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.

Community Modeling and Analysis System Conference
Chapel Hill, NC

October 6, 2008

Amy J. Nail

Jacqueline M. Hughes-Oliver
John F. Monahan

Outline

1. Scientific and regulatory context → goals
2. The data
3. The models

   Original, Newmean, Metcov
4. Two different predictors for two different purposes
5. Model performance and CMAQ comparison
6. Discussion and future work

Outline

1. **Scientific and regulatory context → goals**
2. The data
3. The models

   Original, Newmean, Metcov
4. Two different predictors for two different purposes
5. Model performance and CMAQ comparison
6. Discussion and future work

Goals based on regulatory needs Formulate and assess the

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

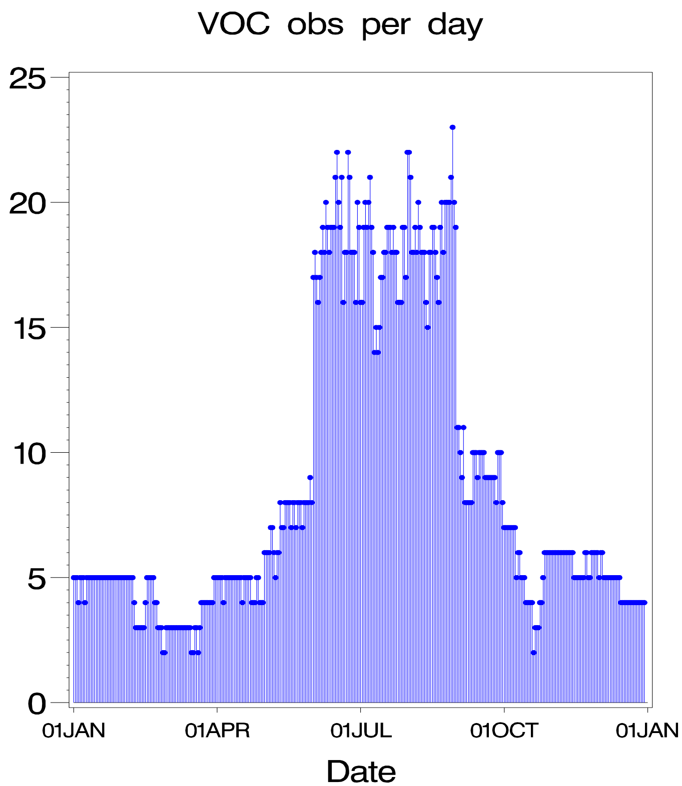
1. Decomposition into local creation vs. regional transport
2. Space-time predictions backward in time (at *any* space-time point)
3. Assessment of past emission control programs
4. Assessment of future emission control programs
5. Quantification of uncertainty

Outline

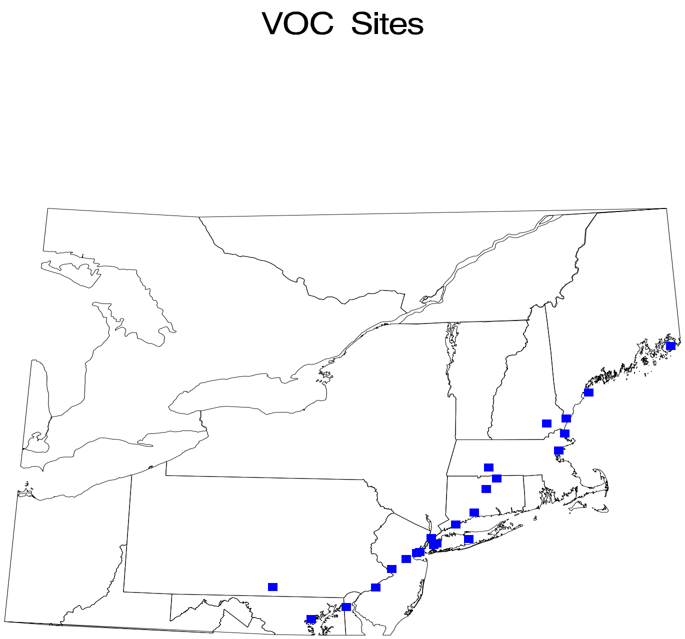
1. Scientific and regulatory context → goals
2. **The data**
3. The models

   Original, Newmean, Metcov
4. Two different predictors for two different purposes
5. Model performance and CMAQ comparison
6. Discussion and future work

VOC: N=3k dataset



(a)

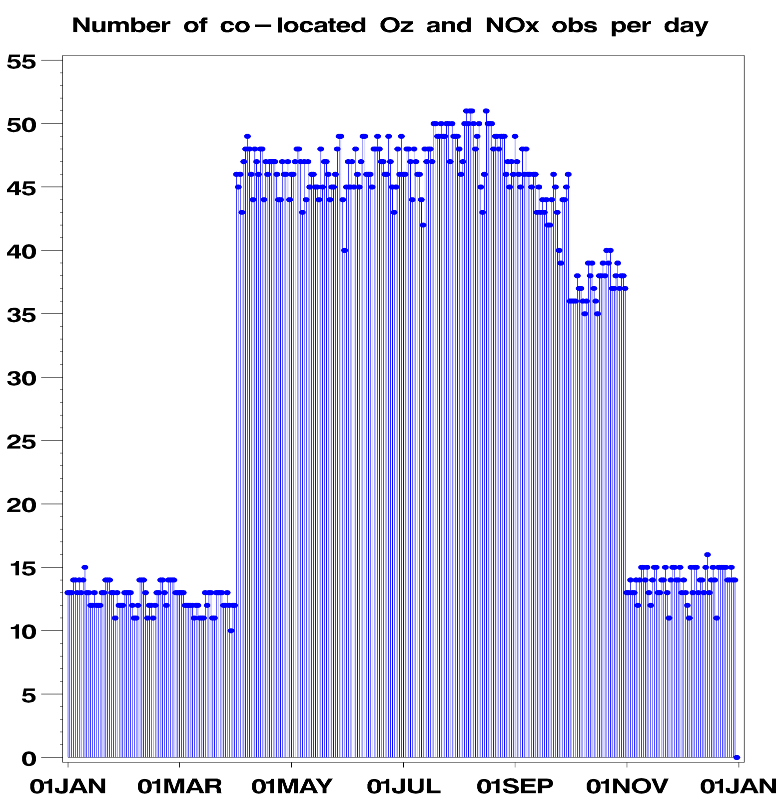


(b)

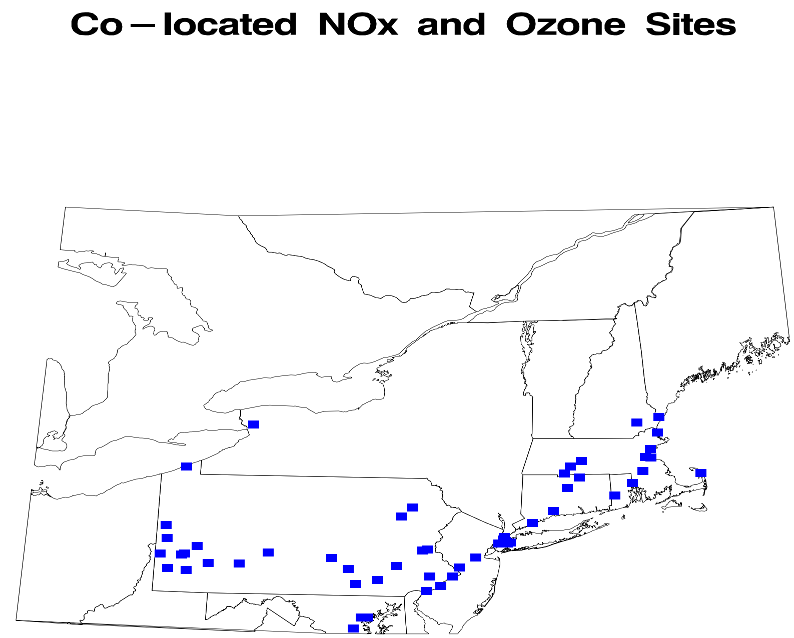
VOC Emissions data resolution before and after

	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

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



(c)



(d)

Outline

1. Scientific and regulatory context → goals
2. The data
3. **The models**

   Original, Newmean, Metcov
4. Two different predictors for two different purposes
5. Model performance and CMAQ comparison
6. Discussion and future work

The model (using N=11k dataset):

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

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

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

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

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

N=54k dataset

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

"Newmean" model improvement

1. Separate NO_x into NO and NO_2
2. Include pressure
3. Include dewpoint as a measure of humidity
4. Include yesterday's ozone at the same site (nighttime transport)
5. Better transformation of maximum temperature

“Metcov” model improvement

Before: covariance parameters based on “season”

Now: covariance parameters based on windspeed and temperature

“Metcov” model:

$$Y_{t,\,i} = Y^C_{t,\,i} + Y^T_{t,\,i} + \nu_{t,\,i}, \qquad \boldsymbol{\nu}_t \sim \begin{cases} N\{\,\mathbf{0}\,,\ V_t(\boldsymbol{\phi}_1^*)\,\} & ws \geq 4.6 \text{ m/s} \\ N\{\,\mathbf{0}\,,\ V_t(\boldsymbol{\phi}_2^*)\,\} & maxtemp < 55 \\ N\{\,\mathbf{0}\,,\ V_t(\boldsymbol{\phi}_3^*)\,\} & 55 \leq maxtemp < 75 \\ N\{\,\mathbf{0}\,,\ V_t(\boldsymbol{\phi}_4^*)\,\} & maxtemp \geq 75 \end{cases}$$

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

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

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

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

N=54k dataset

$$\boldsymbol{\omega}_t \sim \begin{cases} N\{\,\mathbf{0}\,,\ W_t(\boldsymbol{\psi}_1^*)\,\} & ws \geq 4.6 \text{ m/s} \\ N\{\,\mathbf{0}\,,\ W_t(\boldsymbol{\psi}_2^*)\,\} & maxtemp < 55 \\ N\{\,\mathbf{0}\,,\ W_t(\boldsymbol{\psi}_3^*)\,\} & 55 \leq maxtemp < 75 \\ N\{\,\mathbf{0}\,,\ W_t(\boldsymbol{\psi}_4^*)\,\} & maxtemp \geq 75 \end{cases}$$

Outline

1. Scientific and regulatory context → goals
2. The data
3. The model
4. **Two different predictors for two different purposes**
5. Model performance and CMAQ comparison
6. Discussion and future work

Predicting unobserved ozone conditional on observed

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

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

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

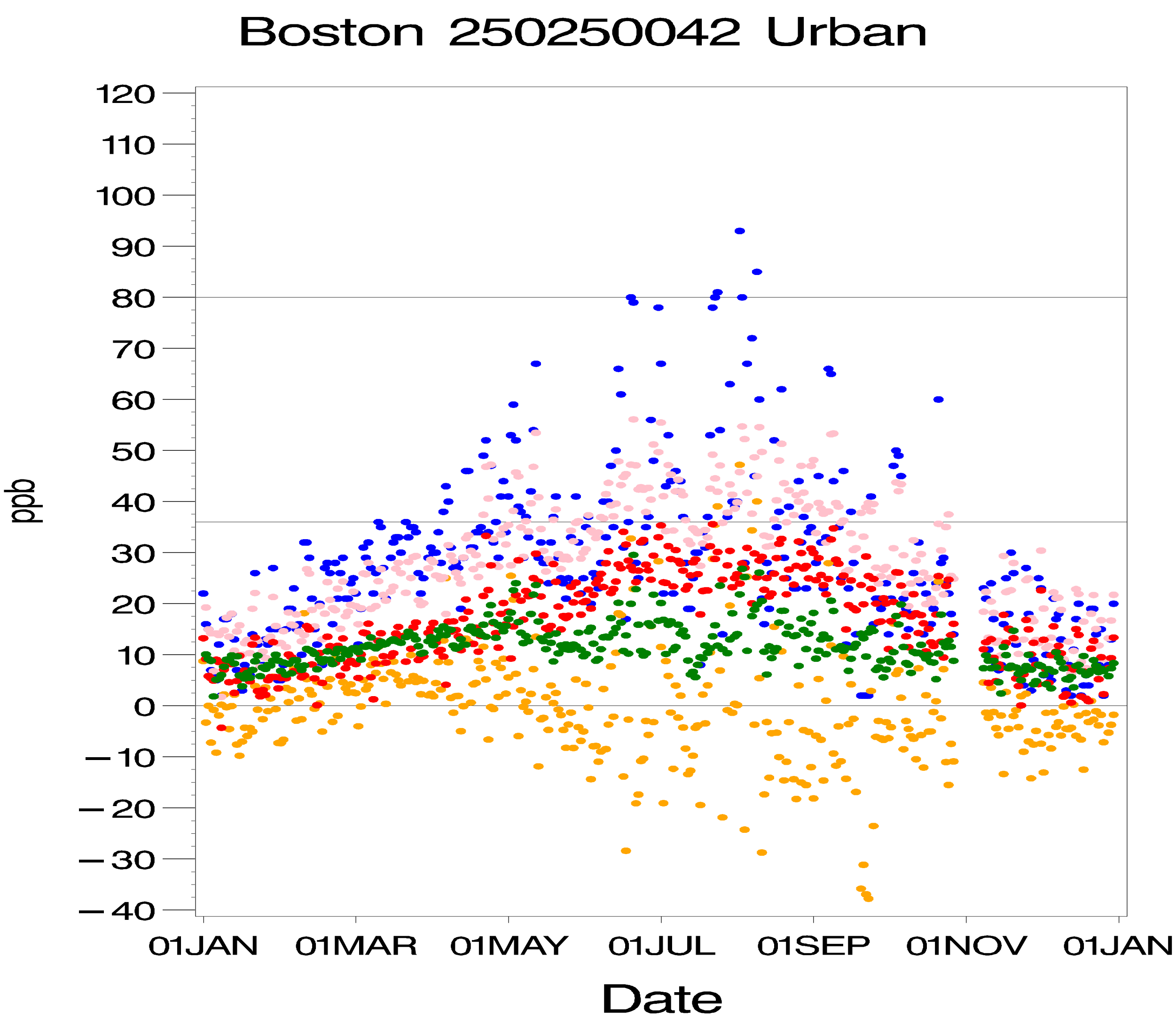
$$N\Big\{ \underbrace{\underbrace{\mu_{Y_t}^u}_{\text{Meanhat}} + \Sigma_{Y_t}^{uo}[\Sigma_{Y_t}^o]^{-1}\big(\boldsymbol{Y}_t^o - \mu_{Y_t}^o\big)}_{\text{Yhat}},\ \Sigma_{Y_t}^u - \Sigma_{Y_t}^{uo}[\Sigma_{Y_t}^o]^{-1}\Sigma_{Y_t}^{ou} \Big\}.$$

Outline

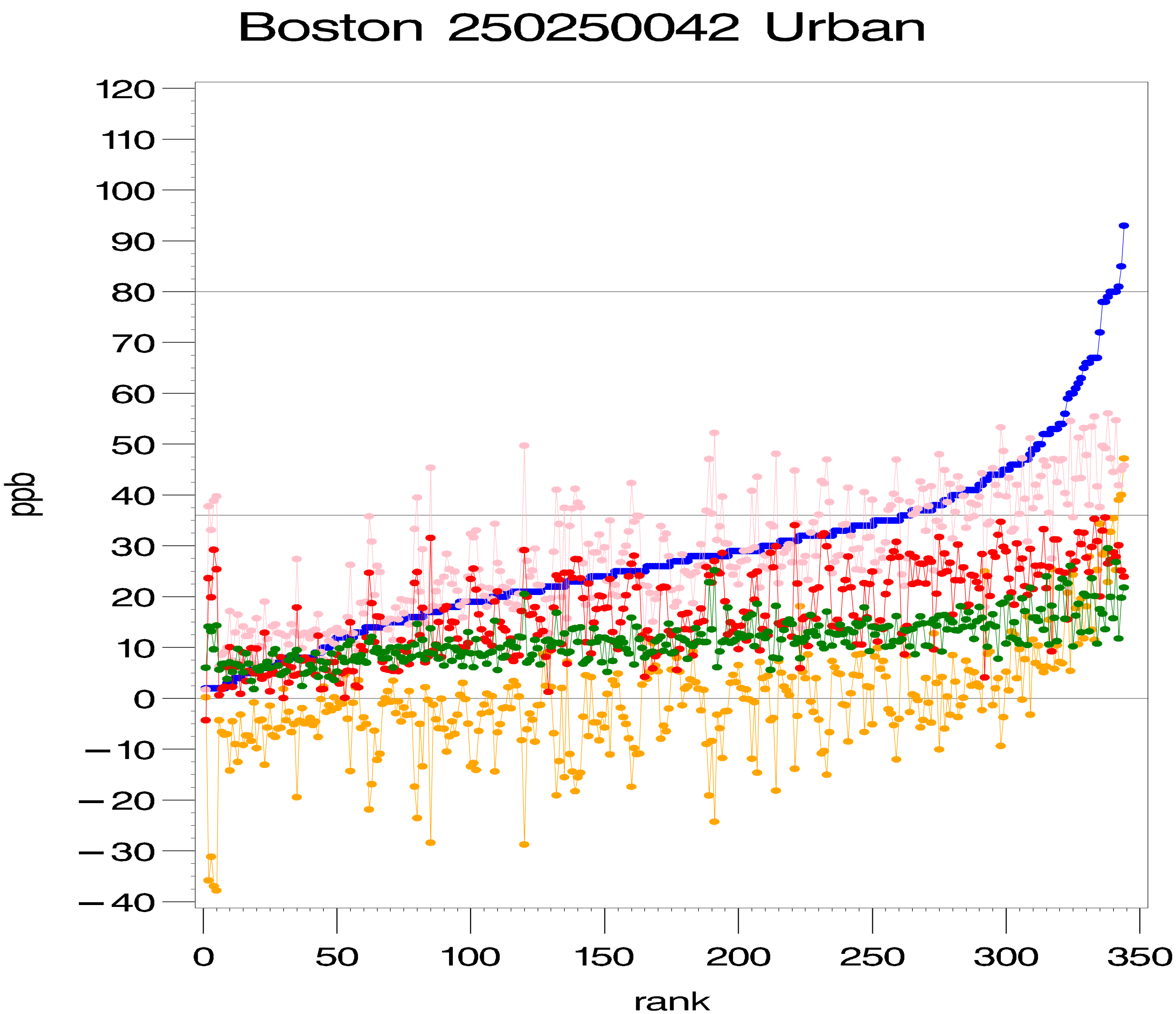
1. Scientific and regulatory context $\rightarrow$ goals
2. The data
3. The models

   Original, Newmean, Metcov
4. Two different predictors for two different purposes
5. **Model performance and CMAQ comparison**
6. Discussion and future work

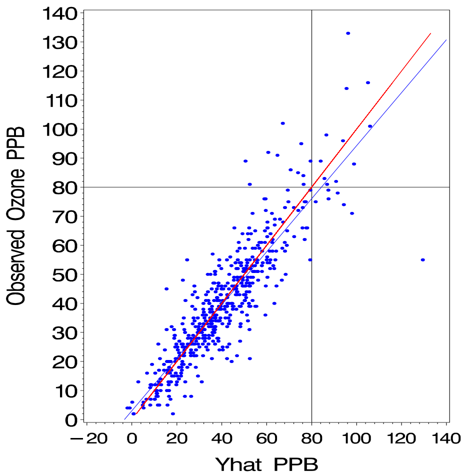
Decomposition of ozone (by day)



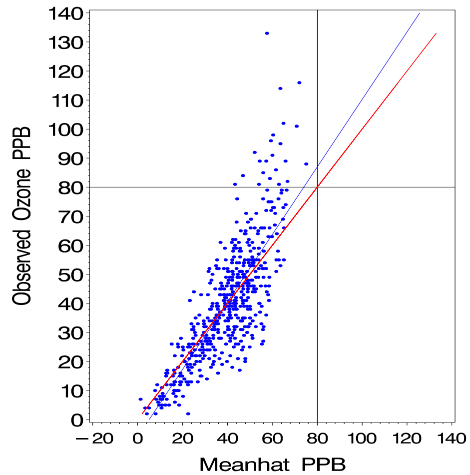
Decomposition of ozone (by rank)



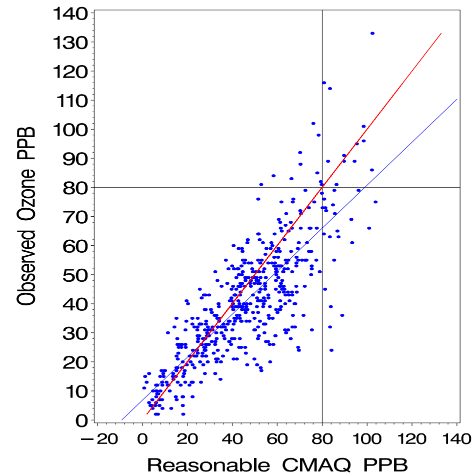
Leave out ten percent scatterplots and residuals



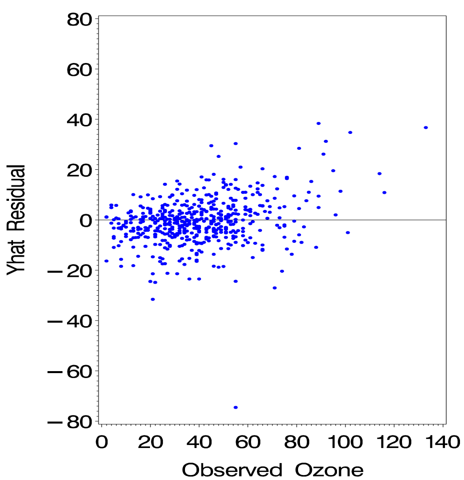
(e)



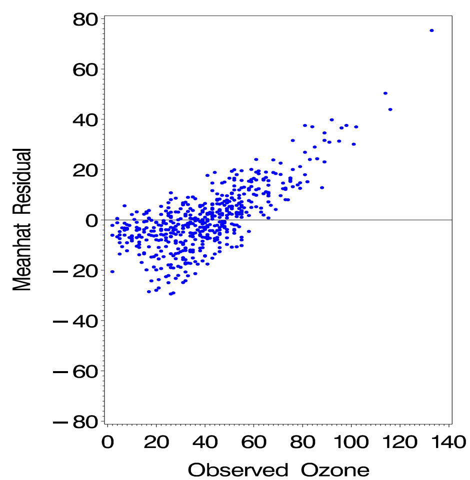
(f)



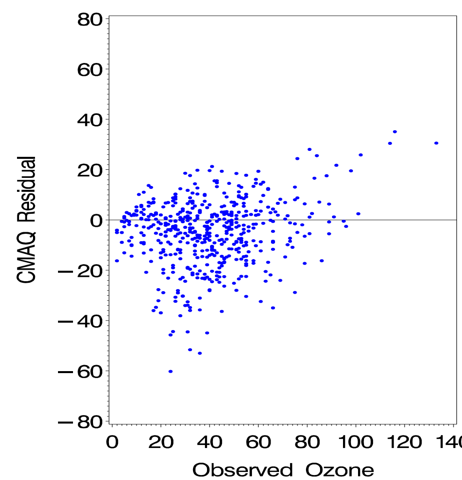
(g)



(h)



(i)



(j)

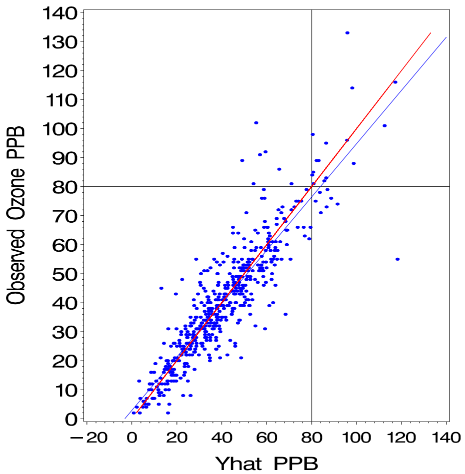
Leave out ten percent regression diagnostics

	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**

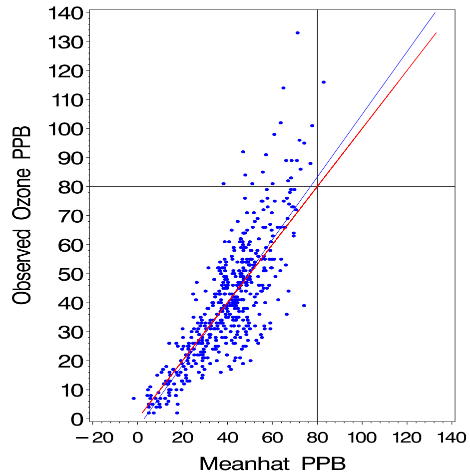
New mean improvements

	N	Model	R^2	RMSE	Slope	Intercept
Yhat	508	Original	.78	9.6	**.91**	**3.0**
					(.022)	**(.98)**
Yhat	508	Newmean	.79	9.4	**.92**	**2.9**
					(.021)	**(.95)**
Meanhat	508	Original	.64	12	**1.2**	**-6.0**
					(.039)	**(1.6)**
Meanhat	508	Newmean	.66	12	**1.1**	**-3.1**
					(.035)	**(1.5)**
Reasonable	508		.64	12	**.74**	**6.8**
CMAQ					**(.025)**	**(1.2)**

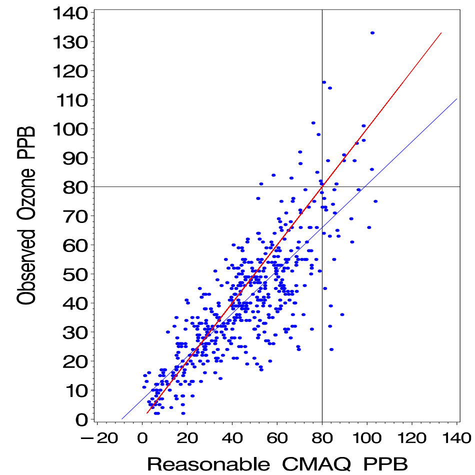
Newmean improvements



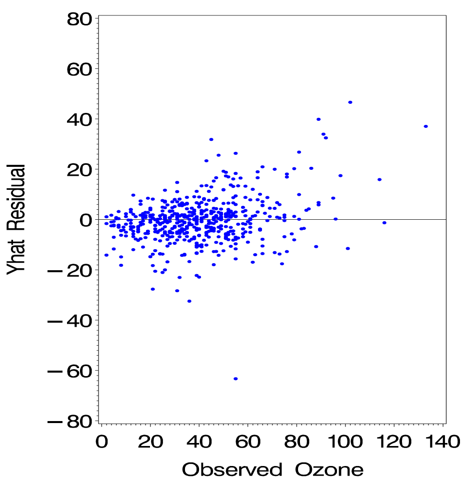
(k)



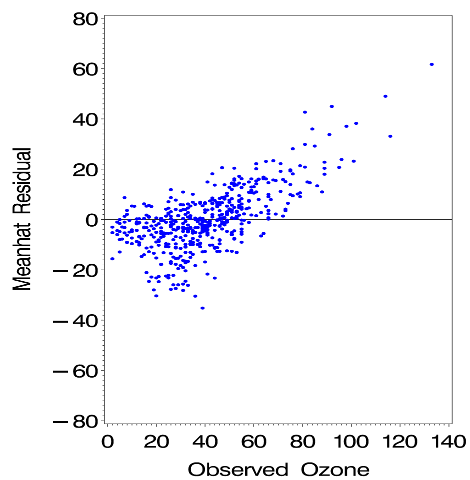
(l)



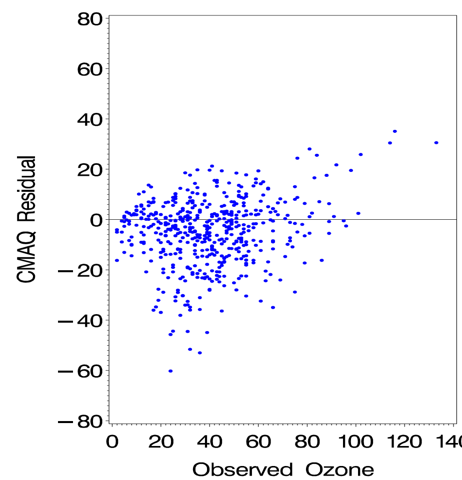
(m)



(n)



(o)

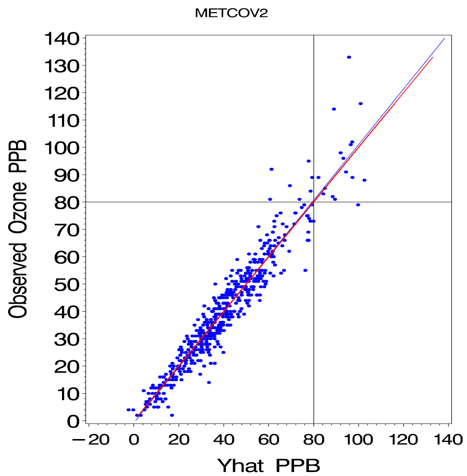


(p)

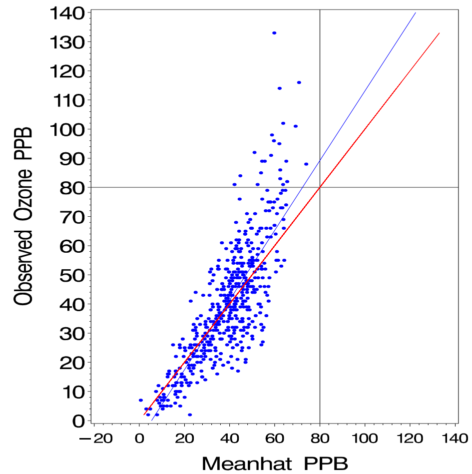
Metcov Improvements

	N	Model	R^2	RMSE	Slope	Intercept
Yhat	508	Original	.78	9.6	**.91**	**3.0**
					(.022)	**(.98)**
Yhat	508	Newmean	.79	9.4	**.92**	**2.9**
					(.021)	**(.95)**
Yhat	508	Metcov	.92	5.9	**1.0**	**-.68**
					(.014)	**(.61)**
Meanhat	508	Original	.64	12	**1.2**	**-6.0**
					(.039)	**(1.6)**
Meanhat	508	Newmean	.66	12	**1.1**	**-3.1**
					(.035)	**(1.5)**
Meanhat	508	Metcov	.65	12	**1.2**	**-6.4**
					(.039)	**(1.6)**
Reasonable	508		.64	12	**.74**	**6.8**
CMAQ					**(.025)**	**(1.2)**

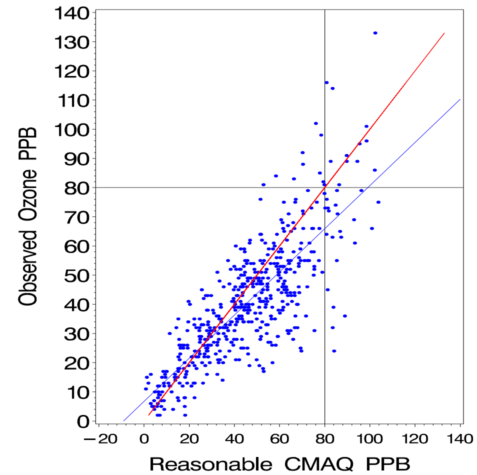
Metcov improvements



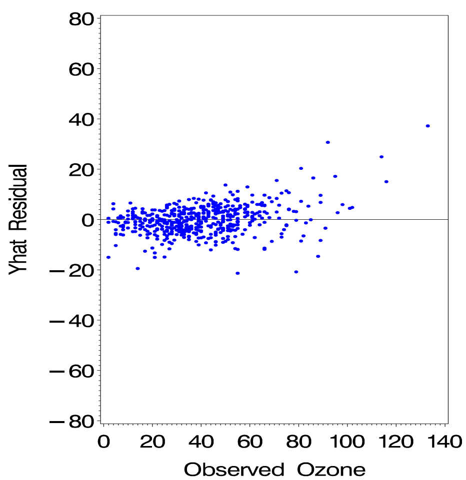
(q)



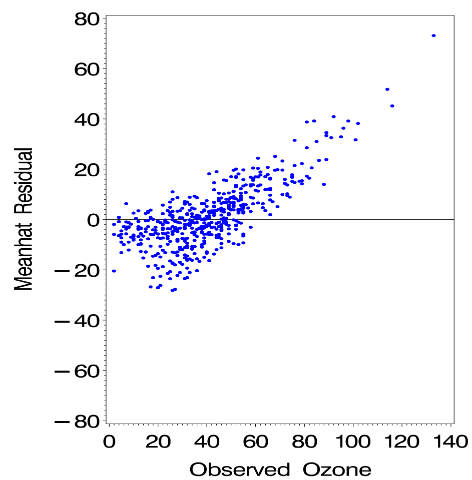
(r)



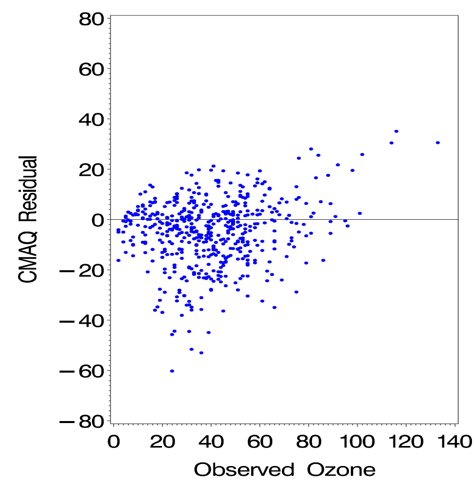
(s)



(t)



(u)



(v)

Outline

1. Scientific and regulatory context → goals
2. The data
3. The model
4. Statistical results in the regulatory context
5. Model performance and CMAQ comparison
6. Two model improvements
7. **Discussion and future work**

Bottom line:

- **The good news**

  Metcov Yhat predictor performs very well for backward-in-time prediction

  Meanhat predictor allows us to decompose ozone into created + transported

- **The bad news**

  We underestimate high ozone values with the meanhat predictor

Future work:

- Different models for different goals:

  Build model from ground up based on ozone season for assessment of future emission programs.

- New explanatory variables? Solar radiation? Cloud cover?

- Explore nonparametric techniques for meanhat predictor.

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

Acknowledgements:

Daniel Tong, Brian Eder

Lance McCluney, Rhonda Thompson, Doug Soloman, Wyatt Appel, Alice Gilleland, Fred Dimmick

Shelly Eberly, Bill Cox, Ellen Baldridge

Tesh Rao, David Mintz, James Hemby, Neil Frank

My email: nail@stat.ncsu.edu

Thank you!