Enhancing EPHT with Satellite-Driven $PM_{2.5}$ Exposure Modeling and Epidemiology – Year 3 Report

Yang Liu (PI), J. Sarnat, M. Klein, H. Chang. Emory Univ.

J. Qualters, A. Vaidyanathan, H. Strosnider. CDC

D. Quattrochi, NASA/MSFC

W. Crosson, M. Al-Hamdan, M. Estes, S. Estes, USRA

September 19, 2012

Newport, RI

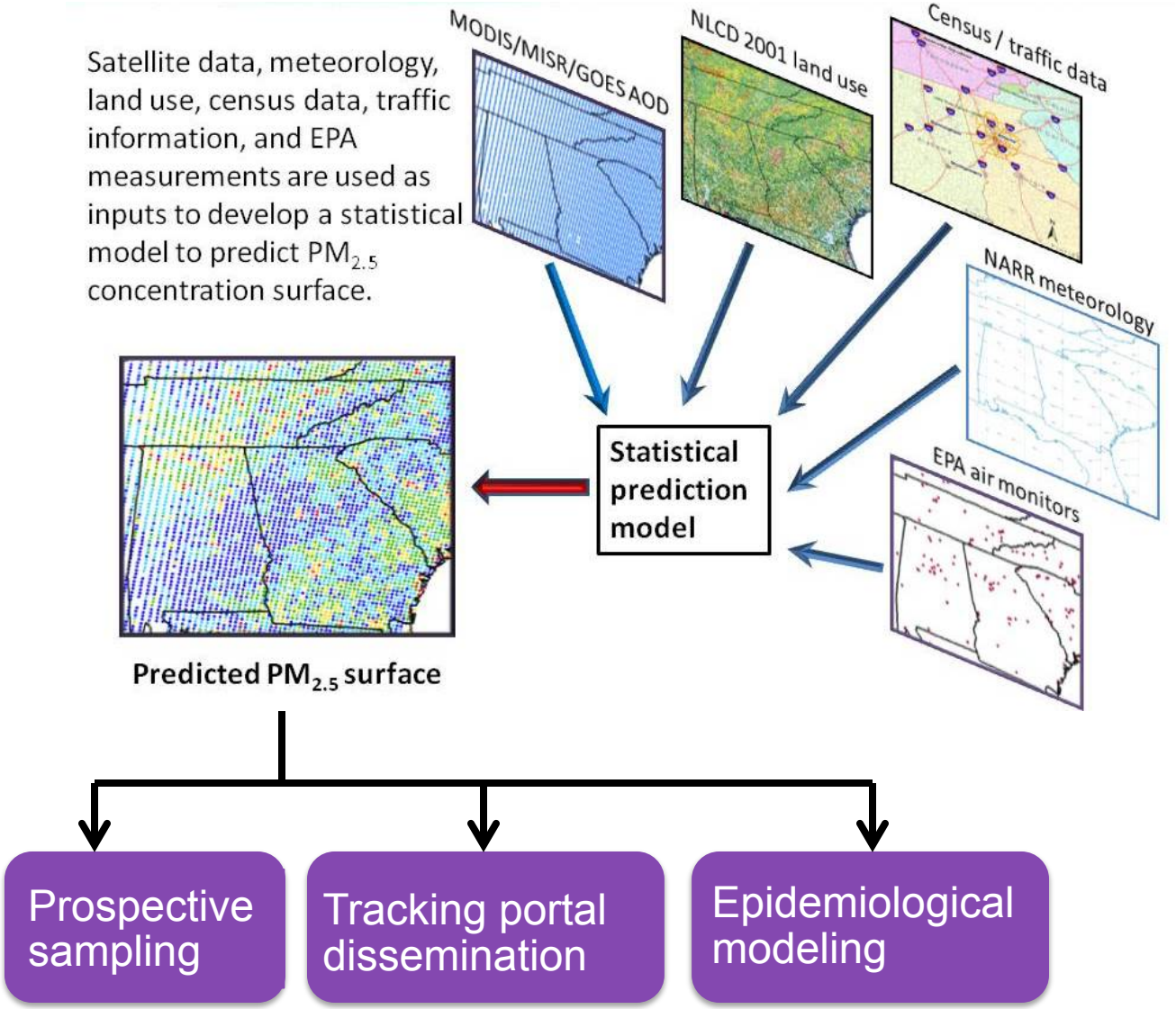
Project Objectives:

- Extend the spatial coverage of the $PM_{2.5}$ indicators in CDC Tracking Network with NASA Earth observations
- Provide timely estimates of county level $PM_{2.5}$ health indicators
- Evaluate satellite $PM_{2.5}$ estimates as a alternative exposure data source in environmental epidemiologic studies

Earth Observations:

- Terra & Aqua MODIS AOD, GASP AOD
- Terra & Aqua MODIS cloud data
- OMI aerosol type
- NLDAS/NARR meteorology

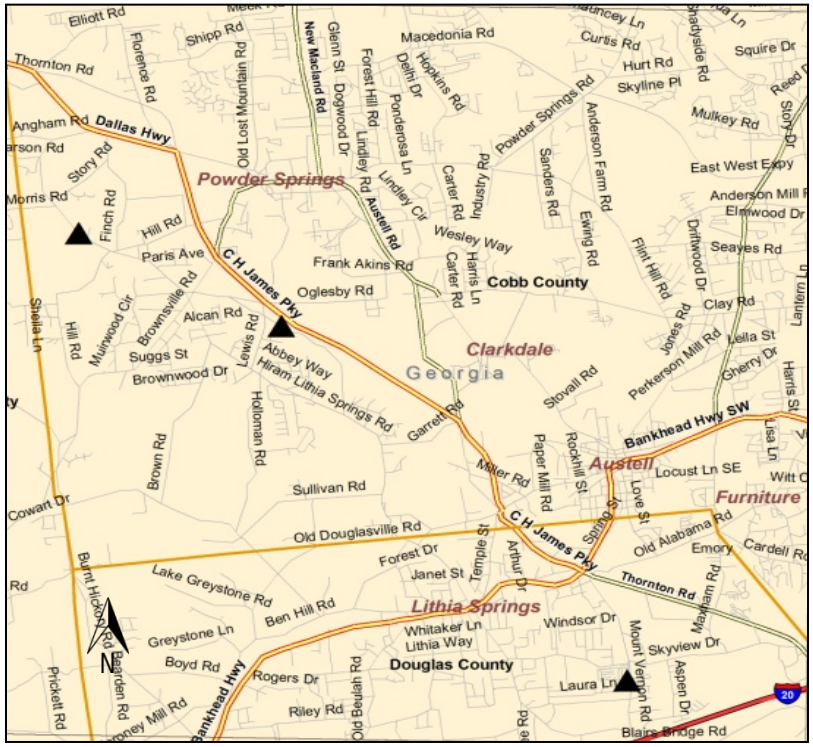
Technical Approach



Milestones

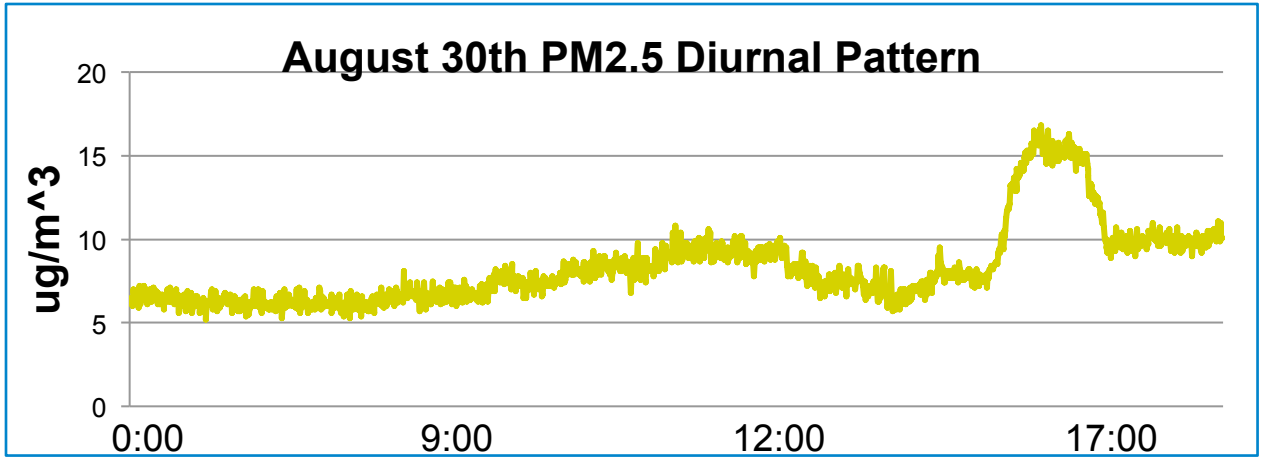
Milestones	Deadline	Team	Status
Model input data sent to Emory	08/2011	MSFC/USRA	Complete
Final exposure model	01/2012	Emory	Complete
Atlanta field sampling and analysis	01/2012	Emory, MSFC/USRA	Complete
Tracking benefit assessment	06/2012	CDC	Complete
Raw data delivery to Tracking	06/2012	Emory	Complete
Tracking format data delivery to Tracking	08/2012	Emory, CDC	Complete
Tracking online data publication	09/2012	CDC	Expected 1/2013
Epidemiological modeling	09/2012	Emory	To be continued during NCE

Field Sampling



- Sampling period: 08/2011- 11/2011, 06/2012 – 08/2012
- 24 hour / real time PM2.5, AOD at 5 wavelengths
- 2011 results were used to validate model predictions

Campus Sampling



- Sampling period: 08/2012 – 10/2012
- 24 hour / real time PM2.5
- Instrument development

PM2.5 Modeling with LME

- Linear Mixture-Effects Model
- Model Structure (day-specific random effects)

$$PM_{2.5} \sim (\beta_0 + \mu) + (\beta_1 + \nu)AOD + (\beta_2 + \kappa)Temperature + (\beta_3 + \sigma)Wind\ Speed$$
$$+ \beta_4 Elevation + \beta_5 Major\ Roads + \beta_6 Forest\ Cover + \beta_7 Point\ Emissions$$

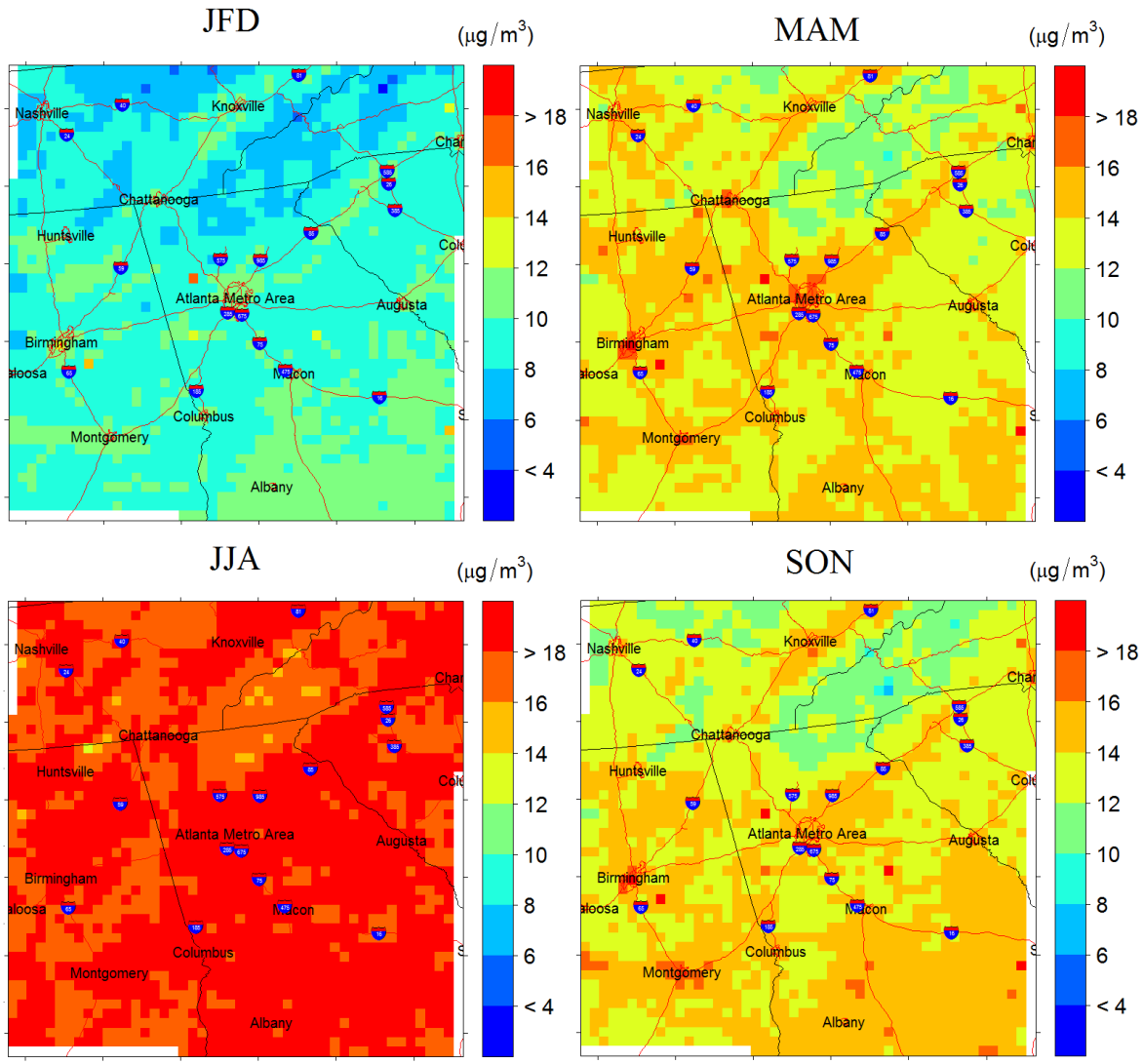
- Cross Validation
- Best Linear Unbiased Prediction (BLUP)

Model Fitting Statistics

	β	P-value
Intercept	-91.08	0.0000
AOD	9.94	0.0000
Temperature	0.36	0.0000
Wind Speed	-0.41	0.0001
Elevation	-0.001	0.0075
Major Roads	0.0002	0.0322
Forest Cover	-2.84	0.0000
Point Emissions	0.03	0.0000

	R^2	RMSE ($\mu g/m^3$)
Model Fitting	0.81	3.01
Cross Validation	0.76	3.36

Domain-wide PM2.5 Mapping



- linear mixed effects model
- Spatial resolution: 12 km on CMAQ grid
- Model year: 2003
- Domain: 600 km x 600 km
- Satellite data: nearest-neighbor filled MODIS AOD
- Meteorology: NLDAS-2

Bayesian PM2.5 Bayesian Downscaler

Goal: provide predicted **daily spatially-resolved** $PM_{2.5}$ concentrations and their associated **uncertainties**.

For day *t* at monitoring location *s*:

Space-time additive bias *Space-time multiplicative bias*

$$PM(s,t) = \boldsymbol{\beta_0} + \boldsymbol{\beta_0}(\boldsymbol{s}) + \boldsymbol{\beta_0}(\boldsymbol{t}) + [\boldsymbol{\beta_1} + \boldsymbol{\beta_1}(\boldsymbol{s}) + \boldsymbol{\beta_1}(\boldsymbol{t})]\, AOD(s,t)$$

$$+ \boldsymbol{\gamma_1 Z_1}(\boldsymbol{s},\boldsymbol{t}) + \ldots + \boldsymbol{\gamma_p Z_p}(\boldsymbol{s},\boldsymbol{t}) + \boldsymbol{\varepsilon}\,(\boldsymbol{s},\boldsymbol{t})$$

Land use and meteorological variables

<u>Temporal Random Effects</u> (first-order random-walk)

For t = 2, 3, …, 1824:

$$\beta_j(t) \sim N(\varphi_j \beta_j(t-1) + \beta_j(t+1)/2,\ \tau_j^2/2)$$

For t = 1:

$$\beta_j(t) \sim N(\varphi_j \beta_j(t+1),\ \tau_j^2)$$

Model Configuration

Spatial Random Effects (Linear model of co-regionalization)

$$\begin{bmatrix} \beta_0(s) \\ \beta_1(s) \end{bmatrix} = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{bmatrix} W_0(s) \\ W_1(s) \end{bmatrix}$$

Two independent Normal latent variables.

$$\mathrm{Cov}\,[W_0(s), W_0(s')] = e^{-\mathrm{distance}[W_0(s),\, W_0(s')]/\rho_0}$$

$$\mathrm{Cov}\,[W_1(s), W_1(s')] = e^{-\mathrm{distance}[W_1(s),\, W_1(s')]/\rho_1}$$

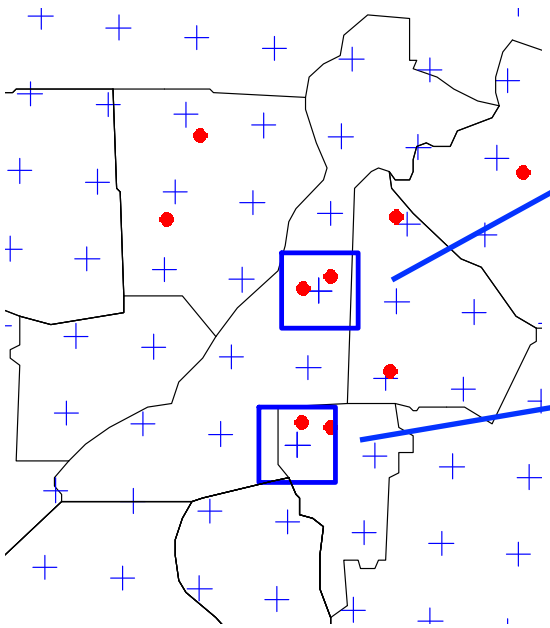
This formulation does not assume that correlations between the intercepts and slopes at different spatial locations are identical.

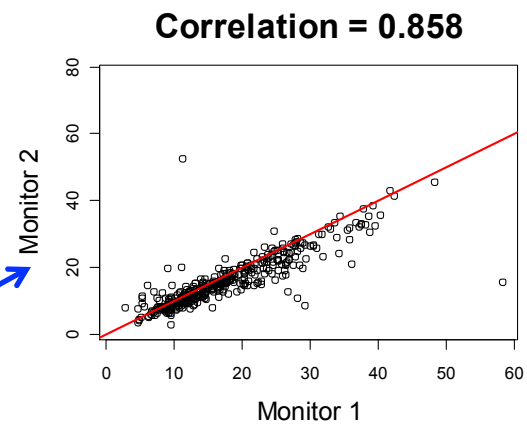
Residual Errors

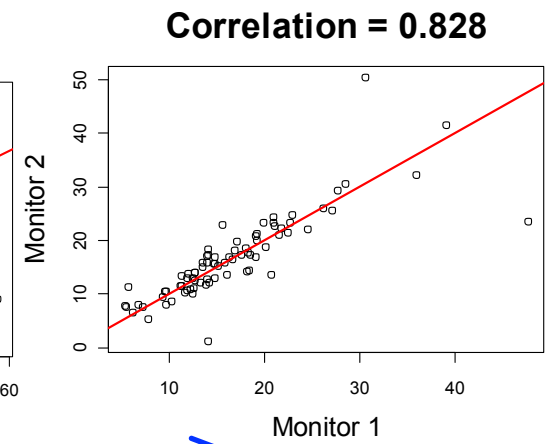
$$\varepsilon(s,t) \sim N(0, \sigma^2)$$

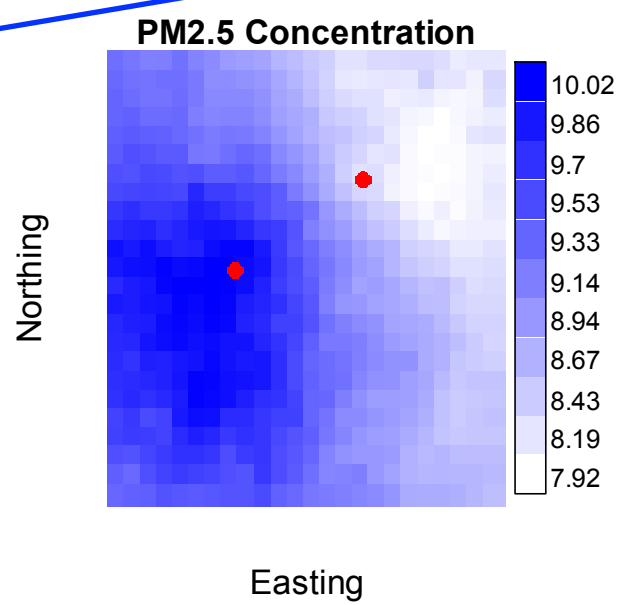
Did not consider space-time interaction … difficult to fit with 5 years of data.

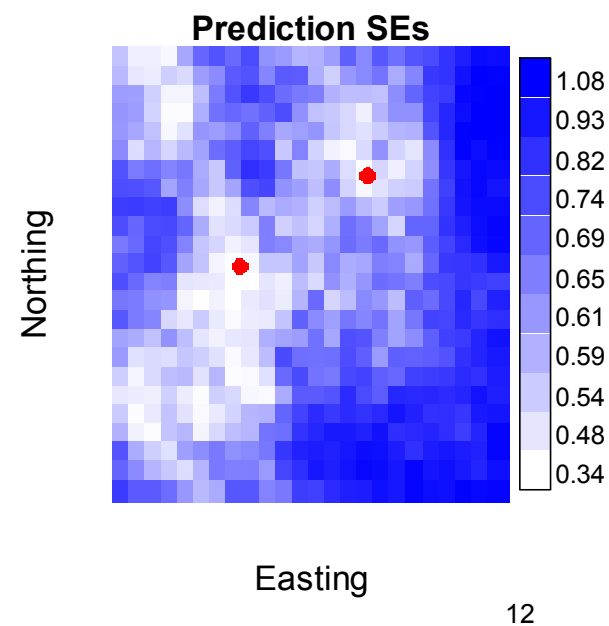
Statistical Downscaling





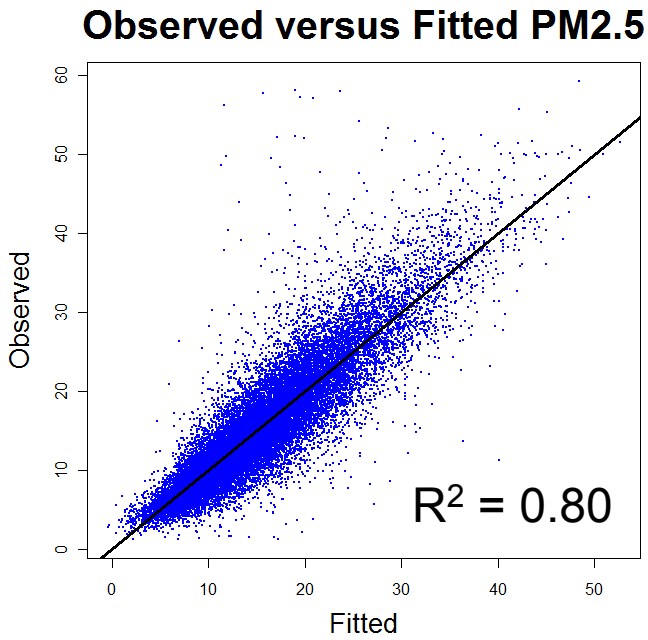


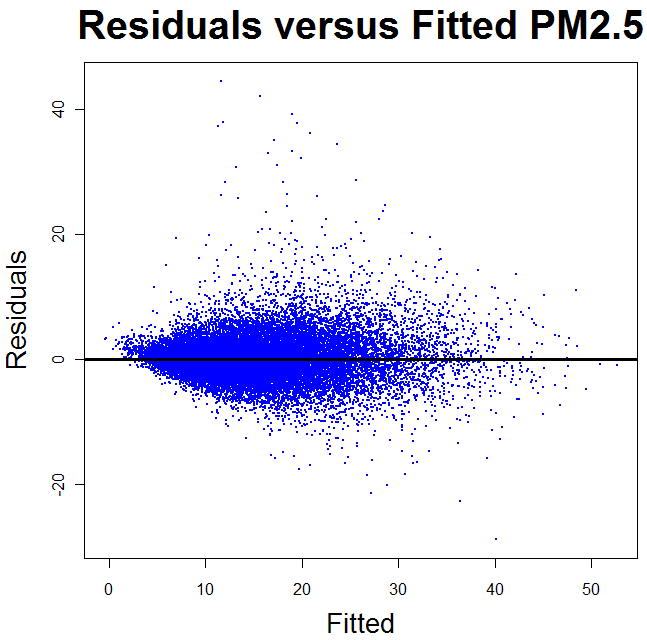


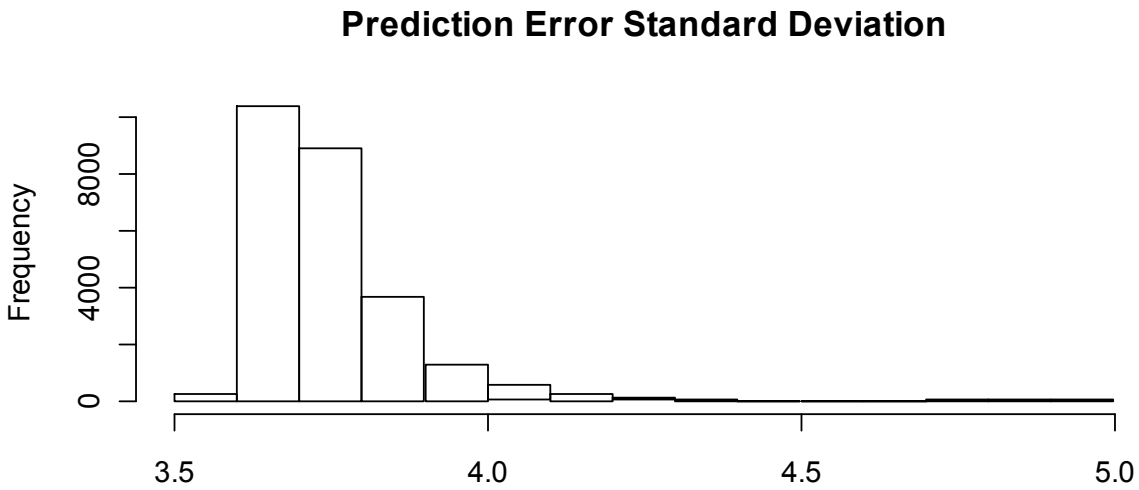


500m X 500m grid within an AOD grid cell

Model Fitting Statistics



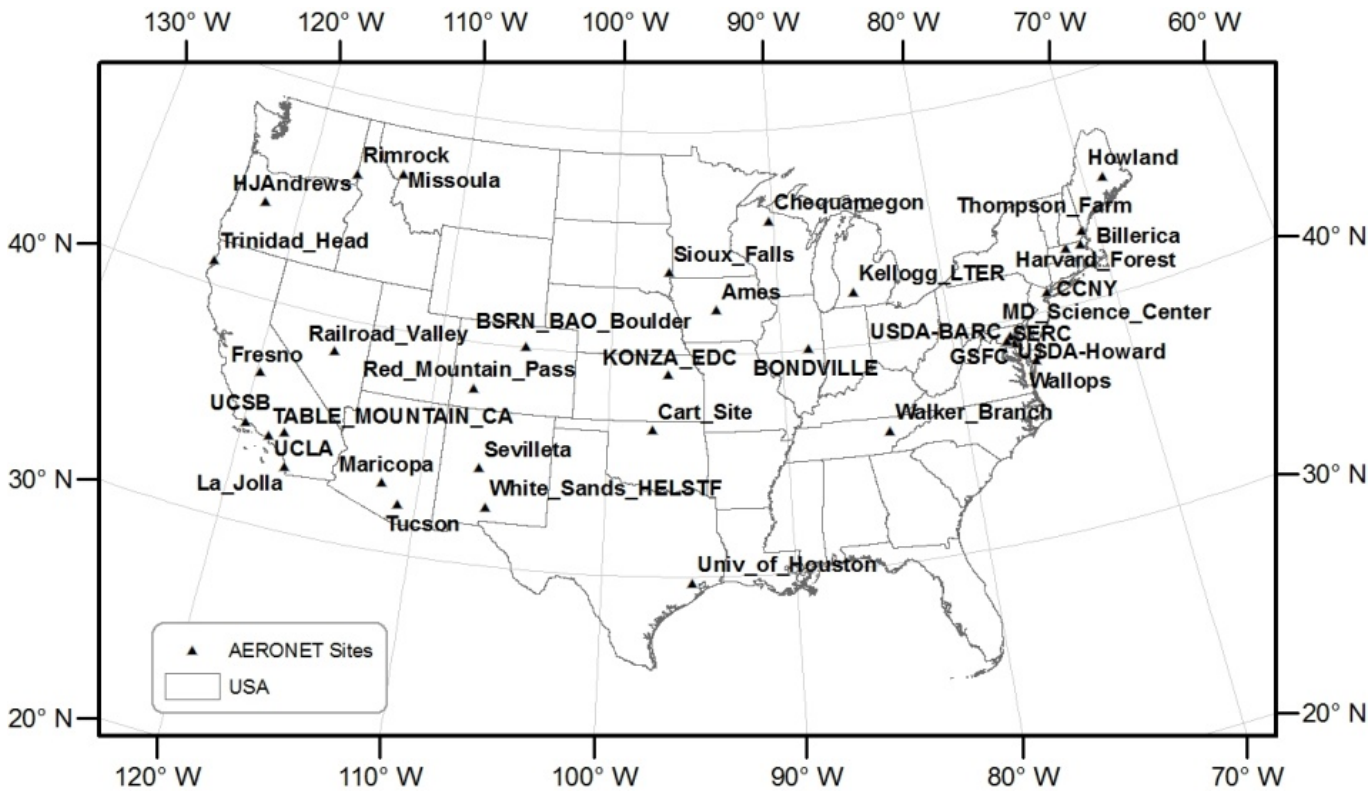




MODIS/GOES Data Fusion

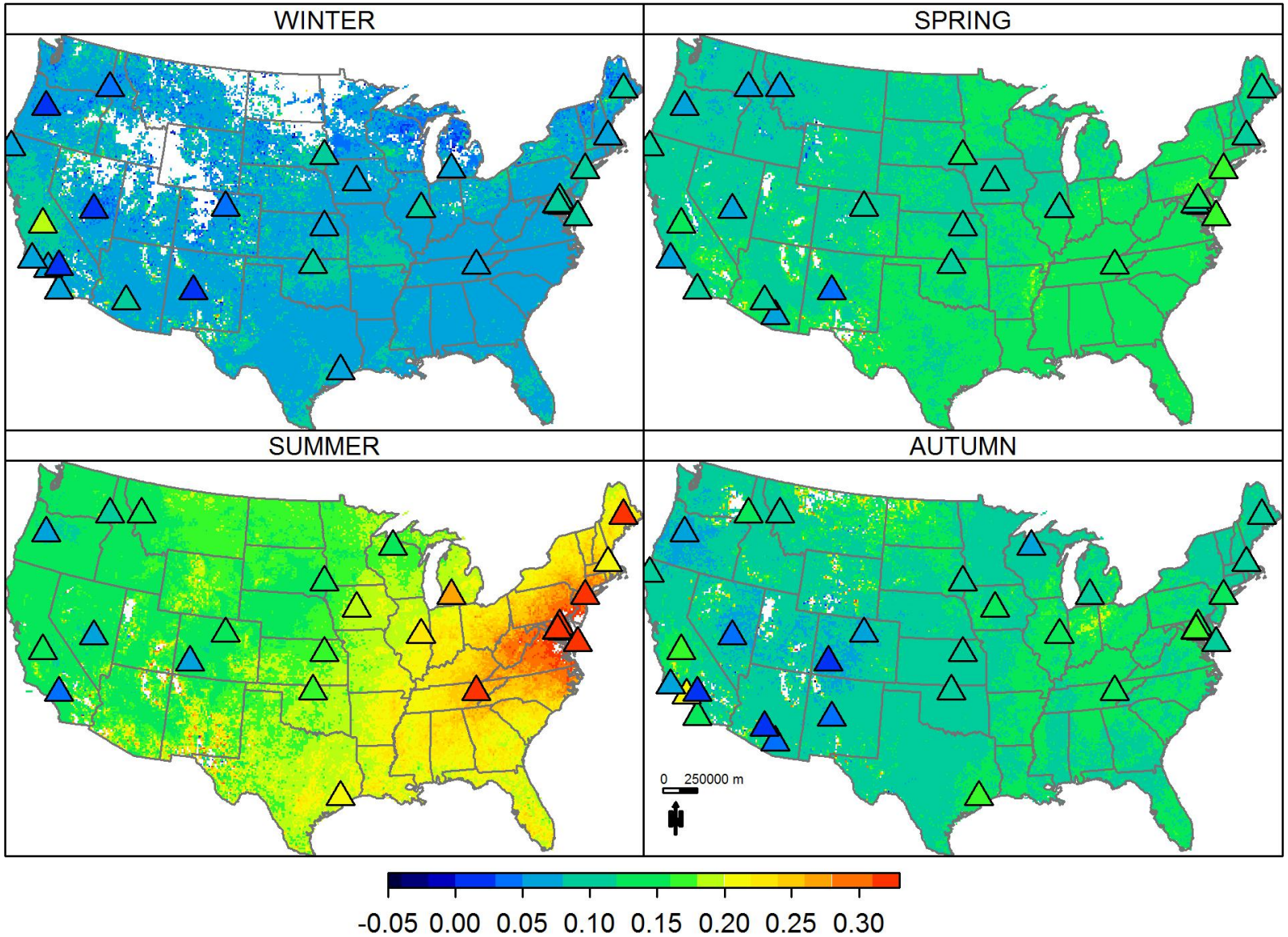
- Motivation
  - Health effects study desires better spatial coverage
  - MODIS / GOES have different value ranges / performances
  - Tracking wants a national dataset
- Universal Kriging Approach
  - Develop daily semi-variograms using MODIS data
  - Use matched AERONET observations to calibrate GOES data in the UK framework
  - Make predictions where GOES AOD is available

Modeling domain

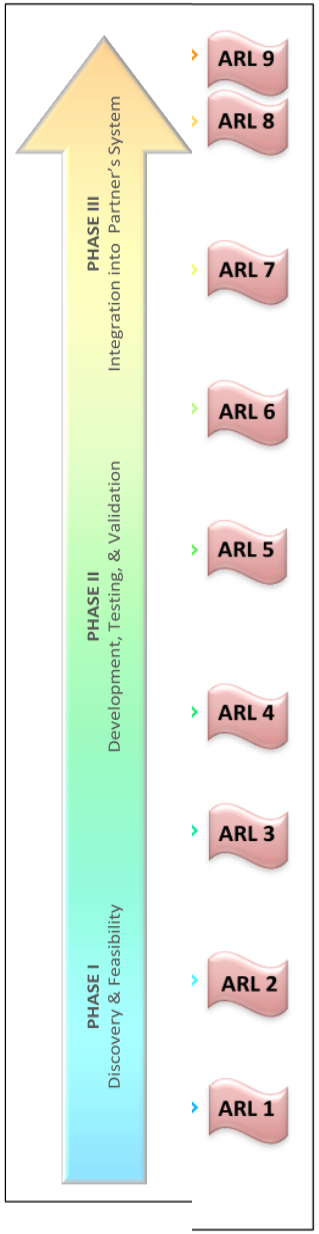


- Study period: 2005 & 2006
- # AERONET sites: 33 (> 6 mo in operation)

AOD Fields Based on GOES



Application Readiness Level Status



- Start of project ARL = 4
- Current ARL = 6
  - Raw and formatted satellite PM2.5 data have been delivered to Tracking
  - Tracking IT team has evaluated the formatted data for potential integration into the national portal
  - Tracking science team's accuracy assessment showed good agreement between satellite data and HBM / EPA observations
- Expected ARL = 7
  - Satellite data are expected to appear in the January, 2013 build of national portal with metadata
- Final ARL with no cost extension = 8
  - Will work with Tracking to finalize data presentation on portal, prepare metadata and guidance document for data use

Budget

- A positive balance was carried over from year 1 and 2 as part of PI's effort has been covered by the CDC ORISE fellowship.
- Surplus used to purchase portable equipments to enhance the sampling components and cover personnel cost for their operation
- Emory projected balance by September 30, 2012 is $57K – 60K, to be used for continued epidemiological modeling, and collaboration with CDC to reach ARL 8

Accomplishments

EMORY UNIVERSITY

Publications:

1. A statistical model to evaluate the effectiveness of PM2.5 emissions control during the Beijing 2008 Olympic Games, *Environmental International,* 44, 100-105, 2012.
2. Estimating Ground-Level PM2.5 Concentrations in the Southeastern U.S. Using Geographically Weighted Regression, in press, *Environmental Research, 2012.*
3. Retrieval of the Haze Optical Thickness in North China Plain using MODIS data, in press, *IEEE Trans.* on *Geoscience and Remote Sensing, 2012.*
4. Statistical Data Fusion of Multi-sensor AOD over the Continental United States, submitted, *Geocarto International, 2012.*
5. Estimating Ground-Level PM2.5 Concentrations in the Southeastern U.S. using MAIAC AOD Retrievals, to be submitted to JGR, 2012.
6. Calibrating Remotely Sensed Aerosol Optical Depth for Predicting Daily Fine Particulate Matter Concentrations via Statistical Downscaling, to be submitted to *ES&T*, 2012
7. The impact of satellite-retrieved cloud properties on ground-level PM2.5 concentrations, to be submitted to *Remote Sensing of Environment, 2012.*

Accomplishments

Selected conference presentations:

- The 90th American Meteorological Society Annual Meeting, Atlanta, GA, January 19, 2010.
- American Thoracic Society International Conference, New Orleans, LA, May 19, 2010.
- American Geophysical Union Fall Meeting, San Francisco, CA, December 15, 2010.
- The 91st Annual Meeting of the American Meteorological Society, Seattle, WA, January 25, 2011.
- Goldschmidt2011, Prague, Czech Republic, August 14-19, 2011.
- American Geophysical Union Fall Meeting, San Francisco, CA, December 4, 2011.

- Multiple abstracts submitted to AGU 2012 fall meeting

Risks and Issues

- Technical Challenges to be addressed during NCE
  - Data missingness due to cloud cover may cause sampling bias
  - More advanced spatial PM2.5 model can better resolve spatial autocorrelation so that model can be extended to a larger region
  - Epidemiological modeling need additional work in order to use zip code level satellite predictions

- Operations Challenges
  - Need to work with Tracking to ensure clear interpretation of satellite data on the portal
  - May need to present in Tracking national meetings

Plan for NCE Period

EMORY UNIVERSITY

- Data missingness due to cloud cover
  - empirical cloud – PM2.5 association
- Epidemiological model
  - Zip code level clustering to reduce spatial autocorrelation
- Data dissemination through Tracking
  - Guidance document, metadata
  - Participation in Tracking workshops/conferences
  - **Pushing for ARL 9 – training Tracking staff for data processing and model maintenance for sustained use?**
    - **Need help from ARSET and CDC personnel commitment**
- Manuscripts / final report preparation