Spatio-temporal Models

- Again point-referenced vs. areal unit data
Spatio-temporal Models

- Again point-referenced vs. areal unit data
- Continuous time vs. discretized time
Spatio-temporal Models

- Again point-referenced vs. areal unit data
- Continuous time vs. discretized time
- Association in space vs. association in time!

For point-referenced data, $t$ continuous, Gaussian data,

$$Y(s, t) = \mu(s, t) + w(s, t) + \epsilon(s, t)$$
Spatio-temporal Models

- Again point-referenced vs. areal unit data
- Continuous time vs. discretized time
- Association in space vs. association in time!
  For point-referenced data, $t$ continuous, Gaussian data,
  \[
  Y(s, t) = \mu(s, t) + w(s, t) + \epsilon(s, t)
  \]
  For non-Gaussian data, instead use appropriate likelihood with link
  \[
  g(E(Y(s, t))) = \mu(s, t) + w(s, t)
  \]
Spatio-temporal Models

- Again point-referenced vs. areal unit data
- Continuous time vs. discretized time
- Association in space vs. association in time!
  For point-referenced data, \( t \) continuous, Gaussian data,

\[
Y(s, t) = \mu(s, t) + w(s, t) + \epsilon(s, t)
\]

- For non-Gaussian data, instead use appropriate likelihood with link \( g(E(Y(s, t))) = \mu(s, t) + w(s, t) \)
- Don’t treat time as a third coordinate – scale issue!

  sensible: \( \text{Cov}(Y(s, t), Y(s', t')) = C(s - s', t - t') \)

  NOT sensible: \( \text{Cov}(Y(s, t), Y(s', t')) = C((s, t) - (s', t')) \)
Spatio-temporal Models

Separable form:

\[ C(s - s', t - t') = \sigma^2 \rho_1(s - s'; \phi_1) \rho_2(t - t'; \phi_2) \]
Spatio-temporal Models

**Separable form:**

\[ C(s - s', t - t') = \sigma^2 \rho_1(s - s'; \phi_1) \rho_2(t - t'; \phi_2) \]

**Nonseparable form:**

- Sum of independent separable processes
- Mixing of separable covariance functions
- Spectral domain approaches
Now suppose time is discretized, i.e. data are $Y_t(s), t = 1, \ldots, T$
Now suppose time is discretized, i.e. data are $Y_t(s), \ t = 1, \ldots, T$

Type of data: time series versus cross-sectional (e.g., real estate sales)
Now suppose time is discretized, i.e. data are $Y_t(s), t = 1, \ldots, T$

Type of data: time series versus cross-sectional (e.g., real estate sales)

For time series data, exploratory analysis:
- Arrange into an $n \times T$ matrix $Y$ with entries $Y_t(s_i)$
- Center by row averages of $Y$ yields $Y_{rows}$
- Center by column averages of $Y$ yields $Y_{cols}$
- Sample spatial covariance matrix: $\frac{1}{T} Y_{rows} Y_{rows}^T$
- Sample autocorrelation matrix: $\frac{1}{n} Y_{cols}^T Y_{cols}$
- $E$, residuals matrix after a regression fitting
Empirical Orthogonal Functions

Can understand the structure of $Y, Y_{rows}, Y_{cols}, E$ using empirical orthogonal functions:
Empirical Orthogonal Functions

- Can understand the structure of $Y, Y_{rows}, Y_{cols}, E$ using empirical orthogonal functions:

- Say for $Y$ and $T < n$, use singular value decomposition,

$$Y = UDV^T = \sum_{j=1}^{T} d_j u_j v_j^T,$$

where $U$ is $n \times n$ orthogonal, $V$ is $T \times T$ orthogonal and $D$ is a $T \times T$ diagonal matrix augmented with $n - T$ rows of 0’s.
Empirical Orthogonal Functions

- Can understand the structure of $Y, Y_{rows}, Y_{cols}, E$ using empirical orthogonal functions:

- Say for $Y$ and $T < n$, use singular value decomposition,

$$Y = UDV^T = \sum_{j=1}^{T} d_j u_j v_j^T,$$

where $U$ is $n \times n$ orthogonal, $V$ is $T \times T$ orthogonal and $D$ is a $T \times T$ diagonal matrix augmented with $n - T$ rows of 0’s

- If we arrange the $d_j$ in decreasing order then $u_j v_j^T$ is the $j$th empirical orthogonal function
Empirical Orthogonal Functions

- Can understand the structure of $Y$, $Y_{rows}$, $Y_{cols}$, $E$ using empirical orthogonal functions:

- Say for $Y$ and $T < n$, use singular value decomposition,

$$Y = UDV^T = \sum_{j=1}^{T} d_j u_j v_j^T,$$

where $U$ is $n \times n$ orthogonal, $V$ is $T \times T$ orthogonal and $D$ is a $T \times T$ diagonal matrix augmented with $n - T$ rows of 0’s

- If we arrange the $d_j$ in decreasing order then $u_j v_j^T$ is the $j$th empirical orthogonal function

- Typically, we only need a few terms in the sum to well approximate $Y$. With just the first term it would suggest approximating $Y(s, t)$ by $d_1 u_1(s)v_1(t)$. 
Spatio-temporal Models

Modeling: \( Y_t(s) = \mu_t(s) + w_t(s) + \epsilon_t(s) \),

or perhaps \( g(E(Y_t(s))) = \mu_t(s) + w_t(s) \)
Spatio-temporal Models

Modeling: \[ Y_t(s) = \mu_t(s) + w_t(s) + \epsilon_t(s), \]
or perhaps \[ g(E(Y_t(s))) = \mu_t(s) + w_t(s) \]

For \( \epsilon_t(s) \), independent \( N(0, \tau_t^2) \)
Spatio-temporal Models

**Modeling:** $Y_t(s) = \mu_t(s) + w_t(s) + \epsilon_t(s)$, 
or perhaps $g(E(Y_t(s))) = \mu_t(s) + w_t(s)$

For $\epsilon_t(s)$, independent $N(0, \tau_t^2)$

For $w_t(s)$
- $w_t(s) = \alpha_t + w(s)$
- $w_t(s)$ independent for each $t$
- $w_t(s) = w_{t-1}(s) + \eta_t(s)$, independent spatial process innovations
Areal unit data

- $Y_i(t)$, temporal process for each unit (rare!)
- $Y_{it}$, a time series for each unit (and occasionally, $Y_{ijt}$), is more common
Areal unit data

- $Y_i(t)$, temporal process for each unit (rare!)
- $Y_{it}$, a time series for each unit (and occasionally, $Y_{ijt}$), is more common
- Again, $Y_{it} = \mu_{it} + \phi_{it} + \epsilon_{it}$, or for non-Gaussian data, $g(E(Y_{it})) = \mu_{it} + \phi_{it}$
Areal unit data

- $Y_i(t)$, temporal process for each unit (rare!)
- $Y_{it}$, a time series for each unit (and occasionally, $Y_{ijt}$), is more common

- Again, $Y_{it} = \mu_{it} + \phi_{it} + \epsilon_{it}$,
  
- or for non-Gaussian data, 
  $$g(E(Y_{it})) = \mu_{it} + \phi_{it}$$

- Again, $\epsilon_{it} \sim N(0, \tau^2_t)$
Areal unit data

- $Y_i(t)$, temporal process for each unit (rare!)
- $Y_{it}$, a time series for each unit (and occasionally, $Y_{ijt}$), is more common

Again, $Y_{it} = \mu_{it} + \phi_{it} + \epsilon_{it}$, or for non-Gaussian data, $g(\mathbb{E}(Y_{it})) = \mu_{it} + \phi_{it}$

- Again, $\epsilon_{it} \sim N(0, \tau_{it}^2)$
- Modeling for $\phi_{it}$?? CAR in space and time!
  - For space nested within time, model $\phi_{it}^{(t)} \sim CAR(\lambda_t)$, with say $\lambda_t^{iid} \sim Gamma(a, b)$
  - $\phi_{it} | \phi_{-(it)}$, space, time neighbors, weight for space, weight for time
  - MCAR, $\phi_i = (\phi_{i1}, \phi_{i2}, \ldots \phi_{iT})$, short series
Neighbors in time and space

\[ t = 1 \]
\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
\end{array}
\]

\[ t = 2 \]
\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
\end{array}
\]

\[ t = 3 \]
\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
\end{array}
\]
Traffic density and pediatric asthma

Map shows $x_{it}$, the traffic density in zip code $i$ for year $t$, San Diego County, CA ($t = 1983$ shown)
Traffic density and pediatric asthma

Map shows $x_{it}$, the traffic density in zip code $i$ for year $t$, San Diego County, CA ($t = 1983$ shown)

Also have $Y_{it}$, the number of discharges from pediatric asthma hospitalizations (children aged $\leq 14$), and
Traffic density and pediatric asthma

Map shows $x_{it}$, the traffic density in zip code $i$ for year $t$, San Diego County, CA ($t = 1983$ shown)

Also have $Y_{it}$, the number of discharges from pediatric asthma hospitalizations (children aged $\leq 14$), and

$n_{it}$, the zip-level population estimates (numbers of residents aged $\leq 14$), all for $t = 1983, \ldots, 1990$. 
Traffic density and pediatric asthma

Traffic density and pediatric asthma

- **Misalignment across years:** the zip code boundaries changed (zips added) in 1984, 1987, 1988, and 1990

- **Spatiotemporal Poisson regression model:**

\[
Y_{it} \mid \mu_{it} \overset{ind}{\sim} Po(E_{it} \exp(\mu_{it})), \ i = 1, \ldots, I_t, \ t = 1, \ldots, T,
\]

where the log-relative risk is modeled as

\[
\mu_{it} = x_{it}\beta_t + \delta_t + \theta_{it} + \phi_{it}.
\]
Traffic density and pediatric asthma

- **Misalignment across years:** the zip code boundaries changed (zips added) in 1984, 1987, 1988, and 1990

- **Spatiotemporal Poisson regression model:**

\[ Y_{it} \mid \mu_{it} \overset{ind}{\sim} Po(E_{it} \exp(\mu_{it})), \ i = 1, \ldots, I_t, \ t = 1, \ldots, T, \]

where the log-relative risk is modeled as

\[ \mu_{it} = x_{it} \beta_t + \delta_t + \theta_{it} + \phi_{it}. \]

- \( E_{it} \): expected count for zip \( i \) in year \( t \), proportional to \( n_{it} \) (internal standardization)
Traffic density and pediatric asthma

- **Misalignment across years:** the zip code boundaries changed (zips added) in 1984, 1987, 1988, and 1990

- **Spatiotemporal Poisson regression model:**

  \[ Y_{it} \mid \mu_{it} \sim \text{Po}(E_{it} \exp(\mu_{it})), \ i = 1, \ldots, I_t, \ t = 1, \ldots, T, \]

  where the log-relative risk is modeled as

  \[ \mu_{it} = x_{it}\beta_t + \delta_t + \theta_{it} + \phi_{it}. \]

- \( E_{it} \): expected count for zip \( i \) in year \( t \), proportional to \( n_{it} \) (internal standardization)

- \( \beta_t \): main effect of traffic density in year \( t \)
Traffic density and pediatric asthma

- Spatiotemporal Poisson regression model:

\[ Y_{it} \mid \mu_{it} \sim \text{Po}(E_{it} \exp(\mu_{it})), \quad i = 1, \ldots, I_t, \quad t = 1, \ldots, T, \]

where the log-relative risk is modeled as

\[ \mu_{it} = x_{it}\beta_t + \delta_t + \theta_{it} + \phi_{it}. \]

- \(E_{it}\): expected count for zip \(i\) in year \(t\), proportional to \(n_{it}\) (internal standardization)
- \(\beta_t\): main effect of traffic density in year \(t\)
- \(\delta_t\): overall intercept for year \(t\)
Traffic density and pediatric asthma

- **Misalignment across years:** the zip code boundaries changed (zips added) in 1984, 1987, 1988, and 1990

- **Spatiotemporal Poisson regression model:**

  \[
  Y_{it} \mid \mu_{it} \sim \text{Po}(E_{it} \exp(\mu_{it})), \ i = 1, \ldots, I_t, \ t = 1, \ldots, T,
  \]

  where the log-relative risk is modeled as

  \[
  \mu_{it} = x_{it} \beta_t + \delta_t + \theta_{it} + \phi_{it}.
  \]

- \(E_{it} \): expected count for zip \(i\) in year \(t\), proportional to \(n_{it}\) (internal standardization)
- \(\beta_t\): main effect of traffic density in year \(t\)
- \(\delta_t\): overall intercept for year \(t\)
- \(\theta_{it}\) and \(\phi_{it}\): zip- and year-specific heterogeneity and clustering random effects
Traffic density and pediatric asthma

Random effect distributions in the spatiotemporal case:

\[
\theta_t \overset{ind}{\sim} N \left( 0, \frac{1}{\tau_t} I \right) \quad \text{and} \quad \phi_t \overset{ind}{\sim} CAR(\lambda_t),
\]

where \( \theta_t = (\theta_1, \ldots, \theta_{I_t})' \) and \( \phi_t = (\phi_1, \ldots, \phi_{I_t})' \)
Traffic density and pediatric asthma

- Random effect distributions in the spatiotemporal case:

\[
\theta_t \sim_{ind} N \left( 0, \frac{1}{\tau_t} I \right) \quad \text{and} \quad \phi_t \sim_{ind} CAR(\lambda_t),
\]

where \( \theta_t = (\theta_1, \ldots, \theta_{I_t})' \) and \( \phi_t = (\phi_1, \ldots, \phi_{I_t})' \)

- We place flat priors on the \( \beta_t \) and \( \delta_t \) (precluding shrinkage), but encourage similarity among the random effects across years by assuming

\[
\tau_t \sim iid G(a, b) \quad \text{and} \quad \lambda_t \sim iid G(c, d)
\]
Traffic density and pediatric asthma

- Random effect distributions in the spatiotemporal case:

\[ \theta_t \overset{ind}{\sim} N \left( 0, \frac{1}{\tau_t} I \right) \quad \text{and} \quad \phi_t \overset{ind}{\sim} CAR(\lambda_t) , \]

where \( \theta_t = (\theta_1, \ldots, \theta_{I_t})' \) and \( \phi_t = (\phi_1, \ldots, \phi_{I_t})' \)

- We place flat priors on the \( \beta_t \) and \( \delta_t \) (precluding shrinkage), but encourage similarity among the random effects across years by assuming

\[ \tau_t \overset{iid}{\sim} G(a, b) \quad \text{and} \quad \lambda_t \overset{iid}{\sim} G(c, d) \]

- Changes in the zip grid over time cloud the \( \theta_{it} \) and \( \phi_{it} \) interpretation (zip \( i \) in year \( t \) may be zip \( j \) in year \( t + 1 \)), but this does not affect the main effect (\( \beta_t \) and \( \delta_t \)) interpretation; analogue of unbalanced longitudinal data
Traffic density and pediatric asthma

We set $a = 1, b = 10$ ($\tau_t$ have prior mean and sd equal to 10) and $c = 0.1, d = 10$ ($\lambda_t$ have prior mean 1, sd $\sqrt{10}$).
Traffic density and pediatric asthma

We set $a = 1, b = 10$ ($\tau_t$ have prior mean and sd equal to 10) and $c = 0.1, d = 10$ ($\lambda_t$ have prior mean 1, sd $\sqrt{10}$).

We ran 3 parallel MCMC chains for 5000 iterations each, following a 500 iteration burn-in period.
Traffic density and pediatric asthma

- We set $a = 1$, $b = 10$ ($\tau_t$ have prior mean and sd equal to 10) and $c = 0.1$, $d = 10$ ($\lambda_t$ have prior mean 1, sd $\sqrt{10}$).

- We ran 3 parallel MCMC chains for 5000 iterations each, following a 500 iteration burn-in period.

- 95% equal-tail Bayesian confidence intervals for $\beta_t$ lie above zero for all years except 1986 $\Rightarrow$ traffic exposure is positively associated with increased pediatric asthma.
Traffic density and pediatric asthma

- We set $a = 1$, $b = 10$ ($\tau_t$ have prior mean and sd equal to 10) and $c = 0.1$, $d = 10$ ($\lambda_t$ have prior mean 1, sd $\sqrt{10}$).
- We ran 3 parallel MCMC chains for 5000 iterations each, following a 500 iteration burn-in period.
- 95% equal-tail Bayesian confidence intervals for $\beta_t$ lie above zero for all years except 1986 $\Rightarrow$ traffic exposure is positively associated with increased pediatric asthma.
- For 1983, next slide provides ARCGIS/INFO maps of
  - crude asthma rate, $r_{it} = Y_{it}/n_{it}$, and
  - fitted asthma rate, $R \exp(\hat{\mu}_{it})$ where $R$ is the grand asthma rate across all zips and years and $\hat{\mu}_{it}$ is model-based posterior mean.

Note shrinkage in thinly populated eastern zips, but continued high rates in urban San Diego (SE side) due to higher sample sizes!
Large datasets

Finally, a few comments on the problem of handling large $n$ in space and large $n$ and/or $T$ in time for point referenced datasets, where handling large matrices ($n \times n$ or $nT \times nT$) is a problem:

- Joint density approximation (Vecchia, Stein)
Large datasets

Finally, a few comments on the problem of handling large $n$ in space and large $n$ and/or $T$ in time for point referenced datasets, where handling large matrices ($n \times n$ or $nT \times nT$) is a problem:

- Joint density approximation (Vecchia, Stein)
- Spectral methods (Whittle likelihood, Fuentes)
Finally, a few comments on the problem of handling large $n$ in space and large $n$ and/or $T$ in time for point referenced datasets, where handling large matrices ($n \times n$ or $nT \times nT$) is a problem:

- Joint density approximation (Vecchia, Stein)
- Spectral methods (Whittle likelihood, Fuentes)
- Lattice modelling instead or lattice approximation to spatial process models (Rue)
Large datasets

Finally, a few comments on the problem of handling large $n$ in space and large $n$ and/or $T$ in time for point referenced datasets, where handling large matrices ($n \times n$ or $nT \times nT$) is a problem:

- Joint density approximation (Vecchia, Stein)
- Spectral methods (Whittle likelihood, Fuentes)
- Lattice modelling instead or lattice approximation to spatial process models (Rue)
- Dimension reduction say through kernel convolution (Higdon)
Finally, a few comments on the problem of handling large $n$ in space and large $n$ and/or $T$ in time for point referenced datasets, where handling large matrices ($n \times n$ or $nT \times nT$) is a problem:

- Joint density approximation (Vecchia, Stein)
- Spectral methods (Whittle likelihood, Fuentes)
- Lattice modelling instead or lattice approximation to spatial process models (Rue)
- Dimension reduction say through kernel convolution (Higdon)
- Create sparsity and fast matrix multiplication (Nychka)