Quantifying local creation and regional transport using a hierarchical space-time model of ozone as a function of observed NO_x, a latent space-time VOC process, emissions, and meteorology.

The 19th Annual Conference
of
The International Environmetrics Society
Kelowna, BC

June 9, 2008

Amy J. Nail

Jacqueline M. Hughes-Oliver
John F. Monahan

Outline

1. Context, goals, and data
2. The model
    - Created ozone
    - Transported ozone
    - Log VOC
    - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

Outline

1. **Context, goals, and data**
2. The model
    - Created ozone
    - Transported ozone
    - Log VOC
    - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

Ozone Regulatory Context

- Ozone causes respiratory problems in humans and damages crops and forests.
- EPA sets National Ambient Air Quality Standard (NAAQS) based on health effects studies.
- Current NAAQS for Ozone says the "three-year average of the annual fourth-highest daily maximum **8-hour average** concentration" must fall beneath **80 ppb** (EPA 2004).
- Ozone is a secondary pollutant.
- NOx + VOC + sunlight $\rightarrow O_3$
- NOx $\Leftarrow$ powerplants, cars, industry
- VOC $\Leftarrow$ cars, industry, TREES!

Original goals based on regulatory needs

Formulate a **process-based** space-time **statistical** model of 8-hour ozone as a function of emissions data and meteorology to allow:

1. Quantification of local creation vs. regional transport
2. Space-time predictions of 8-hour ozone to be used in
   - Health and ecosystem effect studies
   - Attainment designations
3. Assessment of past and future emission control programs
   - Did ozone decrease? Did it decrease because of changes in emissions that actually occurred?
   - Will a proposed emissions control program reduce ozone in the future? How much?
4. Automatic quantification of uncertainty

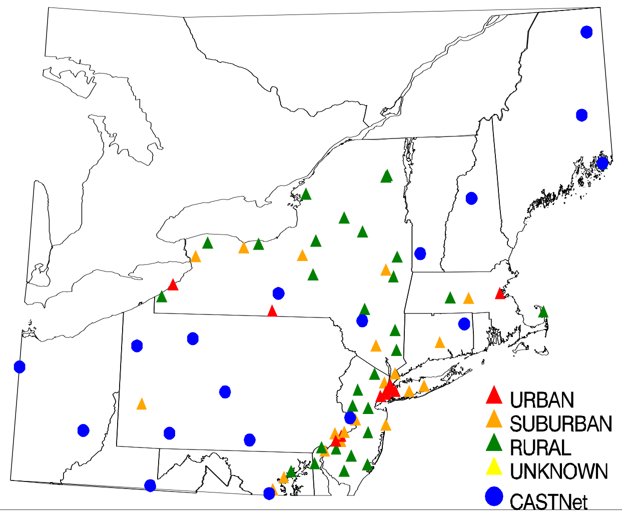
Achieved goals

Formulate a **process-based** space-time **statistical** model of 8-hour ozone as a function of emissions data and meteorology to allow:

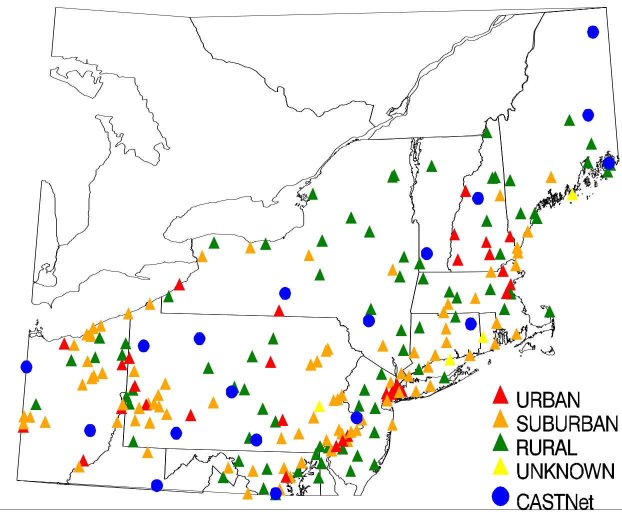
1. **Quantification of local creation vs. regional transport**
2. Space-time predictions of 8-hour ozone to be used in
   - Health and ecosystem effect studies
   - Attainment designations
3. Assessment of past and future emission control programs
   - Did ozone decrease? Did it decrease because of changes in emissions or because of changes in meteorology?
   - Will a proposed emissions control program reduce ozone in the future? How much?
4. **Automatic quantification of uncertainty**

Ozone: N=54k dataset

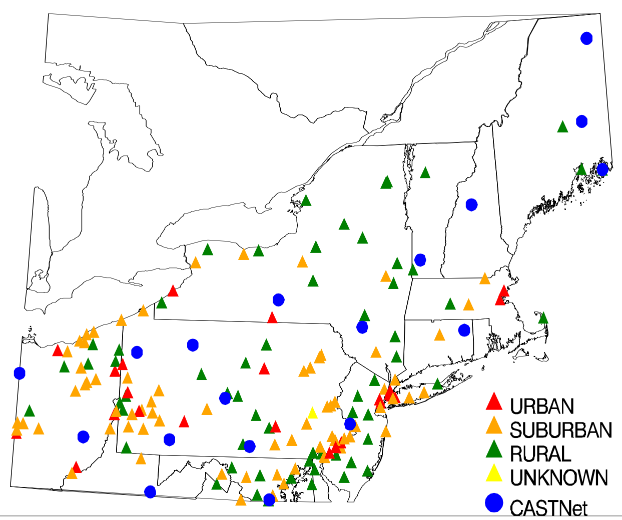
SLAMS NAMS PAMS and CASTNet Sites JAN−APR



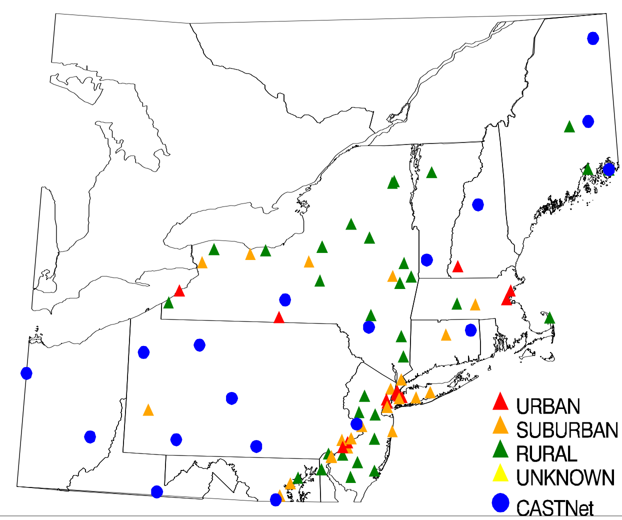
SLAMS NAMS PAMS and CASTNet Sites MAY−SEP



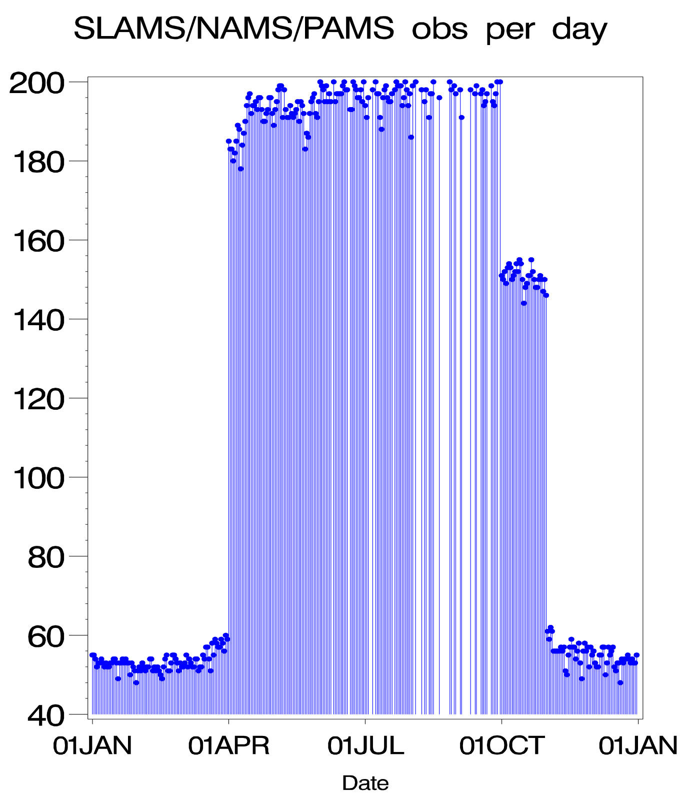
SLAMS NAMS PAMS and CASTNet Sites OCT



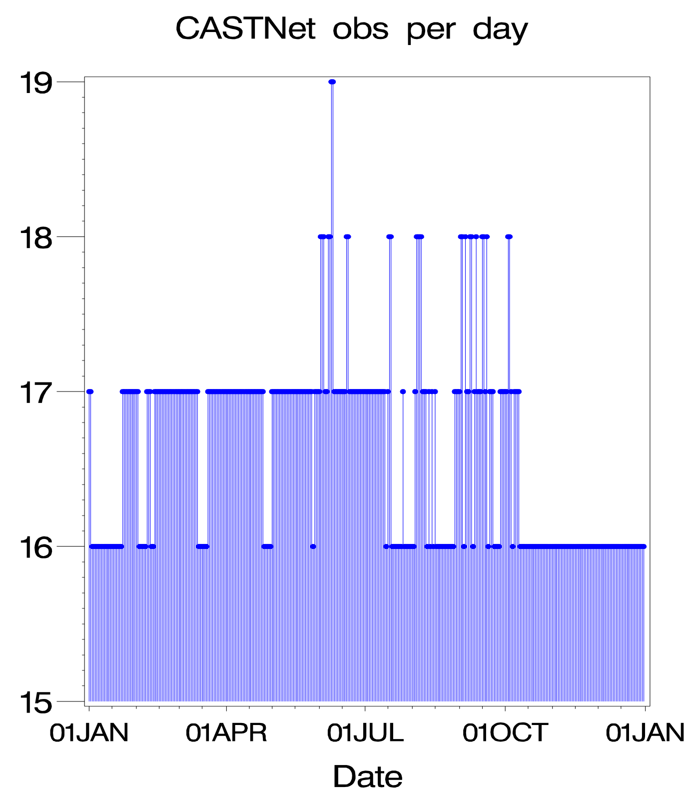
SLAMS NAMS PAMS and CASTNet Sites NOV−DEC



Ozone: N=54k dataset

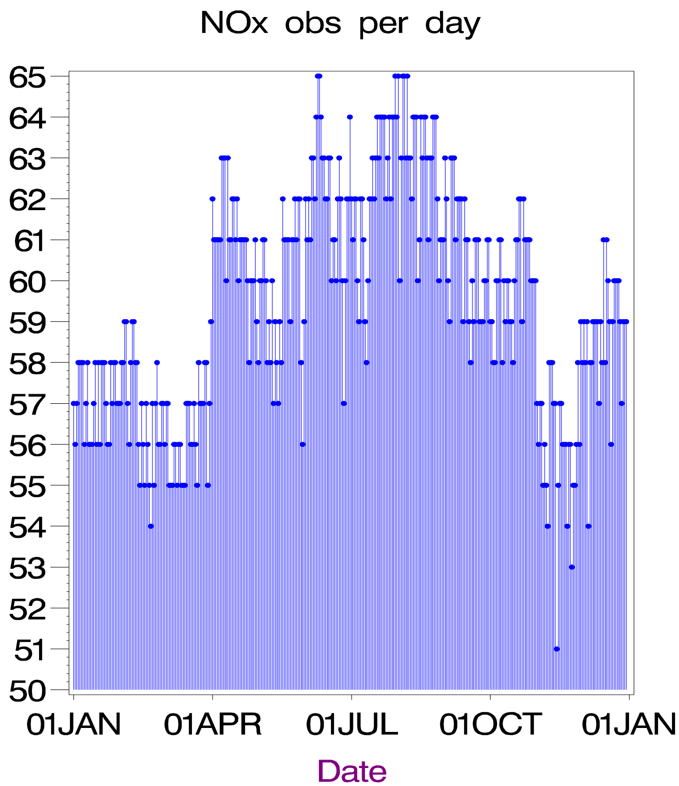


(a)

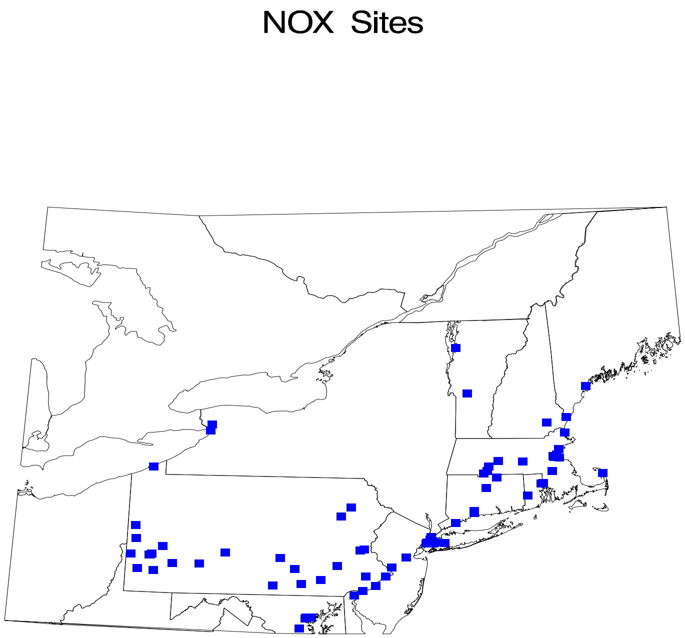


(b)

NO_x : N=21k dataset

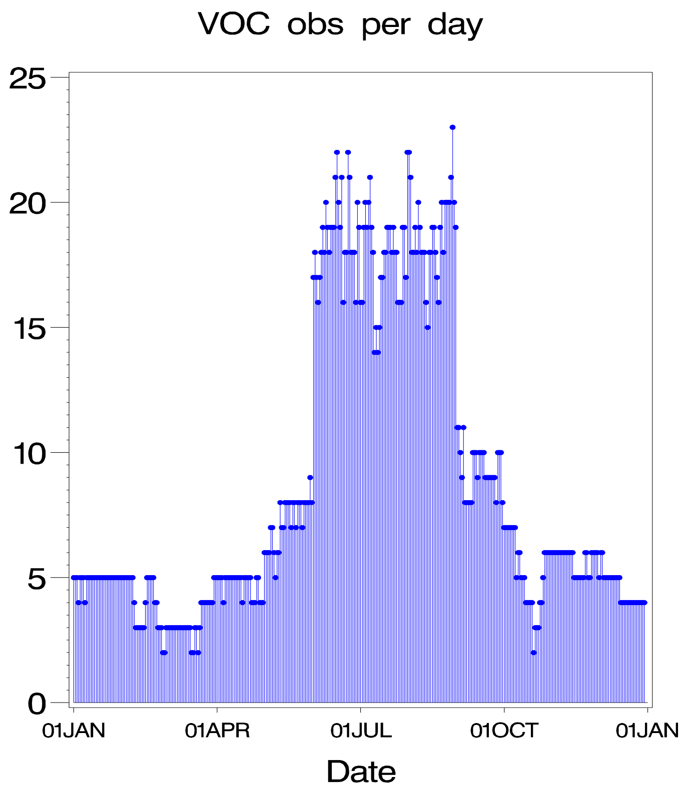


(c)

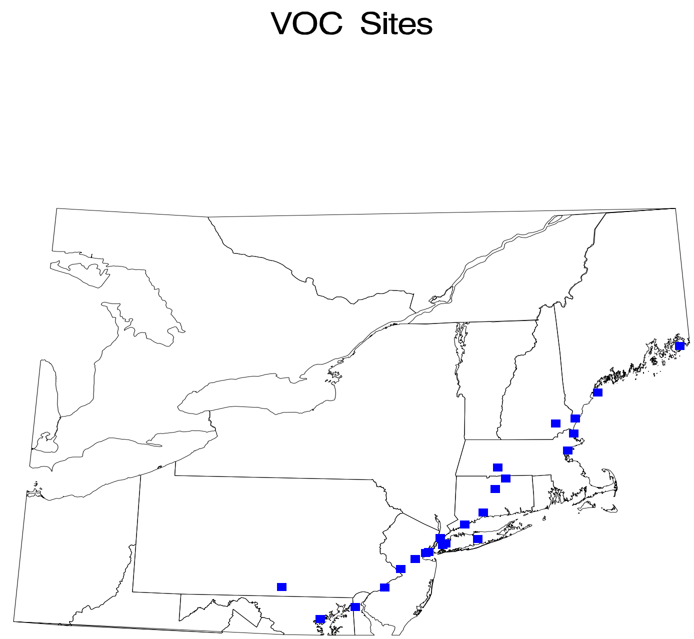


(d)

VOC: N=3k dataset

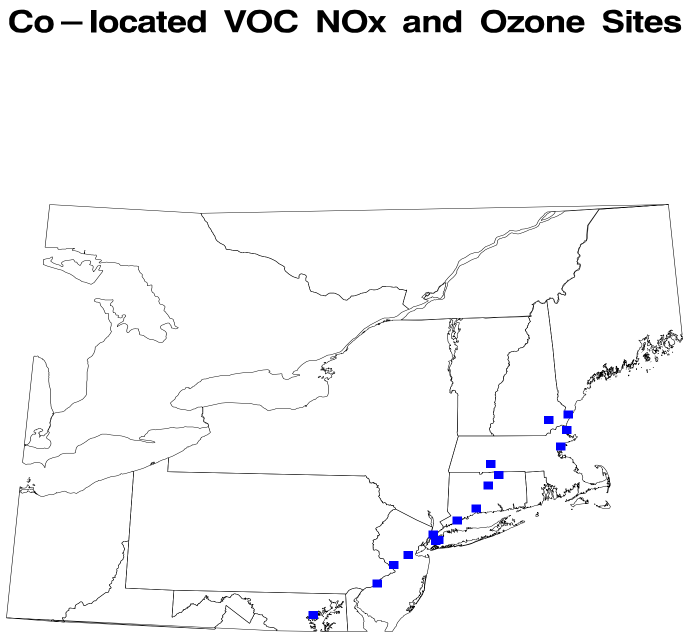


(e)

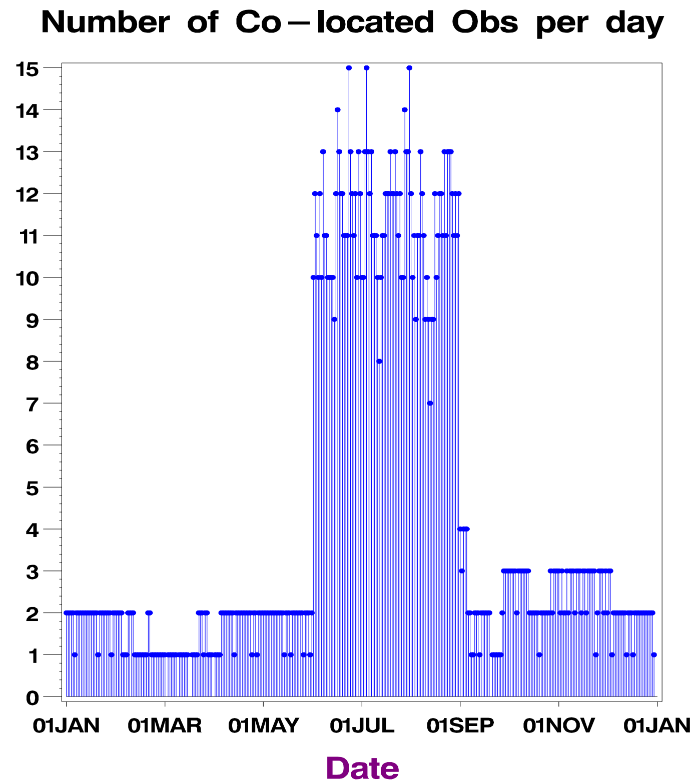


(f)

Co-located O_3, NO_x, and VOC: N=1563 dataset

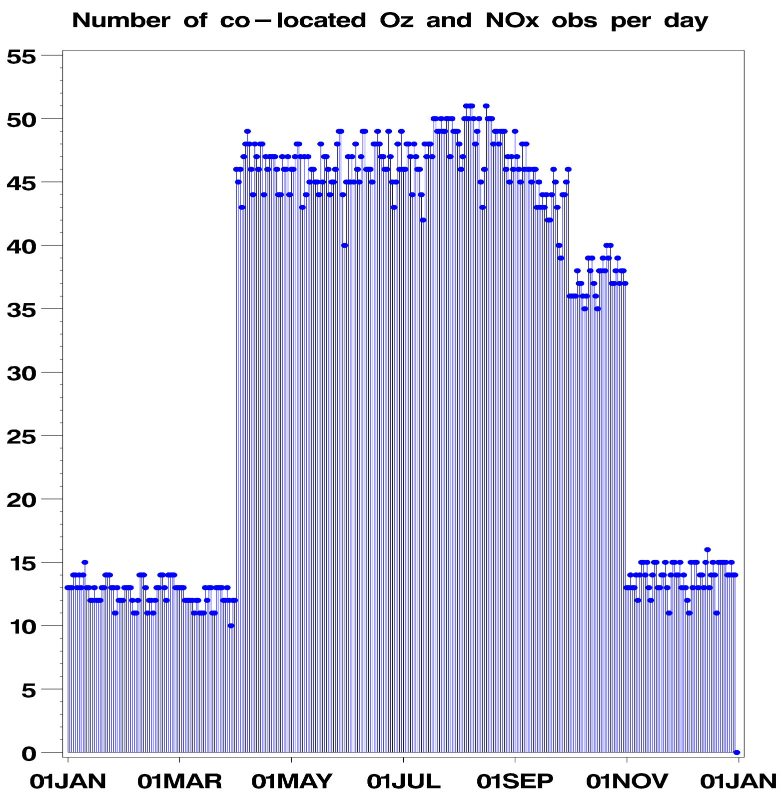


(g)

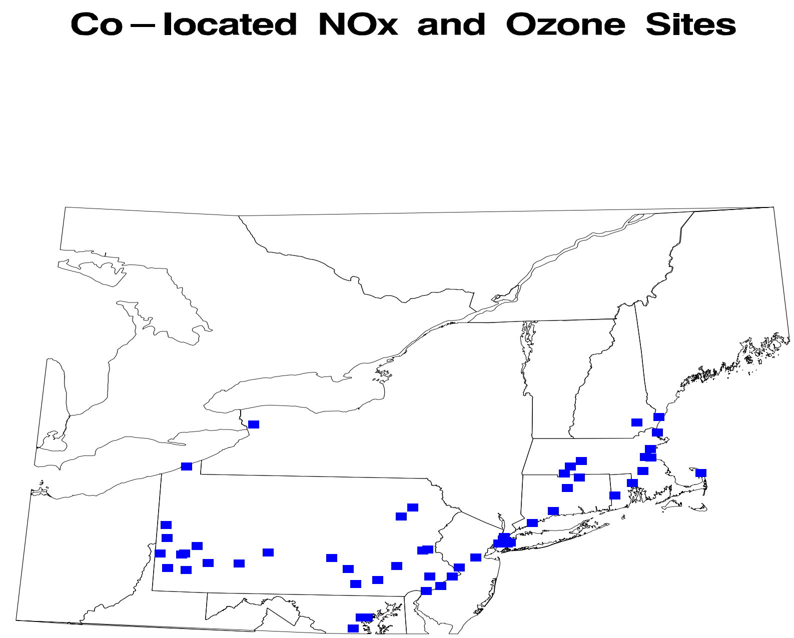


(h)

Co-located O_3 and NO_x: N=11k dataset



(i)



(j)

Ambient data summary

- Ozone data: N=54k dataset

  Will use to model transport

- NO_x data: N=21k dataset
- VOC data: N=3k dataset

  Will use to learn about relationships between emissions data and ambient VOC data

- Ozone $\cap$ NO_x $\cap$ VOC: N=1563 dataset

  Will use to learn about relationships among O_3, NO_x, VOC, and temperature

- Ozone $\cap$ NO_x: N=11k dataset

  Will use in main model

Outline

1. Context, goals, and data
2. **The model**
   - Created ozone
   - Transported ozone
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

The model (using N=11k dataset):

$$Y_{t,\,i} = Y^C_{t,\,i} + Y^T_{t,\,i} + \nu_{t,\,i}, \qquad \boldsymbol{\nu}_t \sim \begin{cases} N\{\,\mathbf{0}\,,\; V_t(\boldsymbol{\phi}^*_1)\,\} & t \text{ in Jan-Apr} \\ N\{\,\mathbf{0}\,,\; V_t(\boldsymbol{\phi}^*_2)\,\} & t \text{ in May-Sept} \\ N\{\,\mathbf{0}\,,\; V_t(\boldsymbol{\phi}^*_3)\,\} & t \text{ in Oct} \\ N\{\,\mathbf{0}\,,\; V_t(\boldsymbol{\phi}^*_4)\,\} & t \text{ in Nov-Dec} \end{cases}$$

$$\beta_1 + \beta_2 \mathcal{N}_{t,\,i} + \beta_3 \mathcal{N}^2_{t,\,i} + \beta_4 \mathcal{N}_{t,\,i}(\mathcal{T}_{t,\,i} - 1.4) +$$
$$\beta_5 \mathcal{N}^2_{t,\,i}(\mathcal{T}_{t,\,i} - 1.4) + \beta_6 \mathcal{N}_{t,\,i}\mathcal{L}_{t,\,i} + \beta_7 \mathcal{N}_{t,\,i}\mathcal{T}_{t,\,i}\mathcal{L}_{t,\,i}$$

$$f_1(\mathcal{L}^N_{m_t, C_i}, \mathcal{L}^{OR}_{C_i}, \mathcal{L}^{NR}_{C_i}, \mathcal{L}^{ST}_{C_i}, \mathcal{L}^{OA}_{C_i},$$
$$\mathcal{M}_{t,i}, \mathcal{W}_{t,i}, \boldsymbol{\gamma}) + \omega_{t,\,i}$$

$$\delta \boldsymbol{\lambda}'_{t-1,i} \boldsymbol{Y}^*_{t-1}$$

$$f_2(ws_{t,\,i}, wd_{t,\,i})$$

N=54k dataset

$$\boldsymbol{\omega}_t \sim \begin{cases} N\{\,\mathbf{0}\,,\; W_t(\boldsymbol{\psi}^*_1)\,\} & t \text{ in Jan-Apr} \\ N\{\,\mathbf{0}\,,\; W_t(\boldsymbol{\psi}^*_2)\,\} & t \text{ in May-Sept} \\ N\{\,\mathbf{0}\,,\; W_t(\boldsymbol{\psi}^*_3)\,\} & t \text{ in Oct} \\ N\{\,\mathbf{0}\,,\; W_t(\boldsymbol{\psi}^*_4)\,\} & t \text{ in Nov-Dec} \end{cases}$$

Outline

1. Context, goals, and data
2. **The model**
   - **Created ozone**
   - Transported ozone
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

How did we learn about created ozone model?

- Atmospheric chemistry results in National Research Council (1991)

- Field study results, e.g., Ryerson et al. (2001)

- N=1563 datatset of co-located ozone, NO_x, and VOC

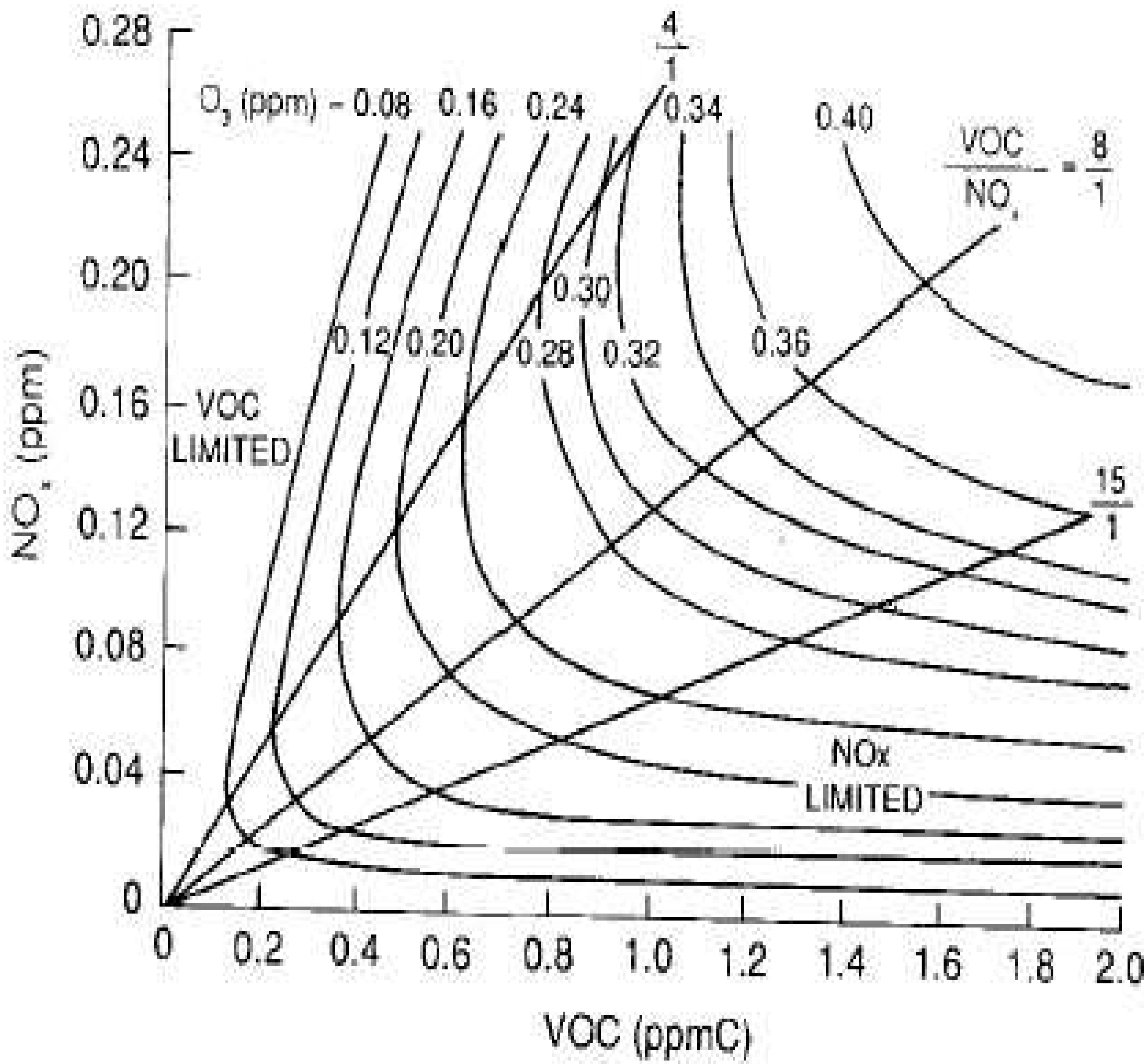
Three Atmospheric Regimes

1. **Low VOC/NO_x ratios**
   - Ozone decreases when NO_x increases.

     Created ozone can be negative!
   - Ozone increases when VOC's increase.
2. **Mid-level VOC/NO_x ratios**
   - Ozone increases when NO_x increases for fixed VOC's.
   - Ozone increases when VOC's increase at fixed NO_x .
   - Ozone increases when both VOC's and NO_x increase.
3. **High VOC/NO_x ratios**
   - Ozone increases when NO_x increases
   - Ozone does not change when VOC's increase.

(National Research Council 1991)

NRC[p.165] SMOG chamber contour plot:



Created ozone

$$Y^C_{t,\,i} = \beta_1 + \beta_2 \mathcal{N}_{t,\,i} + \beta_3 \mathcal{N}^2_{t,\,i} + \beta_4 \mathcal{N}_{t,\,i}(\mathcal{T}_{t,\,i} - 1.4) + \beta_5 \mathcal{N}^2_{t,\,i}(\mathcal{T}_{t,\,i} - 1.4) + \beta_6 \mathcal{N}_{t,\,i}\mathcal{L}_{t,\,i} + \beta_7 \mathcal{N}_{t,\,i}\mathcal{T}_{t,\,i}\mathcal{L}_{t,\,i}$$

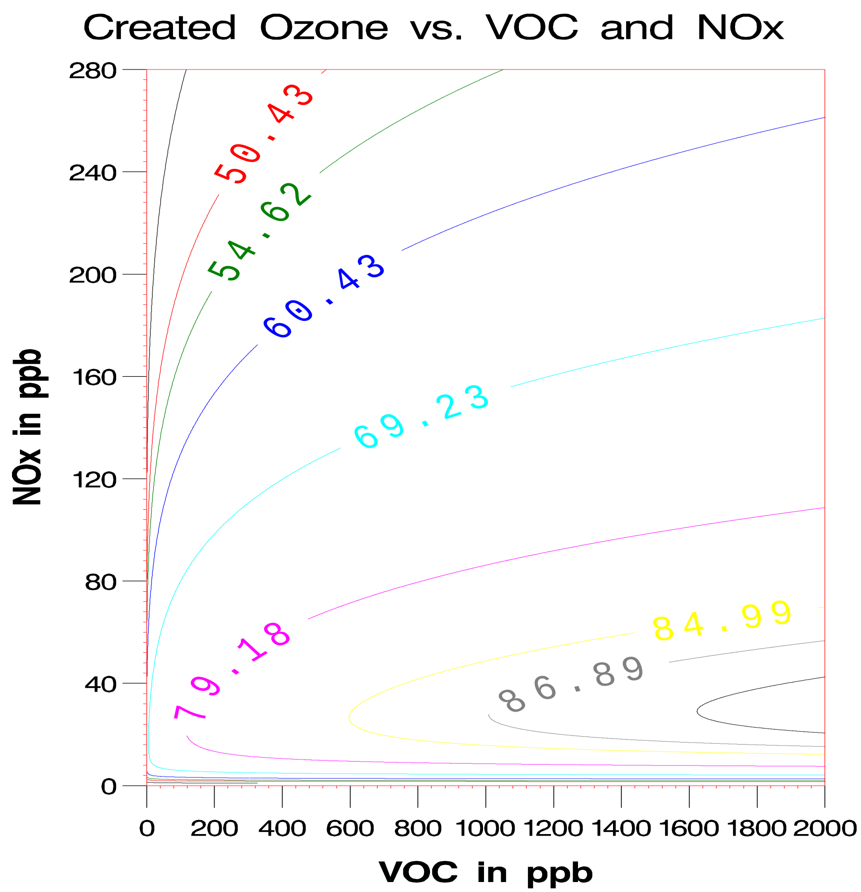
$$\mathcal{L}_{t,\,i} \equiv log(VOC_{t,\,i} + 1)$$

$$\mathcal{N}_{t,\,i} \equiv \log(\mathrm{NO}_x\ _{t,\,i} + 1)$$

$$\mathcal{T}_{t,\,i} \equiv \exp((\mathrm{maxtemperature}_{t,\,i} - 73.9)/14.78)$$

Discovery: We can match the NRC contour plot, but……

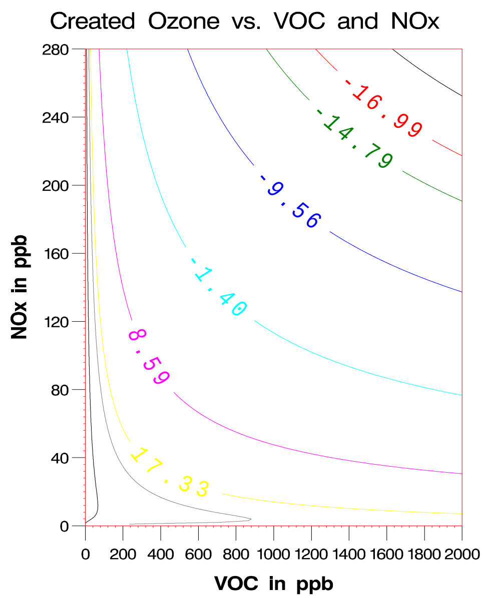
95th percentile



(k)

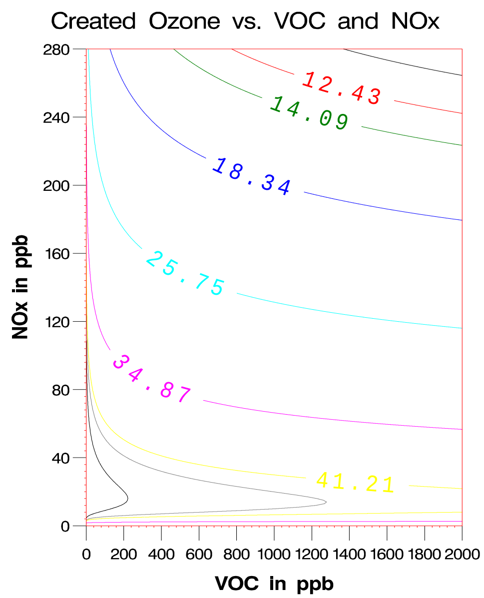
… ratios that demarcate regimes are highly dependent on temperature

5th percentile



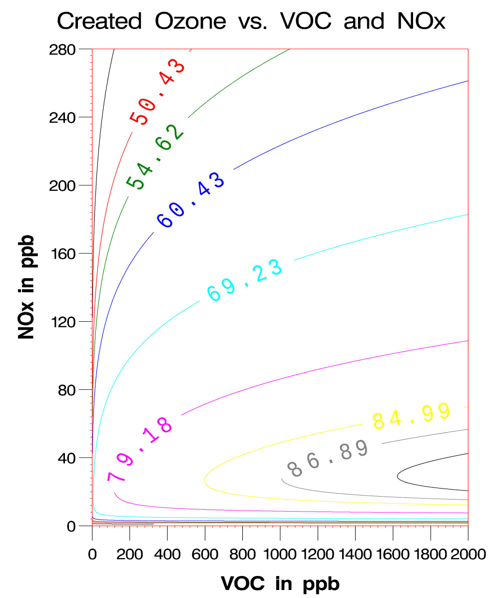
(l)

median



(m)

95th percentile



(n)

Outline

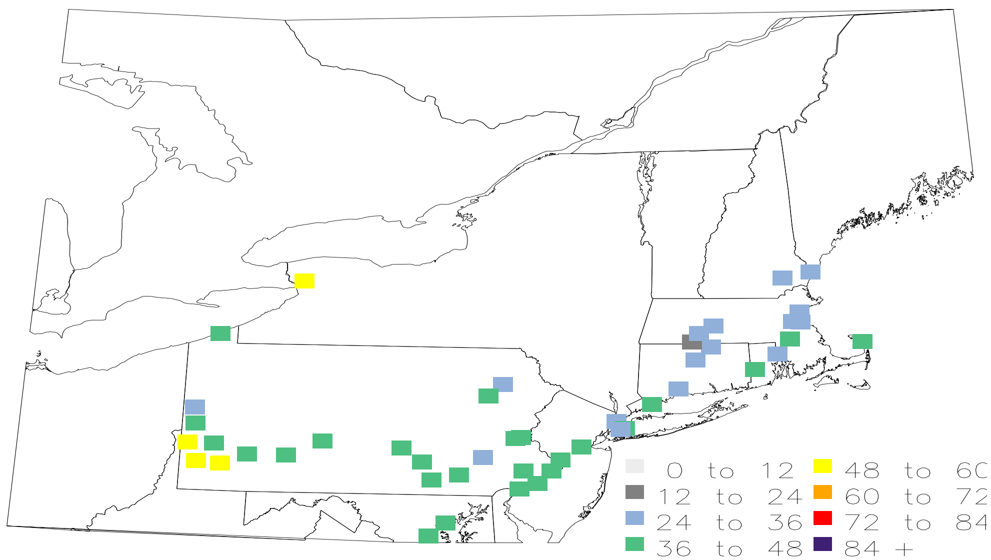
1. Context, goals, and data
2. **The model**
   - Created ozone
   - **Transported ozone**
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

Consider a point in time and space

July 15 Site 420692006

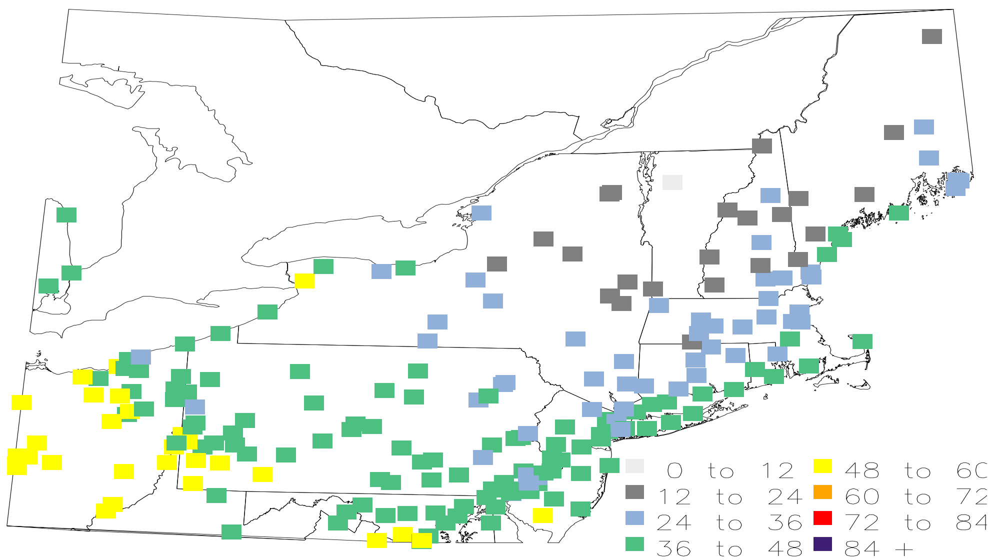
Consider all points yesterday from N=11k dataset

July 14 N=11K Ozone



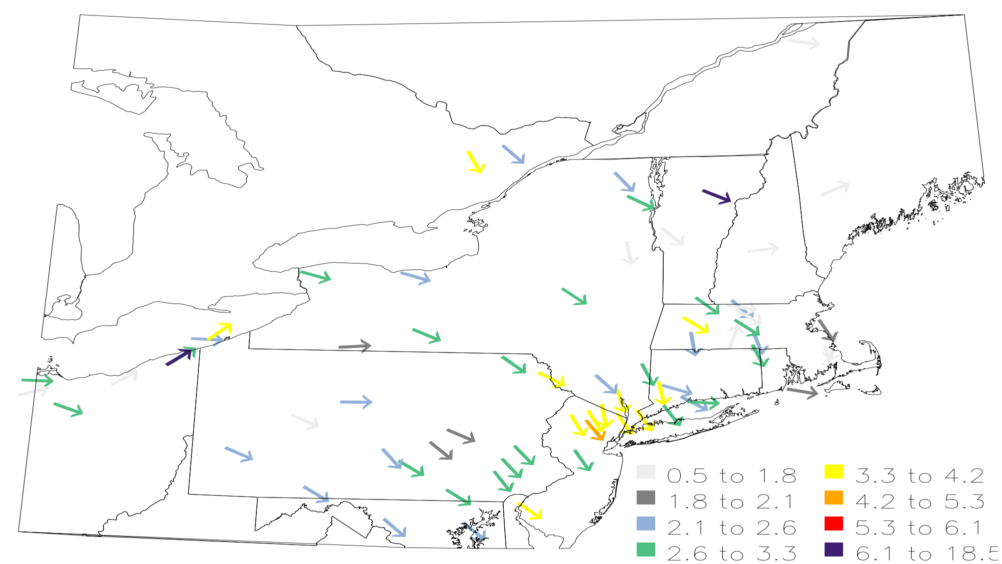
Consider all points yesterday from N=54k dataset

July 14 N=54K Ozone



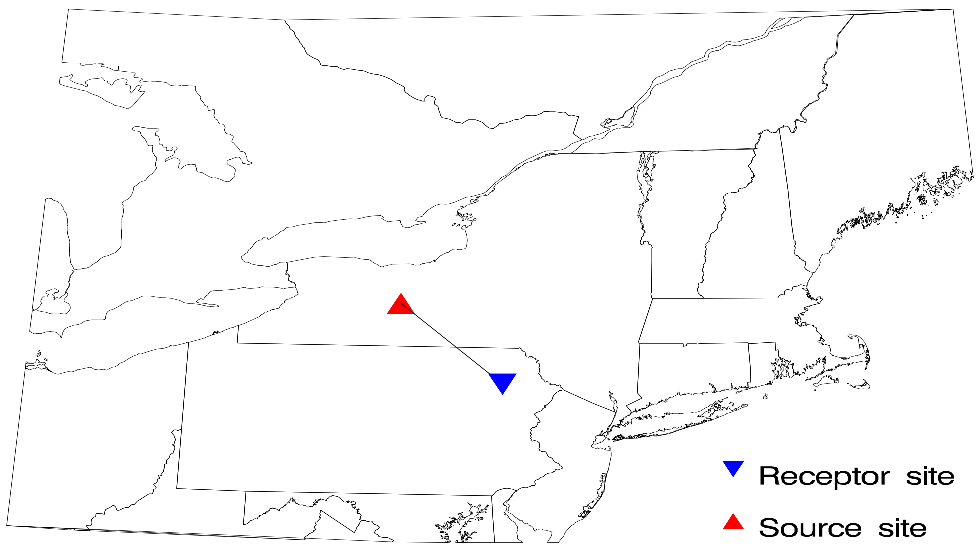
Consider the windfield on July 15

July 15 wind in m/s



Draw the vector of 24 hour travel at windspeed

July 15 Site 420692006 and OSS



Transported ozone model

$$Y^T_{t,\,i} = \delta \boldsymbol{\lambda}' \boldsymbol{Y}^*_{t-1}$$

Uniform windfield, midnight-to-midnight.

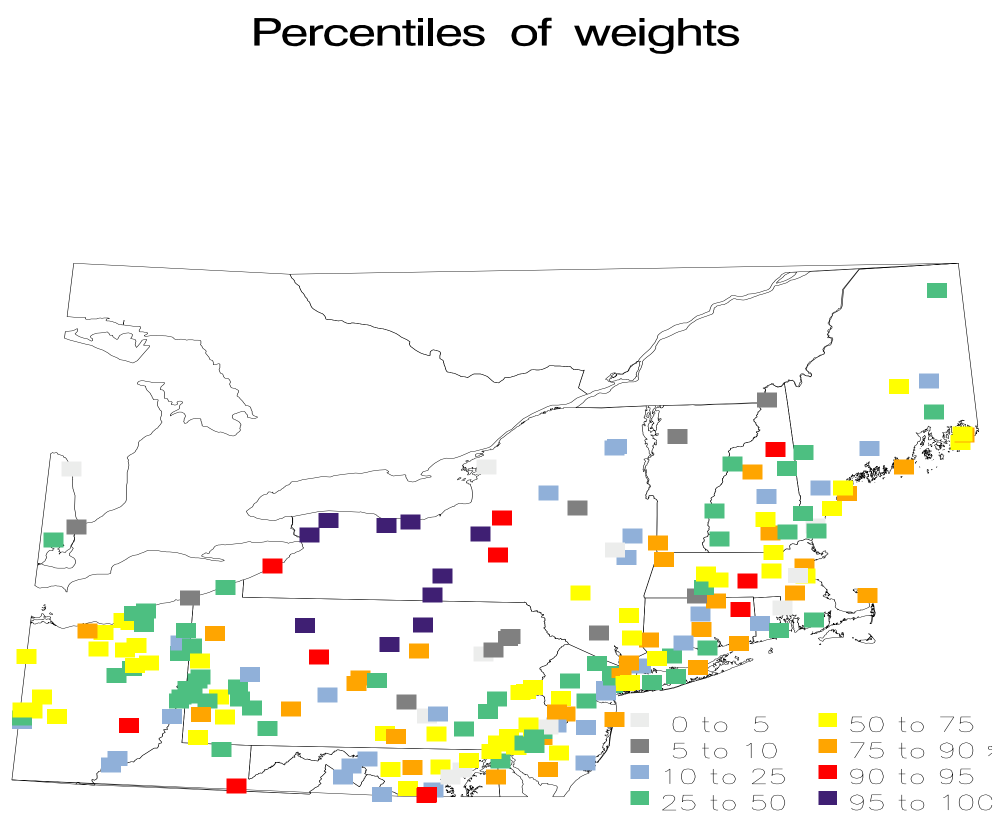
$\boldsymbol{s}_j$ is the "optimal source site" for $Y_{t,\,i}$

$$Y^*_{t-1,\,i} = \mu_{t-1} + \alpha_{t-1}\mathcal{T}_{t-1,i} + \varepsilon_{t-1,i}, \qquad \boldsymbol{\varepsilon}_{t-1} \sim N\{\mathbf{0},\, \Omega_{t-1}\}$$

$$\boldsymbol{\lambda}_{t-1} = \Omega^{-1}_{t-1}[\boldsymbol{c}^{\Omega}_{t-1,j} + X_{t-1}(X'_{t-1}\Omega^{-1}_{t-1}X_{t-1})^{-1}(\boldsymbol{x}_{t-1,j} - X'_{t-1}\Omega^{-1}_{t-1}\boldsymbol{c}^{\Omega}_{t-1,j})]$$

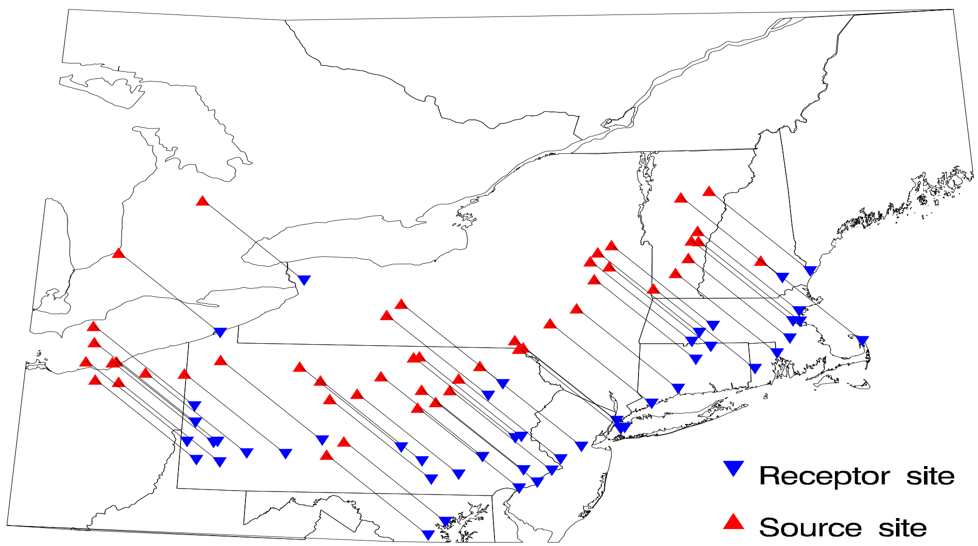
Universal kriging weights for prediction of ozone at oss

Sites near "Optimal source site" get most weight



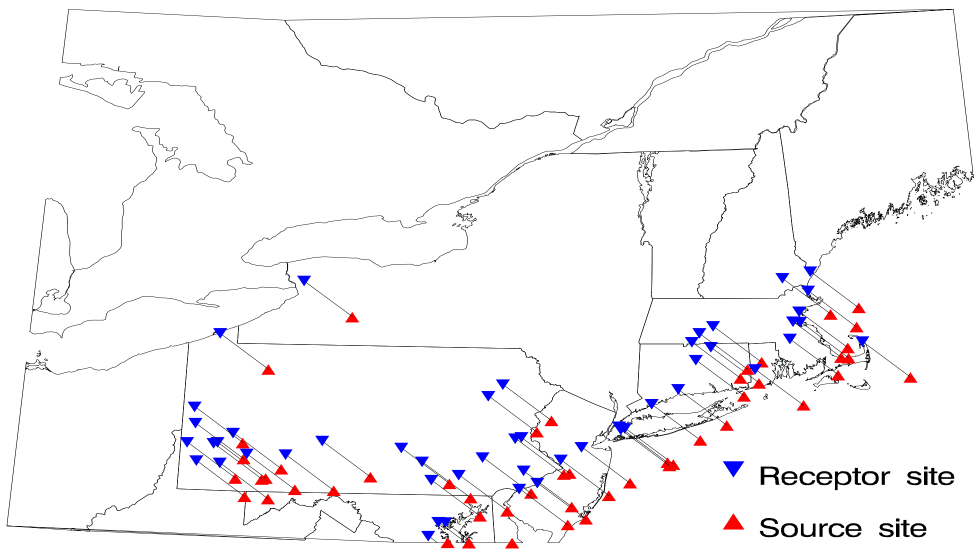
All "sources" and receptors on July 15

July 15 source and receptor sites



All "sources" and receptors on July 28

July 28 source and receptor sites



Ozone process model

$$Y_{t,\,i} = Y^C_{t,\,i} + Y^T_{t,\,i} + \nu_{t,\,i}, \qquad t = 2, \ldots, T$$

$$\begin{aligned}
=&\beta_1 + \beta_2 \mathcal{N}_{t,\,i} + \beta_3 \mathcal{N}^2_{t,\,i} + \beta_4 \mathcal{N}_{t,\,i}(\mathcal{T}_{t,\,i} - 1.4) + \\
&\beta_5 \mathcal{N}^2_{t,\,i}(\mathcal{T}_{t,\,i} - 1.4) + \beta_6 \mathcal{N}_{t,\,i}\mathcal{L}_{t,\,i} + \beta_7 \mathcal{N}_{t,\,i}\mathcal{T}_{t,\,i}\mathcal{L}_{t,\,i} + \\
&\delta \boldsymbol{\lambda}' \boldsymbol{Y}^*_{t-1} + \nu_{t,\,i}, \qquad t = 2, \ldots, T
\end{aligned}$$

$$\boldsymbol{\nu}_t \overset{\text{indep}}{\sim} N\{\mathbf{0}, V_t(\boldsymbol{\phi}_t)\}, \qquad t = 1, \ldots, T$$

$$\boldsymbol{Y}_1 \,|\boldsymbol{\mathcal{L}}_1, \boldsymbol{\beta}, \boldsymbol{\phi}_1 \sim N\{\, X_1(\boldsymbol{\mathcal{L}}_1)\boldsymbol{\beta} \;,\; V_1(\boldsymbol{\phi}_1)\}$$

$$\boldsymbol{Y}_t \,|\boldsymbol{\mathcal{L}}_t, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t \overset{\text{indep}}{\sim} N\{\, X_t(\boldsymbol{\mathcal{L}}_t)\boldsymbol{\beta} + \delta \Lambda_{t-1} \boldsymbol{Y}^*_{t-1} \;,\; V_t(\boldsymbol{\phi}_t)\}, \qquad t = 2, \ldots, T$$

Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - **Log VOC**
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

VOC Emissions data resolution before and after

	Resolution			
	In the data		**In the model**	
Dataset	**Time**	**Space**	**Time**	**Space**
Onroad	Year	County	Day	Lon, lat
Nonroad	Year	County	Day	Lon, lat
Storage & Transport	Year	County	Day	Lon, lat
Other area	Year	County	Day	Lon, lat
Biogenic	Month	County	Day	Lon, lat

VOC process model:

$$\mathcal{L}_{t,\,i} =$$

$$\gamma_1 + \gamma_2 \mathcal{M}_{t,\,i} +$$
$$\gamma_3 \mathcal{L}_{C_i}^{N} + \gamma_4 \mathcal{L}_{C_i}^{OR} + \gamma_5 \mathcal{L}_{C_i}^{NR} + \gamma_6 \mathcal{L}_{C_i}^{ST} + \gamma_7 \mathcal{L}_{C_i}^{OA} +$$
$$\gamma_8 \mathcal{L}_{C_i}^{N} \mathcal{M}_{t,\,i} + \gamma_9 \mathcal{L}_{C_i}^{OR} \mathcal{M}_{t,\,i} + \gamma_{10} \mathcal{L}_{C_i}^{NR} \mathcal{M}_{t,\,i} + \gamma_{11} \mathcal{L}_{C_i}^{ST} \mathcal{M}_{t,\,i} + \gamma_{12} \mathcal{L}_{C_i}^{OA} \mathcal{M}_{t,\,i} +$$
$$\gamma_{13} \mathcal{L}_{C_i}^{OR} \mathcal{W}_t + \omega_{t,\,i}\ , \qquad t = 1, \ldots, T$$

$$\boldsymbol{\omega}_t \overset{\text{indep}}{\sim} N\{\mathbf{0}, W_t(\boldsymbol{\psi}_t)\} \qquad t = 1, \ldots, T$$

$$\boldsymbol{\mathcal{L}}_t \mid \boldsymbol{\gamma}, \boldsymbol{\psi} \overset{\text{indep}}{\sim} N\{Z_t \boldsymbol{\gamma}, W_t(\boldsymbol{\psi}_t)\}, \qquad t = 1, \ldots, T$$

Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - Log VOC
   - **Covariance models**
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

Covariance models: exponential covariance function

$$\boldsymbol{\phi}_t \equiv (\sigma_t^2, \rho_t, \sigma_{n_t}^2)' \qquad t = 1, \ldots, T$$

$$\boldsymbol{\psi}_t \equiv (\tau_t^2, \eta_t, \tau_{n_t}^2)' \qquad t = 1, \ldots, T,$$

$$V_{t,j,k} = \begin{cases} \sigma_{n_t}^2 + \sigma_t^2 & \text{if } \boldsymbol{s}_j = \boldsymbol{s}_k \\ \sigma_t^2 \exp(-d_{jk}/\rho_t) & \text{otherwise} \end{cases}$$

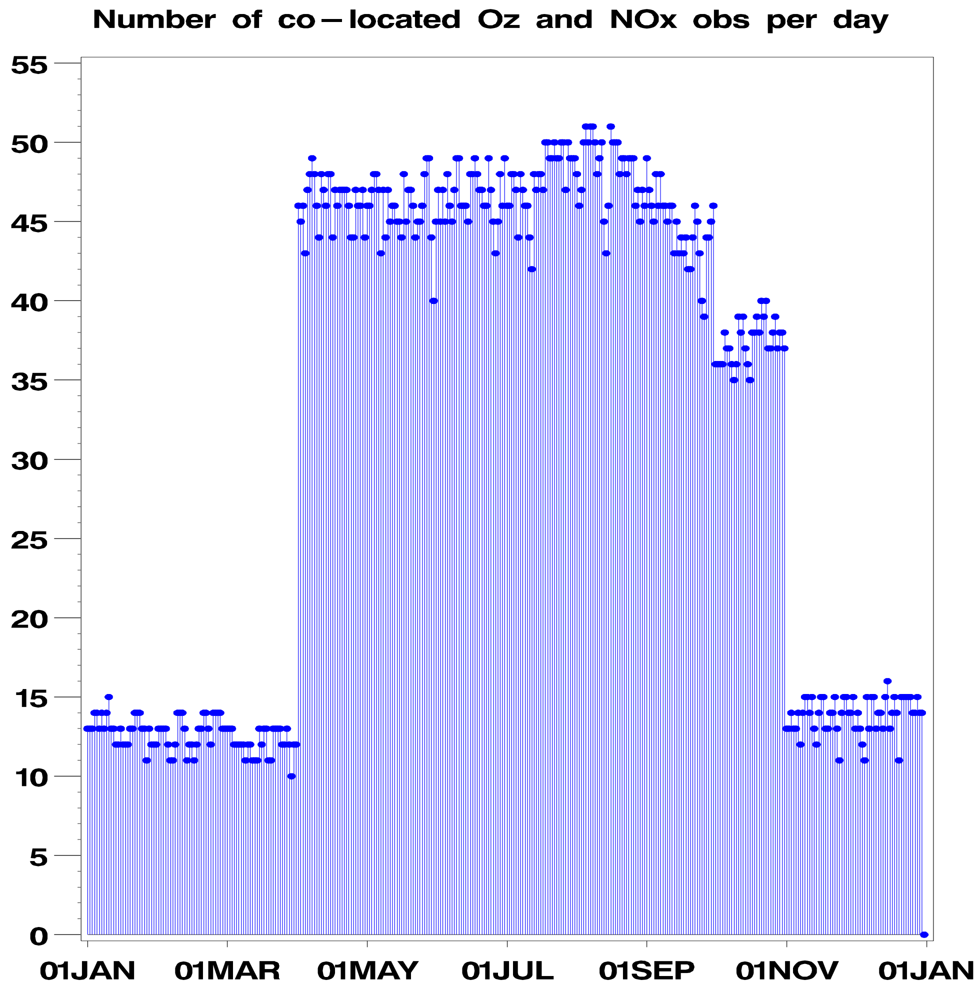
$$W_{t,j,k} = \begin{cases} \tau_{n_t}^2 + \tau_t^2 & \text{if } \boldsymbol{s}_j = \boldsymbol{s}_k \\ \tau_t^2 \exp(-d_{jk}/\eta_t) & \text{otherwise} \end{cases}$$

Covariance parameters: seasonally varying

$$\boldsymbol{\phi}_t = \begin{cases} \boldsymbol{\phi}_1^* \equiv \left(\sigma_1^{2*}, \rho_1^*, \sigma_{n_1}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 1 \\ \boldsymbol{\phi}_2^* \equiv \left(\sigma_2^{2*}, \rho_2^*, \sigma_{n_2}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 2 \\ \boldsymbol{\phi}_3^* \equiv \left(\sigma_3^{2*}, \rho_3^*, \sigma_{n_3}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 3 \\ \boldsymbol{\phi}_4^* \equiv \left(\sigma_4^{2*}, \rho_4^*, \sigma_{n_4}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 4, \end{cases}$$

$$\boldsymbol{\psi}_t = \begin{cases} \boldsymbol{\psi}_1^* \equiv \left(\tau_1^{2*}, \eta_1^*, \tau_{n_1}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 1 \\ \boldsymbol{\psi}_2^* \equiv \left(\tau_2^{2*}, \eta_2^*, \tau_{n_2}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 2 \\ \boldsymbol{\psi}_3^* \equiv \left(\tau_3^{2*}, \eta_3^*, \tau_{n_3}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 3 \\ \boldsymbol{\psi}_4^* \equiv \left(\tau_4^{2*}, \eta_4^*, \tau_{n_4}^{2*}\right)' & \text{if } t \in \ \text{timeperiod } 4. \end{cases}$$

Time periods/seasons process and frequency based



Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - Log VOC
   - Covariance models
3. **Results (the theoretical type)**
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

Derive the likelihood: integrate out $\mathcal{L}$

$$[\boldsymbol{Y} \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\psi}] =$$

$$\int [\boldsymbol{Y} \mid \boldsymbol{\mathcal{L}}, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}][\boldsymbol{\mathcal{L}} \mid \gamma, \boldsymbol{\psi}] d\boldsymbol{\mathcal{L}} =$$

$$\int \prod_{t=1}^{T} [\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{L}}_t, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t] \prod_{t=1}^{T} [\boldsymbol{\mathcal{L}}_t \mid \gamma, \boldsymbol{\psi}_t] d\boldsymbol{\mathcal{L}} =$$

$$\prod_{t=1}^{T} \left\{ \int [\boldsymbol{Y}_t \mid \boldsymbol{\mathcal{L}}_t, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t][\boldsymbol{\mathcal{L}}_t \mid \boldsymbol{\gamma}, \boldsymbol{\psi}_t] d\boldsymbol{\mathcal{L}}_t \right\} =$$

Can perform integration separately for each day!

$$\prod_{t=1}^{T} [\boldsymbol{Y}_t \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t].$$

Can write unconditional likelihood as product of daily likelihoods!

Daily distributions of $\boldsymbol{Y}$ unconditional on $\mathcal{L}$

$$\boldsymbol{Y}_t \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim$$

$$N\{\; X_t^A \boldsymbol{\beta}^A + M_t Z_t \boldsymbol{\gamma} + \delta \Lambda_{t-1} \boldsymbol{Y}^*_{t-1} \;,$$

$$V_t(\boldsymbol{\phi}_t) + M_t\, W_t(\boldsymbol{\psi}_t)\, M_t \;\}$$

$$\begin{aligned}
X_t^A &\equiv \begin{pmatrix} \mathbf{1} & \boldsymbol{\mathcal{N}}_t & \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{N}}_t & \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_C & \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_C \end{pmatrix} \\
\boldsymbol{\beta}^A &\equiv (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)' \\
M_t &\equiv \beta_6 diag(\boldsymbol{\mathcal{N}}_t) + \beta_7 diag(\boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_t) \\
Z_t &\equiv \text{design matrix for latent log VOC process} \\
\Lambda_{t-1} \boldsymbol{Y}^*_{t-1} &\equiv \text{lag ozone at the optimal source site,} \\
&\quad\; \text{predicted offline; treated as an explanatory variable}
\end{aligned}$$

Daily distributions of Y unconditional on $\mathcal{L}$

$$\boldsymbol{Y}_t \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim$$

$$N\{\ X_t^A \boldsymbol{\beta}^A + M_t Z_t \boldsymbol{\gamma} + \delta \Lambda_{t-1} \boldsymbol{Y}^*_{t-1}\ ,$$

$$\underbrace{V_t(\boldsymbol{\phi}_t) + M_t\, W_t(\boldsymbol{\psi}_t)\, M_t}_{\text{non-isotropic and non-stationary}} \qquad \}$$

$$X_t^A \equiv \begin{pmatrix} \mathbf{1} & \boldsymbol{\mathcal{N}}_t & \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{N}}_t & \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_C & \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_C \end{pmatrix}$$

$$\boldsymbol{\beta}^A \equiv (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$$

$$M_t \equiv \beta_6 diag(\boldsymbol{\mathcal{N}}_t) + \beta_7 diag(\boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_t)$$

$$Z_t \equiv \text{design matrix for latent log VOC process}$$

$$\Lambda_{t-1} \boldsymbol{Y}^*_{t-1} \equiv \text{lag ozone at the optimal source site,}$$

predicted offline; treated as an explanatory variable

$-2 \log \boldsymbol{L}$

$$
\begin{aligned}
&-2\log[\boldsymbol{L}(\boldsymbol{Y} \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\psi})] = \\
&\quad \text{constant} + \sum_{t=1}^{T} \log\left(|V_t(\boldsymbol{\phi}_t) + M_t W_t(\boldsymbol{\psi}_t) M_t|\right) \\
&\quad + [\boldsymbol{Y}_1 - X_1^A \boldsymbol{\beta}^A - M_1 Z_1 \boldsymbol{\gamma}]' [V_1(\boldsymbol{\phi}_1) + M_1 W_1(\boldsymbol{\psi}_1) M_1]^{-1} \\
&\qquad [\boldsymbol{Y}_1 - X_1^A \boldsymbol{\beta}^A - M_1 Z_1 \boldsymbol{\gamma}] \\
&\quad + \sum_{t=2}^{T} [\boldsymbol{Y}_t - X_t^A \boldsymbol{\beta}^A - \delta \Lambda_{t-1} \boldsymbol{Y}_{t-1}^* - M_t Z_t \boldsymbol{\gamma}]' [V_t(\boldsymbol{\phi}_t) + M_t W_t(\boldsymbol{\psi}_t) M_t]^{-1} \\
&\qquad [\boldsymbol{Y}_t - X_t^A \boldsymbol{\beta}^A - \delta \Lambda_{t-1} \boldsymbol{Y}_{t-1}^* - M_t Z_t \boldsymbol{\gamma}].
\end{aligned}
$$

Non-identifiability

Daily mean vector, long version

$$X_t^A \boldsymbol{\beta}^A + [\beta_6 diag(\boldsymbol{\mathcal{N}}_t) + \beta_7 diag(\boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_t)] Z_t \boldsymbol{\gamma} + \delta \Lambda_{t-1} \boldsymbol{Y}^*_{t-1}.$$

Daily covariance matrix, long version

$\sigma^2_{n_t} I + \sigma^2_t H(\rho) +$

$[\beta_6 diag(\boldsymbol{\mathcal{N}}_t) + \beta_7 diag(\boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_t)] \left[\tau^2_{n_t} I + \tau^2_t H(\eta)\right] [\beta_6 diag(\boldsymbol{\mathcal{N}}_t) + \beta_7 diag(\boldsymbol{\mathcal{N}}_t \# \boldsymbol{\mathcal{T}}_t)].$

Equivalent paramter vectors:

$$(\beta_6,\ \beta_7,\ \boldsymbol{\gamma},\ \tau^2_t,\ \tau^2_{n_t}) \equiv (\beta_6/k,\ \beta_7/k,\ k\boldsymbol{\gamma},\ k^2\tau^2_t,\ k^2\tau^2_{n_t})$$

Solution:

Fix $\beta_6 = -1$.

Predicting unobserved ozone conditional on observed

$$\begin{pmatrix} \boldsymbol{Y}_t^o \\ \boldsymbol{Y}_t^u \end{pmatrix} \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim N\left\{ \begin{pmatrix} \mu_{Y_t}^o \\ \mu_{Y_t}^u \end{pmatrix}, \begin{pmatrix} \Sigma_{Y_t}^o & \Sigma_{Y_t}^{ou} \\ \Sigma_{Y_t}^{uo} & \Sigma_{Y_t}^u \end{pmatrix} \right\}$$

$$\begin{aligned}
\mu_{Y_t}^o &\equiv X_t^{Ao}\boldsymbol{\beta}^A + M_t^o Z_t^o \boldsymbol{\gamma} + \delta \Lambda_{t-1}^o \boldsymbol{Y}_{t-1}^* \\
\mu_{Y_t}^u &\equiv X_t^{Au}\boldsymbol{\beta}^A + M_t^u Z_t^u \boldsymbol{\gamma} + \delta \Lambda_{t-1}^u \boldsymbol{Y}_{t-1}^* \\
\Sigma_{Y_t}^o &\equiv V_t^o(\phi_t) + M_t^o W_t^o(\phi_t) M_t^o \\
\Sigma_{Y_t}^u &\equiv V_t^u(\phi_t) + M_t^u W_t^u(\phi_t) M_t^u \\
\Sigma_{Y_t}^{ou} &\equiv V_t^{ou}(\phi_t) + M_t^o W_t^{ou}(\phi_t) M_t^u.
\end{aligned}$$

$$\boldsymbol{Y}_t^u \mid \boldsymbol{Y}_t^o, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim$$

$$N\left\{ \mu_{Y_t}^u + \Sigma_{Y_t}^{uo}[\Sigma_{Y_t}^o]^{-1}\left(\boldsymbol{Y}_t^o - \mu_{Y_t}^o\right), \ \Sigma_{Y_t}^u - \Sigma_{Y_t}^{uo}[\Sigma_{Y_t}^o]^{-1}\Sigma_{Y_t}^{ou} \right\}.$$

Predicting unobserved ozone conditional on observed

$$\begin{pmatrix} \boldsymbol{Y}_t^o \\ \boldsymbol{Y}_t^u \end{pmatrix} \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim N\left\{ \begin{pmatrix} \mu_{Y_t}^o \\ \mu_{Y_t}^u \end{pmatrix}, \begin{pmatrix} \Sigma_{Y_t}^o & \Sigma_{Y_t}^{ou} \\ \Sigma_{Y_t}^{uo} & \Sigma_{Y_t}^u \end{pmatrix} \right\}$$

$$\begin{aligned}
\mu_{Y_t}^o &\equiv X_t^{Ao} \boldsymbol{\beta}^A + M_t^o Z_t^o \boldsymbol{\gamma} + \delta \Lambda_{t-1}^o \boldsymbol{Y}_{t-1}^* \\
\mu_{Y_t}^u &\equiv X_t^{Au} \boldsymbol{\beta}^A + M_t^u Z_t^u \boldsymbol{\gamma} + \delta \Lambda_{t-1}^u \boldsymbol{Y}_{t-1}^* \\
\Sigma_{Y_t}^o &\equiv V_t^o(\phi_t) + M_t^o W_t^o(\phi_t) M_t^o \\
\Sigma_{Y_t}^u &\equiv V_t^u(\phi_t) + M_t^u W_t^u(\phi_t) M_t^u \\
\Sigma_{Y_t}^{ou} &\equiv V_t^{ou}(\phi_t) + M_t^o W_t^{ou}(\phi_t) M_t^u .
\end{aligned}$$

$$\boldsymbol{Y}_t^u \mid \boldsymbol{Y}_t^o, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim$$

$$N\Bigg\{ \underbrace{\underbrace{\mu_{Y_t}^u}_{\text{Meanhat}} + \Sigma_{Y_t}^{uo} [\Sigma_{Y_t}^o]^{-1} \left(\boldsymbol{Y}_t^o - \mu_{Y_t}^o \right)}_{\text{Yhat}}, \; \Sigma_{Y_t}^u - \Sigma_{Y_t}^{uo} [\Sigma_{Y_t}^o]^{-1} \Sigma_{Y_t}^{ou} \Bigg\}.$$

Predicting the latent log VOC process 1

$$
\begin{pmatrix} \boldsymbol{Y}_t^o \\ \boldsymbol{\mathcal{L}}_t^o \\ \boldsymbol{\mathcal{L}}_t^u \end{pmatrix} \mid \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim
$$

$$
N\left\{ \begin{pmatrix} \mu_{Y_t}^o \\ Z_t^o \boldsymbol{\gamma} \\ Z_t^u \boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \Sigma_{Y_t}^o & M_t W_t^o(\boldsymbol{\psi}) & M_t W_t^{ou}(\boldsymbol{\psi}) \\ W_t^o(\boldsymbol{\psi}) M_t & W_t^o(\boldsymbol{\psi}) & W_t^{ou}(\boldsymbol{\psi}) \\ W_t^{uo}(\boldsymbol{\psi}) M_t & W_t^{uo}(\boldsymbol{\psi}) & W_t^u(\boldsymbol{\psi}) \end{pmatrix} \right\}.
$$

Predicting the latent log VOC process 2

$$
\begin{pmatrix} \boldsymbol{\mathcal{L}}_t^o \\ \boldsymbol{\mathcal{L}}_t^u \end{pmatrix} \mid \boldsymbol{Y}_t^o, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim
$$

$$
N\Bigg\{ \begin{pmatrix} Z_t^o \boldsymbol{\gamma} \\ Z_t^u \boldsymbol{\gamma} \end{pmatrix} + \begin{pmatrix} W_t^o(\boldsymbol{\psi}) M_t \\ W_t^{uo}(\boldsymbol{\psi}) M_t \end{pmatrix} [\Sigma_{Y_t}^o]^{-1} (\boldsymbol{Y}_t^o - \mu_{Y_t}^o),
$$

$$
\begin{pmatrix} W_t^o(\boldsymbol{\psi}) & W_t^{ou}(\boldsymbol{\psi}) \\ W_t^{uo}(\boldsymbol{\psi}) & W_t^u(\boldsymbol{\psi}) \end{pmatrix}
$$

$$
- \begin{pmatrix} W_t^o(\boldsymbol{\psi}) M_t \\ W_t^{uo}(\boldsymbol{\psi}) M_t \end{pmatrix} [\Sigma_{Y_t}^o]^{-1} \Big(M_t W_t^o(\boldsymbol{\psi}) \quad M_t W_t^{ou}(\boldsymbol{\psi}) \Big) \Bigg\}.
$$

Predicting the latent log VOC process 2

$$
\begin{pmatrix} \boldsymbol{\mathcal{L}}_t^o \\ \boldsymbol{\mathcal{L}}_t^u \end{pmatrix} \mid \boldsymbol{Y}_t^o, \boldsymbol{\beta}, \delta, \boldsymbol{\phi}_t, \boldsymbol{\gamma}, \boldsymbol{\psi}_t \sim
$$

$$
N\Bigg\{ \underbrace{\underbrace{\begin{pmatrix} Z_t^o \boldsymbol{\gamma} \\ Z_t^u \boldsymbol{\gamma} \end{pmatrix}}_{\text{Z gammahat}} + \begin{pmatrix} W_t^o(\boldsymbol{\psi}) M_t \\ W_t^{uo}(\boldsymbol{\psi}) M_t \end{pmatrix} [\Sigma_{Y_t}^o]^{-1} (\boldsymbol{Y}_t^o - \mu_{Y_t}^o)}_{\text{Lhat}},
$$

$$
\begin{pmatrix} W_t^o(\boldsymbol{\psi}) & W_t^{ou}(\boldsymbol{\psi}) \\ W_t^{uo}(\boldsymbol{\psi}) & W_t^u(\boldsymbol{\psi}) \end{pmatrix}
$$

$$
- \begin{pmatrix} W_t^o(\boldsymbol{\psi}) M_t \\ W_t^{uo}(\boldsymbol{\psi}) M_t \end{pmatrix} [\Sigma_{Y_t}^o]^{-1} \begin{pmatrix} M_t W_t^o(\boldsymbol{\psi}) & M_t W_t^{ou}(\boldsymbol{\psi}) \end{pmatrix} \Bigg\}.
$$

Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. **Simulation / Parametric Bootstrap**
5. Results (of the application)
6. Model validation and CMAQ comparison
7. Discussion and future work

Simulation / bootstrap questions

- If our model were true, would our estimation method recover true parameter values?
- Method of estimation = minimizing $-2\log L$ via SAS IML nlpnra
- Do we (I) have any coding errors?
- Are the inverse-Hessian standard errors valid?

Simulation / bootstrap method

- Used our ML estimates as true parameter values
- Generated 1000 datasets according to stated model
- Fit all 1000 datasets using the same method we used to fit the model
- **Computationally expensive:** 37 computers running simultaneously for 10 days
- Between 5 and 15 hours per run

Simulation / bootstrap answers

- If our model were true, would our estimation method recover true parameter values?
- Yes! We had enough estimates that were unbiased so that we believe our methods. Where the bootstrap mean did not match truth, we believe there was no signal in the data, as evidenced by high standard error estimates in both the ML fit and the bootstrap.
- Do we (I) have any coding errors?
- Well, we did discover some, but they're all fixed now.
- Are the inverse-Hessian standard errors valid?
- Most of the inverse-Hessian standard errors were underestimates. If we replace them with the bootstrap standard deviation, a few parameters that were significant become insignificant.

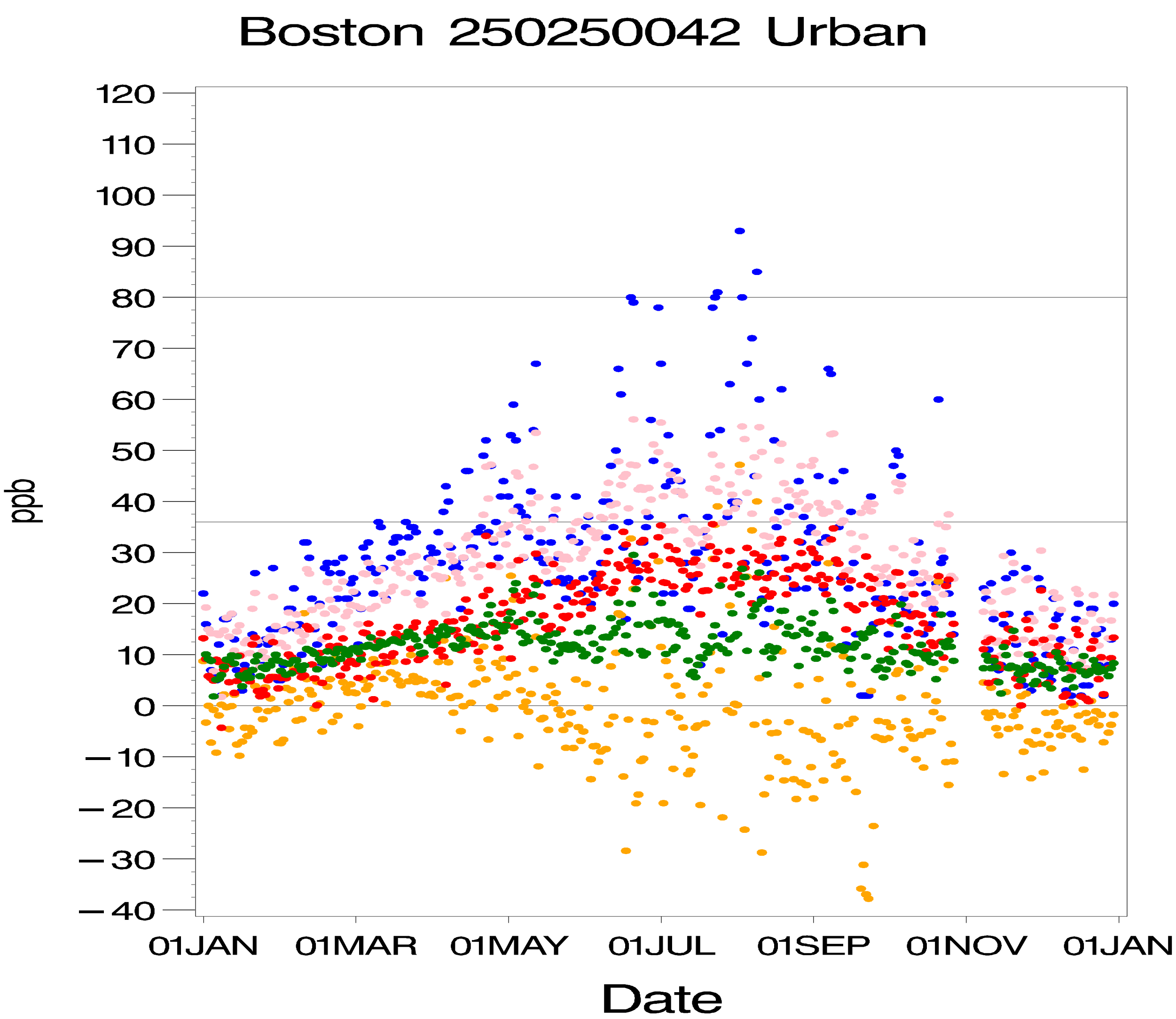
Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. **Results (of the application)**
6. Model validation and CMAQ comparison
7. Discussion and future work

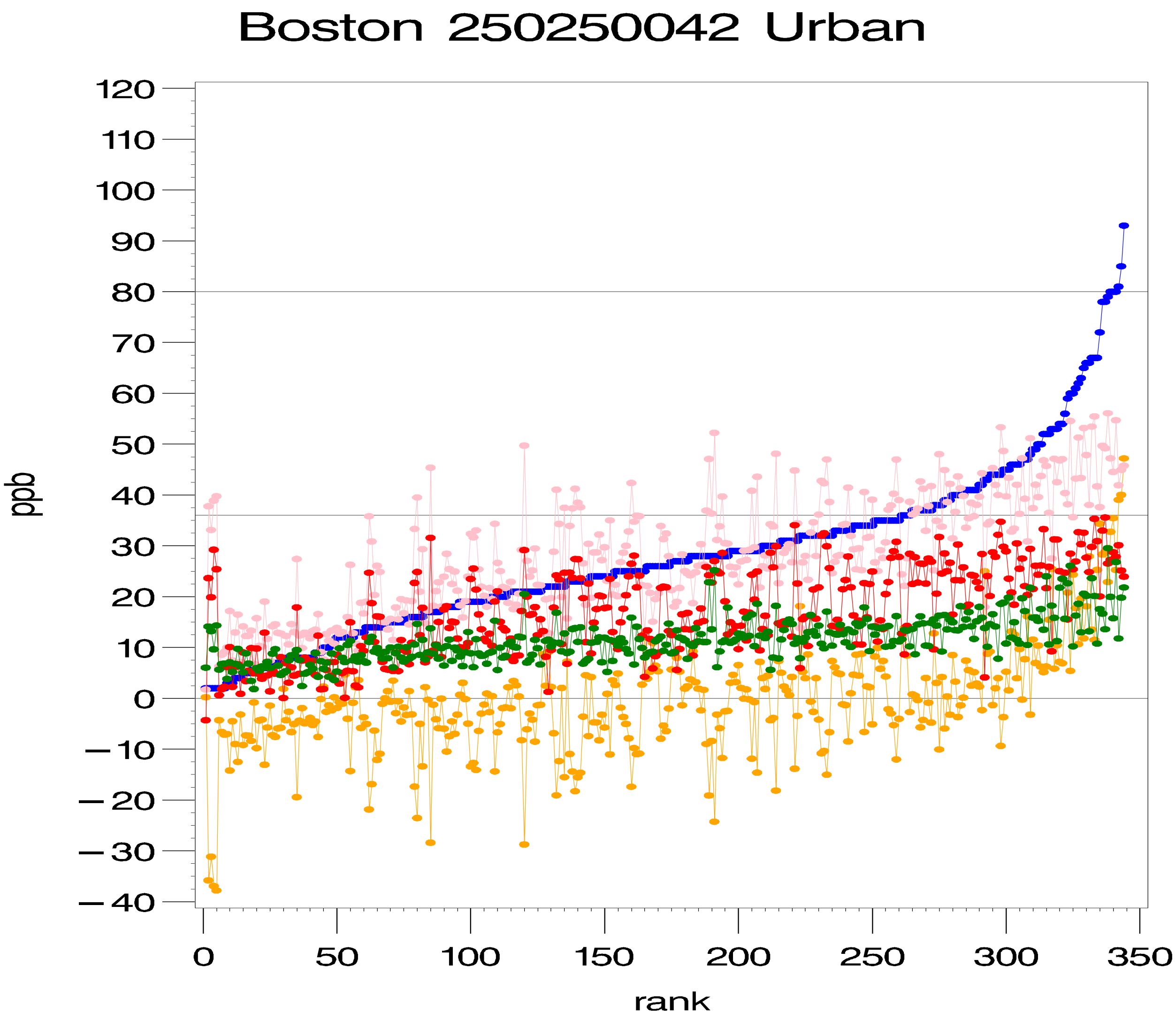
Ozone process mean trend parameters

Parameter	Effect	Estimate	Std err	Lower	Upper
β_1	*intercept*	**36**	**.76**	**35**	**38**
β_2	$\mathcal{N}$	-2.2	3.4	-8.9	4.5
β_3	$\mathcal{N}^2$	**-1.3**	**.077**	**-1.4**	**-1.1**
β_4	$\mathcal{N}\mathcal{T}$	**5.4**	**2.1**	**1.3**	**9.6**
β_5	$\mathcal{N}^2\mathcal{T}$	**-.68**	**.051**	**-.78**	**-.58**
β_6	$\mathcal{L}\mathcal{N}$	-1	-	-	-
β_7	$\mathcal{L}\mathcal{N}\mathcal{T}$	**-3.7**	**.31**	**-4.3**	**-3.1**
δ	*transport*	**.29**	**.013**	**.27**	**.32**

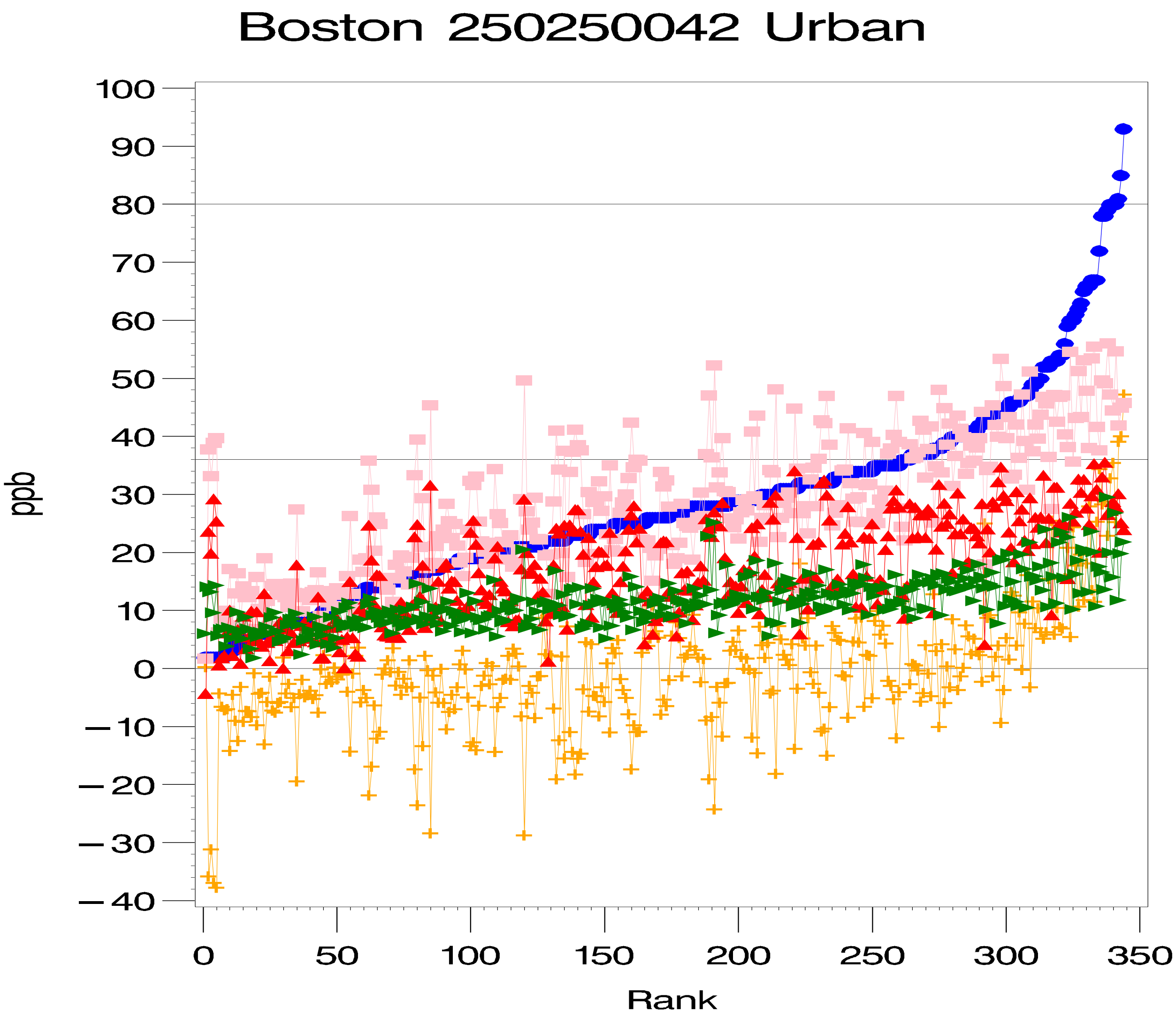
Decomposition of ozone (by day)



Decomposition of ozone (by rank)



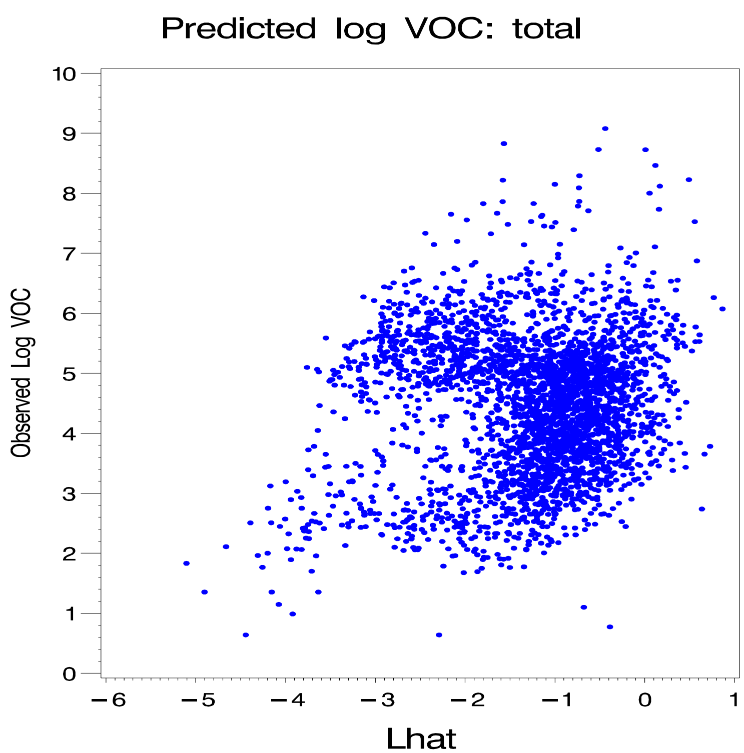
Decomposition of ozone (by ozone)



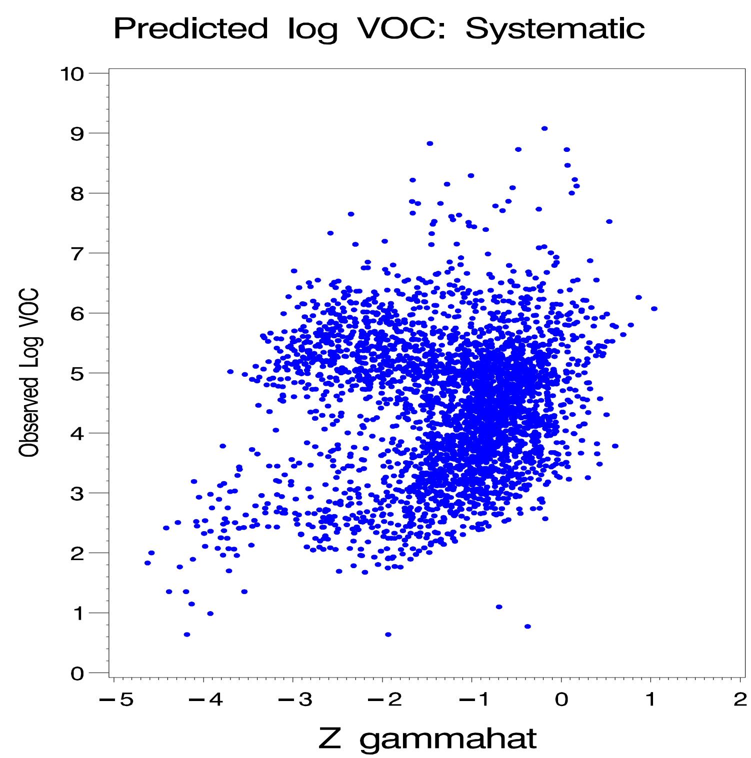
Latent log VOC process mean trend parameters

Parameter	Effect	Estimate	Std err	Lower	Upper
γ_1	*intercept*	**-1.3**	**.56**	**-2.4**	**-.20**
γ_2	$\mathcal{M}$	**.072**	**.0085**	**.055**	**.088**
γ_3	$\mathcal{L}^N$	**-.062**	**.023**	**-.11**	**-.017**
γ_4	$\mathcal{L}^{OR}$	.023	.016	-.0088	.055
γ_5	$\mathcal{L}^{NR}$	**-.064**	**.0091**	**-.082**	**-.046**
γ_6	$\mathcal{L}^{ST}$	**.36**	**.027**	**.31**	**.41**
γ_7	$\mathcal{L}^{OA}$	**-.069**	**.018**	**-.10**	**-.034**
γ_8	$\mathcal{L}^N\mathcal{M}$	.00012	.0018	-.0035	.0037
γ_9	$\mathcal{L}^{OR}\mathcal{M}$	**-.0029**	**.0012**	**-.0053**	**-.00058**
γ_{10}	$\mathcal{L}^{NR}\mathcal{M}$	**.0038**	**.00067**	**.0025**	**.0051**
γ_{11}	$\mathcal{L}^{ST}\mathcal{M}$	**-.0028**	**.0013**	**-.0053**	**-.0003**
γ_{12}	$\mathcal{L}^{OA}\mathcal{M}$	**-.0035**	**.0013**	**-.0061**	**-.00092**
γ_{13}	$\mathcal{L}^{OR}\mathcal{W}$	**.023**	**.0077**	**.0077**	**.038**

Latent log VOC process vs. observed log VOC



(o)



(p)

Ozone covariance parameters

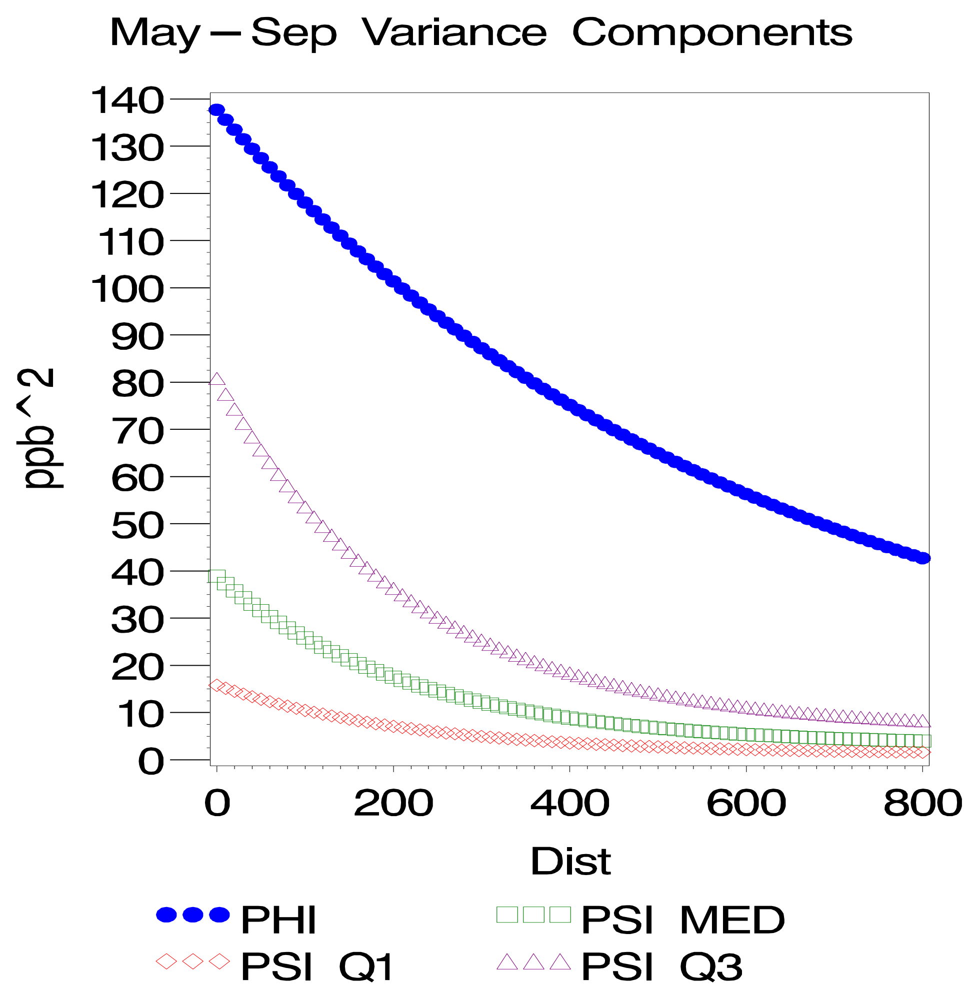
Parameter	Estimate	Std err	Lower	Upper
σ_1^2	**36**	**2.5**	**31**	**41**
ρ_1^*	**360**	**39**	**280**	**430**
$\sigma_{n_1}^{2*}$	.58	.77	0	2.1
σ_2^2	**130**	**7.2**	**120**	**150**
ρ_2^*	**610**	**42**	**530**	**700**
$\sigma_{n_2}^{2*}$	**7.7**	**.42**	**6.9**	**8.6**
σ_3^2	**75**	**12**	**52**	**98**
ρ_3^*	**1500**	**350**	**780**	**2100**
$\sigma_{n_3}^{2*}$	**8.6**	**.83**	**7.0**	**10**
σ_4^2	**55**	**6.2**	**43**	**67**
ρ_4^*	**920**	**180**	**580**	**1300**
$\sigma_{n_4}^{2*}$	**4.3**	**.86**	**2.6**	**6.0**

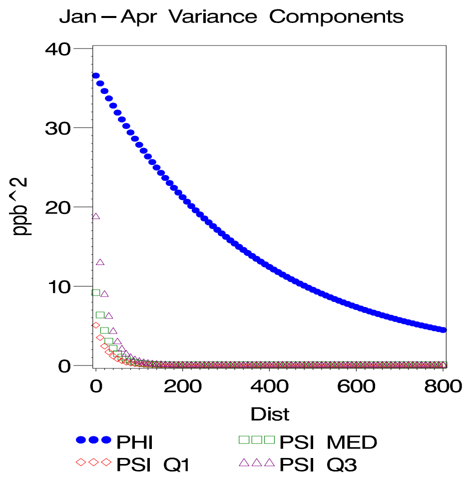
Log VOC process covariance parameters

Parameter	Estimate	Std err	Lower	Upper
τ_1^2	**.24**	**.042**	**.16**	**.33**
η_1^*	**27**	**6.6**	**14**	**40**
$\tau_{n_1}^{2*}$	.0020	.030	0	.062
τ_2^2	**.13**	**.023**	**.083**	**.17**
η_2^*	**220**	**27**	**170**	**280**
$\tau_{n_2}^{2*}$	**.011**	**.0022**	**.0068**	**.015**
τ_3^2	**.10**	**.021**	**.063**	**.15**
η_3^*	**60**	**13**	**35**	**86**
$\tau_{n_3}^{2*}$	.0026	.0046	0	.012
τ_4^2	**.11**	**.027**	**.059**	**.16**
η_4^*	**52**	**20**	**12**	**92**
$\tau_{n_4}^{2*}$	.0011	.013	0	.027

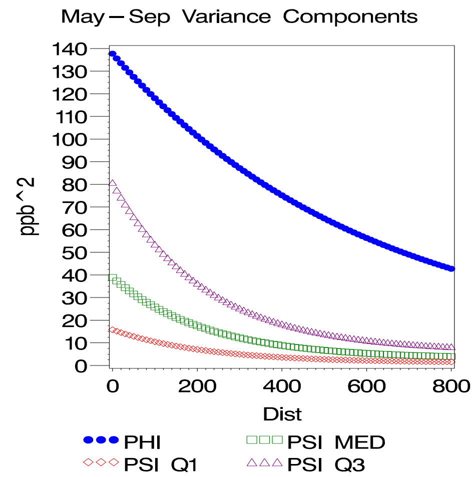
Decomposition of covariance: time period 2

$V_t(\boldsymbol{\phi}_t) + M_t W_t(\boldsymbol{\psi}_t) M_t$

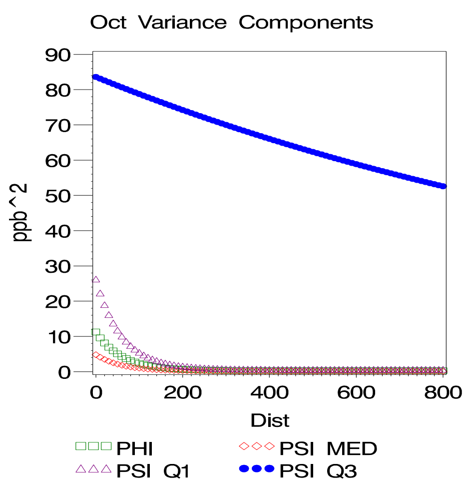



Jan – Apr Variance Components
ppb^2
40
30
20
10
0
0
200
400
600
800
Dist
PHI
PSI MED
PSI Q1
PSI Q3

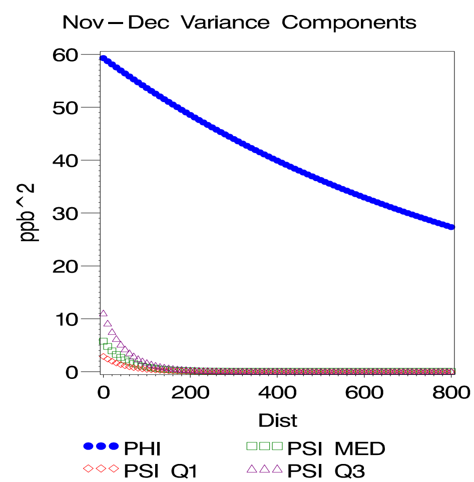
(q)


May – Sep Variance Components
ppb^2
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
0
200
400
600
800
Dist
PHI
PSI MED
PSI Q1
PSI Q3

(r)


Oct Variance Components
ppb^2
90
80
70
60
50
40
30
20
10
0
0
200
400
600
800
Dist
PHI
PSI MED
PSI Q1
PSI Q3

(s)

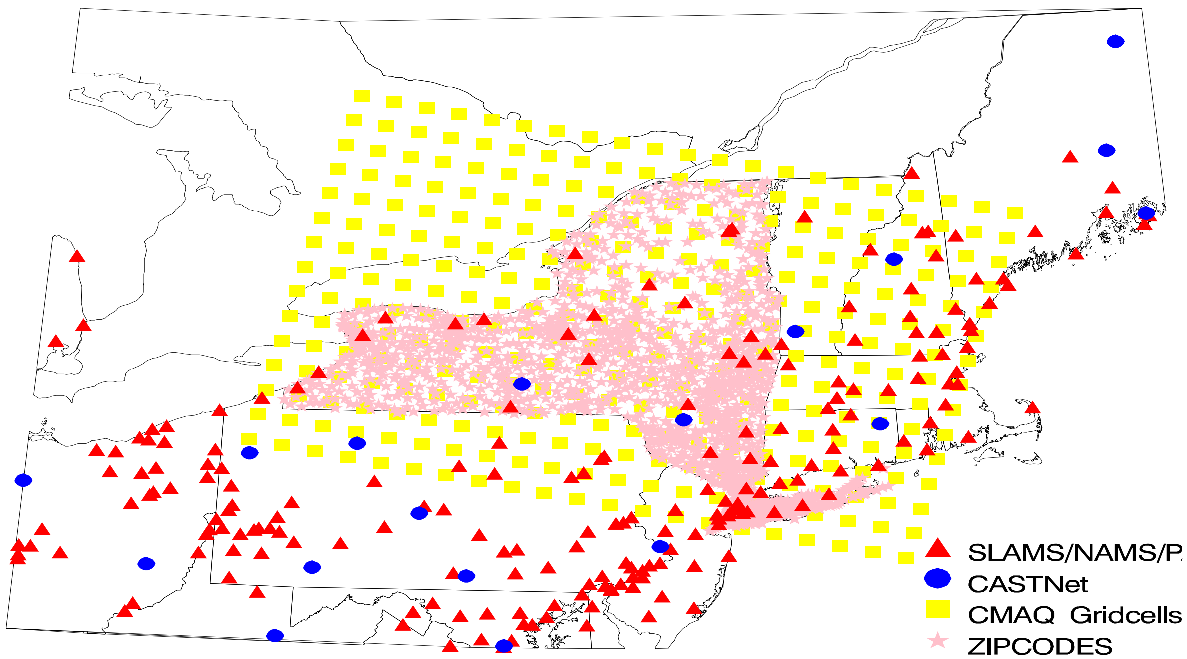

Nov – Dec Variance Components
ppb^2
60
50
40
30
20
10
0
0
200
400
600
800
Dist
PHI
PSI MED
PSI Q1
PSI Q3

(t)

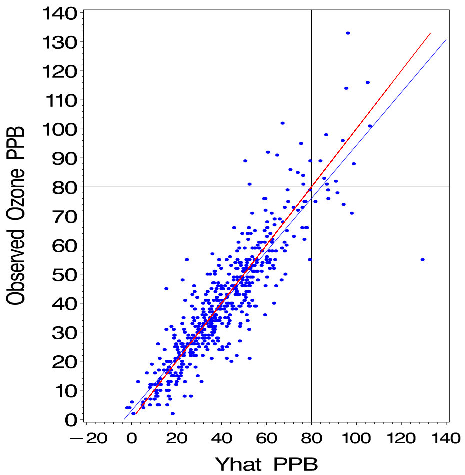
Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. **Model validation and CMAQ comparison**
7. Discussion and future work

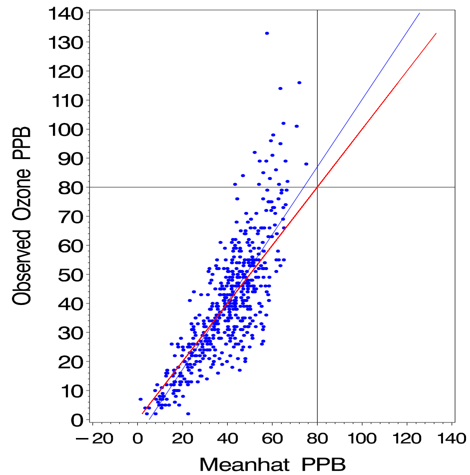
CMAQ comparison: candidate sites



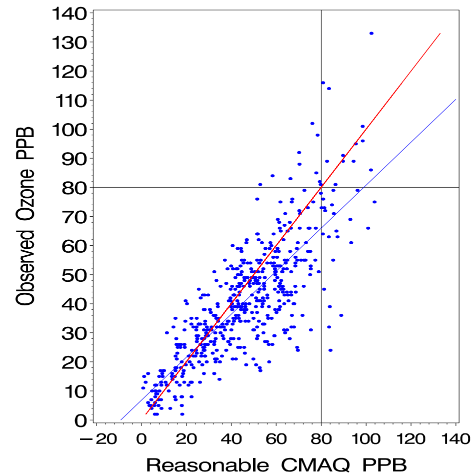
Leave out ten percent scatterplots and residuals



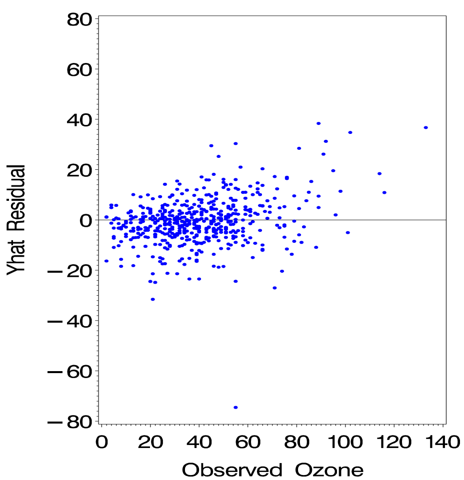
(u)



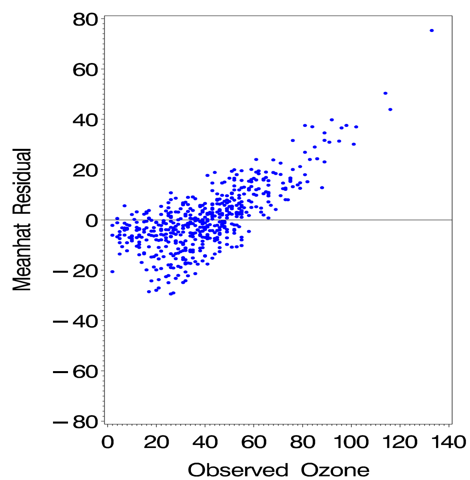
(v)



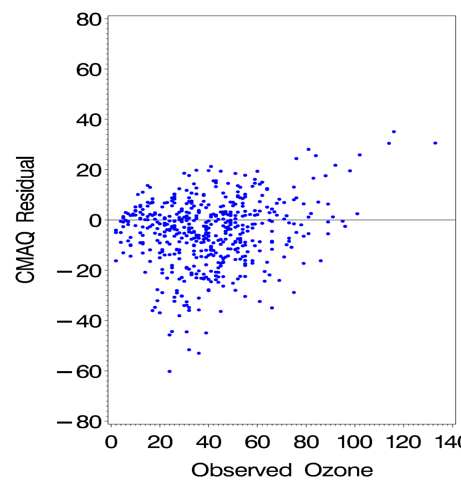
(w)



(x)



(y)



(z)

Leave out ten percent regression diagnostics

	N	R^2	RMSE	Slope	Intercept
Yhat	508	.78	9.6	**.91**	**3.0**
				.022	**.98**
Meanhat	508	.64	12	**1.2**	**-6.0**
				.039	**1.6**
Reasonable	508	.64	12	**.74**	**6.8**
CMAQ				**.025**	**1.2**
CMAQ	532	2.0E-4	21	-3.8E-5	**40**
				1.1E-4	**.91**

Outline

1. Context, goals, and data
2. The model
   - Created ozone
   - Transported ozone
   - Log VOC
   - Covariance models
3. Results (the theoretical type)
4. Simulation / Parametric Bootstrap
5. Results (of the application)
6. Model validation and CMAQ comparison
7. **Discussion and future work**

Bottom line:

- **The good news**

  This model allows us to decompose ozone into created + transported

- **The bad news**

  We underestimate extremely high ozone values with our mean trend.

- **The plan**

  Work with atmospheric scientists to improve mean trend

  Expand model to two latent space-time fields (VOC and NO_x) via Bayesian framework

References

EPA (2004), The ozone report: Measuring progress through 2003, Technical Report EPA 454/K-04-001, Environmental Protection Agency.

Nail, A. J. (2007), Quantifying local creation and regional transport using a hierarchical space-time model of ozone as a function of observed NOx, a latent space-time VOC process, emissions, and meteorology, Dissertation, North Carolina State University.

National Research Council (1991), Rethinking the ozone problem in urban and regional air pollution, Technical report, National Academy of Sciences, Washington, D.C.

Ryerson, T., Buhr, M. P., Frost, G. J., Goldan, P. D., Holloway, J. S., Hübler, G., Jobson, B. T., Kuster, W. C., McKeen, S. A., Parrish, D. D., Roberts, J. M., Sueper, D. T., Trainer, M., Williams, J. &

Fehsenfeld, F. C. (2001), 'Observations of ozone formation in power plant plumes and implications for ozone control strategies', *Science* **292**, 719–723.