Enhancing EPHT with Satellite-Driven $PM_{2.5}$ Exposure Modeling and Epidemiology – Year 2 Report

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

NASA Public Health Review Meeting
Santa Fe, NM
September 14-16, 2011

Project Team

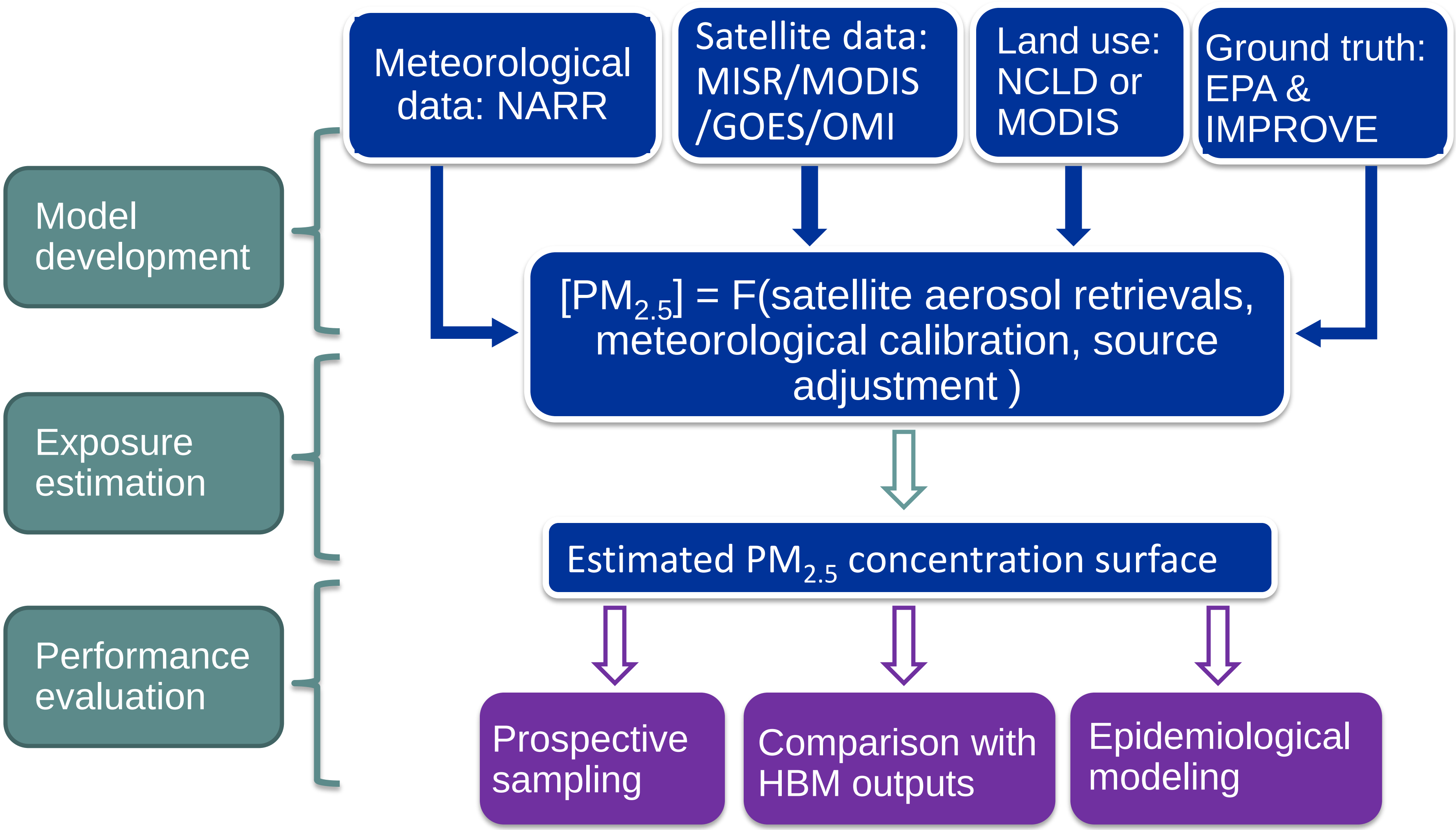
- Emory/RSPH: Yang Liu (PI), Jeremy Sarnat, Mitch Klein, Xuefei Hu, Heather Strosnider, and Erika Rees
- MSFC/USRA: Dale Quattrochi, Bill Crosson, Mohammad Al-Hamdan, Maury Estes, Sue Estes, Sarah Hemmings, and Gina Wade
- CDC/NCEH: Judy Qualters, Paul Garbe, Helen Flowers, and Ambarish Vaidyanathan

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

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

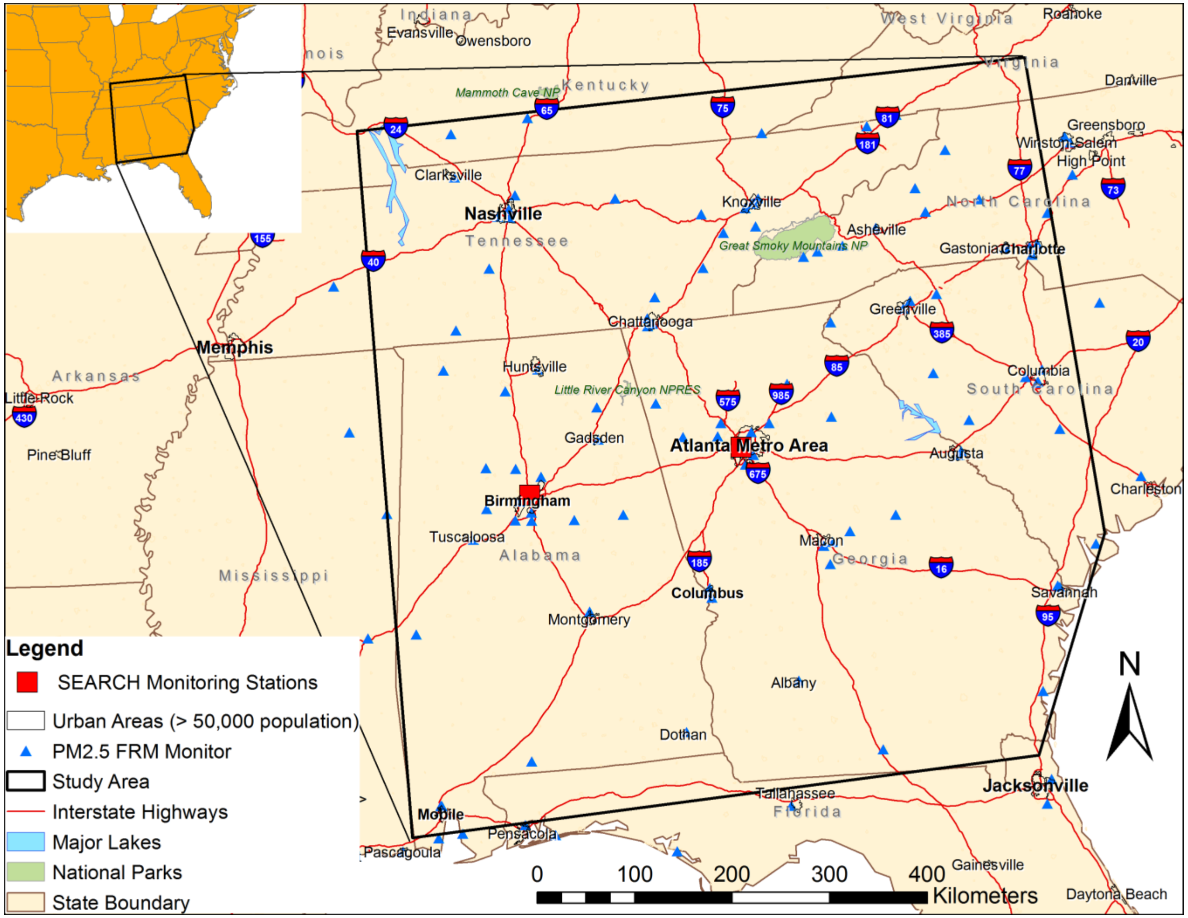
Technical Approach



Year 2 Progress Summary

- Proposed Tasks:
  1. Spatial model development and comparison of NARR and NLDAS (Emory, manuscript submitted)
  2. Initiation of prospective sampling (Emory)
  3. AOD calibration with AERONET (Emory, MSFC)
  4. Nearest neighbor approach development (MSFC)
- Ahead of Schedule:
  5. Initiation of epidemiological analysis (Emory)
- Need More Work
  6. Comparison with HBM

Study Domain



- Number of monitoring sites: 119
- Exposure modeling domain: 700 x 700 km^2
- SEARCH sites: 2 independent validation sites

Geographically Weighted Regression Model

GWR allows model parameters to vary in space to better capture spatially varying AOD-PM relationship – major advantage over global regression models.

Model Structure

$$[PM_{2.5}]_{(x,y)} \sim \beta_{0(x,y)} + \beta_{1(x,y)} \times AOD + \beta_{2\,(x,y)} \times PBL + \beta_{3(x,y)} \times RH + \beta_{4(x,y)} \times Temp + \beta_{5\,(x,y)} \times Wind\_Speed + \beta_{6\,(x,y)} \times Forest\_Cover$$

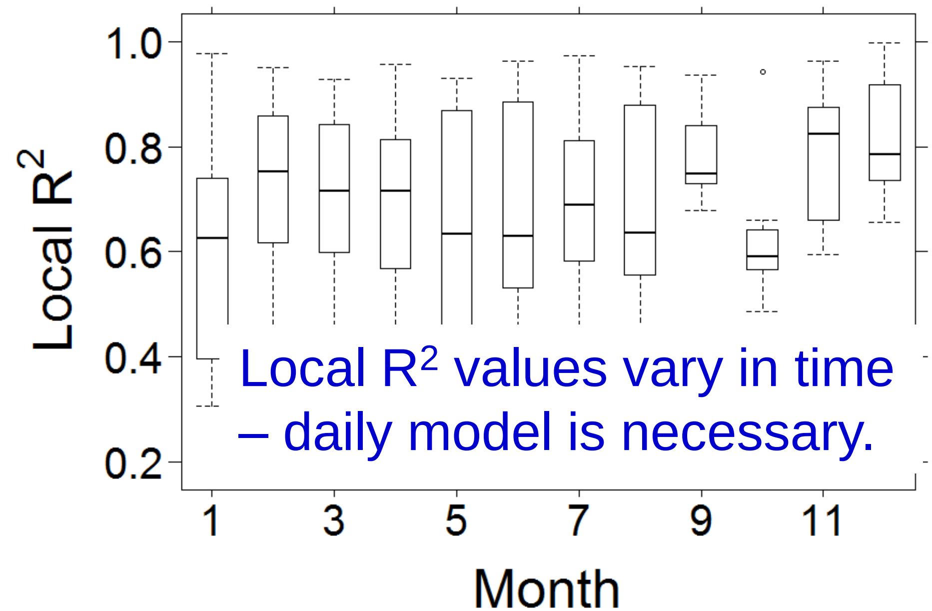
Datasets (2003):

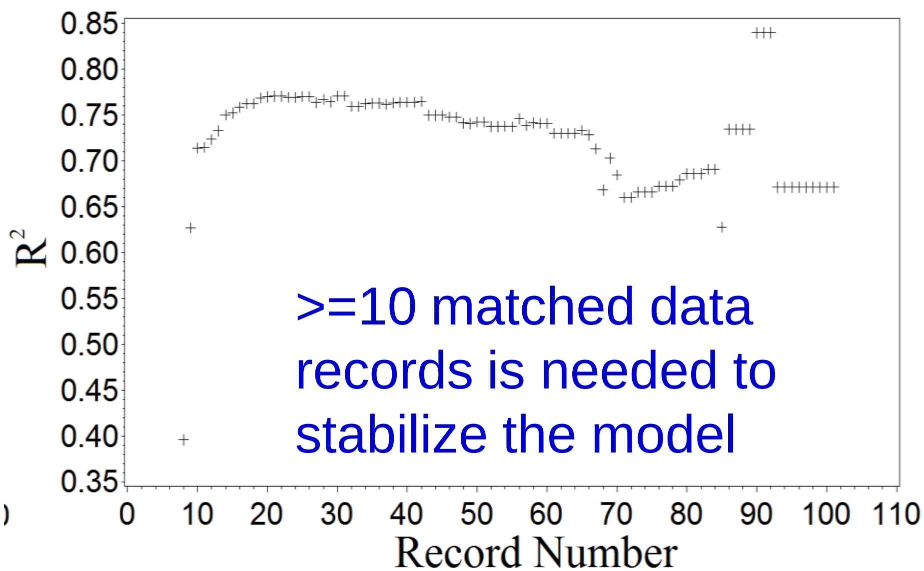
- $PM_{2.5}$ – EPA / IMPROVE daily measurements
- AOD – MODIS collection 5 (10 km) or GASP (4 km)
- Meteorology – NLDAS-2 (14 km) or NARR (32 km)
- Land use: NLCD 2001

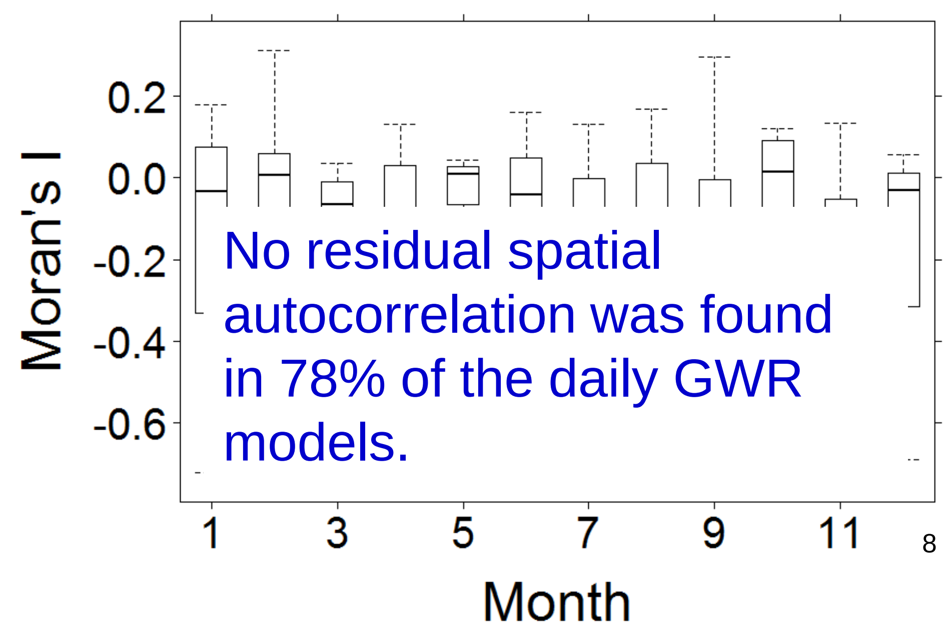
Model is fitted at daily level

Model Fitting Results

Max Obs. Per Day	101
Model Days	137 (37.5%)
Total Obs.	4,477







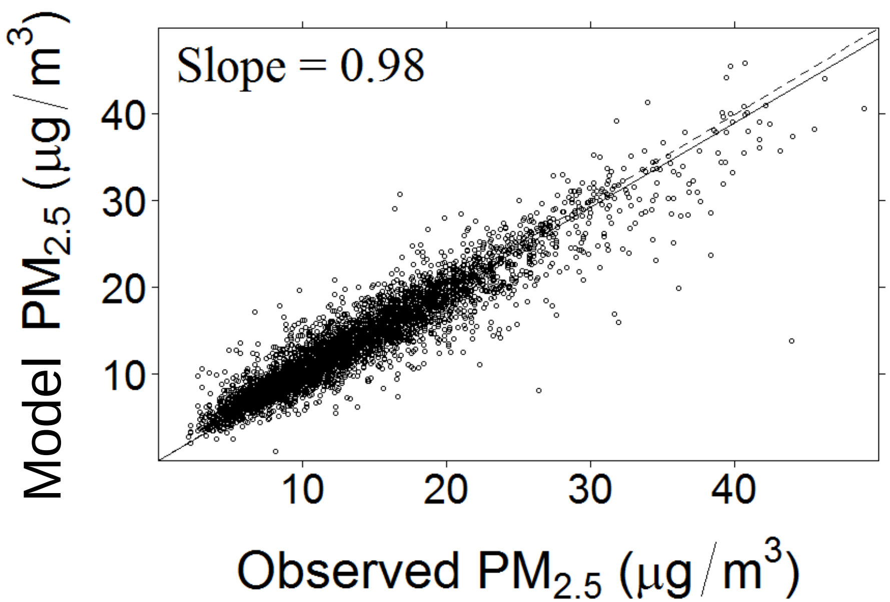
Model Performance Evaluation

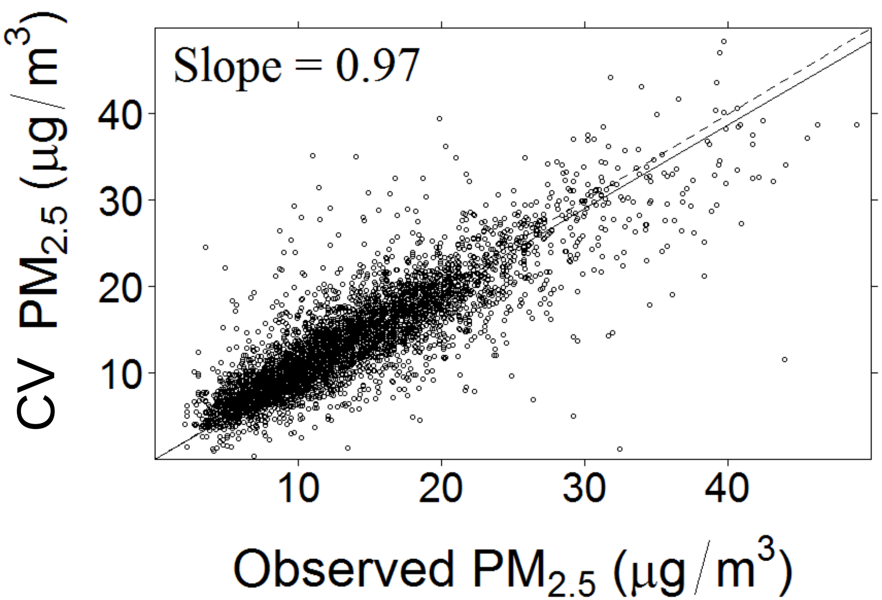
	Mean	Min	Max
Model R^2	0.86	0.56	0.92
CV R^2	0.70	0.22	0.85

SEARCH site predictions

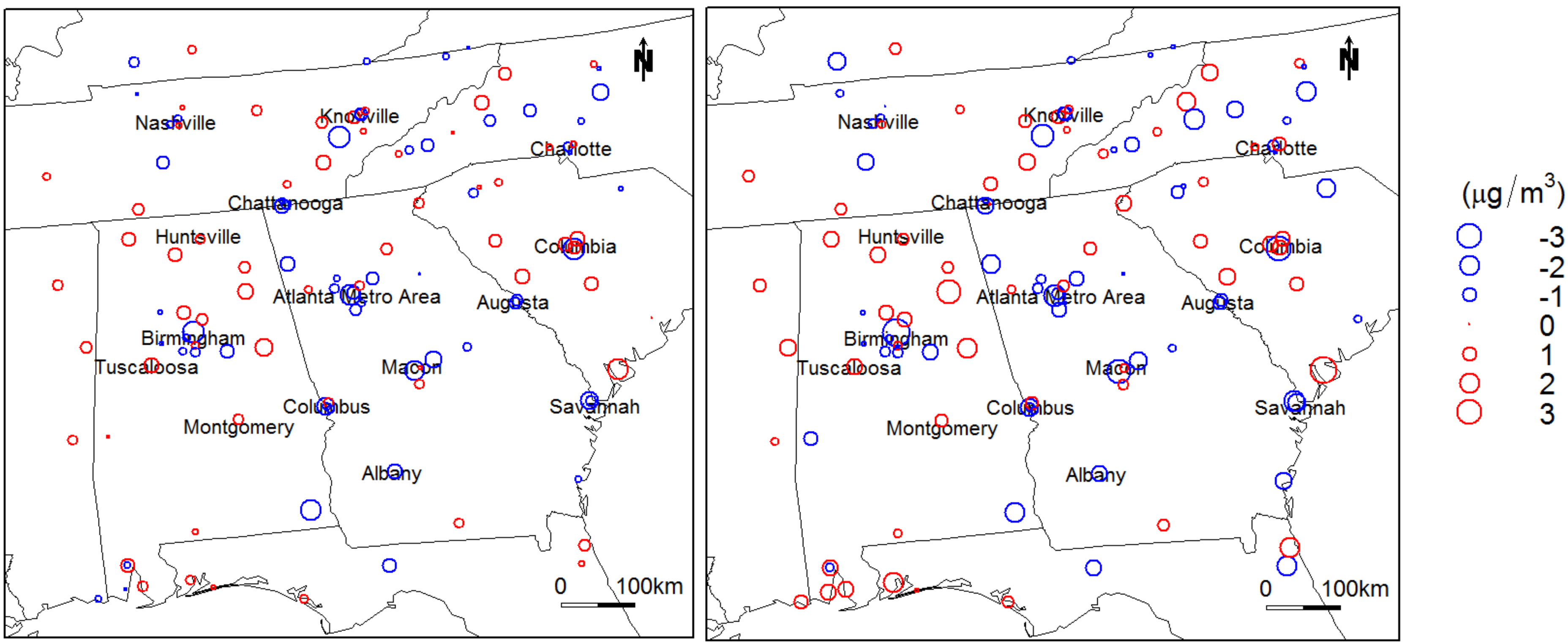
Site	N	Annual $PM_{2.5}$	r
BHM	85	19.1 $\mu g/m^3$	0.90
JST	87	15.3 $\mu g/m^3$	0.82

Putting all the data points together, we see unbiased estimates





Spatial Pattern of Model Bias

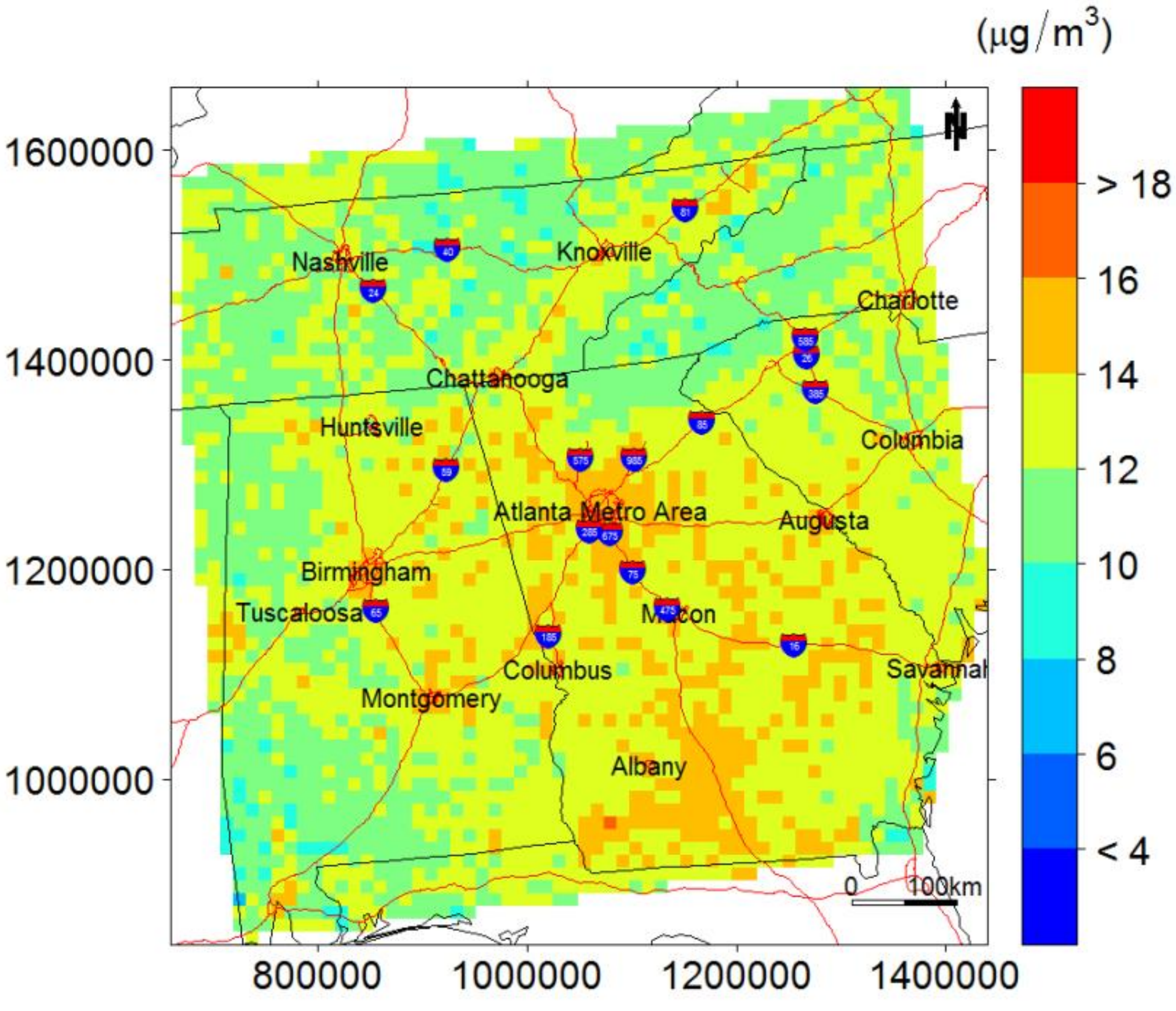


Model Fitting

Cross Validation

Negative and positive model / CV residuals are randomly distributed.

Model Predicted Mean $PM_{2.5}$ Surface



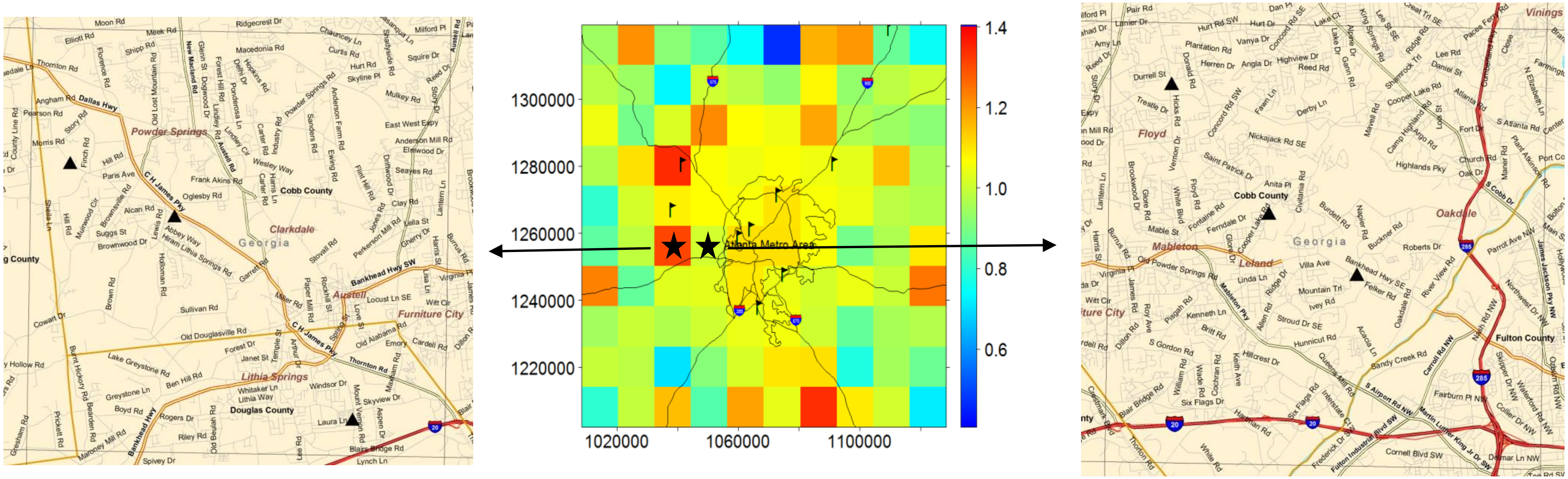
Note: annual mean calculated with137 days

Comparison with Other Models

- Pros:
  - Better performance than global regression models
  - Better reflection of temporal variability than LUR models
  - Stronger physical base than kriging models
  - Simpler and faster than air quality models
- Cons:
  - Integration with air quality models?
  - Statistical data filling is under study
  - Higher resolution data will become available soon

Manuscript submitted to *Environmental Research*

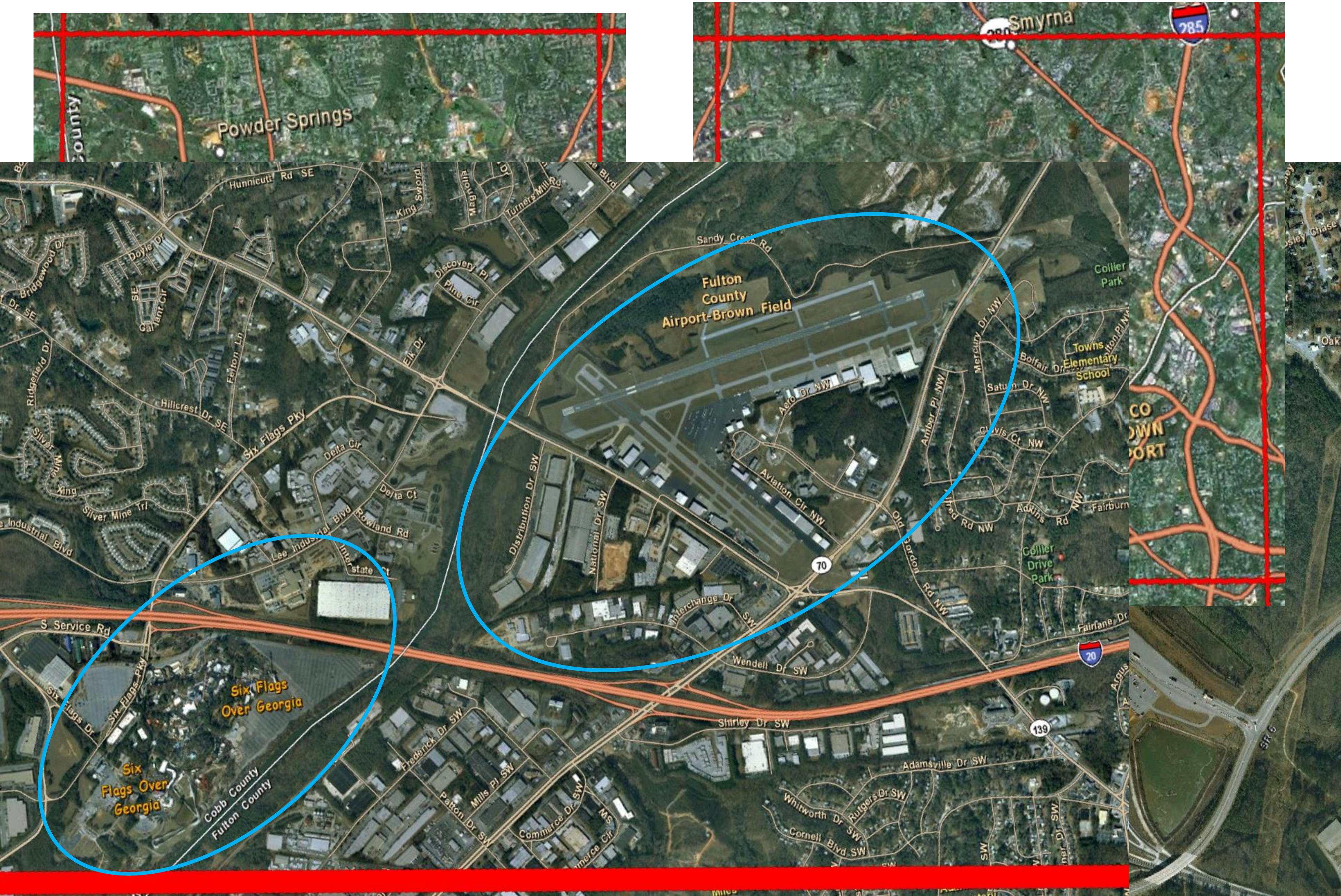
Strategy



- Identify a "hot" and a "cool" pixel based on ratios of GWR daily $PM_{2.5}$ concentrations over regional mean.
- 3 sampling locations > 3 km apart in each 12 km pixel
- ~20 24-hr samples in the next 6-9 months

So far, 3 sites located, portable samplers tested, made 2 sampling trips.

A Closer Look



Rational and Approach

For satellite data to be considered a reliable source of exposure estimates in health studies, both the spatial pattern and absolute levels of predicted $PM_{2.5}$ concentrations are important.

General calibration model structure (fitted annually)

$$AERONET\ AOD = \alpha + \beta_1 \times satellite\ AOD + season + \beta_2 \times satellite\ AOD \times season$$

Caveat: without calibration, MODIS can't be used for seasonal trend analysis, GOES can't be used for either seasonal or interannual trend analysis

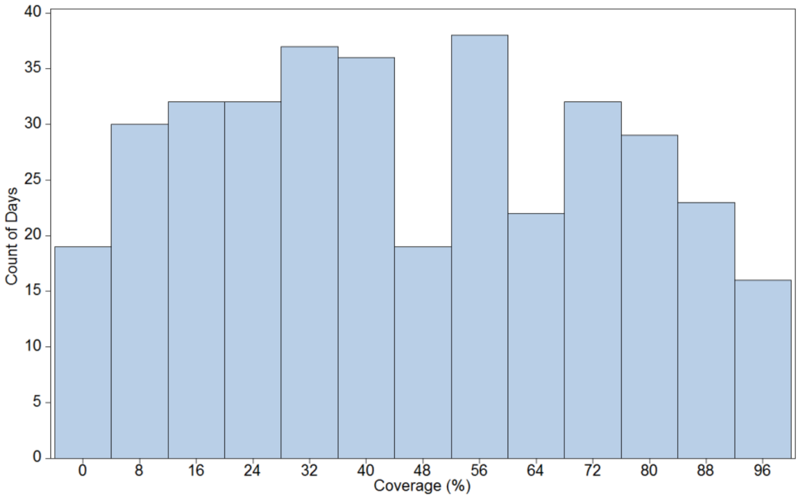
Rational and Approach

Problem: cloud cover causes a lot of data missingness. Without any treatment, best possible coverage is ~ 50%.

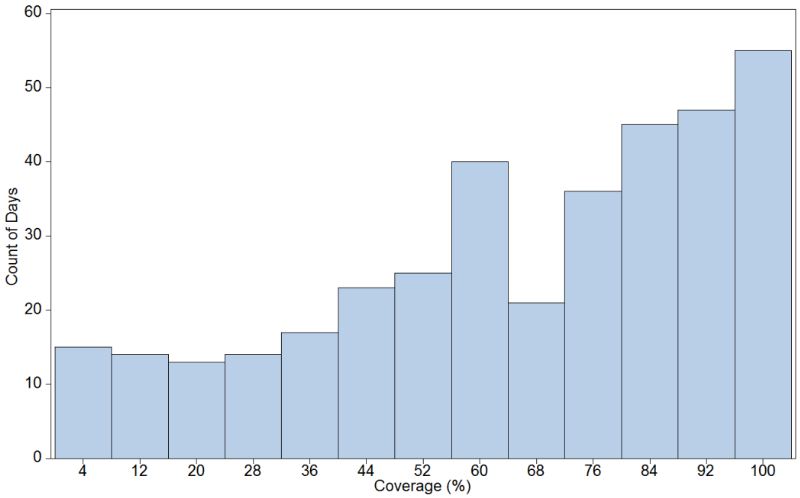
Hypothesis: missing AOD values due to small clouds can be filled with its nearest neighbors without significantly disturbing the predicted $PM_{2.5}$ surface.

Method: maximum distance over which nearby observations may be used to fill in missing grid cell values = 20 km

Preliminary Results



Raw MODIS, 2007



NN filled MODIS, 2007

Coverage (%)	N_days	Mean
Raw	365	46.04
NN	365	65.53

	RMSE ($\mu g/m^3$)	Relative Accuracy (%)
Raw_NARR	5.61	60.4
NN_NARR	4.82	66.8

NN filling: (1) improve coverage (2) improve model performance

Plan of epidemiological analysis

1. Communicate with epidemiologists on data format, structure, and modeling needs
2. Generate daily $PM_{2.5}$ estimates using calibrated, nearest neighbor-filled MODIS AOD for 2000 – 2007
3. Spatially join with zip code level patient addresses
4. Work with epidemiologists to develop space-time model
5. Evaluate resulted exposure-response functions

Year 3 Tasks

- Emory
  - MODIS/GOES data fusion
  - Final GWR $PM_{2.5}$ modeling
  - Development of new model structure
  - Field sampling and sample analysis
  - Health effects modeling and evaluation
- MSFC
  - Further study of gap filling techniques
  - Finalization of gridded aerosol data
- CDC
  - Comparison between HBM and satellite
  - Project benefit assessment