Flexible Spatial Latent Variable Modeling for Proxy Data with Systematic Discrepancy

Chris Paciorek
Department of Statistics; University of California, Berkeley
and
Department of Biostatistics; Harvard School of Public Health

www.biostat.harvard.edu/~paciorek

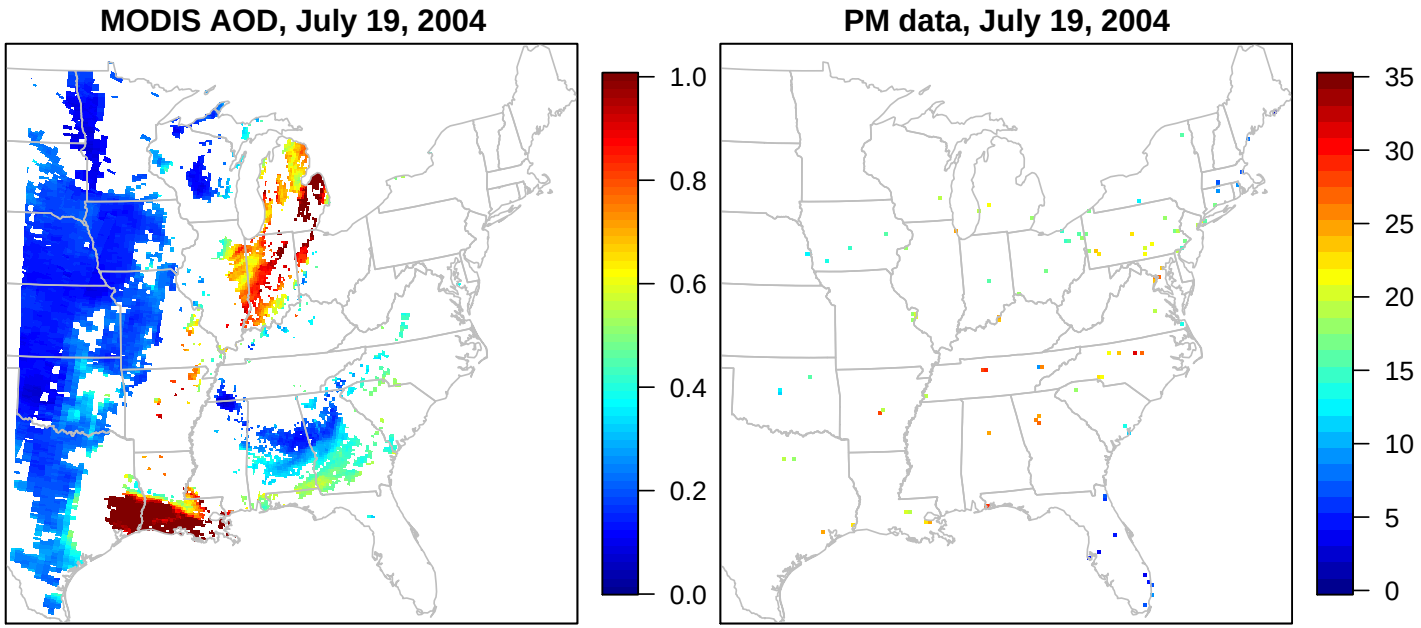
Research supported by HEI 4746-RFA05-2/06-7

August 11, 2010

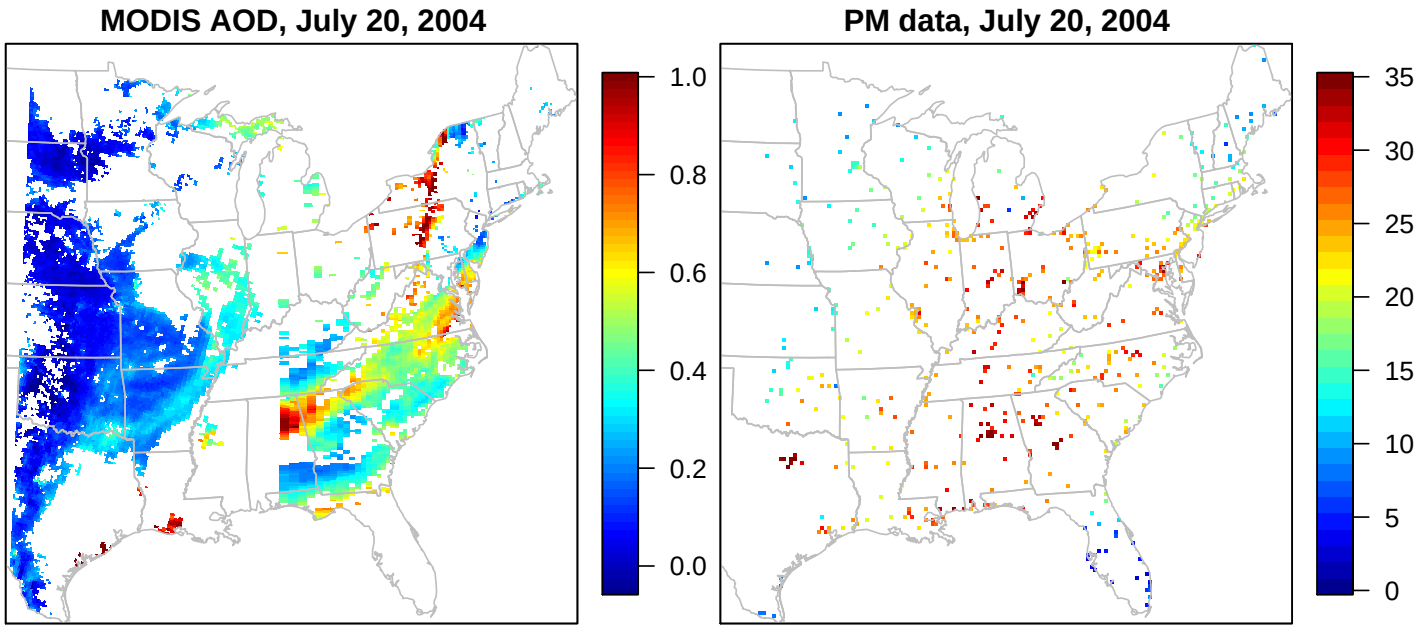
Proxy Information in Environmental Applications

- Proxy information is increasingly common in environmental science and other applications
- Deterministic model output
  - Climate models
  - Atmospheric chemistry models
  - Meteorological models
  - Hydrologic and subsurface models
- Remote sensing information
  - Pollutant concentrations
  - Meteorological variables
  - Land use, land change
  - (Seismic data)

Combining Information



Combining Information



Challenges of Proxy Information

- Systematic spatial (and temporal) discrepancy between proxy and truth
  - White noise error structure often implausible
  - This impacts predictions, prediction uncertainty, and assessment of proxy usefulness
  - Ignoring the discrepancy leads to overinterpreting patterns in the proxy
  - Proxy may not directly quantify the process of interest, hence 'discrepancy' rather than 'error' or 'bias'
- Spatial misalignment of gridded proxy information and point-level observations
  - Temporal misalignment can also be an issue
- Proxy datasets are usually very large
  - Standard GP modeling is infeasible

Prediction of Fine Particulate Matter (PM)

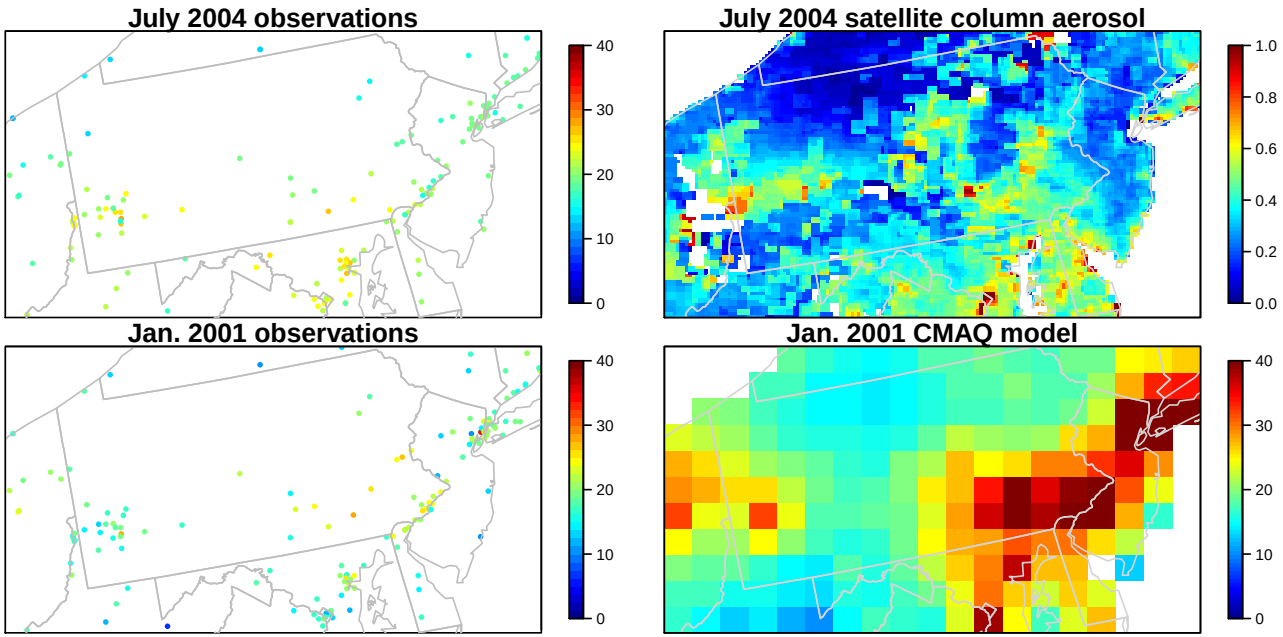
Proxy sources:

- Satellite-derived Aerosol Optical Depth (AOD)
  - Integrated vertical column measurement based on light reflecting off the earth surface
  - Gridded
  - Lots of missing data
- Atmospheric chemistry model output (CMAQ)
  - Gridded, no missing data

Gold standard:

- Ground monitoring network
  - Point-level observations
  - Influenced by local heterogeneity in PM

PM Information



A Basic Data Fusion Model

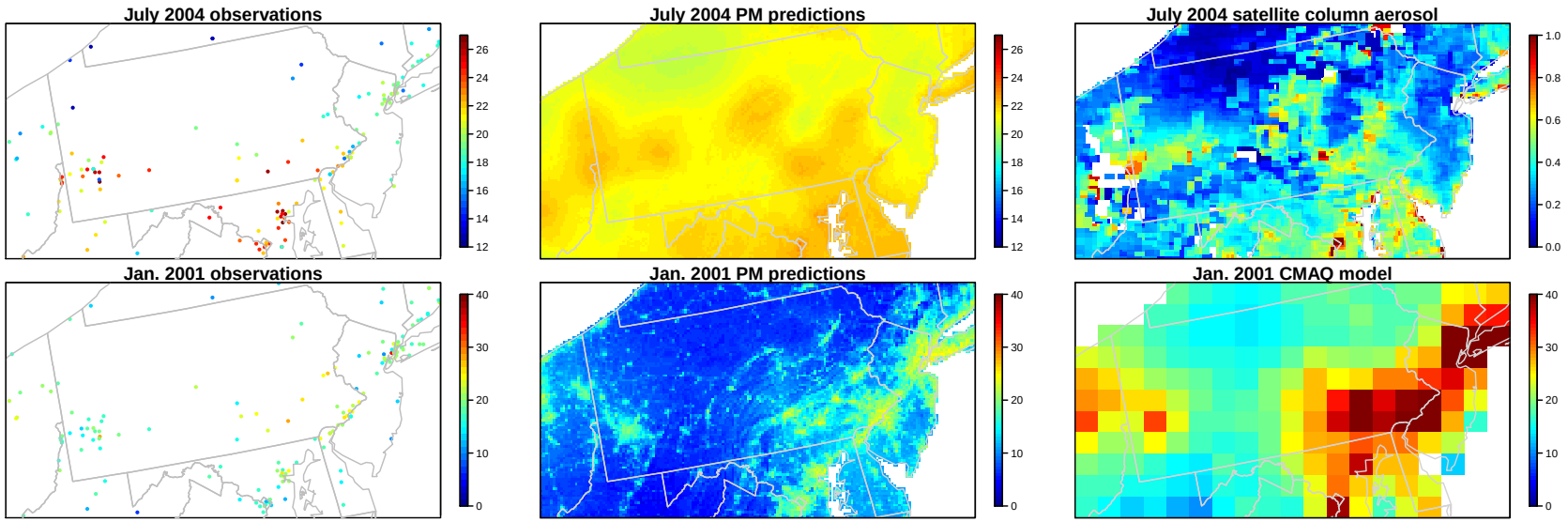
- Fuentes and Raftery (2005, Biometrics) proposed treating the proxy as a second data source.
- A basic model:

$$\begin{aligned} Y_i &\sim \mathcal{N}(L(s_i), \sigma_y^2) \\ A_m &\sim \mathcal{N}(\beta_0(s) + \beta_1 L(s_m), \sigma_a^2) \\ L(\cdot) &\sim \mathcal{GP}(\mu(\cdot), C(\cdot,\cdot)) \end{aligned}$$

  where Y is the gold-standard data, A is the proxy information source, and $L(\cdot)$ is the latent process of interest.
- This model treats the proxy as reflecting the latent process with additive bias, $\beta_0(s)$, and multiplicative bias, β_1, plus white noise error.
  - The additive bias, $\beta_0(s)$, in Fuentes and Raftery (2005) was polynomial in s.

Implications of Simple Bias Structures

Predictions Based on Non-spatial Bias



Predictions of the process of interest appear to be distorted by unrelated patterns in the proxy.

Flexible Spatial Discrepancy Modeling

- Consider additive bias as a spatial discrepancy process, $D(\cdot)$:

$$\begin{aligned}
\mathbf{Y} &\sim \mathcal{N}(\mu_y(\mathbf{x}) + \mathbf{K}_y\mathbf{L}, \sigma_y^2\mathbf{I}) \\
\mathbf{A} &\sim \mathcal{N}(\mathbf{K}_A\mathbf{D} + \beta_1\mathbf{K}_A\mathbf{L}, \sigma_a^2\mathbf{I}) \\
\mathbf{L} &\sim \text{MRF}(\mu_L(\mathbf{x}), \mathbf{Q}_L) \\
\mathbf{D} &\sim \text{MRF}(\mu_D(\mathbf{x}), \mathbf{Q}_D)
\end{aligned}$$

- Latent processes, $L(\cdot)$ and $D(\cdot)$, are represented on a fine grid.
- We can explore the relationship of the proxy and gold standard through analysis of the spatial scales of $D(\cdot)$.
- $\mu_y(x)$ involves the effect of covariates that explain sub-grid scale variation in the point measurements, while $\mu_L(x)$ and $\mu_D(x)$ are covariate effects on the grid-scale process and the discrepancy term, respectively.

Discrepancy Scenarios

- $D(\cdot)$ very smooth (large-scale variation only):
  - Proxy and gold standard show similar patterns at small and moderate scales, but there is a large-scale discrepancy that causes an offset between proxy and gold standard.
  - $D(\cdot)$ is a large-scale bias correction term that should be estimable with a moderate amount of gold standard data.
- $D(\cdot)$ 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 gold standard.
  - $D(\cdot)$ is small-scale discrepancy, or equivalently, spatially-correlated error in the proxy.
  - Without dense data, discrepancy cannot be corrected for; model treats it as error that is uninformative about the latent process.
- $D(\cdot)$ 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.

A Markov Random Field Model

- Rue and Held (2005) and Yue and Speckman (2010; JCGS) describe a MRF that approximates a thin plate spline (TPS).
  - The weights of the (sparse) precision matrix are determined from the discrete approximation to the TPS penalty:

$$J(g) = \int \int_{\Re^2} \left[\left(\frac{\partial^2 g(s_1,s_2)}{\partial s_1^2} \right)^2 + 2 \left(\frac{\partial^2 g(s_1,s_2)}{\partial s_1 \partial s_2} \right)^2 + \left(\frac{\partial^2 g(s_1,s_2)}{\partial s_2^2} \right)^2 \right] ds_1 ds_2.$$

Standard CAR

$$\begin{array}{ccc} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{array}$$

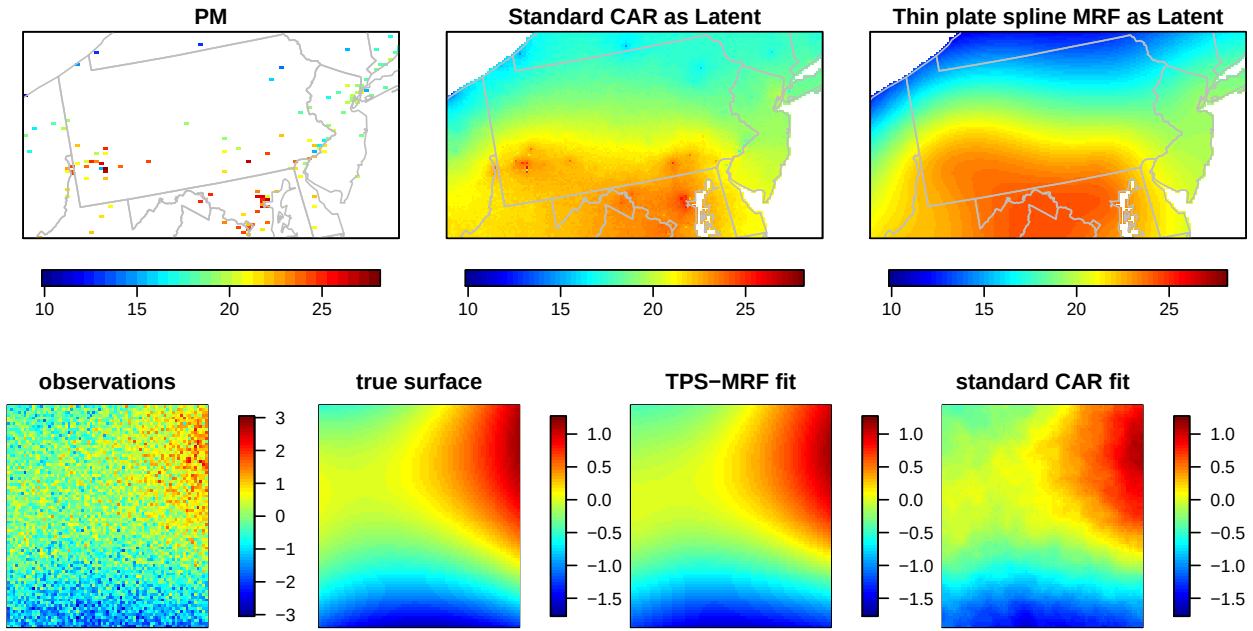
Thin plate spline MRF approximation

$$\begin{array}{ccccc} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & 20 & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{array}$$

Precision matrix elements for one row of $\mathbf{Q}$, oriented spatially (with respect to that row's focal grid cell) to indicate neighborhood structure.

Comparing the TPS-MRF with a traditional CAR model

- The asymptotic limit of a traditional intrinsic Gaussian CAR model is two-dimensional Brownian motion (the de Wijs process) (Besag and Mondal 2005), with continuous but not differentiable realizations.
- TPS-MRF realizations can be either globally smooth or just locally smooth.



Features of the MRF Approach

- Advantages
  - The TPS approximation can capture smoothly-varying large-scale variation as well as fine-scale spatial patterns.
  - Misalignment is handled through:
    - Weighted averages of grid cells as approximation to integral
    - Assignment of grid cell value to points, with offset regression terms
  - Sparse prior precision matrix provides computational efficiency.
- Disadvantage?
  - Splines can behave badly in gaps and edges of the domain.

MCMC Considerations

- Cross-level dependence between latent processes and hyperparameters controlling their dependence are a major impediment to good MCMC performance.
  1. Here, with conjugacy, we can marginalize over the latent process values to avoid this dependence.
  2. Joint proposals are another possibility.
- In thinking about marginalization and subsequent MCMC, one might avoid marginalization that will introduce off-diagonal structure in large covariance matrices.

Computational Strategy: Exploiting Sparsity

1. Integrate first over $\{\mathbf{D}, \mathbf{L}\}$, then over $\{\mu_y, \mu_L, \mu_D\}$ so that resulting marginal posterior still involves sparse matrices.
   - I.e., exploit matrix identities to get matrix representations that retain sparsity and avoid $\mathbf{Q}^{-1}$.
2. In marginal posterior computations, exploit the sparse structure appropriately.

$$P(\theta, \delta|\mathbf{A}, \mathbf{Y}) \propto |\mathbf{\Lambda}|^{-\frac{1}{2}}|\mathbf{V}_Y|^{-\frac{1}{2}}|\mathbf{\Sigma}_A|^{-\frac{1}{2}}|\mathbf{V}_b|^{\frac{1}{2}}P(\delta)P(\theta)\cdot$$
$$\exp(-\tfrac{1}{2}((\mathbf{Y}-\mathbf{K}_\delta\delta)^T\mathbf{V}_Y^{-1}(\mathbf{Y}-\mathbf{K}_\delta\delta)+\mathbf{A}^T\mathbf{\Sigma}_A^{-1}\mathbf{A}-\mathbf{M}_b^T\mathbf{V}_b^{-1}\mathbf{M}_b))$$
$$\mathbf{V}_b = (\mathbf{Z}_Y^T\mathbf{V}_Y^{-1}\mathbf{Z}_Y + \mathbf{Z}_A^T\mathbf{\Sigma}_A^{-1}\mathbf{Z}_A + \mathbf{\Lambda}^{-1})^{-1}$$
$$\mathbf{\Sigma}_A^{-1} = \mathbf{V}_A^{-1} - \mathbf{V}_A^{-1}\mathbf{K}_D\mathbf{V}_D\mathbf{K}_D^T\mathbf{V}_A^{-1}$$
$$\mathbf{V}_D = (\mathbf{K}_D^T\mathbf{V}_A^{-1}\mathbf{K}_D + \kappa\mathbf{Q})^{-1}$$

Avoid integrating over δ, site-specific random effects, as this would introduce off-diagonal structure.

A Scale-specific Discrepancy Diagnostic

- Jun and Stein (2004; Atmos. Env.) consider scales of model error ($\mathbf{Y} - \mathbf{A}$) relative to observations ($\mathbf{Y}$) and model output ($\mathbf{A}$):

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

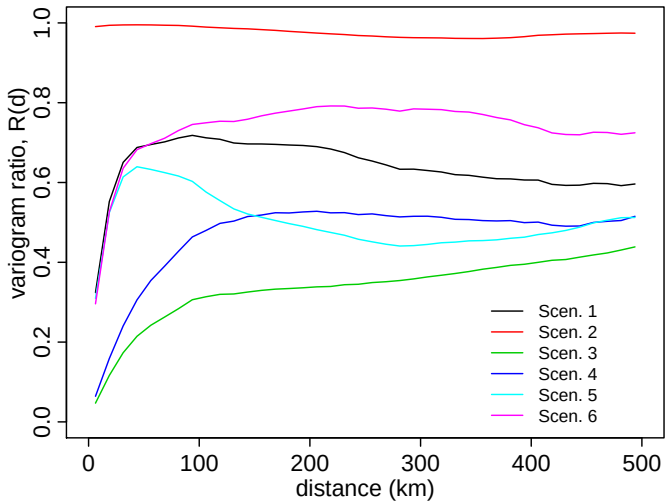
  where $R(d) = 1$ if the model output captures none of the variability in the observations at scale (distance) d.

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

$$R(d) = \frac{\text{Variog}(\mathbf{D})}{\text{Variog}(\beta_1 \mathbf{L}) + \text{Variog}(\mathbf{D} + \beta_1 \mathbf{L})}$$

  $R(d)$ is the spatial discrepancy variability as a proportion of the explained variation in the proxy, at scale d.

Simulations: Trying Out the Diagnostic



1. Large- and small-scale discrepancy; no signal in proxy
2. Sparse observations ($n = 40$), with some small-scale discrepancy and a minor amount of large-scale discrepancy, plus signal
3. Large- and small-scale discrepancy present, plus signal
4. Small-scale discrepancy only, plus signal
5. Large-scale discrepancy only, plus signal
6. No spatial discrepancy, plus signal (+ white noise discrepancy)

Simulations: Prediction Results

- Proxy = signal + systematic discrepancy:
  - Prediction R^2 decreases from c. 0.8 to c. 0.7 when including proxy.
  - Ignoring spatial discrepancy leads to much worse predictive performance.
  - Using the proxy as a regressor improves prediction (to 0.82-0.90).
- Proxy = signal + white noise discrepancy:
  - Prediction R^2 decreases from c. 0.8 to c. 0.7 when including proxy: model attributes some of the signal to discrepancy.
  - Assuming no spatial discrepancy improves prediction (c. 0.95)
  - Using the proxy as a regressor improves prediction (c. 0.90)
- Proxy = discrepancy: Model correctly discounts proxy.
- With sparse data:
  - Model with proxy outperforms model without proxy.
  - Using the proxy as a regressor outperforms the dual likelihood model.

Core Spatial Models with AOD

- For monthly PM and AOD in the mid-Atlantic in 2004, I fit spatial models for each month

$$
\begin{aligned}
\mathbf{Y} &\sim \mathcal{N}_{n_Y}(\mathbf{Z}_y\mathbf{b}_y + \mathbf{K}_Y\mathbf{L} + \mathbf{K}_\delta\boldsymbol{\delta}, \mathbf{V}_Y) \\
\mathbf{A} &\sim \mathcal{N}_{n_A}(\mathbf{K}_A\mathbf{D} + \mathbf{Z}_a\mathbf{b}_a + \beta_1\mathbf{K}_A\mathbf{L}, \mathbf{V}_A) \\
\boldsymbol{\delta} &\sim \mathcal{N}(\mathbf{0}, \sigma_h^2\mathbf{I}) \\
\mathbf{D} &\sim \mathrm{MRF}_{17500-3}(\mathbf{0}, \kappa\mathbf{Q}) \\
\mathbf{b} = \{\mathbf{b}_y, \mathbf{b}_L, \mathbf{b}_a\} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})
\end{aligned}
$$

  - $\mathbf{L} = \mathbf{Z}_L\mathbf{b}_L + \mathbf{g}$
  - $\mathbf{Z}_L\mathbf{b}_L$ represents grid-resolved covariates (population density, road density, elevation, area emissions) while $\mathbf{Z}_Y\mathbf{b}_Y$ represents local effects (distance to major roads and point sources)
  - $\mathbf{g}$ represented as a penalized spline.
  - Our core grid is a regular 4 km grid with 17500 cells.

Additional Models

- Spatio-temporal model with CMAQ output
  - For monthly PM and CMAQ PM in the mid-Atlantic in 2001, I fit spatio-temporal models with an autoregressive structure on the spline coefficients of $\mathbf{g}$ and an exchangeable structure,

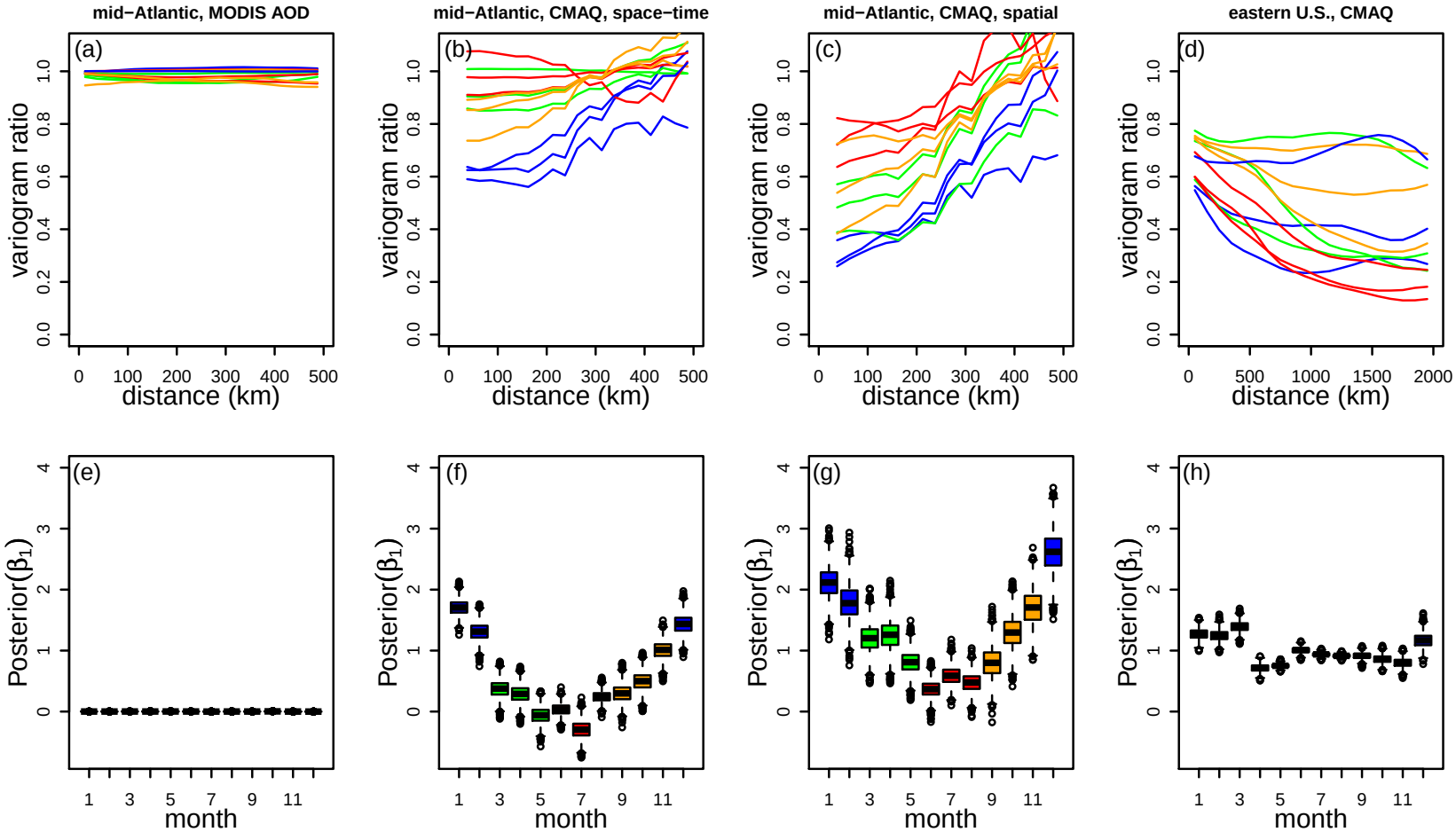
$$\mathbf{D}_t \sim \text{MRF}_{219-3}(\mathbf{D}, \kappa_1 \mathbf{Q})$$
$$\mathbf{D} \sim \text{MRF}_{219-3}(\mathbf{0}, \kappa_2 \mathbf{Q}).$$

    Adding an AR(1) structure is straightforward.
- Spatial models with CMAQ output in the eastern US
  - For monthly PM and CMAQ PM in the eastern U.S. in 2001, the model is similar to the core model but with $\mathbf{g}$ and $\mathbf{D}$ both represented as TPS-MRFs on a 36 km grid with 5621 cells.

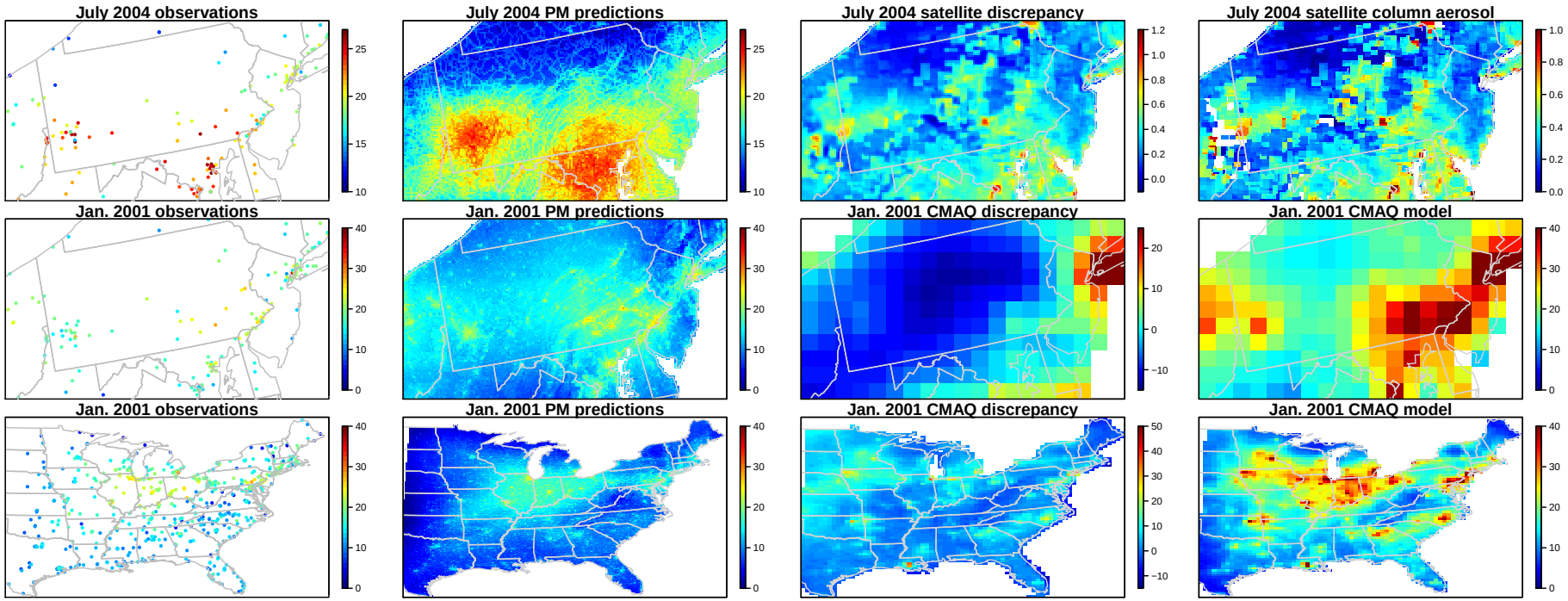
Results

- Satellite AOD:
  - The model fitting suggests there is little common spatial pattern to PM and AOD observations.
    - The discrepancy term, $D(\cdot)$, varies at both small and large scales.
  - As a result the model discounts AOD in predicting PM.
- Atmospheric Chemistry Model (CMAQ):
  - More apparent relationship between CMAQ output and latent PM.
    - The discrepancy term also varies at small and large scales, but more of the variation in the proxy appears to be signal than for AOD.
  - Statistical model still heavily discounts the proxy.

Discrepancy Diagnostic



Predicted PM



Cross-validation predictive ability, R^2 (RMSPE)

Time scale	Proxy?	mid-Atlantic, 2004, MODIS AOD	mid-Atlantic, 2001, CMAQ, space-time	mid-Atlantic, 2001, CMAQ, spatial models	eastern U.S., 2001, CMAQ
Monthly[1]	w/ proxy	0.806 (1.80)	0.640 (2.60)	0.755 (2.14)	0.827 (1.71)
	no proxy	0.808 (1.79)	0.686 (2.42)	0.777 (2.04)	0.826 (1.72)
	as regr.				0.849 (1.60)
Yearly[2]	w/ proxy	0.668 (1.00)[3]	<0[4] (1.97)[3]	0.503 (1.32)[3]	0.800 (1.21)
	no proxy	0.650 (1.03)[3]	0.169 (1.70)[3]	0.584 (1.20)[3]	0.835 (1.09)
	as regr.				0.849 (1.05)

[1] Including monthly averages based on at least five daily observations.
[2]Including yearly averages (averages of monthly values) based on at least nine months with at least five daily observations.
[3] Excludes one site outside Pittsburgh just downwind of a major industrial facility.
[4]Squared correlation of held-out data and predictions is 0.473, but observations vs. predictions are not centered on the one to one line, so error sum of squares exceeds total sum of squares.

Conclusions (1)

- We need to be more explicit about our assumptions about the error structure of proxies.
  - White noise error, while convenient, is generally not appropriate.
  - Modeling the discrepancy can help to enhance simple deterministic model assessment.
    - Standard validation relies on scatterplots and R^2 calculations.
    - Modeling the discrepancy allows us to consider scales of concordance and discordance.

Conclusions (2)

- Distinguishing spatio-temporal signal from spatio-temporal noise is difficult and likely sensitive to modeling assumptions.
  - Additivity assumptions, error structures, spatial field representations.
  - Is there useful information in the proxies that the current model structure is not exploiting?
- Here we had relatively abundant gold standard data, but often this won't be the case and prior assumptions about the correlation structure of the error will be critical.
  - One prominent application is in climate model uncertainty quantification.
  - What can be said about uncertainty in regional climate projections?