Bayesian inference & process convolution models
Dave Higdon, Statistical Sciences Group, LANL
MOVING AVERAGE SPATIAL MODELS
Kernel basis representation for spatial processes \( z(s) \)

Define \( m \) basis functions \( k_1(s), \ldots, k_m(s) \).

Here \( k_j(s) \) is normal density centered at spatial location \( \omega_j \):

\[
k_j(s) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}(s - \omega_j)^2 \right\}
\]

set \( z(s) = \sum_{j=1}^{m} k_j(s)x_j \) where \( x \sim N(0, I_m) \).

Can represent \( z = (z(s_1), \ldots, z(s_n))^T \) as \( z = Kx \) where

\[
K_{ij} = k_j(s_i)
\]
$x$ and $k(s)$ determine spatial processes $z(s)$

$$k_j(s)x_j$$

Continuous representation:

$$z(s) = \sum_{j=1}^{m} k_j(s)x_j$$

where $x \sim N(0, I_m)$.

Discrete representation: For $z = (z(s_1), \ldots, z(s_n))^T$, $z = Kx$ where $K_{ij} = k_j(s_i)$
Estimate $z(s)$ by specifying $k_j(s)$ and estimating $x$

$$k_j(s)x_j$$

Discrete representation: For $z = (z(s_1), \ldots, z(s_n))^T$, $z = Kx$ where $K_{ij} = k_j(s_i)$

$\Rightarrow$ standard regression model: $y = Kx + \epsilon$
Formulation for the 1-d example

Data $y = (y(s_1), \ldots, y(s_n))^T$ observed at locations $s_1, \ldots, s_n$. Once knot locations $\omega_j, j = 1, \ldots, m$ and kernel choice $k(s)$ are specified, the remaining model formulation is trivial:

Likelihood:

$$L(y|x, \lambda_y) \propto \lambda_y^n \exp \left\{ -\frac{1}{2} \lambda_y (y - Kx)^T(y - Kx) \right\}$$

where $K_{ij} = k(\omega_j - s_i)$.

Priors:

$$\pi(x|\lambda_x) \propto \lambda_x^m \exp \left\{ -\frac{1}{2} \lambda_x x^T x \right\}$$

$$\pi(\lambda_x) \propto \lambda_x^{-1} \exp \{ -b_x \lambda_x \}$$

$$\pi(\lambda_y) \propto \lambda_y^{-1} \exp \{ -b_y \lambda_y \}$$

Posterior:

$$\pi(x, \lambda_x, \lambda_y|y) \propto \lambda_y^{n/2} \exp \left\{ -\lambda_y \left[ b_y + .5(y - Kx)^T(y - Kx) \right] \right\} \times \lambda_x^{m/2} \exp \left\{ -\lambda_x \left[ b_x + .5x^T x \right] \right\}$$
Posterior and full conditionals

**Posterior:**

\[
\pi(x, \lambda_x, \lambda_y|y) \propto \lambda_y^{a_y+n-1} \exp\{-\lambda_y[by + .5(y - Kx)^T(y - Kx)]\} \times \\
\lambda_x^{a_x+m-1} \exp\{-\lambda_x[bx + .5x^T x]\}
\]

**Full conditionals:**

\[
\pi(x|\cdots) \propto \exp\{-\frac{1}{2}[\lambda_y x^T K^T K x - 2\lambda_y x^T K^T y + \lambda_x x^T x]\}
\]
\[
\pi(\lambda_x|\cdots) \propto \lambda_x^{a_x+m-1} \exp\{-\lambda_x[bx + .5x^T x]\}
\]
\[
\pi(\lambda_y|\cdots) \propto \lambda_y^{a_y+n-1} \exp\{-\lambda_y[by + .5(y - Kx)^T(y - Kx)]\}
\]

**Gibbs sampler implementation**

\[
x|\cdots \sim N((\lambda_y K^T K + \lambda_x I_m)^{-1}\lambda_y K^T y, (\lambda_y K^T K + \lambda_x I_m)^{-1})
\]
\[
\lambda_x|\cdots \sim \Gamma(a_x + \frac{m}{2}, bx + .5x^T x)
\]
\[
\lambda_y|\cdots \sim \Gamma(a_y + \frac{n}{2}, by + .5(y - Kx)^T (y - Kx))
\]
Gibbs sampler: intuition

Gibbs sampler for a bivariate normal density

$$\pi(z) = \pi(z_1, z_2) \propto \left| \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\}$$

Full conditionals of $\pi(z)$:

- $z_1 | z_2 \sim N(\rho z_2, 1 - \rho^2)$
- $z_2 | z_1 \sim N(\rho z_1, 1 - \rho^2)$

- Initialize chain with
  $$z^0 \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

- Draw $z_1^1 \sim N(\rho z_0^0, 1 - \rho^2)$
  Now $(z_1^1, z_2^0)^T \sim \pi(z)$
Gibbs sampler: intuition

Gibbs sampler gives $z^0, z^2, \ldots, z^T$ which can be treated as dependent draws from $\pi(z)$.

If $z^0$ is not a draw from $\pi(z)$, then the initial realizations will not have the correct distribution. In practice, the first 100?, 1000? realizations are discarded. The draws can be used to make inference about $\pi(z)$:

- Posterior mean of $z$ is estimated by:
  \[
  \left( \hat{\mu}_1, \hat{\mu}_2 \right) = \frac{1}{T} \sum_{k=1}^{T} \left( z^k_1, z^k_2 \right)
  \]

- Posterior probabilities:
  \[
  \hat{P}(z_1 > 1) = \frac{1}{T} \sum_{k=1}^{T} I[z^k_1 > 1]
  \]
  \[
  \hat{P}(z_1 > z_2) = \frac{1}{T} \sum_{k=1}^{T} I[z^k_1 > z^k_2]
  \]

- 90% interval: $(z_1^{[5\%]}, z_1^{[95\%])}$. 

1-d example

$m = 6$ knots evenly spaced between $- .3$ and $1.2$.

$n = 5$ data points at $s = .05, .25, .52, .65, .91$.

$k(s)$ is $N(0, \text{sd} = .3)$

$a_y = 10, b_y = 10 \cdot (.25^2)$ ⇒ strong prior at $\lambda_y = 1/.25^2$; $a_x = 1, b_x = .001$
1-d example

From posterior realizations of knot weights $x$, one can construct posterior realizations of the smooth fitted function $z(s) = \sum_{j=1}^{m} k_j(s)x_j$.

Note strong prior on $\lambda_y$ required since $n$ is small.
1-d example

\( m = 20 \) knots evenly spaced between \(-2\) and \(12\).
\( n = 18 \) data points evenly spaced between \(0\) and \(10\).
\( k(s) \) is \( N(0, \text{sd} = 2) \)
1-d example

\( m = 20 \) knots evenly spaced between \(-2\) and \(12\).
\( n = 18 \) data points evenly spaced between 0 and 10.
\( k(s) \) is \( N(0, \text{sd} = 1) \)
Basis representations for spatial processes $z(s)$

Represent $z(s)$ at spatial locations $s_1, \ldots, s_n$.

$$z = (z(s_1), \ldots, z(s_n))^T \sim N(0, \Sigma_z).$$

Recall

$$z = Kx, \text{ where } KK^T = \Sigma_z \text{ and } x \sim N(0, I_n).$$

Gives a discrete representation of $z(s)$ at locations $s_1, \ldots, s_n$.

**discrete representation**

$$z(s_i) = \sum_{j=1}^{n} K_{ij} x_j$$

Columns of $K$ give basis functions.

Can use a subset of these basis functions to reduce dimensionality.

**continuous representation**

$$z(s) = \sum_{j=1}^{n} k_j(s) x_j$$
decomposition

Cholesky (w/piv)

SVD

kernels
How many basis kernels?

Define $m$ basis functions $k_1(s), \ldots, k_m(s)$.

$m = 20$?

$m = 10$?

$m = 6$?
Moving average specifications for spatial models $z(s)$

- **Smoothing kernel** $k(s)$

- **White noise process** $x = (x_1, \ldots, x_m)$ at spatial locations $\omega_1, \ldots, \omega_m$.

  $$x \sim N(0, \frac{1}{\lambda_x} I_m)$$

- **Spatial process** $z(s)$ constructed by convolving $x$ with smoothing kernel $k(s)$

  $$z(s) = \sum_{j=1}^{m} x_j k(\omega_j - s)$$

  \[ \Rightarrow z(s) \text{ is a Gaussian process with mean 0 and covariance given by} \]

  $$\text{Cov}(z(s), z(s')) = \frac{1}{\lambda_x} \sum_{j=1}^{m} k(\omega_j - s) k(\omega_j - s')$$
USES

"Non-parametric" Covariance Function

$\begin{align*}
  k(\cdot) & \quad \leftrightarrow \quad C(d) \\
  0 & \quad \leftrightarrow \quad d
\end{align*}$

Multivariate Specifications

$Z_1 = k_{11} \ast X_1 + k_{12} \ast X_2$

$Z_2 = k_{22} \ast X_2 + k_{23} \ast X_3$

Edges

Non-stationarity
Example: constructing 1-d models for \( z(s) \)

\( m = 20 \) knot locations

\( \omega_1, \ldots, \omega_m \) equally spaced between \(-2\) and \(12\).

\( x = (x_1, \ldots, x_m)^T \sim N(0, I_m) \)

\( z(s) = \sum_{k=1}^{m} k(\omega_k - s)x_k \)

\( k(s) \) is a normal density with sd \( = \frac{1}{4}, \frac{1}{2}, \) and \( 1 \).

4th frame uses \( k(s) = 1.4 \exp\{-1.4|s|\} \).

General points:

• smooth kernels required
• spacing depends on kernel width
  – knot spacing \( \leq 1 \) sd for normal \( k(s) \)
• kernel width is equivalent to scale parameter in GP models
## Kernels and Induced Covariance Functions

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Covariance Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k(s) \propto \exp{-\frac{1}{2} |s|^2}$</td>
<td>$c(d) \propto \exp{-\frac{1}{2} |d|}$</td>
</tr>
<tr>
<td>$c(d) \propto \exp{-\frac{1}{2} |\sqrt{d}|^2}$</td>
<td></td>
</tr>
<tr>
<td>$k(s) \propto I[|s| &lt; \frac{1}{2}]$</td>
<td>$c(d) \propto (1 - \frac{3}{2} |d| + \frac{1}{2} |d|^3)I[d &lt; 1]$</td>
</tr>
<tr>
<td>$k(s) \propto \left(1 - \frac{|s|^3}{r^3}\right)^3 I[s \leq r]$</td>
<td></td>
</tr>
</tbody>
</table>
MRF formulation for the 1-d example

Data \( y = (y(s_1), \ldots, y(s_n))^T \) observed at locations \( s_1, \ldots, s_n \). Once knot locations \( \omega_j, j = 1, \ldots, m \) and kernel choice \( k(s) \) are specified, the remaining model formulation is trivial:

Likelihood:

\[
L(y|x, \lambda_y) \propto \lambda_y^n \exp \left\{ -\frac{1}{2}\lambda_y (y - Kx)^T(y - Kx) \right\}
\]

where \( K_{ij} = k(\omega_j - s_i)x_j \).

Priors:

\[
\pi(x|\lambda_x) \propto \lambda_x^m \exp \left\{ -\frac{1}{2}\lambda_x x^T W x \right\}
\]

\[
\pi(\lambda_x) \propto \lambda_x^{a_1 - 1} \exp \{ -b_x \lambda_x \}
\]

\[
\pi(\lambda_y) \propto \lambda_y^{a_2 - 1} \exp \{ -b_y \lambda_y \}
\]

Posterior:

\[
\pi(x, \lambda_x, \lambda_y|y) \propto \lambda_y^{a_2 + \frac{n}{2} - 1} \exp \left\{ -\lambda_y [b_y + 0.5(y - Kx)^T(y - Kx)] \right\} \times \\
\lambda_x^{a_1 + \frac{m}{2} - 1} \exp \left\{ -\lambda_x [b_x + 0.5x^T W x] \right\}
\]
Posterior and full conditionals (MRF formulation)

Posterior:
\[
\pi(x, \lambda_x, \lambda_y | y) \propto \lambda_y^{a_y + \frac{n}{2} - 1} \exp\{-\lambda_y [b_y + .5(y - Kx)^T(y - Kx)]\} \times \\
\lambda_x^{a_x + \frac{m}{2} - 1} \exp\{-\lambda_x [b_x + .5x^TWx]\}
\]

Full conditionals:
\[
\pi(x | \cdots) \propto \exp\{-\frac{1}{2}[\lambda_y x^TK^TKx - 2\lambda_y x^TK^Ty + \lambda_x x^TWx]\}
\]
\[
\pi(\lambda_x | \cdots) \propto \lambda_x^{a_x + \frac{m}{2} - 1} \exp\{-\lambda_x [b_x + .5x^TWx]\}
\]
\[
\pi(\lambda_y | \cdots) \propto \lambda_y^{a_y + \frac{n}{2} - 1} \exp\{-\lambda_y [b_y + .5(y - Kx)^T(y - Kx)]\}
\]

Gibbs sampler implementation
\[
x | \cdots \sim N((\lambda_y K^TK + \lambda_x W)^{-1}\lambda_y K^Ty, (\lambda_y K^TK + \lambda_x W)^{-1})
\]
\[
\lambda_x | \cdots \sim \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x)
\]
\[
\lambda_y | \cdots \sim \Gamma(a_y + \frac{n}{2}, b_y + .5(y - Kx)^T(y - Kx))
\]
1-d example - using MRF prior for $x$

post mean & 80% ci for $z(s)$

post mean for $x$'s

spatial location $s$

estimated process $z(s)$

estimated $x$ & $Kx$

(a) iid $x$'s; $k(s)$ skinny

(b) rw $x$'s; $k(s)$ fat

(c) iid $x$'s; $k(s)$ skinny

(d) gp model
8 hour max for ozone on a summer day in the Eastern US

\[ n = 510 \] ozone measurements

\( \omega_k \)'s laid out on a hexagonal lattice as shown.

\( k(s) \) is circular normal, with sd = lattice spacing.

Choice of width for \( k(s) \):

• could look at empirical variogram

• could estimate using ML or cross-validation

• could treat as additional parameter in posterior
A multiresolution spatial model formulation

\[ z(s) = z_{\text{coarse}}(s) + z_{\text{fine}}(s) \]

Coarse process:

- \( m_c = 27 \) locations \( \omega_1^c, \ldots, \omega_{m_c}^c \) on a hexagonal grid.
- \( x_c = (x_{c1}, \ldots, x_{cm_c})^T \sim N(0, \frac{1}{\lambda_c} I_{m_c}) \)
- coarse smoothing kernel \( k_c(s) \) is normal with sd = coarse grid spacing.

Fine process:

- \( m_f = 87 \) locations \( \omega_1^f, \ldots, \omega_{m_f}^f \) on a hexagonal grid.
- \( x_f = (x_{f1}, \ldots, x_{fm_f})^T \sim N(0, \frac{1}{\lambda_f} I_{m_f}) \)
- fine smoothing kernel \( k_f(s) \) is normal with sd = fine grid spacing.

note: coarse kernel width is twice the fine kernel width.
Multiresolution formulation and full conditionals

Model:

\[ y = K_c x_c + K_f x_f + \epsilon \]
\[ y = K x + \epsilon \]

where

\[ K = \begin{pmatrix} K_c & K_f \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} x_c \\ x_f \end{pmatrix} \]

Define

\[ W_x = \begin{pmatrix} \lambda_c I_{m_c} & 0 \\ 0 & \lambda_f I_{m_f} \end{pmatrix} \]

Gibbs sampler implementation then becomes

\[ x|\cdots \sim N((\lambda_y K^T K + W_x)^{-1} \lambda_y K^T y, (\lambda_y K^T K + W_x)^{-1}) \]
\[ \lambda_c|\cdots \sim \Gamma(a_x + \frac{m_c}{2}, b_x + 0.5 x_c^T x_c) \]
\[ \lambda_f|\cdots \sim \Gamma(a_x + \frac{m_f}{2}, b_x + 0.5 x_f^T x_f) \]
\[ \lambda_y|\cdots \sim \Gamma(a_y + \frac{n}{2}, b_y + 0.5(y - K x)^T (y - K x)) \]
Multiresolution model for 8 hour max ozone

coarse formulation

coarse + fine formulation

[Map showing pollution patterns]
• binary spatial process $z^*(s)$

• spatial area partitioned into two regions: $z^*(s) = 1$ and $z^*(s) = 0$.

• $n = 10$ measurements $y = (y_1, \ldots, y_n)^T$ taken at spatial locations $s_1, \ldots, s_n$.

\[ y_i = z^*(s_i) + \epsilon_i, \quad i = 1, \ldots, n; \quad \epsilon_i \overset{iid}{\sim} N(0, 1), \quad i = 1, \ldots, n \]
Constructing a binary spatial process $z^*(s)$

$$x \text{ convolved with } k(s) \text{ gives } z(s)$$

$x \sim N(0, I_m)$ where each $x_j$ is located at $\omega_j$; $z(s) = \sum_{j=1}^{m} x_j k(s - \omega_j)$

Now define the binary field $z^*(s)$ by “clipping” $z(s)$: $z^*(s) = I[z(s) > 0]$. 
Model Formulation

Define $n = 10$ observations $y = (y(s_1), \ldots, y(s_n))^T$.

Define $z^* = (z^*(s_1), \ldots, z^*(s_n))^T$.

Define $x = (x(\omega_1), \ldots, x(\omega_m))^T$ to be the $m = 25$-vector of white noise knot values at spatial grid sites $\omega_1, \ldots, \omega_m$.

Recall $z^*(s)$ and the vector $z^*$ are determined by the knot values $x$.

Likelihood

$$L(y|z^*) \propto \exp\left\{-\frac{1}{2}(y - z^*)^T(y - z^*)\right\}$$

Independent normal prior for $x$

$$\pi(x) \propto \exp\left\{-\frac{1}{2}x^Tx\right\}$$

Posterior distribution

$$\pi(x|y) \propto \exp\left\{-\frac{1}{2}(y - z^*)^T(y - z^*) - \frac{1}{2}x^Tx\right\}$$
Sampling the posterior via Metropolis

Full conditional distributions
\[ \pi(x_j|x_{-j}, y) \propto \exp\left\{-\frac{1}{2}(y - z^*)^T(y - z^*) - \frac{1}{2}x_j^2\right\}, \quad j = 1, \ldots, m \]

Metropolis implementation for sampling from \( \pi(x|y) \):
Initialization: \( x \) at \( x = 0 \).
Cycle thru full conditionals updating each \( x_j \) according to Metropolis rules.

- generate proposal \( x_j^* \sim U[x_j - r, x_j + r] \).
- compute acceptance probability
  \[ \alpha = \min \left\{ 1, \frac{\pi(x_j^*|x_{-j}, y)}{\pi(x_j|x_{-j}, y)} \right\} \]
- update \( x_j \) to new value:
  \[ x_j^{\text{new}} = \begin{cases} x_j^* & \text{with probability } \alpha \\ x_j & \text{with probability } 1 - \alpha \end{cases} \]

Here we ran for \( T = 1000 \) scans, giving realizations \( x^1, \ldots, x^T \) from the posterior. Discarded the first 100 for burn in.

Note: proposal width \( r \) tuned so that \( x_j^* \) is accepted about half the time.
Sampling from non-standard multivariate distributions

Nick Metropolis – Computing pioneer at Los Alamos National Laboratory

— inventor of the Monte Carlo method

— inventor of Markov chain Monte Carlo:


Originally implemented on the MANIAC1 computer at LANL

Algorithm constructs a Markov chain whose realizations are draws from the target (posterior) distribution.

Constructs steps that maintain detailed balance.
Gibbs Sampling and Metropolis for a bivariate normal density

\[
\pi(z_1, z_2) \propto \left| \begin{array}{cc}
1 & \rho \\
\rho & 1 \\
\end{array} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} z_1 & z_2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\
\rho & 1 \\
\end{pmatrix}^{-1} \begin{pmatrix} z_1 \\
z_2 \end{pmatrix} \right\}
\]

sampling from the full conditionals

\[
\begin{align*}
\pi(z_1 | z_2) & \sim N(\rho z_2, 1 - \rho^2) \\
\pi(z_2 | z_1) & \sim N(\rho z_1, 1 - \rho^2)
\end{align*}
\]

also called heat bath

Metropolis updating:

generate \(z_1^* \sim U[z_1 - r, z_1 + r]\)
calculate \(\alpha = \min\{1, \frac{\pi(z_1^*, z_2)}{\pi(z_1, z_2)} = \frac{\pi(z_1^* | z_2)}{\pi(z_1 | z_2)}\}\)

set \(z_1^{\text{new}} = \begin{cases} z_1^* \text{ with probability } \alpha \\
z_1 \text{ with probability } 1 - \alpha \end{cases}\)
Posterior realizations of $z^*(s) = I[z(s) > 0]$
Posterior mean of $z^*(s) = I[z(s) > 0]$
Data at $16 \times 16 - 9$ locations over 10 $m^2$ area

Assume $y(s_i)|z^*(s_i) \sim N(\mu(z^*(s_i)), 1) - \mu(0) = 1, \mu(1) = 2$.

Recall $z(s) = \sum_{j=1}^{m} x_j k(s - \omega_j)$ and $z^*(s) = I[z(s) > 0]$.

$m = 8 \times 8$ lattice of knot locations $\omega_1, \ldots, \omega_m$;

sd of $k(s)$ equal to knot spacings.
Posterior mean and realizations for $z^*(s) = I[z(s) > 0]$
Bayesian formulation

\[ L(y|\eta(z)) \propto |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y - \eta(z))^T \Sigma^{-1} (y - \eta(z)) \right\} \]

\[ \pi(z|\lambda_z) \propto \lambda_z^m \exp \left\{ -\frac{1}{2} z^T W_z z \right\} \]

\[ \pi(\lambda_z) \propto \lambda_z^{a_z-1} \exp \{ b_z \lambda_z \} \]

\[ \pi(z, \lambda_z|y) \propto L(y|\eta(z)) \times \pi(z|\lambda_z) \times \pi(\lambda_z) \]
Posterior realizations of $z$ under MRF and moving average priors

Well Data

MRF Realization

MRF Posterior Mean

GP Realization

GP Posterior Mean
References


NON-STATIONARY SPATIAL CONVOLUTION MODELS
A convolution-based approach for building non-stationary spatial models

\[ z(s) = \sum_{j=1}^{m} x_j k(\omega_j - s) = \sum_{j=1}^{m} x_j k_s(\omega_j) \]

where each \( x_j \) is located at \( \omega_j \) over a regular 2-d grid.

\( x \sim N(0, I_m) \)
Convolutions for constructing non-stationary spatial models

\[ z(s) = \sum_{j=1}^{m} x_j k_s(\omega_j) \]

\[ x \sim N(0, \lambda_x^{-1} I_m) \text{ at regular 2-d lattice locations } \omega_1, \ldots, \omega_m. \]

\[ \Rightarrow z(s) \sim GP(0, C(s_1, s_2)) \text{ where } C(s_1, s_2) = \lambda_x^{-1} \sum_{j=1}^{m} k_{s_1}(\omega_j) k_{s_2}(\omega_j) \]

smoothing kernel \( k_s(\cdot) \) changes smoothly over spatial location
Define smoothly varying kernels via basis kernels

\[ k_s(\cdot) = w_1(s)k_1(\cdot - s) + w_2(s)k_2(\cdot - s) + w_3(s)k_3(\cdot - s) \]

\( k_s(\cdot) \) is a weighted combination of kernels centered at \( s \).

Define weights that change smoothly over space.
Define iid, mean 0 Gaussian Processes $\phi_1(s)$, $\phi_2(s)$ and $\phi_3(s)$

Set

$$w_i(s) = \frac{\exp\{\phi_i(s)\}}{\exp\{\phi_1(s)\} + \exp\{\phi_2(s)\} + \exp\{\phi_3(s)\}}$$

Estimate $\phi_i(s)$’s like any other parameter in the analysis.
An application to Piazza Road Superfund site
MOVING AVERAGE/BASIS SPACE-TIME MODELS
A convolution-based approach for building space-time models

Define space-time domain $S \times T$

Define discrete knot process $x(s, t)$ on $\{(\omega_1, \tau_1), \ldots, (\omega_m, \tau_m)\}$ within $S \times T$

Define smoothing kernel $k(s, t)$

Construct space-time process $z(s, t)$

$$z(s, t) = \sum_{j=1}^{m} k((s, t) - (\omega_j, \tau_j))x(\omega_j, \tau_j) \quad \text{or with varying kernel}$$

$$z(s, t) = \sum_{j=1}^{m} k_{st}(\omega_j, \tau_j)x_j$$
A Space-time model for ocean temperatures

Data:

\[ y = (y_1, \ldots, y_n)^T \]

at space-time locations:

\[ (s_1, t_1), \ldots, (s_n, t_n) \]

Times: 1910–1988

assume data are centered (\( \bar{y} = 0 \))
Knot locations and kernels

$m$ knot locations over a space-time grid

\[(\omega_1, \tau_1), \ldots, (\omega_m, \tau_m)\]

Spatial knot locations shown;

Temporal knot spacing $\sim$ 7 years; 1900-1995;

Kernels vary with spatial location

\[k_{st}(\omega, \tau) = k_s(\omega, \tau)\]

use spatial mixture of normal “basis” kernels.
Formulation for the ocean example

Likelihood:

\[ L(y|x, \lambda_y) \propto \lambda_y^{n_y} \exp \left\{ -\frac{1}{2} \lambda_y (y - Kx)^T (y - Kx) \right\} \]

where \( K_{ij} = k_{s_i t_i}(\omega_j, \tau_j)x_j \).

Priors:

\[ \pi(x|\lambda_x) \propto \lambda_x^{m/2} \exp \left\{ -\frac{1}{2} \lambda_x x^T x \right\} \]

\[ \pi(\lambda_x) \propto \lambda_x^{a_x-1} \exp\{-b_x \lambda_x\} \]

\[ \pi(\lambda_y) \propto \lambda_y^{a_y-1} \exp\{-b_y \lambda_y\} \]

Posterior:

\[ \pi(x, \lambda_x, \lambda_y|y) \propto \lambda_y^{a_y+n_y-1} \exp \left\{ -\lambda_y [b_y + .5(y - Kx)^T (y - Kx)] \right\} \times \lambda_x^{a_x+m_x-1} \exp \left\{ -\lambda_x [b_x + .5x^T x] \right\} \]
Full conditionals for ocean formulation

Full conditionals:

\[
\pi(x|\cdots) \propto \exp\{ -\frac{1}{2}[\lambda_y x^T K^T K x - 2\lambda_y x^T K^T y + \lambda_x x^T x] \} \\
\pi(\lambda_x|\cdots) \propto \lambda_x^{a_x+m-1} \exp\{-\lambda_x[b_x + .5x^T x]\} \\
\pi(\lambda_y|\cdots) \propto \lambda_y^{a_y+n-1} \exp\{-\lambda_y[b_y + .5(y - Kx)^T(y - Kx)]\}
\]

Gibbs sampler implementation

\[
\begin{align*}
x|\cdots & \sim N((\lambda_y K^T K + \lambda_x I_m)^{-1}\lambda_y K^T y, (\lambda_y K^T K + \lambda_x I_m)^{-1}) \\
x_j|\cdots & \sim N\left( \frac{\lambda_y r_j^T k_j}{\lambda_y k_j^T k_j + \lambda_x}, \frac{1}{\lambda_y k_j^T k_j + \lambda_x} \right) \\
\lambda_x|\cdots & \sim \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x) \\
\lambda_y|\cdots & \sim \Gamma(a_y + \frac{n}{2}, b_y + .5(y - Kx)^T(y - Kx))
\end{align*}
\]

where \( k_j \) is \( j \)th column of \( K \) and \( r_j = y - \sum_{j' \neq j} k_{j'} x_{j'} \).
Posterior mean of the space-time temperature field
Deviations from time-averaged mean temperature field
Posterior probabilities of differing from time-averaged mean field
Alternative approaches for building space-time models

Define space-time domain $\mathcal{S} \times \mathcal{T}$ with $\mathcal{T} = \{1, \ldots, n_t\}$

Discrete knot process $x(s, t)$ on $\{\omega_1, \ldots, \omega_{n_s}\} \times \{1, \ldots, n_t\}$, $n_t \cdot n_s = m$

Define spatial smoothing kernel $k(s)$

Construct space-time process $z(s, t)$

$$z(s, t) = \sum_{j=1}^{n_s} k(s - \omega_j)x(\omega_j, t) \quad \text{or with varying kernel}$$

$$z(s, t) = \sum_{j=1}^{n_s} k_{st}(\omega_j)x_{jt}$$
8 hour max for ozone over summer days in the Eastern US

$n = 510$ ozone measurements each of 30 days

$\omega_k$'s laid out on same hexagonal lattice
Temporally evolving latent $x(s, t)$ process

Times: $t \in T = \{1, \ldots, n_t\}$, $n_t = 30$

Spatial knot locations: $W = \{\omega_1, \ldots, \omega_{n_s}\}$, $n_s = 27$

$x(s, t)$ defined on $W \times T$

Space-time process $z(s, t)$ obtained by convolving $x(s, t)$ with $k(s)$:

$$z(s, t) = \sum_{j=1}^{n_s} k(\omega_j - s)x(\omega_j, t) = \sum_{j=1}^{n_s} k_s(\omega_j)x_{jt}$$

Specify locally linear MRF priors for each $x_j = (x_{j1}, \ldots, x_{jn_t})^T$

$$\pi(x_j|\lambda_x) \propto \lambda_x^{n_t/2} \exp\left\{ -\frac{1}{2} \lambda_x x_j^T W x_j \right\}$$

where

$$W_{ij} = \begin{cases} -1 & \text{if } |i - j| = 1 \\ 1 & \text{if } i = j = 1 \text{ or } i = j = n_t \\ 2 & \text{if } 1 < i = j < n_t \\ 0 & \text{otherwise} \end{cases}$$

So for $x = (x_{11}, x_{21}, \ldots, x_{n_s1}, x_{12}, \ldots, x_{n_s2}, \ldots, x_{1n_t}, \ldots, x_{n_sn_t})^T$

$$\pi(x|\lambda_x) \propto \lambda^{n_t n_s/2} \exp\left\{ -\frac{1}{2} \lambda_x x^T (W \otimes I_{n_s}) x \right\}$$
Formulation for temporally evolving $z(s, t)$

Data: at each time $t$, observe $n$-vector $y_t = (y_{1t}, \ldots, y_{nt})^T$ at sites $s_1, \ldots, s_n$.

Likelihood for data observed at time $t$:

$$L(y_t|x_t, \lambda_y) \propto \lambda_y^{n_t} \exp \left\{-\frac{1}{2} \lambda_y (y_t - K^t x_t)^T(y_t - K^t x_t) \right\}$$

where $K^t_{ij} = k(\omega_j - s_i)$

Define $n \cdot n_t$-vector $y = (y_1^T, \ldots, y_{n_t}^T)^T$

Likelihood for entire data $y$:

$$L(y|x, \lambda_y) \propto \lambda_y^{nm_t} \exp \left\{-\frac{1}{2} \lambda_y (y - K x)^T(y - K x) \right\}$$

where $K = \text{diag}(K^1, \ldots, K^{n_t})$.

Priors:

$$\pi(x|\lambda_x) \propto \lambda_x^m \exp \left\{-\frac{1}{2} \lambda_x x^T W x \right\}$$

$$\pi(\lambda_x) \propto \lambda_x^{a_x-1} \exp\{-b_x \lambda_x\}$$

$$\pi(\lambda_y) \propto \lambda_y^{a_y-1} \exp\{-b_y \lambda_y\}$$
Posterior and full conditionals

\[ \pi(x, \lambda_x, \lambda_y|y) \propto \lambda_y^{a_y+\frac{m sn_t}{2}-1} \exp \{-\lambda_y[b_y + .5(y - Kx)^T(y - Kx)]\} \times \lambda_x^{a_x+\frac{n sn_t}{2}-1} \exp \{-\lambda_x[b_x + .5x^T Wx]\} \]

Full conditionals:

\[ \pi(x|\cdots) \propto \exp \{-\frac{1}{2}[\lambda_y x^T K^T K x - 2\lambda_y x^T K^T y + \lambda_x x^T W x]\} \]

\[ \pi(\lambda_x|\cdots) \propto \lambda_x^{a_x+\frac{m}{2}-1} \exp \{-\lambda_x[b_x + .5x^T Wx]\} \]

\[ \pi(\lambda_y|\cdots) \propto \lambda_y^{a_y+\frac{n}{2}-1} \exp \{-\lambda_y[b_y + .5(y - Kx)^T(y - Kx)]\} \]

Gibbs sampler implementation

\[ x|\cdots \sim N((\lambda_y K^T K + \lambda_x W)^{-1} \lambda_y K^T y, (\lambda_y K^T K + \lambda_x W)^{-1}) \]

\[ x_{jt}|\cdots \sim N \left( \frac{\lambda_y r_{tj}^T k_{tj} + n_j \bar{x}_{\partial j}}{\lambda_y k_{tj}^T k_{tj} + n_j \lambda_x}, \frac{1}{\lambda_y k_{tj}^T k_{tj} + n_j \lambda_x} \right) \]

\[ \lambda_x|\cdots \sim \Gamma(a_x + \frac{m}{2}, b_x + .5x^T x) \]

\[ \lambda_y|\cdots \sim \Gamma(a_y + \frac{n}{2}, b_y + .5(y - Kx)^T(y - Kx)) \]

where \( k_{tj} \) is \( j \)th column of \( K^t \), \( r_{tj} = y_t - \sum_{j' \neq j} k_{tj'} x_{tj'} \), \( n_j \) = number of neighbors of \( x_{jt} \), and \( x_{\partial jt} \) = mean of neighbors of \( x_{jt} \).
DLM setup for ozone example

Given latent process $x_t = (x_{1,t}, \ldots, x_{27,t})^T$, $t = 1, \ldots, 30$

$y_t = (y_{1t}, \ldots, y_{ny,t})^T$ at sites $s_{1t}, \ldots, s_{ny,t}$

\[ y_t = K^t x_t + \epsilon_t \]
\[ x_t = x_{t-1} + \nu_t \]

where $K^t$ is the $ny \times 27$ matrix given by:

\[ K^t_{ij} = k(s_{it} - \omega_j), \quad t = 1, \ldots, 30, \]
\[ \epsilon_t \overset{iid}{\sim} N(0, \sigma^2_\epsilon), \quad t = 1, \ldots, 30, \]
\[ \nu_t \overset{iid}{\sim} N(0, \sigma^2_\nu), \quad t = 1, \ldots, 30, \text{ and} \]
\[ x_1 \sim N(0, \sigma^2_x I_{27}). \]

can use dynamic linear model (DLM)/Kalman filter machinery

single site MCMC works too

See Stroud et.al. (1999) for alternative model.
Posterior mean for first 9 days
Posterior mean of selected $x_j$'s

Graphs showing latent process values $x(s,t)$ over time. The left graph demonstrates values that are independent over time, while the right graph shows values that follow a random walk over time.
References


