Spatial bias modeling with application to assessing remotely-sensed aerosol as a proxy for particulate matter

Chris Paciorek
Department of Biostatistics
Harvard School of Public Health

application joint with Yang Liu, HSPH Environmental Health

October 23, 2008

Research supported by HEI 4746RFA052/067.

Abstract

Recent research highlights the potential of remotely-sensed aerosol optical depth (AOD) to serve as a proxy for ground-level fine particulate matter, $PM_{2.5}$. Standard models for the use of spatio-temporal proxy data to predict a latent process of interest assume a simple bias structure for the relationship of the proxy to the gold standard variable. However, in many contexts, including this one, there are a variety of reasons why the nature of the bias of the proxy is likely to be complicated spatially. We present a simple bivariate spatial model with an additive spatial bias process. We consider the implications of different amounts of spatial variability in the bias process with regard to identifiability and effects on our ability to predict the process of interest. In the application, use of this modeling structure suggests that the bias process is very heterogeneous spatially, and predictions of long-term average ground-level PM essentially ignore AOD. The model attributes spatial variability in AOD to the additive bias process rather than to the latent process of interest. The results suggest that AOD is of little use in helping to predict PM. More generally, we suggest consideration of spatially-correlated error processes in such models, avoiding the assumption that all of error in the proxy is spatial white noise.

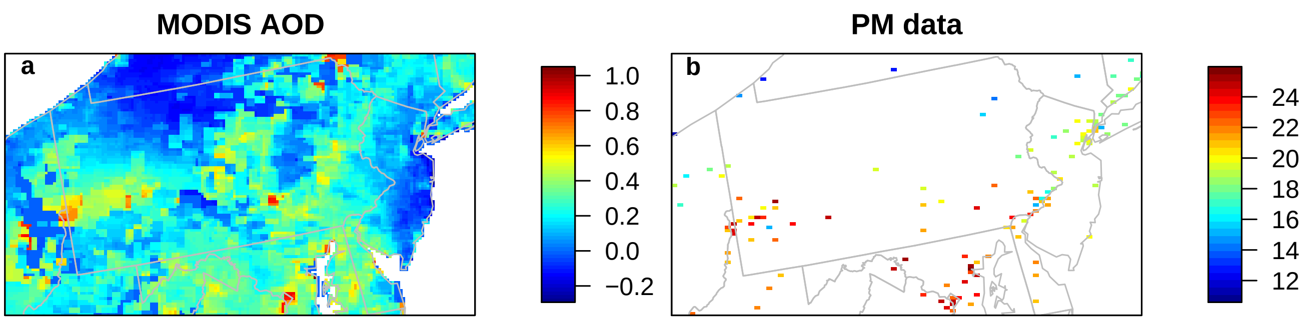
Setting

- Proxy information is increasingly common in environmental science and other applications
  - Deterministic model output
    - Climate models
    - Atmospheric chemistry models
    - Meteorological models
  - Remote sensing information
    - Pollutant concentrations
    - Meteorological variables
    - Land use
  - Proxy data such as biomarkers

- Understanding the discrepancies (biases) between the proxy and the process of interest is critical, but not adequately explored scientifically or statistically.

Particulate Matter Case Study

- AOD (aerosol optical depth) measurements have been suggested as a proxy for ground-level PM ($PM_{2.5}$).
- However, AOD is a noisy and biased proxy, and AOD retrievals are often missing.
- Correlations of daily matched AOD and $PM_{2.5}$ (reflecting both spatial and temporal associations) are on the order of 0.4-0.6.
  - Correlations of daily matched AOD and $PM_{2.5}$ stratified by day (reflecting only spatial associations) are only on the order of 0.2-0.4.
- Key applied question: Can AOD help inform spatial patterns of $PM_{2.5}$ and how does the spatial association vary by scale?



(a) Monthly average MODIS AOD and (b) monthly average $PM_{2.5}$ from ground-level monitors, both for July 2004 in our mid-Atlantic study region of the U.S.

A Basic Data Fusion Model

- Fuentes and Raftery (2005, Biometrics) proposed treating the proxy as a second data source.

- A basic model might have the form:

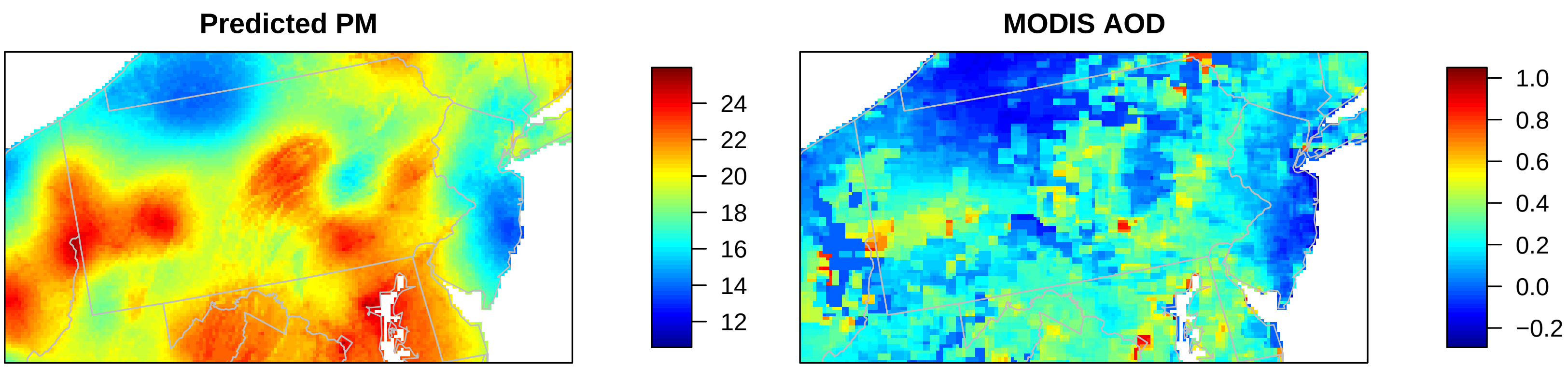
$$
\begin{aligned}
Y(s_i) &\sim \mathcal{N}(P(s_i), \sigma_y^2) \\
A(s_i) &\sim \mathcal{N}(\beta_0 + \beta_1 P(s_i), \sigma_a^2) \\
P(\cdot) &\sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))
\end{aligned}
$$

  where Y is the gold-standard data, A is the proxy information source, and $P(\cdot)$ is the latent process of interest, assuming a Gaussian process structure.

- This model treats the proxy as reflecting the latent process with additive bias, β_0, and multiplicative bias, β_1, plus white noise error.

Implications of Simple Bias Structures

Assuming simple bias with no spatial structure is likely to cause the estimated latent process to strongly reflect the proxy, such as here in the AOD-PM case study. Notice how the patterns in predicted PM follow the patterns in AOD at moderate scales, and the contrast with the patterns in the PM measurements (seen on Slide #3).



Key applied question: Is our prediction of PM distorted by patterns in AOD unrelated to PM?

Flexible Spatial Bias Modeling

- Instead, we propose a model with the additive bias as a spatial process, $\phi(\cdot)$:

$$
\begin{aligned}
Y(s_i) &\sim \mathcal{N}(P(s_i), \sigma_y^2) \\
A(s_i) &\sim \mathcal{N}(\beta_0 + \phi(s_i) + \beta_1 P(s_i), \sigma_a^2) \\
P(\cdot) &\sim \mathcal{GP}(\mu_p(\cdot), C_p(\cdot, \cdot)) \\
\phi(\cdot) &\sim \mathcal{GP}(\mu_\phi(\cdot), C_\phi(\cdot, \cdot))
\end{aligned}
$$

- This extends the polynomial bias in Fuentes and Raftery (2005, Biometrics) and the spline-based bias in McMillan et al. (2008, in submission) by specifying a general spatial bias process, $\phi(\cdot)$.

- The modeling framework allows exploration of the relationship of the proxy and gold standard.
  - We can distinguish small-scale correspondence from large-scale correspondence from no correspondence by analyzing the spatial scales of $\phi(\cdot)$.

Additional Comments on the Flexible Model

- Treating the multiplicative bias, β_1, as a spatial process rather than the additive bias, is another possibility, but treating both spatially leads to identifiability issues.

- One can easily include covariate effects in the various mean terms.

- One can view the model in the form of a factor analysis (or a coregionalization model):

$$\begin{pmatrix} Y \\ A \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \beta_1 & 1 \end{pmatrix} \begin{pmatrix} P \\ \phi \end{pmatrix} + \begin{pmatrix} \epsilon_y \\ \epsilon_a \end{pmatrix}$$

  with spatial factors, P and ϕ, and constrained loadings that reflect our intepretation of the Y measurements as the gold standard. Our interest lies in understanding the partitioning of variability between the two factors.

Bias Scenarios

- $\phi(s)$ very smooth (large-scale variation only):
  - Proxy and gold standard show similar patterns at small and moderate scales, but there is a large-scale bias that causes an offset between proxy and gold standard.
  - $\phi(s)$ is a large-scale bias correction term that should be estimable with a moderate amount of gold standard data.

- $\phi(s)$ wiggly but with little large-scale variation (small-scale variation only):
  - Proxy and gold standard show similar large-scale patterns but small-scale variation in proxy unrelated to the process of interest.
  - $\phi(s)$ is small-scale bias, or equivalently, spatially-correlated error in the proxy.
  - Without dense data, bias cannot be corrected for; rather the model treats this as error that is uninformative about the process of interest.

- $\phi(s)$ with both large- and small-scale variation, $\beta_1 \approx 0$:
  - Little correspondence between proxy and process of interest at any scale.
  - Proxy best described by a separate latent process from the one of interest.

Bias Diagnostics

Jun and Stein (2004; Atmos. Env.) consider scales of model error ($Y - A$) relative to observations (Y) and model output (A) using a ratio of variograms:

$$R(d) = \frac{\text{Variog}(Y - A)}{\text{Variog}(Y) + \text{Variog}(A)}$$

where $R(d) = 1$ if the model output captures none of the variability in the observations at scale d.

We propose a similar diagnostic in the model-based framework as

$$R(d) = \frac{\text{Variog}(\phi)}{\text{Variog}(\beta_1 P) + \text{Variog}(\phi + \beta_1 P)}$$

Interpretation is that $R(d)$ is the spatial bias variability as a proportion of the explained variation in the proxy (disregarding the proxy residual, assumed to be uncorrelated), at scale d.

Computational Issues

- Proxy information from deterministic models and remote sensing is often massive in size.

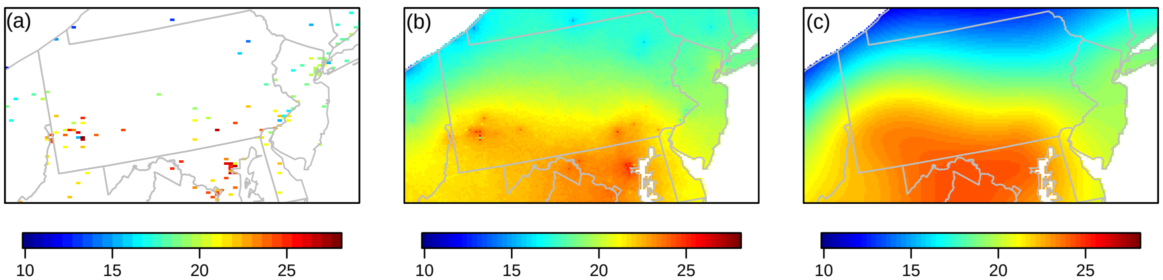
- The model proposed has a likelihood term for the proxy information, which may be computationally prohibitive.

- Some potential approaches include:
  - If ϕ is smooth, reduced rank approaches such as penalized thin plate splines work well (Ruppert et al. 2003 book; Wood 2006 book)
  - If ϕ is wiggly and can be represented on a regular grid, Markov random field approximations to a thin plate spline may be effective, with reliance on sparse matrix routines (Rue and Held 2005 book, Yue and Speckman, in submission)
  - Other techniques for large spatial datasets may also be useful: covariance tapering (Furrer et al. 2006, JCGS; Kaufman et al. in press, JASA), approximate likelihoods (Stein et al. 2004, JRSSB; Fuentes 2007, JASA)

Sidenote: MRF Approximation to the Thin Plate Spline

Standard CAR models used for point data using a fine grid process representation show disturbing behavior, unable to estimate smooth processes. In contrast, MRF thin plate spline approximations can model point data that exhibit either large or small scale spatial variation.

In the example of July 2004 $PM_{2.5}$ (panel a), a standard CAR model on a fine grid shows bulls-eyes around the data and smoothing to the overall mean (panel b), while the MRF thin plate spline approximation has better behavior (panel c):



The key is to use higher order neighbors and to have oscillatory weights motivated by the thin plate spline roughness penalty (Rue and Held, 2005; page 114) (panel a) rather than the standard CAR weight structure (panel b).

(a)

	−1	
−1	4	−1
	−1	

(b)

		1		
	2	−8	2	
1	−8	20	−8	1
	2	−8	2	
		1		

Effects of Bias Structure in the Case Study

For July 2004 for MODIS AOD, the figure (slide #13) shows model-based predictions of PM and estimates of $\phi(\cdot)$, allowing different amounts of complexity in $\phi(\cdot)$ as indicated below. Note that the models include other covariates that help to predict PM.

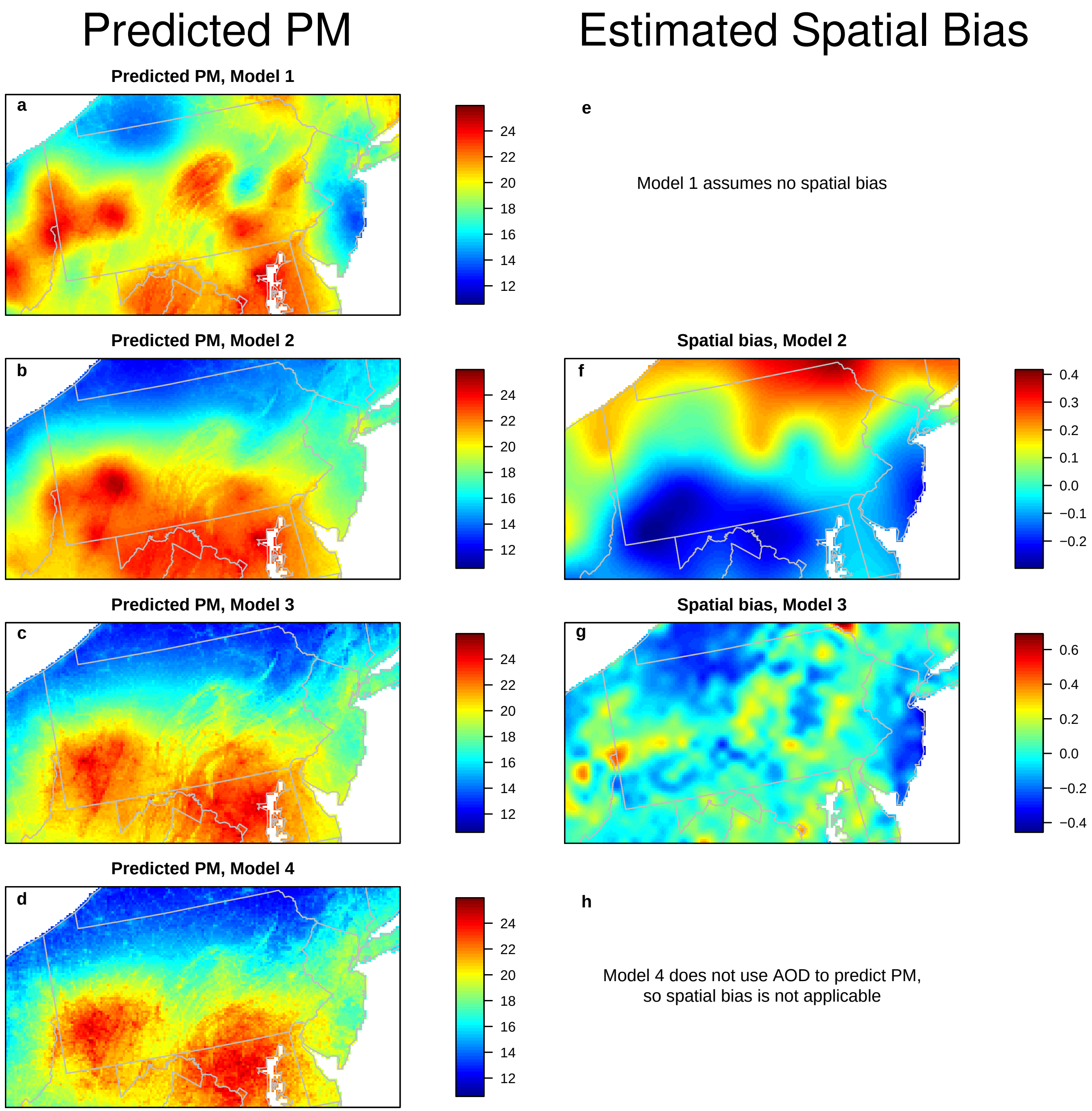
- Model 1: Model omits spatial bias, representing AOD as reflecting PM up to simple additive and multiplicative bias, so predictions of PM strongly track AOD spatial patterns (see Slide #3).

- Model 2: Spatial bias is introduced with $\phi(\cdot)$ as a penalized spline with 55 knots. Predictions track PM observations (see Slide #3) more closely.

- Model 3: Even more flexibility in $\phi(\cdot)$ is allowed (p.s. with 755 knots), with predictions closely following the PM observations and closely following the predictions of Model 4 in which AOD is not included in the model. Note the fine-scale variability in the estimated $\phi(\cdot)$.

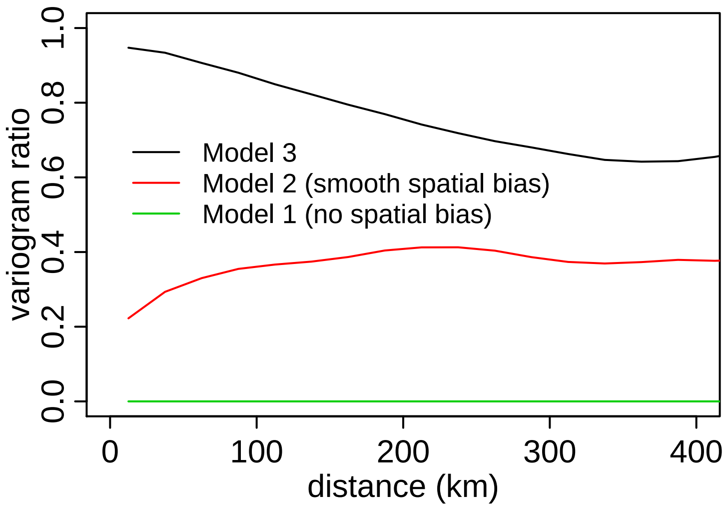
- Model 4: AOD not used.

Note that ideally I would have used the MRF approach for $\phi(\cdot)$, to allow for full flexibility unconstrained by the number of knots, but I have not yet implemented this.


Predicted PM
Estimated Spatial Bias
Predicted PM, Model 1
a
24
22
20
18
16
14
12
e
Model 1 assumes no spatial bias
Predicted PM, Model 2
b
24
22
20
18
16
14
12
Spatial bias, Model 2
f
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
Predicted PM, Model 3
c
24
22
20
18
16
14
12
Spatial bias, Model 3
g
0.6
0.4
0.2
0.0
−0.2
−0.4
Predicted PM, Model 4
d
24
22
20
18
16
14
12
h
Model 4 does not use AOD to predict PM,
so spatial bias is not applicable

Spatial Scales of Variation

The proposed variogram ratio is: $R(d) = \frac{\text{Variog}(\phi)}{\text{Variog}(\beta_1 P) + \text{Variog}(\phi + \beta_1 P)}$



- Under Model 1, none of the proxy variability is explained by the spatial bias, since $\phi(s) \equiv 0$.

- Under Model 2, the bias explains relatively little of the small-scale variability and a moderate amount of the larger-scale variability.

- Under Model 3, the bias explains most of the small-scale variability and much of the larger-scale variability, indicating that AOD is a poor proxy at all scales.
  - Note that even Model 3 constrains the flexibility of $\phi(\cdot)$ artificially because of computational limitations as the number of knots increases, so a fully flexible specification of $\phi(\cdot)$ might indicate a variogram ratio even closer to 1.

Implications for Using AOD to Predict $PM_{2.5}$

- The results suggest there is little common spatial pattern to PM and AOD observations, reflected in Model 3 where the model essentially disregards AOD in predicting PM and attributes most of the variability in AOD to $\phi(\cdot)$.

- Systematic error is considerable and critical to include, and predictions are very sensitive to assumptions about this error.

- It appears that we are in the scenario in which the bias term, $\phi(\cdot)$, varies at both small- and large-scales, with little apparent relationship with the gold standard (β_1 is estimated to be near zero and $R(d)$ is near 1).

- Results for the other 11 months and using GOES AOD or meteorology-adjusted AOD give similar conclusions.

- This suggests that despite raw daily correlations between AOD and $PM_{2.5}$, spatial patterns in AOD provide little useful information for predicting spatial patterns in PM.

Further Assessment of AOD: Raw Correlations

Correlations between AOD and PM, differentiating spatial from temporal association.

Correlations of raw and calibrated daily AOD with matched 24-h PM in 2004 for the eastern U.S. (top portion) and correlations of raw and calibrated yearly-average AOD with yearly-average PM (sites with at least 100 daily PM observations, matched in space to AOD) for our mid-Atlantic focal region in 2004 (bottom portion). Yearly averages reflect all available AOD retrievals and all available 24-h average PM concentrations. Calibrated AOD has been adjusted to account for the effects of planetary boundary layer (PBL) height, relative humidity (RH), season, and regional variation in modifying the relationship between daily AOD and PM. Yearly results exclude one site outside Pittsburgh with high PM levels that is downwind of a major industrial facility.

	Raw AOD			Calibrated AOD		
	MODIS	MISR	GOES	MODIS	MISR	GOES
	Daily values, eastern U.S.					
Temporal plus spatial variation: Overall correlation of daily values across all sites and days.	0.60	0.50	0.38	0.64	0.57	0.40
Spatial variation only: Average of daily spatial correlations (only days with at least 20 matched sites) .	0.35	0.30	0.23	0.45	0.32	0.29
	Yearly averages, mid-Atlantic focal region					
Spatial variation only: Correlation of yearly averages.	0.09	0.25	-0.07	0.49	0.22	0.53

Further Assessment of AOD: AOD as a covariate for PM

Accuracy of prediction of PM when fitting a model for the PM observations with AOD as a covariate rather than a separate likelihood term

Cross-validation R^2(mean squared prediction error) for yearly average (i.e., space-only variation) and monthly average (i.e., space-time variation) predictions of PM from regression style models with PM data as the gold standard, with and without calibrated AOD and other predictors. The 'population exposure' designation assigned to monitors by EPA indicates that such monitors are not likely to be affected by large, local sources. Results exclude one site outside Pittsburgh with high PM levels that is downwind of a major industrial facility.

Model	Yearly averages		Monthly averages	
	All monitors (n=151)	Population exposure monitors (n=130)	All monitors (n=1793)	Population exposure monitors (n=1542)
	Models including land use, emissions, and meteorological predictors			
No AOD	0.580 (1.04)	0.570 (0.93)	0.827 (2.71)	0.839 (2.48)
With calibrated MODIS AOD	0.573 (1.06)	0.564 (0.94)	0.825 (2.73)	0.839 (2.50)
With calibrated GOES AOD	0.572 (1.06)	0.563 (0.95)	0.825 (2.73)	0.838 (2.50)
	Models without land use, emissions, and meteorological predictors			
No AOD	0.401 (1.49)	0.385 (1.33)	0.784 (3.38)	0.799 (3.11)
With calibrated MODIS AOD	0.467 (1.32)	0.459 (1.17)	0.794 (3.22)	0.810 (2.94)
With calibrated GOES AOD	0.467 (1.33)	0.458 (1.17)	0.794 (3.22)	0.810 (2.94)

Results support the conclusion that AOD provides little benefit in helping predict PM when other information is available.

- Advantages of the covariate approach:
  - Direct estimation of the regression coefficient for AOD
  - Ease of interpretation

- Disadvantages:
  - Doesn't easily handle missing AOD
  - Doesn't address scale issues directly.

Reasons for the Mismatch between AOD and PM$_{2.5}$

- The limitations of AOD seen here contrast somewhat with the empirical correlations seen in previous studies. We attribute this primarily to the difference between temporal association and spatial association. Because temporal associations are not affected by spatial variability in conditions affecting the aerosol retrievals, using AOD as a temporal proxy works better than as a spatial proxy.

- Some potential causes of spatial variability that interfere with spatial association of AOD and PM$_{2.5}$:
  - Spatial variability in surface reflectivity.
  - Spatial variability in aerosol chemical composition and size distributions.
  - Spatially-coherent missingness due to daily cloud cover, with aggregate effects for longer-term averages.
  - Spatial structure in pollution aloft in the atmosphere
  - Spatial structure in pollution at times not captured by the satellite (night-time and hours with no satellite coverage)

General Conclusions

- We need to be more explicit about our assumptions about proxy information and potential bias/spatially-correlated error.
- We know there is error in proxy data sources and in many situations this is likely to be spatially-correlated, and at fine scales, so white noise error, while convenient, may not be appropriate.
- Modeling as bias/spatially-correlated error provides for careful assessment of the variation in the proxy, considering scales of concordance and discordance.
- This approach can enhance simple deterministic model validation, which is often done via scatterplots and R^2 calculations.

Introduction

Model Framework

Case Study

Final Thoughts