld for revised model development in year 3

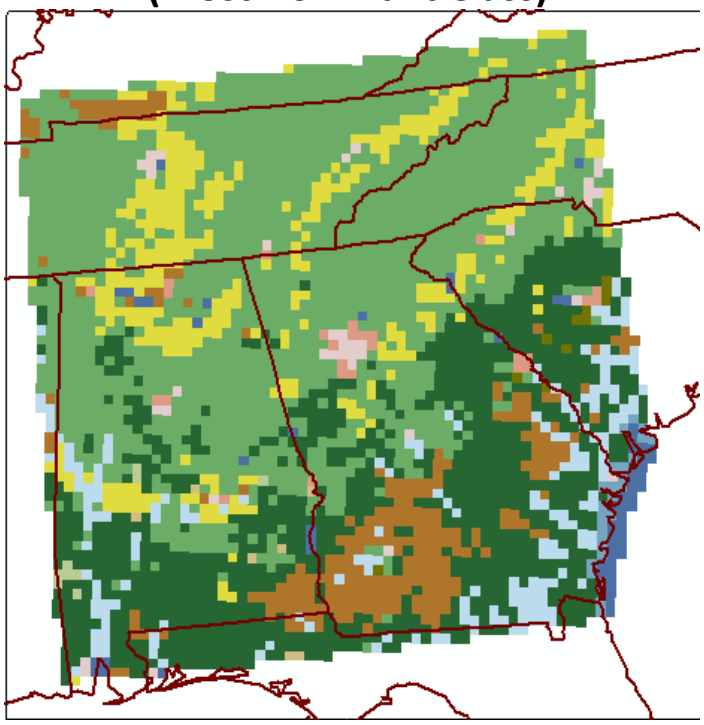
4. Land Use Data Analysis (3)

Resampling of NLCD data: identify dominant land use class in each 12 km grid cell

NLCD2001 (30 m)



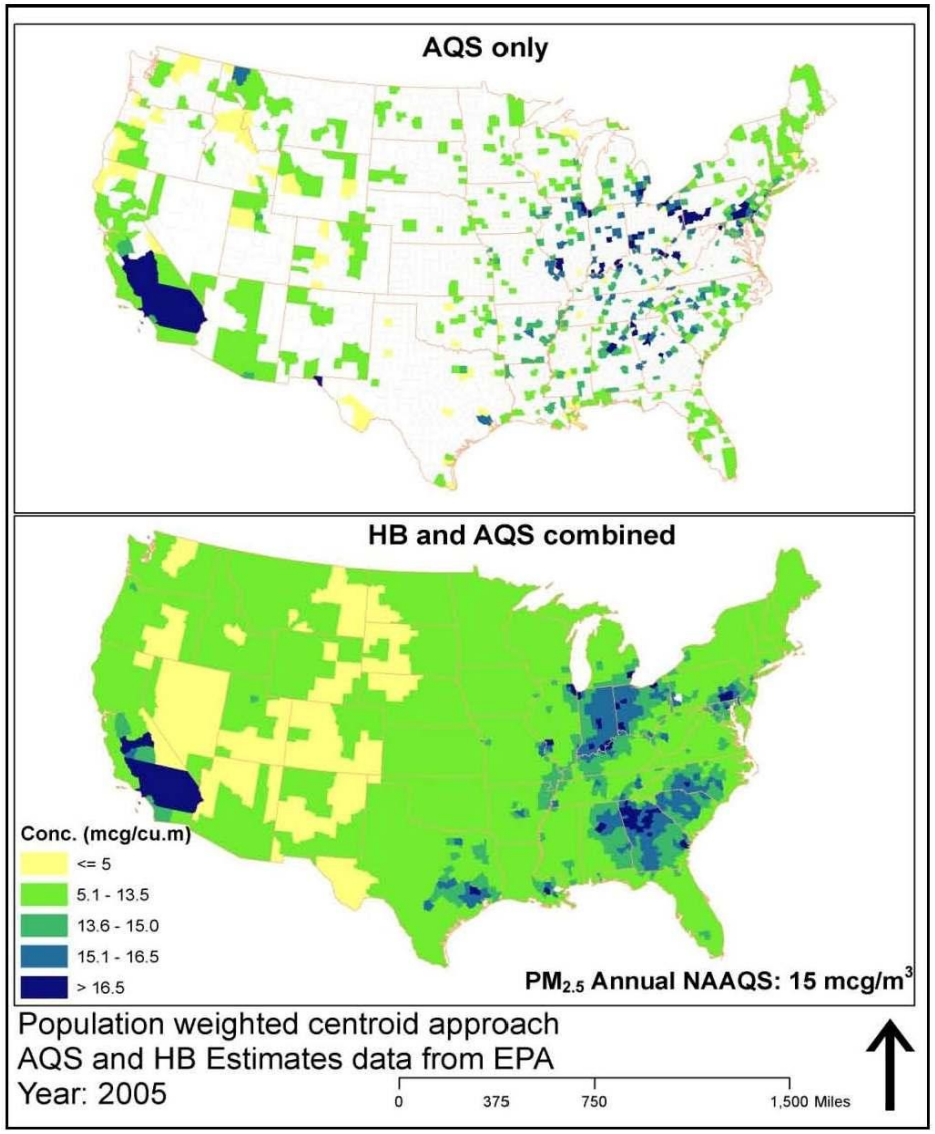
Resampled to the 12-km Grid (Most Dominant Class)



- Open Water
- Developed, Open Space
- Developed, Low Intensity
- Developed, Medium Intensity
- Developed, High Intensity
- Barren Land (Rock/Sand/Clay)
- Deciduous Forest
- Evergreen Forest
- Mixed Forest
- Shrub/Scrub
- Grassland/Herbaceous
- Pasture/Hay
- Cultivated Crops
- Woody Wetlands
- Emergent Herbaceous Wetlands

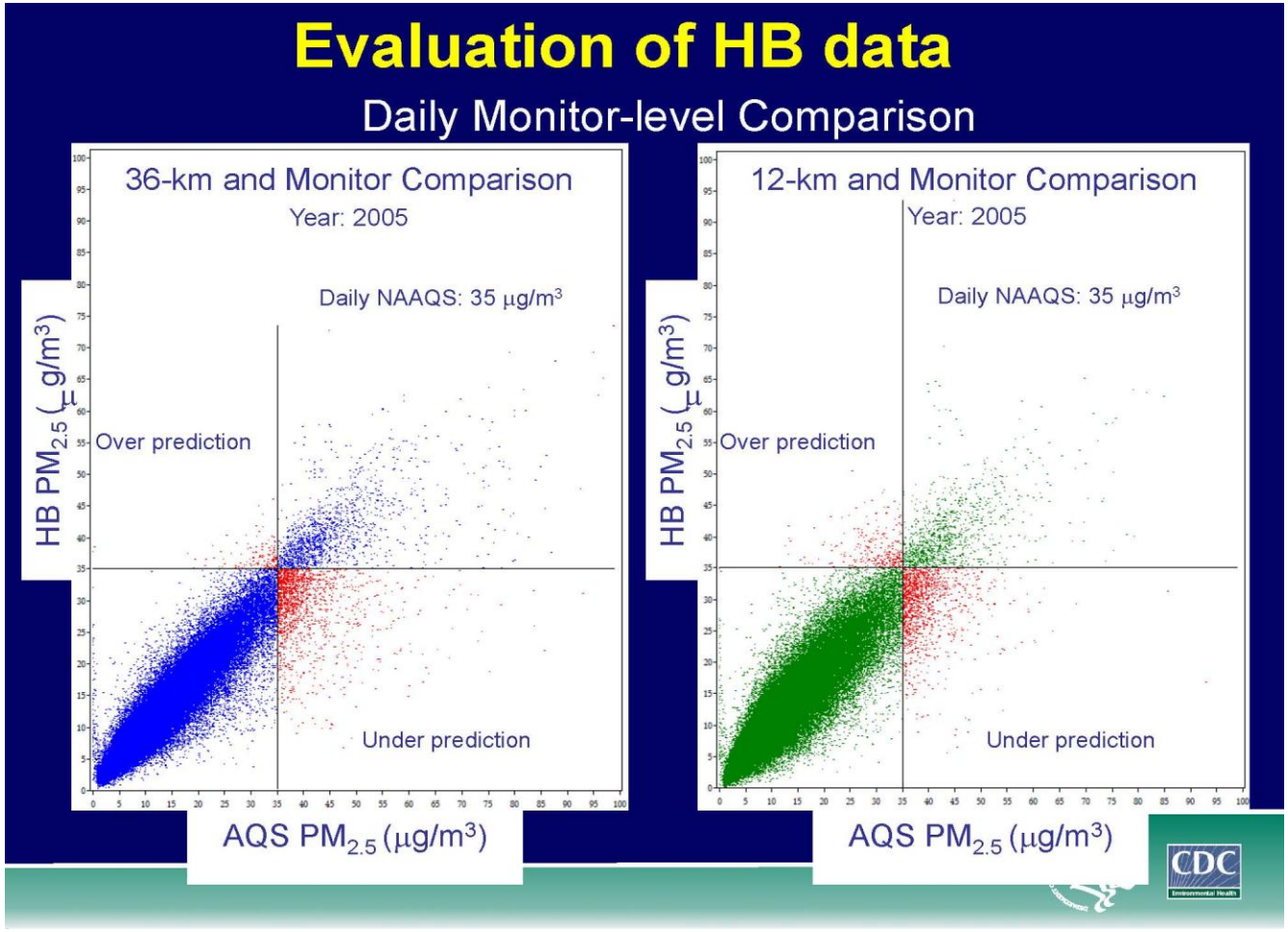
Potential for factor analysis will be evaluated later.

5. HBM Data Analysis (1)



- EPA generated HBM $PM_{2.5}$ daily concentrations, 2001 – 2006
- Raw data at 12 km resolution (east), reprojected to county level using population weighted centroids.

5. HBM Data Analysis (2)



Overall, HBM $PM_{2.5}$ compares well with EPA observations. Underestimation occurs at high concentrations.

6. Meteorological Data Analysis

- NARR data, 32 km, 2001 – 2007
- Mapped to 12 km CMAQ grid using nearest neighborhood method
- Parameters of interest: humidity, temperature, wind speed and direction, PBL height

7. Initial Data Integration

MODIS/MISR/GOES, OMI, NLCD, NARR, and census data have all been linked to the CMAQ 12 km grid ready for exposure model development.

Year 1 Presentations (1)

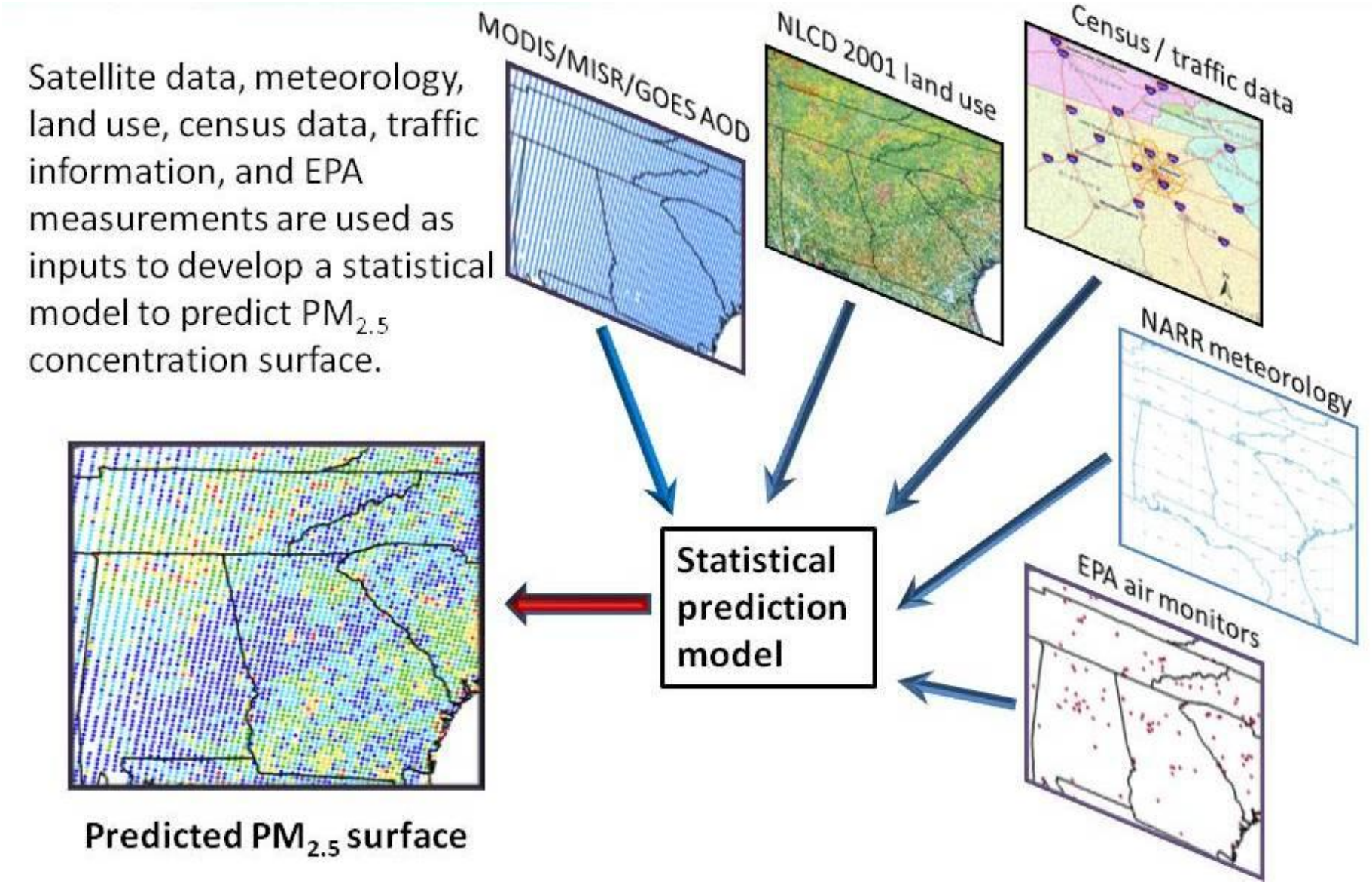
- W. Crosson, et al., Examining the use of satellite aerosol remote sensing as a potential means to extend the coverage of the CDC NEPHT Network, **American Thoracic Society International Conference,** New Orleans, LA, May 19, 2010.
- Y. Liu, et al., Enhancing Environmental Public Health Tracking With Satellite-driven Particle Exposure Modeling And Epidemiology, **American Meteorological Society Annual Meeting,** Atlanta, GA, January 19, 2010.
- Y. Liu, Applications of Satellite Remote Sensing Data in Air Pollution and Public Health Research. **Tsinghua University**, October 15, and **Institute of Remote Sensing Applications, Chinese Academy of Sciences**, October 18, 2009, Beijing, China.

Year 1 Presentations (2)

- We created a project factsheet for the NASA's Earth Day display in March 2010.
- We contributed to the Applied Science Accomplishments presentation to the NASA Headquarter in April 2010.
- Our project was mentioned in a NASA blog on April 23, 2010. http://nasa-satellites.blogspot.com/2010/04/nasa-satellite-data-helps-everyone.html

Year 2 Plan

Satellite data, meteorology, land use, census data, traffic information, and EPA measurements are used as inputs to develop a statistical model to predict $PM_{2.5}$ concentration surface.



Year 2 Tasks and Timeline

- Emory
  - MODIS/GOES validation (Mar 2011)
  - Preliminary $PM_{2.5}$ modeling (Mar 2011)
  - Final $PM_{2.5}$ modeling (Sep 2011)
  - Pilot sampling (Mar 2011)
  - Prospective sampling and sample analysis (Mar 2012)
- MSFC
  - Revision of gridded aerosol data (Apr 2011)
  - Final data integration (aerosol, cloud, weather, land use, etc.) (Jul 2011)
- CDC
  - Comparison between HBM and satellite (Sep 2011)

Proposed Exposure Modeling - GAM

First stage GAM: temporal variables

$$Y_{(t,s)} \sim \mu_1 + f_t(t) + f_{AOD}(AOD_t) + f_{cloud}(cloud) + OMI + \sum_k f_{met_k}(met_k)$$

Second stage GAM: spatial variables

$$Y_s = \overline{Y_{t,s} - \hat{Y}_{t,s}} \sim \mu_2 + f_{AOD}(AOD_s) + f(population) + f_{x,y}(x,y) + f(land\ use)$$

Final prediction

$$[PM2.5]_{t,s} = \hat{Y}_{t,s} + \hat{Y}_s = \hat{Y}_{t,s} + \overline{Y_{t,s} - \hat{Y}_{t,s}}$$

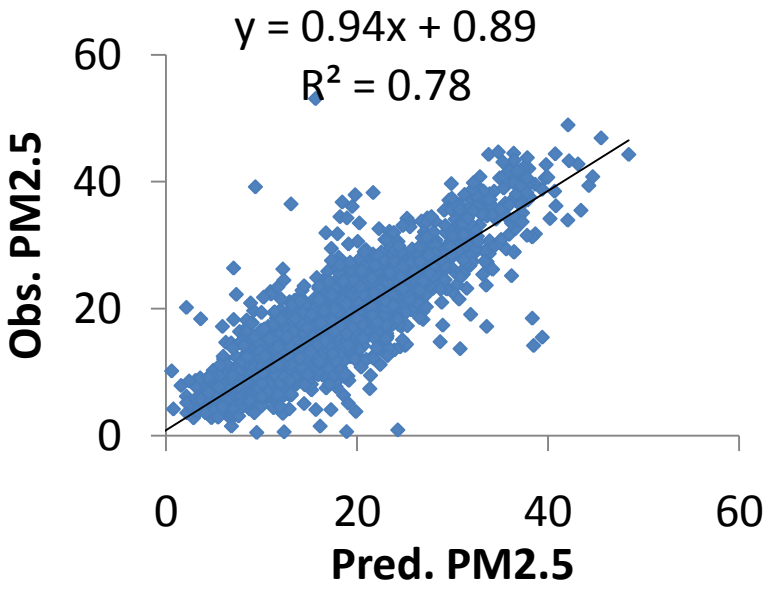
Additional Model - GWR

- Geographically Weighted Regression allows regression coefficients to vary in space

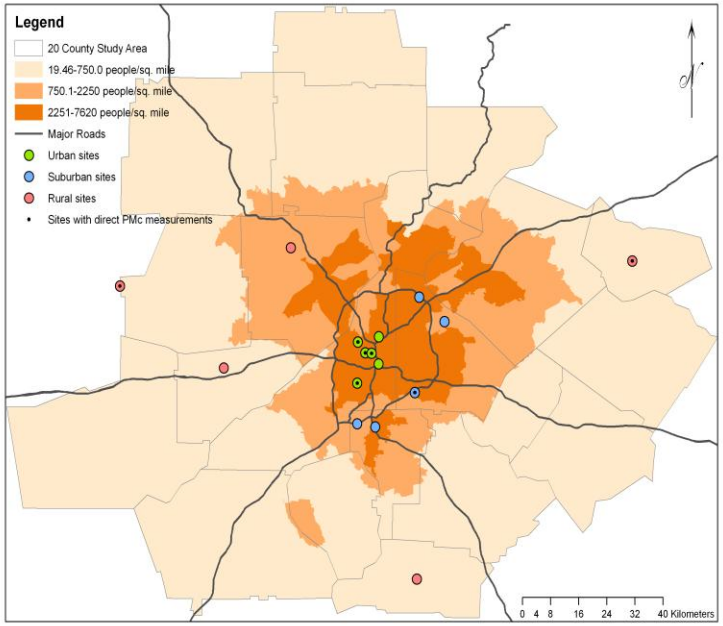
- $PM_{2.5}(g) \sim \beta_0(g) + \beta_1\ PBL(g) + \beta_2\ RH(g) + \beta_3\ T(g) + \beta_4\ Wind + \beta_5\ AOD(g) + \beta_6\ Forest\%(g)$

  $\beta(g) = (X^T W(g) X)^{-1} X^T W(g) Y$

	Model	CV
R^2	0.89	0.78
RMSE	2.5 μg/m³	3.6 μg/m³



Prospective Validation

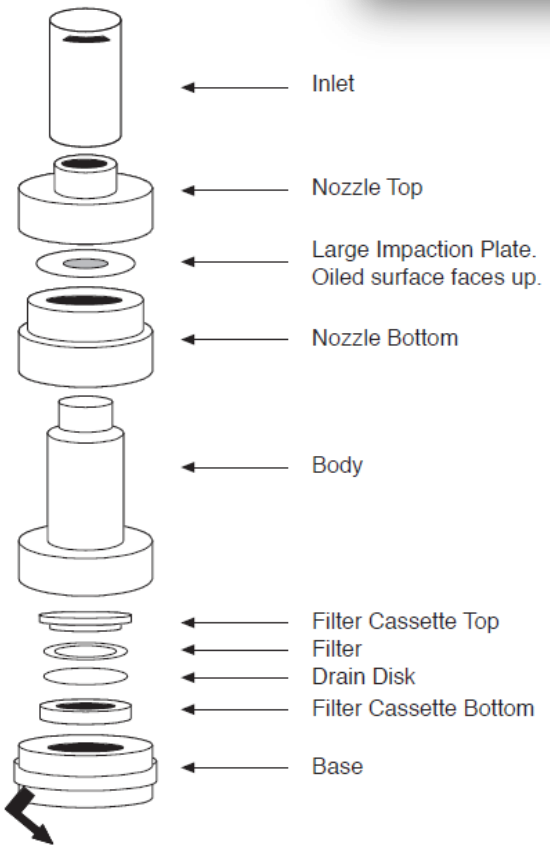


Existing network of PM monitors in 20-county Atlanta area

- Background: Areas in domain exist as heterogeneous spikes in $PM_{2.5}$ concentration due to local or point sources
- Objective: Predict hot spots and conduct ground level monitoring to verify.
- Results from the field sampling will be compared to the model predictions to assess the accuracy of the satellite-driven model.

Sampling Methods

- Harvard Impactors
  - 10 l/min
  - 2.5 μm cut points
  - 37mm Teflon filters
- Quantified via gravimetric analysis at Emory



Harvard Impactor diagram