Fusing point and areal level space-time data with application to wet deposition

Alan Gelfand
Duke University

Joint work with Sujit Sahu and David Holland
Chemical Deposition

- Combustion of fossil fuel produces various chemicals including sulfate and nitrate gases.
- In the eastern U.S., most SO$_2$, and NO$_x$ release attributed to power plants.
- Emitted to the air; wet deposition and dry deposition; interest in total deposition.
- Deposition means return to the earth’s surface by means of precipitation (rain or snow) for example.
- Wet Deposition = Precipitation $\times$ Concentration.
- Wet deposition is responsible for damage to lakes, forests, and streams.
Chemical Deposition

- Combustion of fossil fuel produces various chemicals including sulfate and nitrate gases.
- In the eastern U.S., most $\text{SO}_2$, and $\text{NO}_x$ release attributed to power plants.
- Emitted to the air; wet deposition and dry deposition; interest in total deposition.
- Deposition means return to the earth’s surface by means of precipitation (rain or snow) for example.
- Wet Deposition $= \text{Precipitation} \times \text{Concentration}$.
- Wet deposition is responsible for damage to lakes, forests, and streams.
Combustion of fossil fuel produces various chemicals including sulfate and nitrate gases.
In the eastern U.S., most SO$_2$ and NO$_x$ release attributed to power plants.
Emitted to the air; wet deposition and dry deposition; interest in total deposition.
Deposition means return to the earth’s surface by means of precipitation (rain or snow) for example.
Wet Deposition = Precipitation $\times$ Concentration.
Wet deposition is responsible for damage to lakes, forests, and streams.

Alan E. Gelfand  Fusing point and areal level space-time data
Chemical Deposition

- Combustion of fossil fuel produces various chemicals including sulfate and nitrate gases.
- In the eastern U.S., most $\text{SO}_2$, and $\text{NO}_x$ release attributed to power plants.
- Emitted to the air; wet deposition and dry deposition; interest in total deposition.
- Deposition means return to the earth’s surface by means of precipitation (rain or snow) for example.
  - Wet Deposition = Precipitation $\times$ Concentration.
  - Wet deposition is responsible for damage to lakes, forests, and streams.
Combustion of fossil fuel produces various chemicals including sulfate and nitrate gases.

In the eastern U.S., most $\text{SO}_2$, and $\text{NO}_x$ release attributed to power plants.

Emitted to the air; wet deposition and dry deposition; interest in total deposition.

Deposition means return to the earth’s surface by means of precipitation (rain or snow) for example.

**Wet Deposition = Precipitation $\times$ Concentration.**

Wet deposition is responsible for damage to lakes, forests, and streams.
Chemical Deposition

- Combustion of fossil fuel produces various chemicals including sulfate and nitrate gases.
- In the eastern U.S., most SO$_2$, and NO$_x$ release attributed to power plants.
- Emitted to the air; wet deposition and dry deposition; interest in total deposition.
- Deposition means return to the earth’s surface by means of precipitation (rain or snow) for example.
- **Wet Deposition** = Precipitation $\times$ Concentration.
- Wet deposition is responsible for damage to lakes, forests, and streams.
NADP collects point-referenced data at several sites. They then use simple interpolation to produce maps.
Community Multi-scale Air Quality Model (CMAQ)

- A computer simulation model which produces “averaged” output on 36km, 12 km (used here), and now 4 km grid cells.
- Uses variables such as power station emission volumes, meteorological data, land-use, etc. with atmospheric science (appropriate differential equations) to predict deposition levels. Not driven by monitoring station data.
- Predictions are biased but no missing data; monitoring data provide more accurate deposition but “missingness”
The CMAQ model

- Community Multi-scale Air Quality Model (CMAQ)
- A computer simulation model which produces “averaged” output on 36km, 12 km (used here), and now 4 km grid cells
- Uses variables such as power station emission volumes, meteorological data, land-use, etc. with atmospheric science (appropriate differential equations) to predict deposition levels. Not driven by monitoring station data.
- Predictions are biased but no missing data; monitoring data provide more accurate deposition but “missingness”
Community Multi-scale Air Quality Model (CMAQ)

A computer simulation model which produces “averaged” output on 36km, 12 km (used here), and now 4 km grid cells

Uses variables such as power station emission volumes, meteorological data, land-use, etc. with atmospheric science (appropriate differential equations) to predict deposition levels. Not driven by monitoring station data.

Predictions are biased but no missing data; monitoring data provide more accurate deposition but “missingness”
Community Multi-scale Air Quality Model (CMAQ)

A computer simulation model which produces “averaged” output on 36km, 12 km (used here), and now 4 km grid cells.

Uses variables such as power station emission volumes, meteorological data, land-use, etc. with atmospheric science (appropriate differential equations) to predict deposition levels. Not driven by monitoring station data.

Predictions are biased but no missing data; monitoring data provide more accurate deposition but “missingness”
Our Contribution

- A fully model-based framework for fusing the NADP and CMAQ wet deposition data
- To accommodate the point masses at 0, i.e., no wet deposition if no precipitation
- To accommodate misalignment between NADP data at points and CMAQ data at grid cells in a computational feasible way across space and time
- To provide spatial interpolation and temporal aggregation
- Both sulfate and nitrate deposition
Our Contribution

• A fully model-based framework for fusing the NADP and CMAQ wet deposition data
• To accommodate the point masses at 0, i.e., no wet deposition if no precipitation
• To accommodate misalignment between NADP data at points and CMAQ data at grid cells in a computational feasible way across space and time
• To provide spatial interpolation and temporal aggregation
• Both sulfate and nitrate deposition
Our Contribution

- A fully model-based framework for fusing the NADP and CMAQ wet deposition data
- To accommodate the point masses at 0, i.e., no wet deposition if no precipitation
- To accommodate misalignment between NADP data at points and CMAQ data at grid cells in a computational feasible way across space and time
- To provide spatial interpolation and temporal aggregation
- Both sulfate and nitrate deposition
Our Contribution

- A fully model-based framework for fusing the NADP and CMAQ wet deposition data
- To accommodate the point masses at 0, i.e., no wet deposition if no precipitation
- To accommodate misalignment between NADP data at points and CMAQ data at grid cells in a computational feasible way across space and time
- To provide spatial interpolation and temporal aggregation
- Both sulfate and nitrate deposition
Our Contribution

- A fully model-based framework for fusing the NADP and CMAQ wet deposition data
- To accommodate the point masses at 0, i.e., no wet deposition if no precipitation
- To accommodate misalignment between NADP data at points and CMAQ data at grid cells in a computational feasible way across space and time
- To provide spatial interpolation and temporal aggregation
- Both sulfate and nitrate deposition
Inverse Distance weighting (IDW)

- “Poor person’s” methodology
- Value at a new site = weighted mean of observations,
- Weights inversely proportional to the square of the distance.

Problems

- Not model based!
- Unable to accommodate known covariate - precipitation!
- Unable to handle 0’s, unable to handle missing data
- Can’t fuse with model output data
- With dynamic data, can only do independent weekly or aggregated annually

No associated uncertainty maps!
Inverse Distance weighting (IDW)

- “Poor person’s” methodology
- Value at a new site = weighted mean of observations,
  Weights inversely proportional to the square of the distance.

Problems

- Not model based!
- Unable to accommodate known covariate - precipitation!
- Unable to handle 0’s, unable to handle missing data
- Can’t fuse with model output data
- With dynamic data, can only do independent weekly or aggregated annually

No associated uncertainty maps!
Inverse Distance weighting (IDW)

- “Poor person’s” methodology
- Value at a new site = weighted mean of observations,
- Weights inversely proportional to the square of the distance.

Problems

- Not model based!
- Unable to accommodate known covariate - precipitation!
- Unable to handle 0’s, unable to handle missing data
- Can’t fuse with model output data
- With dynamic data, can only do independent weekly or aggregated annually

No associated uncertainty maps!
Change of support problem

Fuentes and Raftery, 2005

- Need to upscale (block-average) point level $Z(s, t)$ to obtain grid level $Z(A_j, t)$.

$$Z(A_j, t) = \frac{1}{|A_j|} \int_{A_j} Z(s, t) \, ds,$$  \hspace{1cm} (1)

Many more $A$’s than $s$’s

- Use MEM (measurement error model) at point level centred around grid level values
- Make inference at the point level by downscaling
- Huge computational advantages.
Change of support problem

Fuentes and Raftery, 2005

- Need to upscale (block-average) point level $Z(s, t)$ to obtain grid level $Z(A_j, t)$.

\[
Z(A_j, t) = \frac{1}{|A_j|} \int_{A_j} Z(s, t) \, ds,
\]

(1)

Many more $A$'s than $s$'s

- Use MEM (measurement error model) at point level centred around grid level values
- Make inference at the point level by downscaling
- Huge computational advantages

Alan E. Gelfand

Fusing point and areal level space-time data
- Model data weekly.
- Fuse with gridded weekly CMAQ output.
- Use weekly precipitation information, available from other monitoring networks.
- Interpolate in space, predict in time
- Obtain quarterly and annual maps.
- Reveal spatial pattern in deposition.
- Results are illustrative, not definitive.
Our data set

Data

- Use 120 sites to estimate, remaining 8 to validate.
- Weekly CMAQ output from $J = 33,390$ grid cells (about 1.7 million values!)
- Weekly precipitation data from 2827 predictive sites.
Location of the NADP

Figure: A map of the study region; points denote the NADP sites for fitting and A-H denote the eight validation sites.
Figure: Map of annual total precipitation in 2001.
Figure: Boxplot of weekly depositions: (a) sulfate and (b) nitrate.
Exploratory Analyses ...

Figure: Deposition against precipitation (both on the log scale): (a) sulfate and (b) nitrate.
Figure: Deposition at the NADP sites against the CMAQ values in the grid cell covering the corresponding NADP site on the log scale: (a) sulfate and (b) nitrate.
No deposition, $Z(s_i, t)$, without precipitation, $P(s_i, t)$; enforced by a latent atmospheric space-time process $V(s_i, t)$ below, $i = 1, \ldots, n = 120$, and for each week $t$, $t = 1, \ldots, 52$.

Similarly, model CMAQ output, $Q(A_j, t)$ for each grid cell $A_j$, $j = 1, \ldots, J = 33,390$ and for each week $t$, modeled using a latent atmospheric areal process $\tilde{V}(A_j, t)$.

Model everything on the log-scale; latent processes take care of point masses at zero. Avoid $\log(0)$ problems.
First stage models

Precipitation model

\[ P(s_i, t) = \begin{cases} \exp(U(s_i, t)) & \text{if } V(s_i, t) > 0 \\ 0 & \text{otherwise,} \end{cases} \]

Deposition model

\[ Z(s_i, t) = \begin{cases} \exp(Y(s_i, t)) & \text{if } V(s_i, t) > 0 \\ 0 & \text{otherwise.} \end{cases} \]

Model for CMAQ output

\[ Q(A_j, t) = \begin{cases} \exp(X(A_j, t)) & \text{if } \tilde{V}(A_j, t) > 0 \\ 0 & \text{otherwise.} \end{cases} \]
Clarification

- $P$’s, $Z$’s, and $Q$’s are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.

- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.

- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.

- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.

- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.

- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.

- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 

Alan E. Gelfand

Fusing point and areal level space-time data
Clarification

- $P's$, $Z's$, and $Q's$ are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.
- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.
- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.
- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.
- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.
- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.
- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 

Alan E. Gelfand | Fusing point and areal level space-time data
Clarification

- $P$’s, $Z$’s, and $Q$’s are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.

- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.

- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.

- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.

- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.

- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.

- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 

Alan E. Gelfand  
Fusing point and areal level space-time data
Clarification

- $P$’s, $Z$’s, and $Q$’s are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.
- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.
- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.
- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.
- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.
- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.
- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 

Alan E. Gelfand  
Fusing point and areal level space-time data
Clarification

- $P$’s, $Z$’s, and $Q$’s are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.
- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.
- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.
- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.
- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.
- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.
- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 
Clarification

- $P$’s, $Z$’s, and $Q$’s are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.
- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.
- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.
- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.
- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.
- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.
- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 
Clarification

- $P$’s, $Z$’s, and $Q$’s are the observed precipitation, NADP deposition, and CMAQ deposition, respectively.
- $V(s, t)$ is a conceptual point level latent atmospheric process which drives $P(s, t)$ and $Z(s, t)$.
- $P(s, t)$ and $Z(s, t) = 0$ if $V(s, t) \leq 0$.
- $U(s, t)$ and $Y(s, t)$ are log precipitation and deposition, respectively.
- Models below will specify their values when $V(s, t) \leq 0$ or if $P(s, t)$ or $Z(s, t)$ are missing.
- $\tilde{V}(A_j, t)$ is a conceptual areal level latent atmospheric process which drives $Q(A_j, t)$.
- $X(A_j, t)$ is log CMAQ output where modeling below will specify its values when $\tilde{V}(A_j, t) \leq 0$. 

Alan E. Gelfand  
Fusing point and areal level space-time data
The first stage likelihood

\[ f(P, Z, Q | U, Y, X, V, \tilde{V}) = f(P | U, V) \times f(Z | Y, V) \times f(Q | X, \tilde{V}) \]

which takes the form

\[
\prod_{t=1}^{T} \left[ \prod_{i=1}^{n} \left\{ 1 \exp(u(s_i, t)) 1 \exp(y(s_i, t)) I(v(s_i, t) > 0) \right\} \prod_{j=1}^{J} \left\{ 1 \exp(x(A_j, t)) I(\tilde{v}(A_j, t) > 0) \right\} \right]
\]

where \(1_x\) denotes a degenerate distribution with point mass at \(x\) and \(I(\cdot)\) is the indicator function.
Deposition Model

\[ Y(s_i, t) = \beta_0 + \beta_1 U(s_i, t) + \beta_2 V(s_i, t) \\
+ (b_0 + b(s_i)) X(A_{k_i}, t) \\
+ \eta(s_i, t) + \epsilon(s_i, t). \]

- Spatially varying coefficients, \( b = (b(s_1), \ldots, b(s_n))' \) is a Gaussian process (GP).
- Spatio-temporal intercept \( \eta_t = (\eta(s_1, t), \ldots, \eta(s_n, t))' \) is a GP independent in time.
- Allow for spatially varying calibration of CMAQ. Could imagine common \( \eta(s_i). \)
- \( \epsilon(s_i, t) \sim N(0, \sigma^2_\epsilon) \), provides the nugget effect.
The second stage models ...
Specification of latent processes

Measurement Error Model (MEM)

\[ V(s_i, t) \sim N(\tilde{V}(A_{ki}, t), \sigma^2_V), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \]

The process \( \tilde{V}(A_j, t) \) is AR in time and CAR in space

\[ \tilde{V}(A_j, t) = \rho \tilde{V}(A_j, t - 1) + \zeta(A_j, t), \]

\[ \zeta(A_j, t) \sim N\left( \sum_{i=1}^{J} h_{ji} \zeta(A_i, t), \frac{\sigma^2_\zeta}{m_j} \right), \]

Let \( \partial_j \) define the \( m_j \) neighboring grid cells of the cell \( A_j \).

\[ h_{ji} = \begin{cases} \frac{1}{m_j} & \text{if } i \in \partial_j \\ 0 & \text{otherwise.} \end{cases} \]
Assume the initial condition for $\tilde{V}_0$:

$$\tilde{V}(A_j, 0) = \frac{1}{T} \sum_{t=1}^{T} X(A_j, t),$$

giving $\tilde{V}_0$.

Now we can write the CAR in closed form:

$$f(\tilde{V}_t | \tilde{V}_{t-1}, \rho, \sigma^2_\zeta) \propto \exp \left\{-\frac{1}{2} \left(\tilde{V}_t - \rho \tilde{V}_{t-1}\right)' D^{-1} (I - H) \left(\tilde{V}_t - \rho \tilde{V}_{t-1}\right) \right\},$$

$D$ is a diagonal matrix with entries $\sigma^2_\zeta / m_j$.

Note that this is an improper CAR.
Note that we can have $Z > 0$, $Q = 0$ and vice versa. Therefore $V$ and $\tilde{V}$ can have opposite signs. This arises because we are modeling at two different scales - need processes at two different scales.

We can view $V(s, t) - \tilde{V}(A, t)$ as a deviation from the areal average. We assume these realized deviations are independent across space and time.

We have a conditional model for $V$ and $X$ given $\tilde{V}$. The resulting marginal model for $U$ and $Y$ given $\tilde{V}$ is multiscale - additive random effects at two scales.
Note that we can have $Z > 0$, $Q = 0$ and vice versa. Therefore $V$ and $\tilde{V}$ can have opposite signs.

This arises because we are modeling at two different scales - need processes at two different scales.

We can view $V(s, t) - \tilde{V}(A, t)$ as a deviation from the areal average. We assume these realized deviations are independent across space and time.

We have a conditional model for $V$ and $X$ given $\tilde{V}$. The resulting marginal model for $U$ and $Y$ given $\tilde{V}$ is *multiscale* - additive random effects at two scales.
Clarification

- Note that we can have $Z > 0, Q = 0$ and vice versa. Therefore $V$ and $\tilde{V}$ can have opposite signs.
- This arises because we are modeling at two different scales - need processes at two different scales.
- We can view $V(s, t) - \tilde{V}(A, t)$ as a deviation from the areal average. We assume these realized deviations are independent across space and time.
- We have a conditional model for $V$ and $X$ given $\tilde{V}$. The resulting marginal model for $U$ and $Y$ given $\tilde{V}$ is multiscale - additive random effects at two scales.
Note that we can have $Z > 0$, $Q = 0$ and vice versa. Therefore $V$ and $\tilde{V}$ can have opposite signs.

This arises because we are modeling at two different scales - need processes at two different scales.

We can view $V(s, t) - \tilde{V}(A, t)$ as a deviation from the areal average. We assume these realized deviations are independent across space and time.

We have a conditional model for $V$ and $X$ given $\tilde{V}$. The resulting marginal model for $U$ and $Y$ given $\tilde{V}$ is *multiscale* - additive random effects at two scales.
The second stage specification

- **Deposition model:** $f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta)$.
- **Space-time intercept:** $f(\eta_t|\theta)$.
- **Precipitation model:** $f(U_t|V_t, \theta)$.
- **Measurement Error model:** $f(V_t|\tilde{V}_t^{(1)}, \theta)$.
- **CMAQ Model:** $f(X_t|\tilde{V}_t, \theta)$.
- **AR and CAR Model:** $f(\tilde{V}_t|\tilde{V}_{t-1}, \theta)$.

$$\prod_{t=1}^{T} \left[ f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t|\theta) f(U_t|V_t, \theta) \right. \\ \left. \times f(V_t|\tilde{V}_t^{(1)}, \theta) \times f(X_t|\tilde{V}_t, \theta) f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \right] f(b|\theta).$$

$$\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma_\delta^2, \sigma_b^2, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\psi^2, \sigma_v^2, \sigma_\zeta^2).$$
The second stage specification

- **Deposition model:** \( f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \).
- **Space-time intercept:** \( f(\eta_t|\theta) \).
- **Precipitation model:** \( f(U_t|V_t, \theta) \).
- **Measurement Error model:** \( f(V_t|\tilde{V}_t^{(1)}, \theta) \).
- **CMAQ Model:** \( f(X_t|\tilde{V}_t, \theta) \).
- **AR and CAR Model:** \( f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \).

\[
\prod_{t=1}^{T} \left[ f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t|\theta) f(U_t|V_t, \theta) \right. \\
\left. \times f(V_t|\tilde{V}_t^{(1)}, \theta) \times f(X_t|\tilde{V}_t, \theta) f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \right] f(b|\theta).
\]

\( \theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma_\delta^2, \sigma_b^2, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\psi^2, \sigma_V^2, \sigma_\zeta^2) \).
The second stage specification

- **Deposition model**: \( f(Y_t | U_t, V_t, X_t, \eta_t, b, \theta) \).
- **Space-time intercept**: \( f(\eta_t | \theta) \).
- **Precipitation model**: \( f(U_t | V_t, \theta) \).
- **Measurement Error model**: \( f(V_t | \tilde{V}_t^{(1)}, \theta) \).
- **CMAQ Model**: \( f(X_t | \tilde{V}_t, \theta) \).
- **AR and CAR Model**: \( f(\tilde{V}_t | \tilde{V}_{t-1}, \theta) \).

\[
\prod_{t=1}^{T} \left[ f(Y_t | U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t | \theta) f(U_t | V_t, \theta) \times f(V_t | \tilde{V}_t^{(1)}, \theta) \times f(X_t | \tilde{V}_t, \theta) f(\tilde{V}_t | \tilde{V}_{t-1}, \theta) \right] f(b | \theta).
\]

\[\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma_\delta^2, \sigma_b^2, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\psi^2, \sigma_v^2, \sigma_\zeta^2).\]
The second stage specification

- **Deposition model**: \( f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \).
- **Space-time intercept**: \( f(\eta_t|\theta) \).
- **Precipitation model**: \( f(U_t|V_t, \theta) \).
- **Measurement Error model**: \( f(V_t|\tilde{V}^{(1)}_t, \theta) \).
- **CMAQ Model**: \( f(X_t|\tilde{V}_t, \theta) \).
- **AR and CAR Model**: \( f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \).

\[
\prod_{t=1}^T \left[ f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t|\theta)\, f(U_t|V_t, \theta) \times f(V_t|\tilde{V}^{(1)}_t, \theta) \times f(X_t|\tilde{V}_t, \theta)\, f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \right] f(b|\theta).
\]

\[
\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma_\delta^2, \sigma_b^2, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\psi^2, \sigma_V^2, \sigma_\zeta^2).
\]
The second stage specification

- **Deposition model**: \( f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \).
- **Space-time intercept**: \( f(\eta_t|\theta) \).
- **Precipitation model**: \( f(U_t|V_t, \theta) \).
- **Measurement Error model**: \( f(V_t|\tilde{V}_t^{(1)}, \theta) \)
- **CMAQ Model**: \( f(X_t|\tilde{V}_t, \theta) \)
- **AR and CAR Model**: \( f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \).

\[
\prod_{t=1}^{T} \left[ f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t|\theta) f(U_t|V_t, \theta) \right. \\
\left. \times f(V_t|\tilde{V}_t^{(1)}, \theta) \times f(X_t|\tilde{V}_t, \theta) f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \right] f(b|\theta).
\]

\[ \theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma^2_{\delta}, \sigma^2_{\beta}, \sigma^2_{\eta}, \sigma^2_\epsilon, \sigma^2_\psi, \sigma^2_\nu, \sigma^2_\zeta). \]
The second stage specification

- **Deposition model:** \( f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \).
- **Space-time intercept:** \( f(\eta_t|\theta) \).
- **Precipitation model:** \( f(U_t|V_t, \theta) \).
- **Measurement Error model:** \( f(V_t|\tilde{V}_t^{(1)}, \theta) \).
- **CMAQ Model:** \( f(X_t|\tilde{V}_t, \theta) \).
- **AR and CAR Model:** \( f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \).

\[
\prod_{t=1}^{T} \left[ f(Y_t|U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t|\theta) \times f(U_t|V_t, \theta) \times f(V_t|\tilde{V}_t^{(1)}, \theta) \times f(X_t|\tilde{V}_t, \theta) \times f(\tilde{V}_t|\tilde{V}_{t-1}, \theta) \right] f(b|\theta).
\]

\[\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma_\delta^2, \sigma_b^2, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\psi^2, \sigma_v^2, \sigma_\zeta^2).\]
The second stage specification

- **Deposition model**: \( f(Y_t | U_t, V_t, X_t, \eta_t, b, \theta) \).
- **Space-time intercept**: \( f(\eta_t | \theta) \).
- **Precipitation model**: \( f(U_t | V_t, \theta) \).
- **Measurement Error model**: \( f(V_t | \tilde{V}_t^{(1)}, \theta) \).
- **CMAQ Model**: \( f(X_t | \tilde{V}_t, \theta) \).
- **AR and CAR Model**: \( f(\tilde{V}_t | \tilde{V}_{t-1}, \theta) \).

\[
\prod_{t=1}^{T} \left[ f(Y_t | U_t, V_t, X_t, \eta_t, b, \theta) \times f(\eta_t | \theta) f(U_t | V_t, \theta) \times f(V_t | \tilde{V}_t^{(1)}, \theta) \times f(X_t | \tilde{V}_t, \theta) f(\tilde{V}_t | \tilde{V}_{t-1}, \theta) \right] f(b | \theta).
\]

\( \theta = (\alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, b_0, \gamma_0, \gamma_1, \rho, \sigma_\delta^2, \sigma_b^2, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\psi^2, \sigma_v^2, \sigma_\zeta^2) \).
Graphical representation of our model.

- $\delta(s_i, t)$
- $V(s_i, t)$
- $\tilde{V}(A_{k_i}, t)$
- $\tilde{V}(A_{k_i}, t)$
- $P(s_i, t)$ (observed precipitation)
- $Z(s_i, t)$ (observed deposition)
- $Q(A_{k_i}, t)$ (CMAQ model output)

Regional atmospheric driver (point)
Regional atmospheric centering process (areal)

$\eta(s_i, t)$

$U(s_i, t)$
$Y(s_i, t)$

Alan E. Gelfand
Fusing point and areal level space-time data
Alan E. Gelfand  
Fusing point and areal level space-time data
Predictions at new locations

At a new site \( s' \) and time \( t' \) we need \( Z(s', t') \) which depends on \( Y(s', t') \). If \( P(s', t') = 0 \) then \( Z(s', t') = 0 \).

Suppose otherwise.

- Bayesian predictive distributions:

\[
\pi(Z_{\text{pred}}|Z_{\text{obs}}) = \int \pi(Z_{\text{pred}}|\text{par}) \pi(\text{par}|Z_{\text{obs}}) d\text{par}.
\]

- Need to simulate \( Y(s', t') \).
- \( V(s', t') \sim N(\tilde{V}(A', t'), \sigma_v^2) \).
- \( U(s', t') \), \( \eta(s', t') \) and \( b(s') \) are simulated from the conditional distributions at \( s' \) given \( s_1, \ldots, s_n \). Kriging.
- \( X(A', t') = \log Q(A'', t') \) if \( Q(A', t') > 0 \), otherwise updated in the MCMC.
- More details in the paper.
At a new site $s'$ and time $t'$ we need $Z(s', t')$ which depends on $Y(s', t')$. If $P(s', t') = 0$ then $Z(s', t') = 0$.

Suppose otherwise.

- Bayesian predictive distributions:
  \[
  \pi(Z_{\text{pred}}|Z_{\text{obs}}) = \int \pi(Z_{\text{pred}}|\text{par}) \pi(\text{par}|Z_{\text{obs}}) d\text{par}.
  \]

- Need to simulate $Y(s', t')$.
  - $V(s', t') \sim N(\bar{V}(A', t'), \sigma_v^2)$.
  - $U(s', t'), \eta(s', t')$ and $b(s')$ are simulated from the conditional distributions at $s'$ given $s_1, \ldots, s_n$. Kriging.
  - $X(A', t') = \log Q(A'', t')$ if $Q(A', t') > 0$, otherwise updated in the MCMC.

- More details in the paper.
Predictions at new locations

At a new site \( \mathbf{s}' \) and time \( t' \) we need \( Z(\mathbf{s}', t') \) which depends on \( Y(\mathbf{s}', t') \). If \( P(\mathbf{s}', t') = 0 \) then \( Z(\mathbf{s}', t') = 0 \).

Suppose otherwise.

- Bayesian predictive distributions:

\[
\pi(Z_{\text{pred}}|Z_{\text{obs}}) = \int \pi(Z_{\text{pred}}|\text{par}) \pi(\text{par}|Z_{\text{obs}}) d\text{par}.
\]

- Need to simulate \( Y(\mathbf{s}', t') \).
- \( V(\mathbf{s}', t') \sim N(\tilde{V}(\mathbf{A}', t'), \sigma^2_v) \).
- \( U(\mathbf{s}', t'), \eta(\mathbf{s}', t') \) and \( b(\mathbf{s}') \) are simulated from the conditional distributions at \( \mathbf{s}' \) given \( \mathbf{s}_1, \ldots, \mathbf{s}_n \).

Kriging.

- \( X(\mathbf{A}', t') = \log Q(\mathbf{A}'', t') \) if \( Q(\mathbf{A}', t') > 0 \), otherwise updated in the MCMC.

- More details in the paper.
At a new site $s'$ and time $t'$ we need $Z(s', t')$ which depends on $Y(s', t')$. If $P(s', t') = 0$ then $Z(s', t') = 0$.

Suppose otherwise.

- Bayesian predictive distributions:
  \[
  \pi(Z_{\text{pred}} | Z_{\text{obs}}) = \int \pi(Z_{\text{pred}} | \text{par}) \pi(\text{par} | Z_{\text{obs}}) d\text{par}.
  \]

- Need to simulate $Y(s', t')$.
- $V(s', t') \sim N(\tilde{V}(A', t'), \sigma^2_V)$.
- $U(s', t'), \eta(s', t')$ and $b(s')$ are simulated from the conditional distributions at $s'$ given $s_1, \ldots, s_n$. Kriging.
- $X(A', t') = \log Q(A'', t')$ if $Q(A', t') > 0$, otherwise updated in the MCMC.
- More details in the paper.
Choosing the spatial decay parameters

- Estimation is challenging due to weak identifiability of variances and ranges.
- Inconsistent see Stein (1999) and Zhang (2004).
- So, we fix $\phi_\eta$, $\phi_\delta$ and $\phi_b$

Use validation mean-square error

\[
\text{VMSE} = \frac{1}{n_v} \sum_{i=1}^{8} \sum_{t=1}^{52} \left( Z(s_i^*, t) - \hat{Z}(s_i^*, t) \right)^2 I(\text{observed})
\]

$n_v = \text{total number of validation observed} (=407, \text{here})$.

- Optimal ranges were 500, 1000 and 500 kilometers.
- VMSE is not sensitive near these values.
Figure: Validation versus the observed values at the 8 reserved sites. Validation prediction intervals are plotted as vertical lines. (a) sulfate and (b) nitrate.
Figure: (a) Sulfate. (b) Nitrate. (c) The s.d. for sulfate. (d) The s.d. for nitrate.
Spatially varying slopes?

Do we need the spatially varying $b(s)$’s?

- Only a few of the $b(s_i)$ are significant; they are small relative to their standard errors.
- Importance of precipitation and the spatially varying intercept makes it difficult to find spatially varying contribution of CMAQ
- Fusion approaches also have not found spatially varying intercepts
- Still can see space-time bias in CMAQ by comparing model predictions with CMAQ output.
Do we need the spatially varying $b(s)$’s?

Only a few of the $b(s_i)$ are significant; they are small relative to their standard errors.

Importance of precipitation and the spatially varying intercept makes it difficult to find spatially varying contribution of CMAQ

Fusion approaches also have not found spatially varying intercepts

Still can see space-time bias in CMAQ by comparing model predictions with CMAQ output.
Do we need the spatially varying $b(s)$’s?

Only a few of the $b(s_i)$ are significant; they are small relative to their standard errors.

Importance of precipitation and the spatially varying intercept makes it difficult to find spatially varying contribution of CMAQ.

Fusion approaches also have not found spatially varying intercepts.

Still can see space-time bias in CMAQ by comparing model predictions with CMAQ output.
Do we need the spatially varying $b(s)$’s?

Only a few of the $b(s_i)$ are significant; they are small relative to their standard errors.

Importance of precipitation and the spatially varying intercept makes it difficult to find spatially varying contribution of CMAQ

Fusion approaches also have not found spatially varying intercepts

Still can see space-time bias in CMAQ by comparing model predictions with CMAQ output.
Spatially varying slopes?

- Do we need the spatially varying $b(s)$’s?
- Only a few of the $b(s_i)$ are significant; they are small relative to their standard errors.
- Importance of precipitation and the spatially varying intercept makes it difficult to find spatially varying contribution of CMAQ
- Fusion approaches also have not found spatially varying intercepts
- Still can see space-time bias in CMAQ by comparing model predictions with CMAQ output.
## Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sulfate</th>
<th>Nitrate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>sd</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>-0.4497</td>
<td>0.0871</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.1787</td>
<td>0.0379</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1.9414</td>
<td>0.0196</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.9103</td>
<td>0.0067</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0029</td>
<td>0.0062</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.0490</td>
<td>0.0053</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>-3.0768</td>
<td>0.0035</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.8957</td>
<td>0.0034</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.7688</td>
<td>0.0012</td>
</tr>
<tr>
<td>$\sigma^2_{\text{S}}$</td>
<td>2.6438</td>
<td>0.0602</td>
</tr>
<tr>
<td>$\sigma^2_{\text{N}}$</td>
<td>0.2812</td>
<td>0.0101</td>
</tr>
<tr>
<td>$\sigma^2_{\text{S}}$</td>
<td>0.0718</td>
<td>0.0057</td>
</tr>
<tr>
<td>$\sigma^2_{\text{N}}$</td>
<td>2.5062</td>
<td>0.0033</td>
</tr>
<tr>
<td>$\sigma^2_{\text{S}}$</td>
<td>0.8087</td>
<td>0.0259</td>
</tr>
<tr>
<td>$\sigma^2_{\text{N}}$</td>
<td>0.4345</td>
<td>0.0011</td>
</tr>
</tbody>
</table>
Validation PMSE for the annual totals using the 8 holdout sites

- For sulfates, *IDW* PMSE is 20.4, our model PMSE is 8.1
- For nitrates, *IDW* PMSE is 3.5, our model PMSE is 1.3

Would expect improvement given the complexity of our model. However 60% is substantial and perhaps justifies the effort.
IDW vs. our model

- Validation PMSE for the annual totals using the 8 holdout sites
- For sulfates, \textit{IDW} PMSE is 20.4, our model PMSE is 8.1
- For nitrates, \textit{IDW} PMSE is 3.5, our model PMSE is 1.3
- Would expect improvement given the complexity of our model. However 60\% is substantial and perhaps justifies the effort
Validation PMSE for the annual totals using the 8 holdout sites

For sulfates, *IDW* PMSE is 20.4, our model PMSE is 8.1

For nitrates, *IDW* PMSE is 3.5, our model PMSE is 1.3

Would expect improvement given the complexity of our model. However 60% is substantial and perhaps justifies the effort.
IDW vs. our model

- Validation PMSE for the annual totals using the 8 holdout sites
- For sulfates, *IDW* PMSE is 20.4, our model PMSE is 8.1
- For nitrates, *IDW* PMSE is 3.5, our model PMSE is 1.3
- Would *expect* improvement given the complexity of our model. However 60% is substantial and perhaps justifies the effort
Figure: Model predicted map of annual sulfate deposition in 2001. The observed annual totals are labeled; a larger font size is used for the validation sites.
Figure: Model predicted map of annual nitrate deposition in 2001. The observed annual totals are labeled; a larger font size is used for the validation sites.
Map of the length of 95% prediction intervals

Figure: Uncertainty map of annual sulfate deposition.

Figure: Uncertainty map of annual nitrate deposition.
Figure: (a) Jan–Mar, (b) Apr-Jun, (c) Jul–Sep, (d) Oct-Dec.
Quarterly Nitrate Deposition

Figure: (a) Jan–Mar, (b) Apr-Jun, (c) Jul–Sep, (d) Oct-Dec.
Discussion

- Novel spatio-temporal model for fusing point and areal data which validates well.
- Inference can be provided for any spatial or temporal aggregation.
- With yearly data can study trends in deposition with regard to regulatory assessment.
- There are models in between IDW and ours but may sacrifice the features we accommodate.
- Preferable to fusion using block averaging since number of modeled grid cells much greater than number of monitoring sites, even worse if across time.
- Develop model for dry deposition, hence total deposition.