Integrating remote sensing and ground monitoring data to improve estimation of $PM_{2.5}$ concentrations for chronic health studies

Chris Paciorek and Yang Liu
Departments of Biostatistics and Environmental Health
Harvard School of Public Health

May 8, 2007

Outline

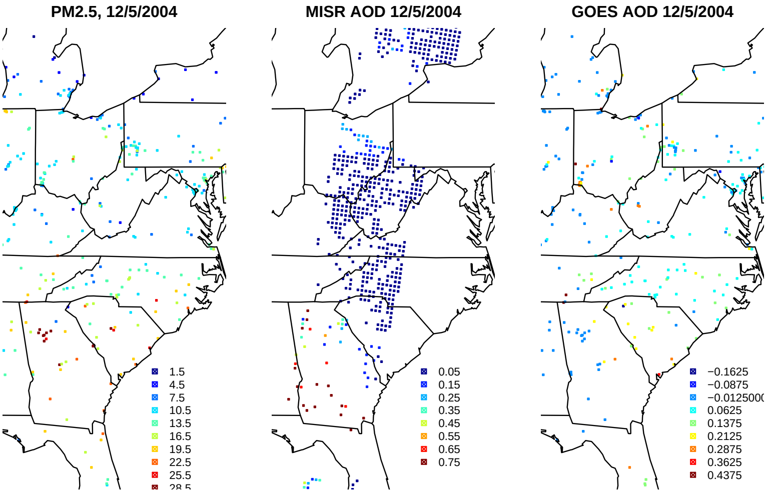
Exposure estimation for $PM_{2.5}$

- For studies of the chronic health effects of PM, estimating spatial heterogeneity is critical.
- Satellite retrievals of AOD may be able to help estimate concentrations at locations far from monitors, particularly in suburban and rural areas.
- But, current retrievals estimate total column aerosol and correlations between AOD and PM are low when considered at high temporal and spatial resolution.
- Statistical modeling can calibrate AOD to ground-level PM and account for the error induced in using AOD as a proxy.
- Our goal: integrate AOD and ground-level PM measurements to estimate ground-level PM on a regular 4 km grid every month, 2000-2006.
- Intended use is as a data product for use in various studies of chronic health effects.

AOD measurements

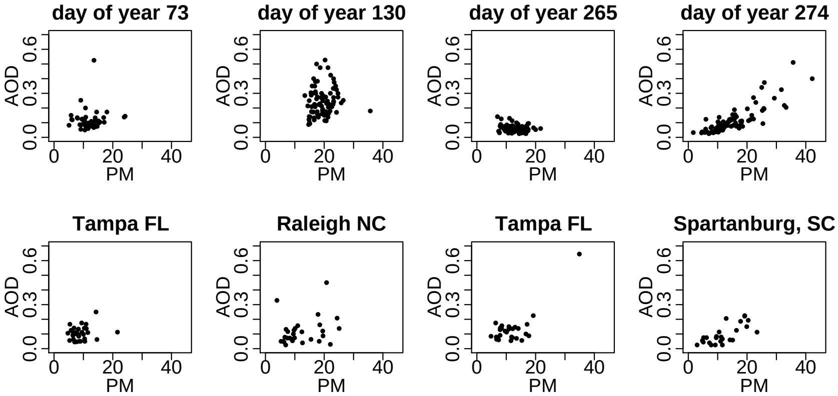
- MISR:
  - once per day
  - nominal 17.6 km resolution, a given location is measured every 4-7 days
  - multi-angle, narrow band, multispectral
- MODIS:
  - once per day
  - nominal 10 km resolution, a given location is measured every 1-2 days
  - narrow band, multispectral
- GOES:
  - every half hour during daylight
  - nominal 4 km resolution
  - broadband

Correspondence of AOD and PM



Correspondence of MISR AOD and PM

Cross-sectional (top row) and longitudinal (bottom row) associations for days and locations with many co-occurring observations.



Statistical calibration of AOD to PM

- Goal: use co-occurring observations of PM and AOD to build a statistical model that accounts for factors that prevent a close relationship of AOD and PM.
- Liu et al. (2005) built a regression model (on the log scale) relating PM to MISR AOD, modified by location, season, RH and PBL.
- Here we extend the approach using more flexible nonparametric regression terms:

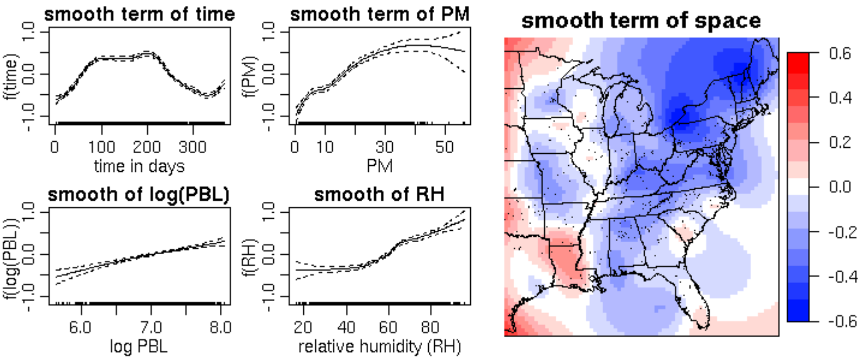
$$\begin{aligned}\log \mathrm{AOD}_{it} &= g(s_i) + f(t) + f(\log \mathrm{PBL}_{it}) + f(\mathrm{RH}_{it}) + \mathrm{PM}_{it}\beta_1 + \epsilon_{it} \\ \log \mathrm{AOD}^*_{it} &= \log \mathrm{AOD}_{it} - \hat{g}(s_i) - \hat{f}(t) - \hat{f}(\log \mathrm{PBL}_{it}) - \hat{f}(\mathrm{RH}_{it})\end{aligned}$$

- After calibration,

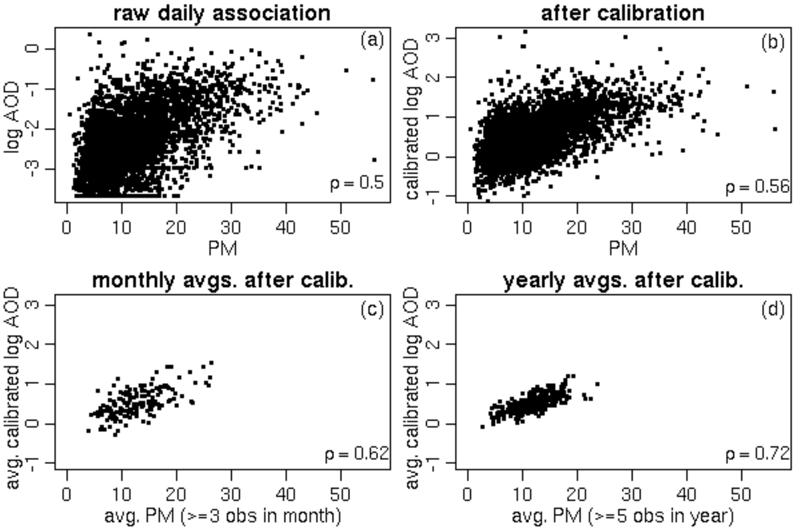
$$\log \mathrm{AOD}^*_{it} \approx \beta_0 + \beta_1 \mathrm{PM}_{it} + \epsilon_{it}$$

Fitted calibration model

$$\log \text{AOD}_{it} = g(s_i) + f(t) + f(\log \text{PBL}_{it}) + f(\text{RH}_{it}) + \text{PM}_{it}\beta_1 + \epsilon_{it}$$



AOD-PM association after calibration



Statistical calibration of GOES AOD

- For each season: summer, spring, fall:

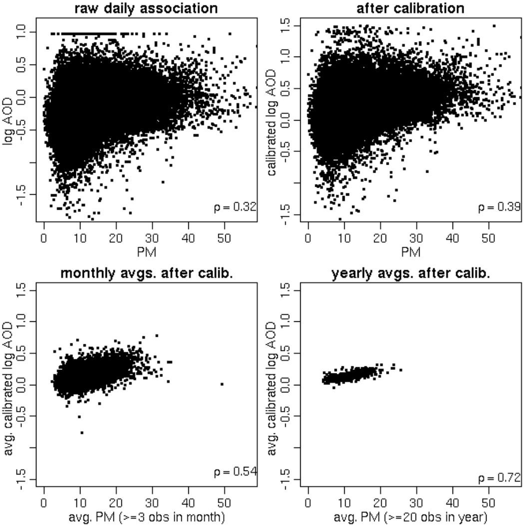
$$
\begin{aligned}
\log \mathrm{AOD}_{it} &= g(s_i) + \mathrm{PM}_{it}\beta_1 + \epsilon_{it} \\
\log \mathrm{AOD}^*_{it} &= \log \mathrm{AOD}_{it} - \hat{g}(s_i)
\end{aligned}
$$

- After calibration,

$$
\log \mathrm{AOD}^*_{it} \approx \beta_0 + \beta_1 \mathrm{PM}_{it} + \epsilon_{it}
$$

- No apparent association of AOD and PM in winter.

Non-winter GOES AOD-PM association after calibration



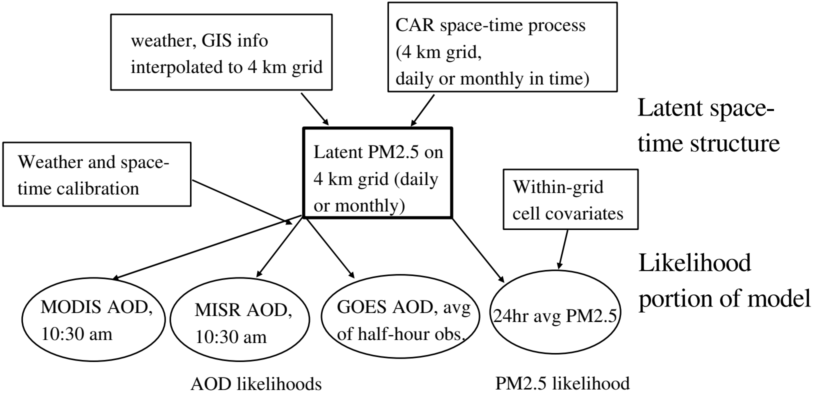
Calibration lessons

- Calibration and temporal averaging both improve associations, which appear reasonably linear at the monthly and yearly resolutions.
- Temporal averaging without calibration shows little improvement.
- GOES AOD in winter appears not to be useful, but negative values and days with only a single retrieval are helpful.

Statistical modeling

- Core idea is a latent spatio-temporal PM process at the daily or monthly scale.
- Treat calibrated AOD and ground measurements as noisy measurements of the process at fixed locations and times.
- Calibrated AOD is considered as an observation, not a regression term, because of missing observations and lack of coverage.
- A Bayesian hierarchical model can integrate the various sources of information and provide for smoothing in space-time to predict at locations and times without any data.
- The model needs to be constructed to take advantage of sparse matrix manipulations to be feasible at the scale of interest (2000-2006, 4 km resolution).
- Model should deal with irregular spatial and temporal sampling and possibility that missing retrievals are correlated with PM.

The basic statistical model



Next steps and open issues

- Calibration:
    - Consider different quality thresholds for GOES retrievals.
    - Make use of CMAQ vertical profiles in calibration.
- Fit the model and make predictions for a subset of the full domain
    - How much improvement is provided by AOD over and above ground measurements, weather variables, and land use type variables?
    - Should the model be fit at the daily level and estimated PM averaged to the month or fit at the monthly level after averaging the observations?

Some perspectives

- Needs of epidemiologists:
  - High spatial (and in some cases spatio-temporal) resolution for exposure assessment and linkage with health outcomes
  - Long-term records for cohort studies
  - Calibration with ground-level measurements to understand and account for error in exposure estimation
  - PM component information
- Opportunities for collaboration of remote sensing scientists and statisticians to improve exposure assessment
  - use of retrievals of varying quality in a statistical model
  - empirical calibration of level 1 retrievals to ground-level PM
    - can more information be extracted by relating level 1 directly to observables of interest, building retrieval algorithm assumptions within an empirical statistical framework?