Statistical and epidemiological considerations in using remote sensing data for exposure estimation

Chris Paciorek
Department of Biostatistics
Harvard School of Public Health

Collaborators:
Yang Liu, Doug Dockery, Francine Laden, Joel Schwartz, Helen Suh
Exposure, Epidemiology and Risk Program
Department of Environmental Health
Harvard School of Public Health

February 21, 2006

www.biostat.harvard.edu/~paciorek

Two-stage Environmental Exposure and Epidemiological Analysis

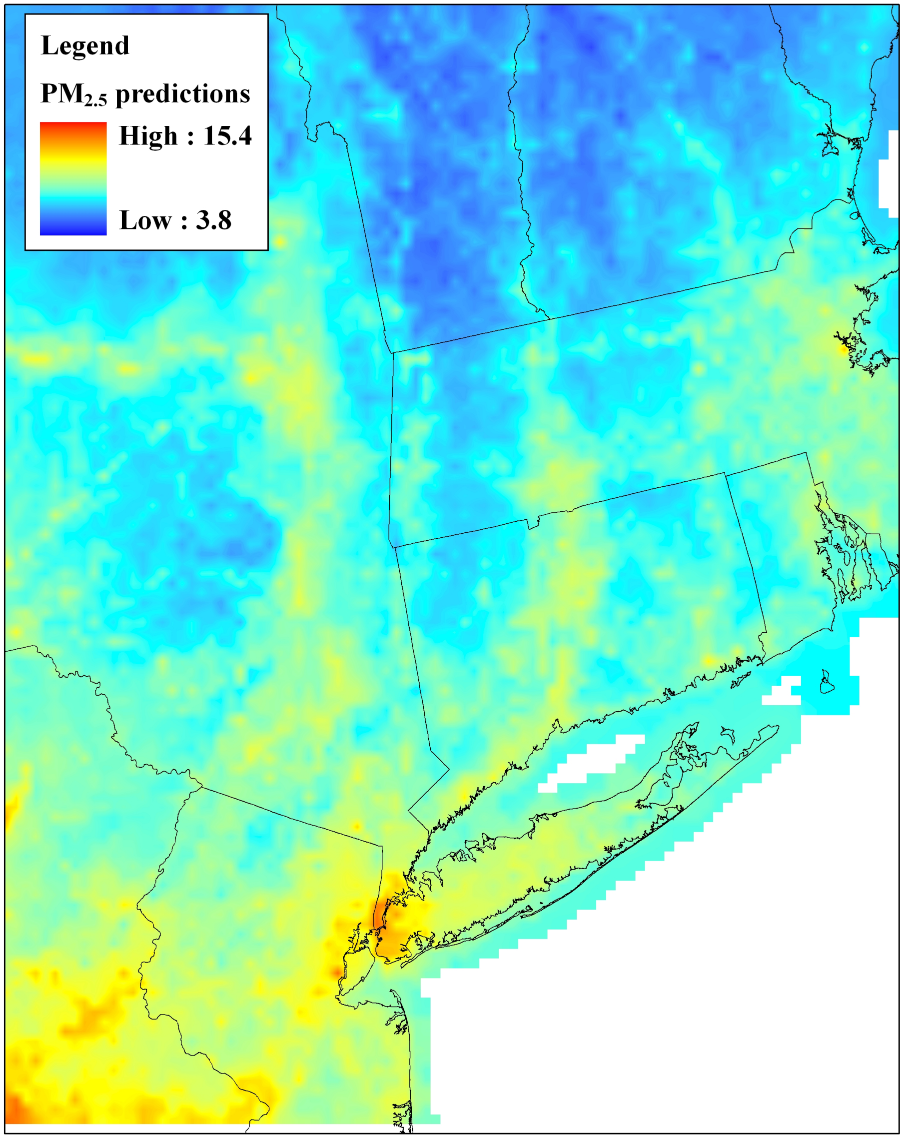
- Spatial exposure estimation
  - Collect monitoring data for pollutant for multiple monitors at different spatial locations
  - Use statistical techniques and GIS-derived information to smooth the data and estimate exposure at residences of study participants
  - [insert remote sensing observations here]

- Health analysis
  - Follow study participants over time and record times of adverse health events
  - Relate estimated exposures to health outcomes, controlling for other personal risk factors (smoking, BMI,...)

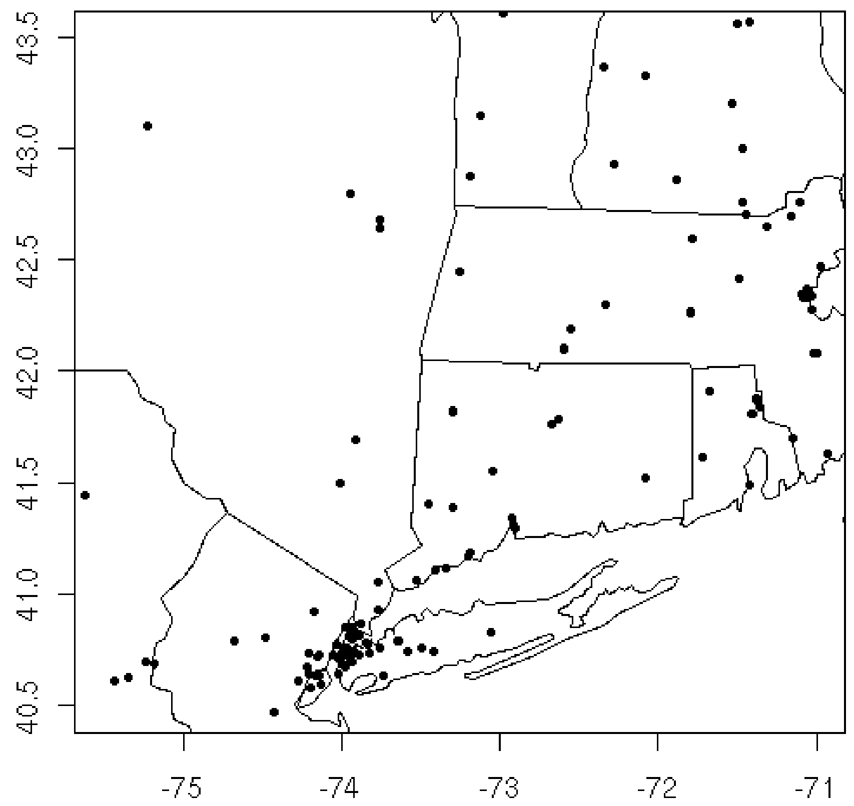
Example analysis: Cardiopulmonary disease and particulate matter in the Nurses' Health Study

- goal: understand long-term health effects of exposure to particulate matter (PM)
- study population: nurses participating in the Nurses' Health Study (1976-...)
- PM estimation
  - statistical model based on monitors and GIS-derived information
    - northeastern U.S., 1990-2002
  - PM exposure estimated for each month at participant residences
- Analysis of association of PM estimates with mortality, heart attacks, stroke
  - 70,000 nurses with detailed personal information to help control for other personal risk factors
  - is PM exposure statistically associated with increased risk of adverse outcomes?

Spatial smoothing of monitoring data

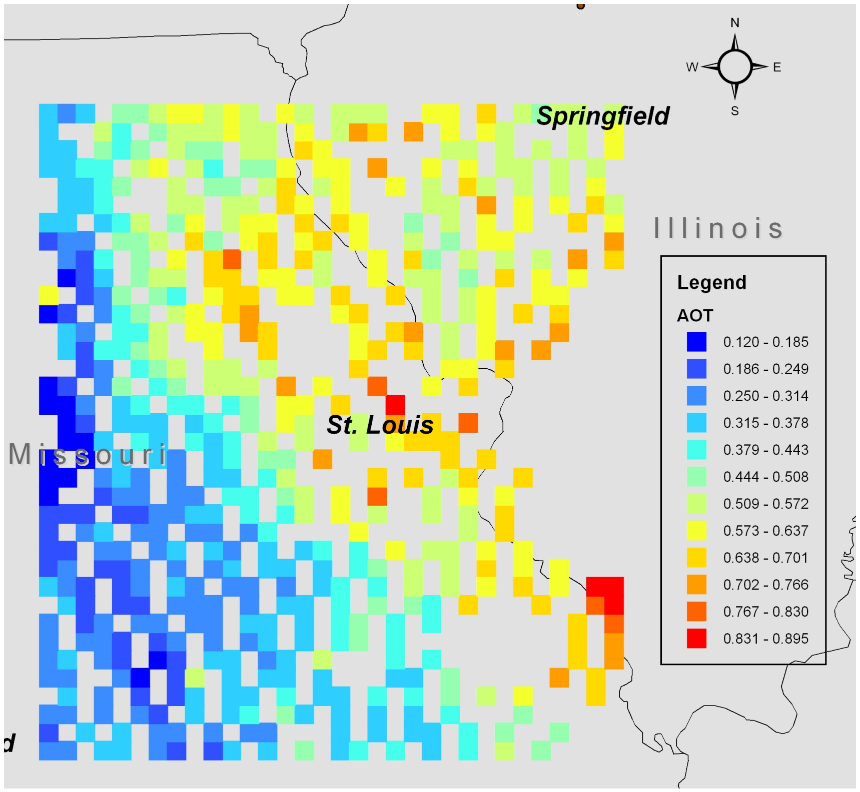


Estimated PM for one month

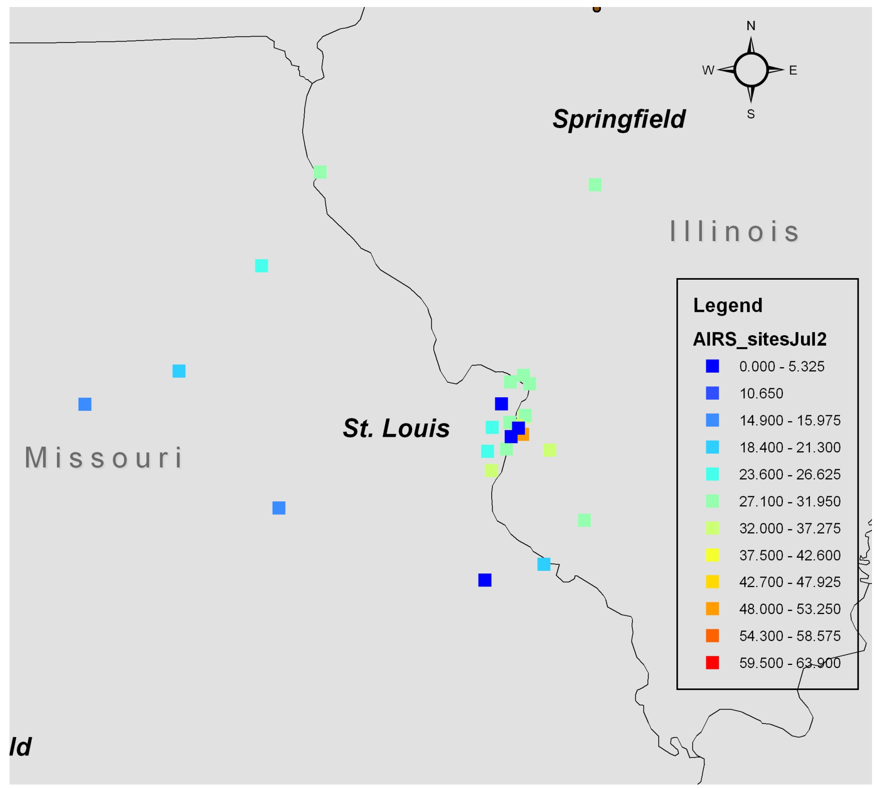


Monitor locations

Remote sensing vs. ground monitoring observations



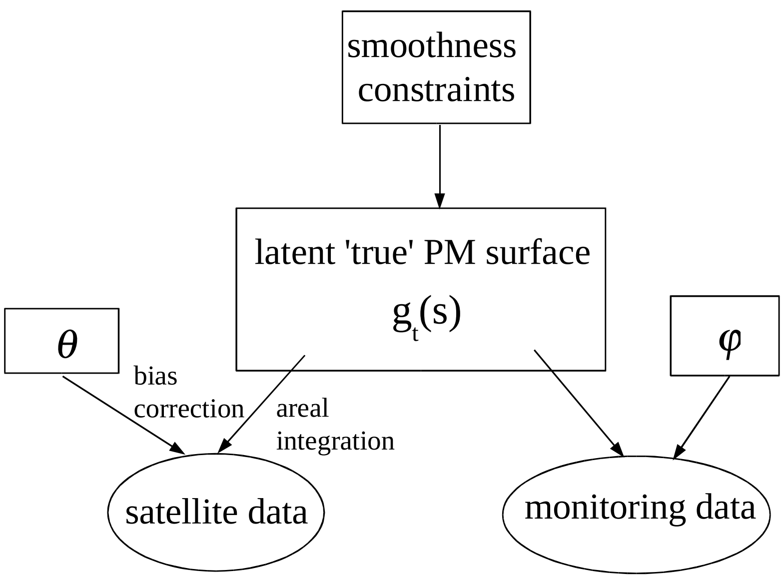
MODIS AOT



$PM_{2.5}$ monitors

Bayesian statistical melding

- statistical technique for combining information sources
- Bayesian statistical models allow for complicated probabilistic relationships and constraints on exposure surfaces
- constraints ensure smooth estimated exposure surfaces and allow estimation in areas with no data



Strengths of statistical melding approach

- estimation of PM surface based on all information
  - ground data: gold standard + higher resolution in urban area
  - remote sensing: broad spatial coverage but coarse resolution
  - other information can be included:
    e.g., GIS information, possible cloud cover biases, vertical profile information from atmospheric chemistry models (Liu et al. 2004)

- $Y_i^g \sim \mathcal{N}(g(s_i), \sigma^2) \qquad Y_A^{rs} \sim \mathcal{N}(a + b \sum_{s \in A} g(s), \tau^2)$
  - internal calibration of remote sensing observations
  - synthesis of differing resolutions of the data sources

- model structure allows for internal validation of remote sensing data

- model provides estimates of uncertainty in estimated PM at every location

Uncertainty considerations

- statistical models can account for uncertainty in a probabilistically rigorous fashion
  - (inputs) weight observations based on certainty
  - (outputs) propagate uncertainty through model to final estimates
- remote sensing data with associated (relative) uncertainty estimates are of great interest
  - higher uncertainty may be a reasonable tradeoff for more useful data (e.g., finely gridded observations)
- uncertainty can be adjusted for based on:
  - quantification of the levels of uncertainty in the observations
  - ground truth against which to internally calibrate (e.g., ground PM, AERONET AOT)
- possibility for involving statisticians earlier in the process to help provide uncertainty levels associated with remote sensing observations

Epidemiological needs and interests

- historical data to match to historical health data (health data reporting lags)
  - e.g., GOES AOT from 1990s
  - other sources for 1990s at high enough resolution?
- higher spatial resolution more important than high time resolution
- short-term data (e.g., daily average) of interest for short-term health effects
  - GOES AOT of interest for acute health
- information on particle types (e.g., black carbon) and sizes
  - attribution and source identification

Conclusions

- epidemiological studies can use AOT and other pollution proxies
- statistical approaches allow for combining ground monitoring and remote sensing data
- uncertainty quantification is a key consideration
- open possibilities for using remote sensing data for exposure assessment for environmental epidemiology?