$p$-spline mixed models for spatio-temporal data

María Durbán

joint work with Dae-Jin Lee

DEPARTMENT OF STATISTICS
UNIVERSIDAD CARLOS III DE MADRID

June 2009
Outline

1. **P-splines**
   - Mixed models approach
   - Multidimensional $P$-splines

2. **P-splines for spatial count data**
   - Spatial smoothing
   - Smooth-CAR model
   - Application: Scottish Lip Cancer data

3. **Spatio-temporal data Smoothing with P-splines**
   - ANOVA-Type Interaction Models
   - Application: Environmental spatio-temporal data

4. **Spatio-temporal Disease Mapping**
Outline

1. **P-splines**
   Mixed models approach
   Multidimensional $P$-splines

2. **P-splines for spatial count data**
   Spatial smoothing
   Smooth-CAR model
   Application: Scottish Lip Cancer data

3. **Spatio-temporal data Smoothing with $P$-splines**
   ANOVA-Type Interaction Models
   Application: Environmental spatio-temporal data

4. **Spatio-temporal Disease Mapping**
Penalized Likelihood splines (Eilers & Marx, 1996):

- Given the data \((x_i, y_i), i = 1, ..., n\)
- Fit a sum of local basis functions:
  \[ y_i = f(x_i) + \epsilon_i, \quad \epsilon \sim N(0, \sigma^2) \]
  where \(f(x_i) = B\theta\) and
  \[ B = B(x) \text{ is a Regression Basis, and} \]
  \[ \theta \text{ is a vector of coefficients.} \]

- Control the fit through a smoothing parameter \((\lambda)\).
**B-splines Basis:**

- \( p + 1 \) Piece-wise polynomials of degree \( p \).
- Connected by knots.
- In general the choice is \( p=3 \), cubic spline.
**P-splines**

“The Flexible Smoother”

**B-splines Basis:**

- \( \hat{y} = f(x_i) = B\hat{\theta} \)

**B-splines Regression:**

\[
\min_{\theta} S(\theta; y) = \| y - B\theta \|^2 \\
\theta = (B'B)^{-1}B'y
\]

- Optimal selection of knots (**Complex**).

- **P-Splines:** add a penalty to control smoothness.

» Methodology
**P-splines**

“The Flexible Smoother”

- **B-splines Basis:**
  \[ \hat{y} = f(x_i) = B\hat{\theta} \]

- **B-splines Regression:**
  \[
  \min \ S(\theta; y) = \|y - B\theta\|^2 \\
  \hat{\theta} = (B'B)^{-1}B'y
  \]

- Optimal selection of knots *(Complex)*.

- **P-Splines:** add a penalty to control smoothness.

---

Example:
**P-splines**

“The Flexible Smoother”

**B-splines Basis:**

- \( \hat{y} = f(x_i) = B\hat{\theta} \)

**B-splines Regression:**

\[
\min_{\theta} S(\theta; y) = \|y - B\theta\|^2
\]

\( \hat{\theta} = (B'B)^{-1}B'y \)

- Optimal selection of knots (Complex).

**P-Splines:** add a penalty to control smoothness.
Methodology:

- Minimize the penalized sum of squares (PSS):

\[ S(\theta; y, \lambda)_p = \|y - B\theta\|^2 + \text{PENALTY} \]

- The PENALTY term, controls the smoothness of the fit by \( \lambda \).

  - Eilers & Marx (1996):
    \( \Rightarrow \) (discrete) Penalty over adjacent coefficients \( \theta \).

  - Lang & Brezger (2004):
    \( \Rightarrow \) “Bayesian P-splines”: random walk priors for \( \theta \), e.g.:

\[
\theta | \theta_{m-1} \sim \mathcal{N}(\theta_{m-1}, \tau^2), \text{ or } \\
\theta | \theta_{m-1}, \theta_{m-2} \sim \mathcal{N}(2\theta_{m-1} - \theta_{m-2}, \tau^2)
\]
Methodology:

- Minimize the **penalized sum of squares (PSS)**:

\[
S(\theta; y, \lambda)_p = \| y - B\theta \|^2 + \text{PENALTY}
\]

- The **PENALTY** term, controls the smoothness of the fit by \( \lambda \).

  - **Eilers & Marx (1996):**
    \( \Rightarrow \)  *(discrete)* **Penalty** over adjacent coefficients \( \theta \).

  - **Lang & Brezger (2004):**
    \( \Rightarrow \) “**Bayesian P-splines**”: random walk priors for \( \theta \), e.g.:
    \[
    \theta | \theta_{m-1} \sim \mathcal{N}(\theta_{m-1}, \tau^2), \text{ or}
    \theta | \theta_{m-1}, \theta_{m-2} \sim \mathcal{N}(2\theta_{m-1} - \theta_{m-2}, \tau^2)
    \]
Methodology:

- Minimize the penalized sum of squares (PSS):

\[ S(\theta; y, \lambda)_p = \| y - B\theta \|^2 + \text{PENALTY} \]

- The PENALTY term, controls the smoothness of the fit by \( \lambda \).

  - Eilers & Marx (1996):
    \[ \Rightarrow \text{(discrete) Penalty over adjacent coefficients } \theta. \]

  - Lang & Brezger (2004):
    \[ \Rightarrow \text{“Bayesian } P\text{-splines”}: \text{ random walk priors for } \theta, \text{ e.g.:} \]
    \[ \theta|\theta_{m-1} \sim \mathcal{N}(\theta_{m-1}, \tau^2), \text{ or} \]
    \[ \theta|\theta_{m-1}, \theta_{m-2} \sim \mathcal{N}(2\theta_{m-1} - \theta_{m-2}, \tau^2) \]
\textbf{P-splines}\\
“The Flexible Smoother”\\

- \textbf{PSS} becomes:

\[ S(\theta; y, \lambda)_p = \|y - B\theta\|^2 + \theta' P\theta \]

- \[ P = \lambda D'D. \]
- \[ \lambda \text{ is the smoothing parameter.} \]
- \[ D \text{ are difference matrices.} \]

- For given \( \lambda \), \( \min\ S(\theta; y, \lambda)_p \)

\[ \hat{\theta} = (B'B + \lambda D'D)^{-1} B'y \]

- \( \lambda \) can be selected by CV, GCV, AIC or BIC.
\textbf{P-splines} \hfill “The Flexible Smoother”

- **PSS** becomes:

\[ S(\theta; y, \lambda)_p = \|y - B\theta\|^2 + \theta' P\theta \]

\[ \Rightarrow P = \lambda D'D. \]

\[ \Rightarrow \lambda \text{ is the smoothing parameter.} \]

\[ \Rightarrow D \text{ are difference matrices.} \]

- For given \( \lambda \), \( \min S(\theta; y, \lambda)_p \)

\[ \hat{\theta} = (B'B + \lambda D'D)^{-1} B'y \]

\[ \Rightarrow \lambda \text{ can be selected by CV, GCV, AIC or BIC.} \]
- P-splines
  “The Flexible Smoother”

- **PSS** becomes:

\[ S(\theta; y, \lambda)_p = \|y - B\theta\|^2 + \theta' P\theta \]

  - \( P = \lambda D'D \).
  - \( \lambda \) is the smoothing parameter.
  - \( D \) are difference matrices.

- For given \( \lambda \), \( \min S(\theta; y, \lambda)_p \)

\[ \hat{\theta} = \left( B'B + \lambda D'D \right)^{-1} B'y \]

  - \( \lambda \) can be selected by **CV**, **GCV**, **AIC** or **BIC**.
**P-splines**

“The Flexible Smoother”

- **1d P-splines:**
  - No penalty over coefficients.
  - Penalty over coefficients.

**Example:**

- $B$-splines basis and $\theta$ without penalty
1d $P$-splines:

- No penalty over coefficients.
- Penalty over coefficients.

Example:

$B$-splines basis and $\theta$ with penalty
Advantages over other smoothers:

- **Low-Rank**: “\( \dim(B) < \dim(\text{data}) \)”.
- **Computationally efficient**: “\# knots \( \leq 40 \)”.
- **Selection of number** and **Location** of knots is **NOT** an issue.
- **Discrete Penalties** over the \( \theta \), not over the fitted curve.
- **Easy extension to**:
  - Mixed models,
  - non-gaussian data (GLM’s) and
  - Multidimensional smoothing.
  - Spatial and Spatio-temporal smoothing.
Reformulate:

- Model $y = B\theta + \epsilon$, into

$$y = X\beta + Z\alpha + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- where $X$ and $Z$ are “fixed” and “random” effects matrices.

- with coefficients $\beta$ and $\alpha \sim \mathcal{N}(0, G)$, and $G = \sigma^2 \Lambda$

- $\lambda = \frac{\sigma^2}{\sigma^2 \alpha}$
Reformulate:

- Model $y = B\theta + \epsilon$, into

$$y = X\beta + Z\alpha + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- where $X$ and $Z$ are “fixed” and “random” effects matrices.

- with coefficients $\beta$ and $\alpha \sim \mathcal{N}(0, G)$, and $G = \sigma^2_\alpha R$

- $\lambda = \frac{\sigma^2}{\sigma^2_\alpha}$

» Reparameterization
Reparameterization:

\[ B \equiv [X : Z] \Rightarrow B\theta = X\beta + Z\alpha \]

We use the Singular Value Decomposition (SVD) on \( D'D \)
**P-splines**
A mixed model approach

**Singular Value Decomposition (SVD)**

\[ D' D = U \Sigma U' \]

- with \( U = [U_n : U_s] \)

\[
\begin{align*}
D' D &= [U_n : U_s] \begin{bmatrix} 0_d & \tilde{\Sigma} \end{bmatrix} \begin{bmatrix} U_n' \\ U_s' \end{bmatrix} \\
\tilde{\Sigma} &\equiv \text{non-null eigenvalues.} \\
U_n &\equiv \text{eigenvectors corresponding to the null eigenvalues.} \\
U_s &\equiv \text{eigenvectors corresponding to the non-null eigenvalues.}
\end{align*}
\]
**P-splines**

A mixed model approach

- The **fix effects** \( \beta \) are **unpenalized** and
- The **Penalty** \( \theta' P \theta \) becomes

\[
\alpha' F \alpha
\]

where \( F = \lambda \tilde{\Sigma} \) is diagonal.

- And the **random effects** \( \alpha \) covariance matrix \( G \):

\[
G = \sigma^2 F^{-1}
\]

- Mixed Model Basis:

\[
X = [ 1 : x ] \\
Z = BU_s
\]
Advantages:

- Flexibility:
  - Easy incorporation of smoothing in complex models ("spatial" random effects and/or correlated errors).

- Mixed Models Theory:
  - Estimation and Inference.

- Software Implementation:
  - R, Splus, MATLAB or SAS.

- Extension to non-gaussian data:
  - Generalized Linear Mixed Models (GLMM)
Multidimensional $P$-splines

Example: 2d-array

- **Data** $Y = y_{ij}$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$

- **Array structure**: $n_1$ rows and $n_2$ columns

$$Y = \begin{bmatrix}
y_{11} & y_{12} & \cdots & y_{1n_2} \\
y_{21} & y_{22} & \cdots & y_{2n_2} \\
\vdots & \vdots & \ddots & \vdots \\
y_{n_11} & \cdots & \cdots & y_{n_1n_2}
\end{bmatrix}$$

- **Regressors**:

$$x_1 = (x_{11}, \ldots, x_{1n_1})'$$

$$x_2 = (x_{21}, \ldots, x_{2n_2})'$$
Multidimensional \( P \)-splines

Use of Tensor Products of \( B \)-splines (Durbán et al, 2002):

Example: 2d-array

- **Marginal Basis:**
  - \( B_1 = B_1(x_1) \), of dim. \( n_1 \times c_1 \).
  - \( B_2 = B_2(x_2) \), of dim. \( n_2 \times c_2 \).

- **2d \( B \)-splines Basis:**
  - Kronecker Product (\( \otimes \)) of marginal basis:
    \[
    B = B_2 \otimes B_1, \quad \text{of dim.} \quad n_1 n_2 \times c_1 c_2
    \]
Multidimensional $p$-splines

Use of Tensor Products of $B$-splines (Durbán et al, 2002):

Example: 2d-array

- **Marginal Basis:**
  - $B_1 = B_1(x_1)$, of dim. $n_1 \times c_1$.
  - $B_2 = B_2(x_2)$, of dim. $n_2 \times c_2$.

- **2d $B$-splines Basis:**
  - Kronecker Product ($\otimes$) of marginal basis:
    
    $$B = B_2 \otimes B_1,$$
    
    of dim. $n_1 n_2 \times c_1 c_2$
Multidimensional $P$-splines

Model:

$$y = f(x_1, x_2) + \epsilon,$$

with $y_{n_1n_2 \times 1}$

- In matrix form, $\hat{y} = B\theta$ can be written as:

  $$\hat{Y} = B_1AB_2, \text{ of dim } n_1 \times n_2$$

  where $A$ is a matrix $c_1 \times c_2$ of coefficients $\theta$ of length $c_1c_2 \times 1$.

IDEA:

- Set penalties over $\Theta$.

  - Row-wise Penalty: $\theta' \left(I_{c_2} \otimes D_1'D_1\right) \theta$

  - Column-wise Penalty: $\theta' \left(D_2'D_2 \otimes I_{c_1}\right) \theta$
Multidimensional $P$-splines

Model:

$$y = f(x_1, x_2) + \epsilon,$$

with $y_{n_1 n_2 \times 1}$

- In matrix form, $\hat{y} = B\theta$ can be written as:

$$\hat{Y} = B_1 A B_2,$$

of dim $n_1 \times n_2$

where $A$ is a matrix $c_1 \times c_2$ of coefficients $\theta$ of length $c_1 c_2 \times 1$.

IDEA:

- Set penalties over $\Theta$.

- **Row-wise** Penalty:

$$\theta' \left( I_{c_2} \otimes D_1' D_1 \right) \theta$$

- **Column-wise** Penalty:

$$\theta' \left( D_2' D_2 \otimes I_{c_1} \right) \theta$$
Multidimensional $P$-splines

- **Penalty Matrix in $2d$:**

\[
P = \lambda_1 I_{c_2} \otimes D_1' D_1 + \lambda_2 D_2' D_2 \otimes I_{c_1}
\]

- $\lambda_1$ and $\lambda_2$ are the smoothing parameters in each dimension.
- **Anisotropy:** ($\lambda_1 \neq \lambda_2$)
As in 1d Case:

Example:

The Mixed Model consists of:

\[ \hat{y} = X\hat{\beta} + Z\hat{\alpha} \]

(Linear/Fixed) + (Non-Linear/Random)
Multidimensional $P$-splines

**Mixed Models Representation:**

- As in 1d case, the aim is:

  \[ B \equiv [X : Z] \implies B\theta = X\beta + Z\alpha \]

- The SVD over $P$ allows the simultaneous diagonalization of $D_1^\prime D_1$ and $D_2^\prime D_2$.

- The penalty $P$ becomes $F$ (block diagonal matrix):

  \[
  F = \begin{pmatrix}
  \lambda_2 \tilde{\Sigma}_2 \otimes I_2 \\
  \lambda_1 I_2 \otimes \tilde{\Sigma}_1 \\
  \lambda_1 I_{c_2-2} \otimes \tilde{\Sigma}_1 + \lambda_2 \tilde{\Sigma}_2 \otimes I_{c_1-2}
  \end{pmatrix}
  \]
Multidimensional $P$-splines

ANOVA-type Decomposition of Smooth Surfaces:

$$\hat{y} = f(x_1) + f(x_2) + f(x_1, x_2)$$

(additive term for $x_1$)  
(additive term for $x_2$)  
(interaction term for $x_1, x_2$)

» Advantages
Multidimensional $P$-splines

Advantages:

- Extension to $d$-dimensions:
  \[ B = B_2 \otimes B_1 \otimes \cdots \otimes B_d \]

- Efficient algorithms:

- Anisotropy (different smoothing for each dimension):

- Complex models: spatial data smoothing
Outline

1. $P$-splines
   - Mixed models approach
   - Multidimensional $P$-splines

2. $P$-splines for spatial count data
   - Spatial smoothing
   - Smooth-CAR model
   - Application: Scottish Lip Cancer data

3. Spatio-temporal data Smoothing with $P$-splines
   - ANOVA-Type Interaction Models
   - Application: Environmental spatio-temporal data

4. Spatio-temporal Disease Mapping
We propose:

- **2d P-splines:**
  - **Geostatistics:** at sampling locations.
  - **Regional/areal:** at the centroids.

Models of the form:

\[ y = f(\text{lon}, \text{lat}) + \epsilon \]

where

- \( f(\text{lon}, \text{lat}) \) is a large-scale spatial smooth trend: \( X\beta + Z\alpha \).
- The mixed model allows the simultaneous estimation of smoothing and spatial correlation.

Spatial count data
We propose:

- 2d $P$-splines:

- Geostatistics: at sampling locations.

- Regional/areal: at the centroids.

Models of the form:

$$y = f(\text{lon}, \text{lat}) + \epsilon$$

where

- $f(\text{lon}, \text{lat})$ is a large-scale spatial smooth trend: $X\beta + Z\alpha$.

- The mixed model allows the simultaneous estimation of smoothing and spatial correlation.
We propose:

• 2d P-splines:
• Geostatistics: at sampling locations.
• Regional/areal: at the centroids.

Models of the form:

\[ y = f(\text{lon}, \text{lat}) + \epsilon \]

where

• \( f(\text{lon}, \text{lat}) \) is a large-scale spatial smooth trend: \( X\beta + Z\alpha \).
• The mixed model allows the simultaneous estimation of smoothing and spatial correlation.
We propose:

- 2d $P$-splines:
- Geostatistics: at sampling locations.
- Regional/areal: at the centroids.

Models of the form:

$$y = f(lon, lat) + \epsilon$$

where

- $f(lon, lat)$ is a large-scale spatial smooth trend: $X\beta + Z\alpha$.
- The mixed model allows the simultaneous estimation of smoothing and spatial correlation.
\( B \)-spline Basis for spatial data:

- Given that data are **NOT** in an array

\[
B = B_2 \otimes B_1 \text{ replace by } B_2 \square B_1
\]

\( \square \) denotes the “Row-wise Kronecker” or **Box-Product**.

\[
B_2 \square B_1 = (B_2 \otimes 1_{c_1}) \odot (1_{c_2} \otimes B_1)
\]

\( \odot \) is the “element-wise” product.
P-splines for spatial count data

In many applications:

- Collect count data **observed in regions or areas**.
  - **E.g.**: # of cases of disease or deaths
- Counts are **Poisson** distributed.

\[ y \sim \mathcal{P}(\mu) \]

**Penalized-GLMM**

- **P-splines as mixed models**:
  - **Linear Predictor**:
    \[ \eta = B\theta \implies X\beta + Z\alpha \]
  - **Penalized log-Likelihood**:
    \[ \ell_p(\beta, \alpha; y) = \ell(\beta, \alpha; y) - \frac{1}{2}\alpha'F\alpha \]

- **Estimation via PQL**
In many applications:

- Collect count data observed in regions or areas.
  - E.g.: # of cases of disease or deaths
- Counts are Poisson distributed.

\[ y \sim \mathcal{P}(\mu) \]

**Penalized-GLMM**

- *P*-splines as mixed models:
  - Linear Predictor:
    \[ \eta = B\theta \implies X\beta + Z\alpha \]
  - Penalized log-Likelihood:
    \[ \ell_p(\beta, \alpha; y) = \ell(\beta, \alpha; y) - \frac{1}{2} \alpha' F \alpha \]
  - Estimation via PQL
Most popular approach:

- Conditional Autoregressive Models (CAR), Besag (1991)
  - Spatial Dependence across "neighbours"
  - Different neighbourhood criteria.
    - Common border.
    - Centroids distance, 4-nearest neighbours.
Smooth-CAR model

Most popular approach:

- Conditional Autoregressive Models (CAR), Besag (1991)
- Spatial Dependence across “neighbours”.
- Different neighbourhood criteria.
  - Common border.
  - Centroids distance, 4-nearest neighbours.
Smooth-CAR model

CAR model

- Most popular approach:
  - Conditional Autoregressive Models (CAR), Besag (1991)
  - Spatial Dependence across “neighbours”.
  - Different neighbourhood criteria.
    - Common border.
    - Centroids distance, 4-nearest neighbours.
Smooth-CAR model

CAR model

- Most popular approach:
  - Conditional Autoregressive Models (CAR), Besag (1991)
  - Spatial Dependence across “neighbours”.
  - Different neighbourhood criteria.
    - Common border.
    - Centroids distance, 4-nearest neighbours.
Smooth-CAR model

Formulation:

\[ y = X\beta + b, \]

where \( b = (b_1, b_2, ..., b_n)' \) is a vector for the spatial effects.

- Impose a spatial dependency structure by a prior distribution for \( b \):
  \[ b \sim \mathcal{N}(0, G_b) \]

where \( G_b \) depends on the “neighbourhood structure”:
- defined by Contiguity matrix (\( Q \))
Smooth-CAR model

CAR model

Formulation:

\[ y = X\beta + b, \]

where \( b = (b_1, b_2, ..., b_n)' \) is a vector for the spatial effects.

- Impose a spatial dependency structure by a prior distribution for \( b \):

  \[ b \sim \mathcal{N}(0, G_b) \]

where \( G_b \) depends on the “neighbourhood structure”:

- Defined by Contiguity matrix \( Q \)
✓ We follow an **Empirical Bayes** approach:

**Intrinsic CAR:**

\[ G_b = \sigma_b^2 Q^- + \kappa^{-1}I \]  
(Besag, 1991)

- Two independent and separate variance components:
  - **Spatially-structured variation:** \( \sigma_b^2 Q^- \)
  - **Unstructured non-spatial correlation:** \( \kappa^{-1}I \)

**Alternative CAR models structures:**

\[ G_b = \sigma_b^2 (\phi Q^- + (1 - \phi)I)^{-1} \]  
(Leroux et al, 1999)

\[ G_b = \sigma_b^2 (\phi Q^- + (1 - \phi)I) \]  
(Dean et al, 2001)

where

- \( \phi \) measures the relative weight between *structured* and *unstructured* variability
- \( 0 \leq \phi \leq 1 \)
✓ We follow an **Empirical Bayes** approach:

**Intrinsic CAR:**

\[ G_b = \sigma_b^2 Q^+ + \kappa^{-1} I \]  

(Besag, 1991)

- Two independent and separate variance components:
  - **Spatially-structured variation:** \( \sigma_b^2 Q^+ \)
  - **Unstructured non-spatial correlation:** \( \kappa^{-1} I \)

**Alternative CAR models structures:**

\[ G_b = \sigma_b^2 (\phi Q^+ + (1 - \phi)I)^{-1} \]  

(Leroux et al, 1999)

\[ G_b = \sigma_b^2 (\phi Q^- + (1 - \phi)I) \]  

(Dean et al, 2001)

where

- \( \phi \) measures the relative weight between *structured* and *unstructured* variability
- \( 0 \leq \phi \leq 1 \)
We propose a “hybrid” model:

- **Spatial P-spline** with **CAR** structure: “Smooth-CAR” model
- **Model:**
  \[ \eta = X\beta + Z\alpha + b , \]
  where \( b \sim \mathcal{N}(0, G_b) \)

**Our approach:**

\[ \eta = \text{Spatial Trend} + \text{Local area-level spatial correlation} \]

\[ X\beta + Z\alpha \quad \text{(Large-scale)} \]

\[ \text{Spatial Random Effects} \quad \text{(Small-scale)} \]
We propose a “hybrid” model:

- **Spatial** $P$-spline with **CAR** structure: “Smooth-CAR” model
- Model:
  \[ \eta = X\beta + Z\alpha + b , \]
  where \( b \sim \mathcal{N}(0, G_b) \)

**Our approach:**

\[ \eta = \underbrace{X\beta + Z\alpha}_{\text{Spatial Trend (Large-scale)}} + \underbrace{\text{Local area-level spatial correlation}}_{\text{Spatial Random Effects (Small-scale)}} \]
**Smooth-CAR model**

**Summary:**

<table>
<thead>
<tr>
<th>Model</th>
<th>Linear Predictor</th>
<th>Area-level var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>$X\beta + Z\alpha$</td>
<td>$-$</td>
</tr>
<tr>
<td>CAR</td>
<td>$X\beta + b$</td>
<td>$b \sim \mathcal{N}(0, G_b)$</td>
</tr>
<tr>
<td>Smooth-CAR</td>
<td>$X\beta + Z\alpha + b$</td>
<td>$b \sim \mathcal{N}(0, G_b)$</td>
</tr>
</tbody>
</table>

**The Smooth-CAR:**

- Allow us model the **spatial trend** ($X\beta + Z\alpha$) along large geographical distances and
- **Local area-level** correlation by a **CAR** component ($b$).
Application: Scottish Lip Cancer data

Example: Scottish Lip Cancer

• Breslow and Clayton (1993)
• Observed \((y)\) and Expected \((e)\)
cases of lip cancer
• 56 counties in Scotland
We fit several models:

- **Smooth $P$-spline models:**

  \[ \eta = \log(e) + X\beta + Z\alpha \]  
  (Poisson)

  \( \log(e) \) is the **offset** term.

- **CAR models:**

  \[ \eta = \log(e) + X\beta + b, \quad b \sim \mathcal{N}(0, G_b), \]  
  (Dean)

  with:

  \[ G_b = \sigma_b^2 (\phi Q^- + (1 - \phi)I) \]

- **Smooth-CAR model:**

  \[ \eta = \log(e) + X\beta + Z\alpha + b, \quad b \sim \mathcal{N}(0, G_b) \]
We fit several models:

- **Smooth P-splines** models:
  
  \[ \eta = \log(e) + X\beta + Z\alpha \]  
  
  \( \log(e) \) is the **offset** term.

- **CAR** models:
  
  \[ \eta = \log(e) + X\beta + b, \quad b \sim \mathcal{N}(0, G_b), \]  
  
  with:

  \[ G_b = \sigma_b^2 (\phi Q^- + (1 - \phi)I) \]  

- **Smooth-CAR** model:
  
  \[ \eta = \log(e) + X\beta + Z\alpha + b, \quad b \sim \mathcal{N}(0, G_b) \]
Application: Scottish Lip Cancer data

Fitted Models

We fit several models:

- **Smooth \( P \)-splines models:**
  \[
  \eta = \log(e) + X\beta + Z\alpha
  \]
  \( (\text{Poisson}) \)

  \( \log(e) \) is the **offset** term.

- **CAR models:**
  \[
  \eta = \log(e) + X\beta + b, \quad b \sim \mathcal{N}(0, G_b),
  \]

  with:
  \[
  G_b = \sigma_b^2 \left( \phi Q^- + (1 - \phi)I \right)
  \]
  \( (\text{Dean}) \)

- **Smooth-CAR model:**
  \[
  \eta = \log(e) + X\beta + Z\alpha + b, \quad b \sim \mathcal{N}(0, G_b)
  \]
In order to compare the proposed models we use:

\[ \text{AIC} = \text{Dev} + 2 \times \text{df} \]
\[ \text{BIC} = \text{Dev} + \log(n) \times \text{df} \]

where:
- \( \text{df} \) is the effective dimension of the model ("degrees of freedom").
- is a measure of the complexity of the fitted model,
- Calculated as the trace(\( H \)).

\[ \hat{y} = Hy \]
In order to compare the proposed models we use:

\[
\text{AIC} = \text{Dev} + 2 \times \text{df}
\]

\[
\text{BIC} = \text{Dev} + \log(n) \times \text{df}
\]

where:

- \( \text{df} \) is the effective dimension of the model (“degrees of freedom”).
- \( \text{df} \) is a measure of the complexity of the fitted model,
- Calculated as the \( \text{trace}(H) \),

\[
\hat{y} = Hy
\]
Application: Scottish Lip Cancer data

Comparisons of fitted models

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>AIC</th>
<th>BIC</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smooth: Poisson</td>
<td>λ_1 = 11.75</td>
<td>λ_2 = 3.63</td>
<td>σ^2_s = -</td>
<td>κ^-1 = -</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>114.04</td>
</tr>
<tr>
<td>CAR: Dean</td>
<td>-</td>
<td>-</td>
<td>σ^2_s = 0.78</td>
<td>κ^-1 = 0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.36</td>
</tr>
<tr>
<td>Smooth-CAR: Dean</td>
<td>λ_1 = 30.11</td>
<td>λ_2 = 16.37</td>
<td>σ^2_s = 0.53</td>
<td>κ^-1 = -</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>87.46</td>
</tr>
</tbody>
</table>

Observations:

- \( \phi \approx 1 \)  
  Overdispersion is due to “structured” spatial correlation (\( \sigma^2_s Q^- \)).
- Smooth-CAR performs better in terms of the selected criteria.
Dean’s CAR model:

(a) Large-scale linear trend: $X\beta$

(b) CAR structured random effects: $b \sim \mathcal{N}(0, G_b)$

(c) $X\beta + b$
Smooth-CAR model:

(a) Smooth Trend
(b) CAR component
(c) Trend + CAR

(a) Smooth large-scale spatial trend: $X\beta + Z\alpha$
(b) CAR structured random effects: $b \sim \mathcal{N}(0, G_b)$
(c) $X\beta + Z\alpha + b$
Outline

1. **P-splines**
   Mixed models approach
   Multidimensional \( P \)-splines

2. **\( P \)-splines for spatial count data**
   Spatial smoothing
   Smooth-CAR model
   Application: Scottish Lip Cancer data

3. **Spatio-temporal data Smoothing with \( P \)-splines**
   ANOVA-Type Interaction Models
   Application: Environmental spatio-temporal data

4. **Spatio-temporal Disease Mapping**
Spatio-temporal data

- Response variable, $y_{ijt}$
  - measured over geographical locations, $s = (x_i, x_j)$, with $i, j = 1, \ldots, n$
  - and over time periods, $x_t$, for $t = 1, \ldots, T$

- ISSUE: huge amount of data available
  - e.g.: Environmental data, epidemiologic studies, disease mapping applications, ...

- Smoothing techniques:
  - Study spatial and temporal trends.
  - Space and time interactions.
  - ✓ 3-dimensional smoothing: $P$-splines and GLAM.
Example of GLAM in $3d$
Currie et. al (2006)

- **3d-case:**
  \[ f(x_1, x_2, x_3) = B\theta \]

- **Basis:** $B = B_1 \otimes B_2 \otimes B_3$

  - $\theta$ can be expressed as a $3d$-array $A = \{\theta\}_{ijk}$ of dim. $c_1 \times c_2 \times c_3$
• **3d-Penalty matrix:**

  • Set penalties over the 3d-array $A$:

    $$ P = \lambda_1 D_1' D_1 \otimes I_{c_2} \otimes I_{c_3} + \lambda_2 I_{c_1} \otimes D_2' D_2 \otimes I_{c_3} + \lambda_t I_{c_1} \otimes I_{c_2} \otimes D_t' D_t $$

    - row-wise
    - column-wise
    - layer-wise

• For spatio-temporal data:

  $$ f( longitude, \text{latitude}, \text{time} ) $$

  - **Spatial anisotropy** ($\lambda_1 \neq \lambda_2$), different amount of smoothing for latitude and longitude.
  - **Temporal smoothing** ($\lambda_t$)
  - Space-time interaction.
For spatio-temporal data, we propose:

\[ B = B_s \otimes B_t, \]

where

\[ B_s \equiv \text{is the spatial } B\text{-spline basis } (B_1 \square B_2) \] and
\[ B_t \equiv \text{is the } B\text{-spline basis for time of dim. } t \times c_3. \]

✓ as GLAM:

Given \( y_{ijt} = Y_{t \times n} \), and \( \theta_{ijt} = A_{ct \times c_s} \), we have

\[ \mathbb{E}[Y] = B_t AB_s' \]

✓ as Mixed models

\[ B\theta = X\beta + Z\alpha \]
Spatio-temporal data Smoothing with P-splines

ANOVA-Type Interaction Models

Smooth-ANOVA decomposition models

- Chen (1993), Gu (2002):
  - “Smoothing-Spline ANOVA” (SS-ANOVA).
  - Interpretation as “main effects” and “interactions”.
  - Models of type:
    \[
    \hat{y} = f(x_1) + f(x_2) + f(x_t) \\
    + f(x_1, x_2) + f(x_1, x_t) + f(x_2, x_t) \\
    + f(x_1, x_2, x_t)
    \]
    “Main/additive effects”
    “2-way interactions”
    “3-way interactions”

- PROBLEMS:
  - identifiability, and
  - basis dimension (“curse of dimensionality”)
Lee and Durbán (2009a), consider:

\[ y = \gamma + f_s(x_1, x_2) + f_s(time) + f_{st}(x_1, x_2, time) + \epsilon, \]

where

- \( f_s(x_1, x_2) \) ≡ Spatial 2d smooth surface
- \( f_t(time) \) ≡ Smooth time trend
- \( f_{st}(x_1, x_2, time) \) ≡ Space-time interaction

We need to construct an identifiable model.

Our approach is based on:

- low-rank basis (\( P \)-splines)
- the mixed model representation and SVD properties.
Basis, Coefficients and Penalty

- For each smooth term $f(\cdot)$, in spatio-temporal ANOVA model we have
  - **$B-$spline basis:**
    \[ B = [1_{nt} : B_s \otimes 1_{t} : 1_{n} \otimes B_t : B_s \otimes B_t] \]
  - **vector of coefficients:**
    \[ \theta = (\gamma, \theta^{(s)}, \theta^{(t)}, \theta^{(st)})' \]
  - and a blockdiagonal **Penalty:**
    \[ P = \begin{pmatrix} 0 & P_s & P_t & P_{st} \\ P_s & P_t & P_{st} & 0 \\ P_t & P_{st} & P_{st} & 0 \\ P_{st} & 0 & 0 & 0 \end{pmatrix} \]
    where
    - $P_s = 2d$-spatial penalty
    - $P_t = 1d$-penalty for time
    - $P_{st} = 3d$ space-time penalty

Uc3m/ Dept. of Statistics
Basis, Coefficients and Penalty

- For each smooth term $f(\cdot)$, in spatio-temporal ANOVA model we have
  - $B$—spline basis:
    $$B = [1_{nt} : B_s \otimes 1_t : 1_n \otimes B_t : B_s \otimes B_t]$$
  - vector of coefficients:
    $$\theta = (\gamma, \theta^{(s)}', \theta^{(t)}', \theta^{(st)}')$$
  - and a blockdiagonal Penalty:
    $$P = \begin{pmatrix} 0 & P_s & P_t & P_{st} \\ P_s & 0 & & \\ P_t & & 0 & \\ P_{st} & & & 0 \end{pmatrix}$$
    where
    - $P_s = 2d$-spatial penalty
    - $P_t = 1d$-penalty for time
    - $P_{st} = 3d$ space-time penalty
Basis, Coefficients and Penalty

- For each smooth term \( f(\cdot) \), in spatio-temporal ANOVA model we have
  - \( B \)-spline basis:
    \[ B = [1_{nt} : B_s \otimes 1_t : 1_n \otimes B_t : B_s \otimes B_t] \]
  - vector of coefficients:
    \[ \theta = (\gamma, \theta^{(s)'}, \theta^{(t)'}, \theta^{(st)'})' \]
- and a blockdiagonal penalty:
  \[ P = \begin{pmatrix} P_{st} & 0 \\ P_{t} & P_s \end{pmatrix} \]
  where
  \[ P_s = 2d\text{-spatial penalty} \]
  \[ P_t = 1d\text{-penalty for time} \]
  \[ P_{st} = 3d\text{ space-time penalty} \]
Basis, Coefficients and Penalty

- For each smooth term \( f(\cdot) \), in spatio-temporal ANOVA model we have
  - \( B \)-spline basis:
    \[
    B = [1_{nt} : B_s \otimes 1_t : 1_n \otimes B_t : B_s \otimes B_t]
    \]
  - vector of coefficients:
    \[
    \theta = (\gamma, \theta^{(s)}', \theta^{(t)}', \theta^{(st)}')'
    \]
  - and a blockdiagonal Penalty:
    \[
    P = \begin{pmatrix}
    0 & P_s \\
    P_s & P_t \\
    P_t & P_{st}
    \end{pmatrix},
    \]
    where
    \[
    P_s = 2d\text{-spatial penalty}
    
    P_t = 1d\text{-penalty for time}
    
    P_{st} = 3d\text{ space-time penalty}
    \]
✓ However, $B$ is **NOT full column-rank** (“linear dependency”)

✓ Model is **NOT identifiable**

**Solution:**

- Reparameterize as a mixed model (using SVD).
- For each term we have:

  Basis $[X : Z]$

\[
\begin{align*}
  f_s(x_1, x_2) &\equiv x_1 : x_2 & (1) \\
  f_t(x_t) &\equiv x_t & (2) \\
  f_{st}(x_1, x_2, x_t) &\equiv x_1 : x_2 : x_t & (3)
\end{align*}
\]

- Some terms in (1) and (2) also appear in (3).
The **mixed model representation**, allow us to **identify the columns to remove** in order to maintain the identifiability of the model.

and obtain a blockdiagonal penalty $F$

$$F = \begin{pmatrix} 0 & F_s & F_t \\ F_s & F_t & F_{st} \end{pmatrix},$$

with $\lambda_1, \lambda_2, \lambda_t, \tau_1, \tau_2, \tau_t$

**In $P$-splines context**, this is equivalent to

✓ apply constraints over **regression coefficients** $\theta_{i,j,k}$
For the **ANOVA spatio-temporal model**, the resultant mixed model reparameterization is equivalent to apply the next constraints:

- **time effect coefficient:**
  \[
  \sum_{t=1}^{c_t} \theta_t^{(t)} = 0,
  \]

- **constraints over the spatio-temporal array of coefficients, \( \Theta^{(st)} \), of dimensions \( c_t \times c_s \):**
  \[
  \sum_{i}^{c_1} \theta_{t,ij}^{(st)} = \sum_{j}^{c_2} \theta_{t,ij}^{(st)} = \sum_{i}^{c_1} \sum_{j}^{c_2} \theta_{t,ij}^{(st)} = 0.
  \]
In practice

We only need to construct the matrices $X$, $Z$ and penalty $F$

\[
\begin{array}{cccc}
  f_s(x_1, x_2) & f_t(x_t) & f_{st}(x_1, x_2, x_t) \\
\end{array}
\]

\[
X \equiv \text{by columns} \quad x_1 : x_2 \quad x_t \quad (x_1, x_2, x_t)
\]

\[
Z \equiv \text{by blocks} \quad '' \quad '' \quad ''
\]

\[
F \equiv \text{blockdiagonal} \quad F_s \quad F_t \quad F_{st} \\
(\lambda_1, \lambda_2) \quad \lambda_t \quad (\tau_1, \tau_2, \tau_t)
\]
Ozone pollution in Europe
Lee and Durbán (2009a)

- Sample of 45 monitoring stations
- Monthly averages of $O_3$ levels (in $\mu g/m^3$ units)
- from January 1999 to December 2005 ($t = 1, \ldots, 84$)

Models:
- **Additive:**
  \[ f_s(x_1, x_2) + f_t(x_t) \]
- **Spatio-temporal Interaction:**
  ✓ **ANOVA:**
  \[ f_s(x_1, x_2) + f_t(x_t) + f_{st}(x_1, x_2, x_t) \]
Spatio-temporal data smoothing with $P$-splines

Application: Environmental spatio-temporal data

Spatial 2d + time

\[ f_s(x_1, x_2) + f_t(x_t) \]

✓ Space-time interaction is not considered
✓ Time smooth trend is additive
Spatio-temporal ANOVA model

\[ \hat{y} = f(\text{space}) + f(\text{time}) + 1999 : 1 + f(\text{space},\text{time}) \]
Comparison of fitted values
Additive VS ANOVA

✓ **Additive model** assumes a spatial smooth surface over all monitoring stations that remains constant over time.

✓ **ANOVA model** captures individual characteristics of the stations throughout time.
## Comparison of Models

**ANOVA and Additive**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>14280.73</td>
<td>366.03</td>
</tr>
<tr>
<td>Additive</td>
<td>16506.28</td>
<td>65.98</td>
</tr>
</tbody>
</table>

► **Observations:**

- Best overall performance of ANOVA in terms of AIC
- **ANOVA model** is more realistic than Additive, and easier to decompose and interpret in terms of the fit.
Outline

1. \( P \)-splines
   - Mixed models approach
   - Multidimensional \( P \)-splines

2. \( P \)-splines for spatial count data
   - Spatial smoothing
   - Smooth-CAR model
   - Application: Scottish Lip Cancer data

3. Spatio-temporal data Smoothing with \( P \)-splines
   - ANOVA-Type Interaction Models
   - Application: Environmental spatio-temporal data

4. Spatio-temporal Disease Mapping
**P-spline ANOVA model for disease mapping**

- \( Y \) and \( E \) are \( t \times n \) arrays of observed and expected cases of disease over \( t \) time periods, and \( M = \log \left( \frac{Y}{E} \right) \).

- Consider an ANOVA model for \( \eta \)

\[
 f_s(x_1, x_2) + f_t(x_t) + f_{st}(x_1, x_2, x_t)
\]
New flexible approach for spatial and spatio-temporal data smoothing:

- based on $P$-splines as mixed models and
- ANOVA decomposition

Methodology also extensible for disease mapping applications.

Computationally efficient algorithms (GLAM)
*Smooth-CAR mixed models for spatial count data.*  
Computational Statistics and Data Analysis 53(8):2968-2979.

*P-spline ANOVA-Type interaction models for spatio-temporal smoothing.*  
Submitted.

*Fast and compact smoothing on large multidimensional grids.*  
Computational Statistics and Data Analysis, 50(1):61-76.

*Generalized linear array models with applications to multidimensional smoothing.*  