Additive Mixed Models for Correlated Functional Data

Fabian Scheipl

Institut für Statistik
Ludwig-Maximilians-Universität München

SuSTaIn Workshop
"High Dimensional and Dependent Functional Data"
September 10, 2012

Joint work with C. Crainiceanu, S. Greven, A. Ivanescu, P. Reiss, A.-M. Staicu

Flexible, multiple regression models with functional predictors for **correlated** functional responses:

$$Y_{ij}(t) = \alpha(t) + \eta(t; \mathbf{x}_{ij}) + \epsilon_{ij}(t)$$

Flexible, multiple regression models with functional predictors for **correlated** functional responses:

$$Y_{ij}(t) = \alpha(t) + \eta(t; \mathbf{x}_{ij}) + \epsilon_{ij}(t)$$

white noise errors: $\epsilon_{ij}(t) \overset{iid}{\sim} N(0, \sigma^2)$

Flexible, multiple regression models with functional predictors for **correlated** functional responses:

$$Y_{ij}(t) = \alpha(t) + \eta(t; \mathbf{x}_{ij}) + \epsilon_{ij}(t)$$

white noise errors: $\epsilon_{ij}(t) \overset{iid}{\sim} N(0, \sigma^2)$

additive predictor $\eta(t; \mathbf{x}_{ij})$ is the sum of:

- (nonlinear, multivariate, index-varying) effects of scalar covariates:
  - $x\beta$, $x\beta(t)$, $f(x)$, $f(x, t)$, $f(x_1, x_2, t)$

Flexible, multiple regression models with functional predictors for **correlated** functional responses:

$$Y_{ij}(t) = \alpha(t) + \eta(t; \mathbf{x}_{ij}) + \epsilon_{ij}(t)$$

white noise errors: $\epsilon_{ij}(t) \overset{iid}{\sim} N(0, \sigma^2)$

additive predictor $\eta(t; \mathbf{x}_{ij})$ is the sum of:

- (nonlinear, multivariate, index-varying) effects of scalar covariates:
  - $x\beta$, $x\beta(t)$, $f(x)$, $f(x, t)$, $f(x_1, x_2, t)$
- effects of functional covariates $X(s)$:
  - $\int X(s)\beta(s, t)ds$
  - for $X_i(s) \approx \sum_{m=1}^{M} \phi_m(s)\xi_{m,i}$: $\sum_{m=1}^{M} \xi_m \beta_m(t)$, $\sum_{m=1}^{M} f_m(\xi_m, t)$

Flexible, multiple regression models with functional predictors for **correlated** functional responses:

$$Y_{ij}(t) = \alpha(t) + \eta(t; \mathbf{x}_{ij}) + \epsilon_{ij}(t)$$

white noise errors: $\epsilon_{ij}(t) \overset{iid}{\sim} N(0, \sigma^2)$

additive predictor $\eta(t; \mathbf{x}_{ij})$ is the sum of:

- (nonlinear, multivariate, index-varying) effects of scalar covariates:
  - $x\beta$, $x\beta(t)$, $f(x)$, $f(x,t)$, $f(x_1, x_2, t)$
- effects of functional covariates $X(s)$:
  - $\int X(s)\beta(s,t)ds$
  - for $X_i(s) \approx \sum_{m=1}^{M} \phi_m(s)\xi_{m,i}$: $\sum_{m=1}^{M} \xi_m \beta_m(t)$, $\sum_{m=1}^{M} f_m(\xi_m, t)$
- scalar or functional random effects for correlated data:
  - b_{0i}, $b_{1i}u_{ij}$, $B_{0i}(t)$, $B_{1i}(t)u_{ij}$

This is just a kind of varying coefficient model...

... effects vary over index of response:

- write as **model for concatenated function values** vec $\left(\underset{n \times T}{\mathbf{Y}} \right)$
- **reformat covariate data** accordingly & add index t as covariate

$\Rightarrow$ can be fitted like **standard AMM for scalar responses** using penalized splines

- effects vary smoothly over $t \Rightarrow$ smoothness of $E(Y(t))$
- sparse/irregular $Y(t)$ possible

Model representation:

- $$\eta(t;\mathbf{x}) = \sum_p f(t, x_p)$$

  effects are weighted sums of basis functions $B(t, x)$ in index of response and covariates:

  $$f(t, x_p) = \sum_d B_d(t, x_p)\theta_d = \mathbf{B}_p\boldsymbol{\theta}_p$$

Model representation:

- $$\eta(t;\mathbf{x}) = \sum_p f(t, x_p)$$

  effects are weighted sums of basis functions $B(t,x)$ in index of response and covariates:

  $$f(t,x_p) = \sum_d B_d(t,x_p)\theta_d = \mathbf{B}_p\boldsymbol{\theta}_p$$

- each associated with penalty term $\text{pen}\left(\boldsymbol{\theta}_p|\boldsymbol{\lambda}_p\right)$ for regularization: enforce smoothness, similarity of correlated units, etc.
- $\text{pen}\left(\boldsymbol{\theta}|\boldsymbol{\lambda}\right) = \boldsymbol{\theta}^T\mathbf{P}(\boldsymbol{\lambda})\boldsymbol{\theta}$ equivalent to distributional assumption:

  $$\boldsymbol{\theta} \sim N\left(\mathbf{0}, \mathbf{P}(\boldsymbol{\lambda})^{-1}\right) \Rightarrow f(t,x) \sim GP(\mathbf{0}, \mathbf{BP}(\boldsymbol{\lambda})^{-1}\mathbf{B}^T)$$

Inference is mostly a solved problem:

- use **penalized splines for smooth effects**, including linear functional effects and functional random effects:
- **standard penalized likelihood inference** with (G)CV or (RE)ML via mixed model representation to solve

$$\min_{\boldsymbol{\theta},\boldsymbol{\lambda}} \left(\left\| \text{vec}(\mathbf{Y}) - \sum_p \mathbf{B}_p \boldsymbol{\theta}_p \right\|_2^2 + \sum_p \text{pen}\left(\boldsymbol{\theta}_p | \boldsymbol{\lambda}_p\right) \right)$$

  - approximate CIs, tests, diagnostics, etc. immediately available

$\Rightarrow$ **refund**'s `pffr()` as wrapper for **mgcv**:

  - optimized, robust, well-tested algorithms
  - versatile library of spline bases ready to use
  - model effects with any given correlation structure via GMRFs

Tensor product basis representation of effects:

$$\underset{nT\times 1}{f(\mathbf{t},\text{covariate})} \approx \left(\underset{n\times K_c}{\mathbf{B}_c} \otimes \underset{T\times K_t}{\mathbf{B}_t} \right) \underset{K_c K_t \times 1}{\boldsymbol{\theta}}$$

$$\text{pen}(\boldsymbol{\theta}|\lambda_t, \lambda_c) = \boldsymbol{\theta}^T \left(\lambda_t \mathbf{I}_{K_c} \otimes \mathbf{P}_t + \lambda_c \mathbf{P}_c \otimes \mathbf{I}_{K_t}\right) \boldsymbol{\theta}$$

- $\mathbf{B}_c$: marginal basis for covariate
- $\mathbf{B}_t$: marginal basis on response's index
- $\mathbf{P}_t, \mathbf{P}_c$: marginal penalty matrices
- $\mathbf{P} \otimes \mathbf{I}$, $\mathbf{I} \otimes \mathbf{P}$: repeated penalties that apply to each subvector of $\boldsymbol{\theta}$ associated with a specific marginal basis function (Wood, 2006)
- construction valid for all types of effects we consider: $\int X(s)\beta(s,t)ds$, $x\beta(t)$, $f(x,t)$, etc.

Tensor product basis representation of effects:

$$\underset{nT\times 1}{f(\mathbf{t}, \text{covariate})} \approx \left(\underset{n\times K_c}{\mathbf{B}_c} \otimes \underset{T\times K_t}{\mathbf{B}_t} \right) \underset{K_c K_t \times 1}{\boldsymbol{\theta}}$$

$$\text{pen}(\boldsymbol{\theta}|\lambda_t, \lambda_c) = \boldsymbol{\theta}^T \left(\lambda_t \mathbf{I}_{K_c} \otimes \mathbf{P}_t + \lambda_c \mathbf{P}_c \otimes \mathbf{I}_{K_t} \right) \boldsymbol{\theta}$$

- e.g. scalar random intercepts $\mathbf{b}_0 \sim N(\mathbf{0}, \frac{1}{\lambda_c}\mathbf{S})$:

$$\mathbf{B}_c = \underset{n\times \# units}{\begin{pmatrix} & 1 & & & \\ & & & 1 & \\ & & 1 & & \\ 1 & & & & \\ & & & & 1 \end{pmatrix}} \text{(unit level dummies)}; \qquad \mathbf{B}_t = \mathbf{1}$$

$$\mathbf{P}_c = \mathbf{S}^{-1}; \qquad \mathbf{P}_t = 0$$

Functional random effects with between-unit correlation:

$$\underset{nT\times 1}{\mathbf{B}_0(\mathbf{t})} = \left(\underset{n\times \#units}{\mathbf{B}_c} \otimes \underset{T\times K_t}{\mathbf{B}_t} \right) \underset{\#units\cdot K_t\times 1}{\boldsymbol{\theta}}$$

$$\text{pen}(\boldsymbol{\theta}|\lambda_t, \lambda_c) = \boldsymbol{\theta}^T \left(\lambda_t \mathbf{P}_t \otimes \mathbf{I}_{\#units} + \lambda_c \mathbf{I}_{K_t} \otimes \mathbf{S}^{-1}\right) \boldsymbol{\theta}$$

- $\mathbf{B}_t$: spline basis over t with smoothness penalty $\mathbf{P}_t$
- $\mathbf{B}_c$: matrix of unit level indicators
- $\mathbf{S}$: correlation of subvectors of coefficients associated with the same marginal basis function in $\mathbf{B}_t$
  $\Rightarrow$ induces similarity of $B_{0i}(t), B_{0j}(t)$ if $\mathbf{S}_{ij}$ is large.
- model effects with any given marginal correlation $\mathbf{S}$, for example:
  - $\mathbf{S} = \mathbf{I}$ for independent random effects
  - $\mathbf{S}^{-1}$: precision matrix of a GMRF, e.g. via adjacency matrix of a graph
  - $\mathbf{S}_{ij} = \rho^{|i-j|}$ for AR(1)-structure

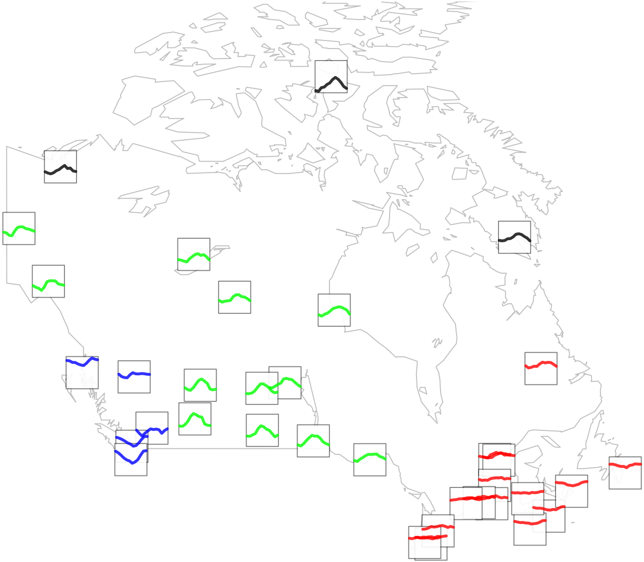
Functional random effects with between-unit correlation:

$$\underset{nT\times 1}{\mathbf{B}_0(\mathbf{t})} = \left(\underset{n\times\#units}{\mathbf{B}_c} \otimes \underset{T\times K_t}{\mathbf{B}_t} \right) \underset{\#units\cdot K_t\times 1}{\boldsymbol{\theta}}$$

$$\text{pen}(\boldsymbol{\theta}|\lambda_t, \lambda_c) = \boldsymbol{\theta}^T \left(\lambda_t \mathbf{P}_t \otimes \mathbf{I}_{\#units} + \lambda_c \mathbf{I}_{K_t} \otimes \mathbf{S}^{-1}\right) \boldsymbol{\theta}$$

- $\Rightarrow$ **fit correlated functional random effects with arbitrary, fixed inter-unit correlation structure**
- $\Rightarrow$ fit correlated smooth residual functions with arbitrary, fixed inter-observation correlation structure

Example: Canadian Weather Data



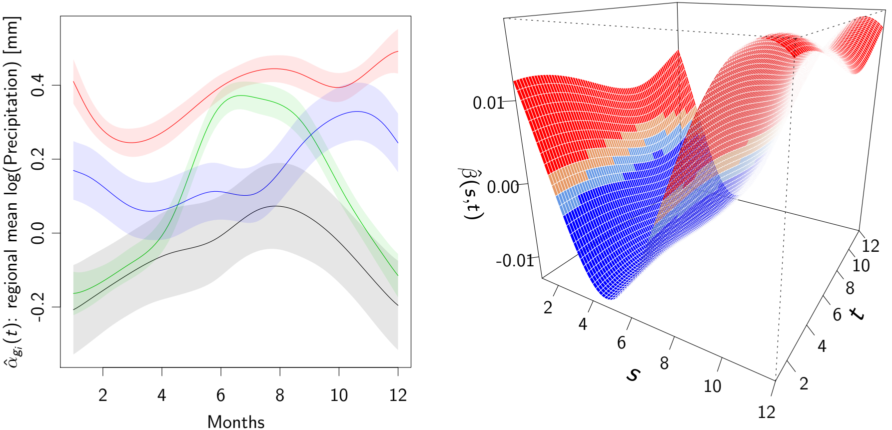
Canadian Weather Data: Matèrn-correlated smooth residuals

$$\log_{10}(\texttt{precipitation}_i(t)) =$$

$$\alpha_{\texttt{region}(i)}(t) + \int \texttt{temperature}_i(s)\beta(t,s)ds + E_i(t) + \epsilon_i(t),$$

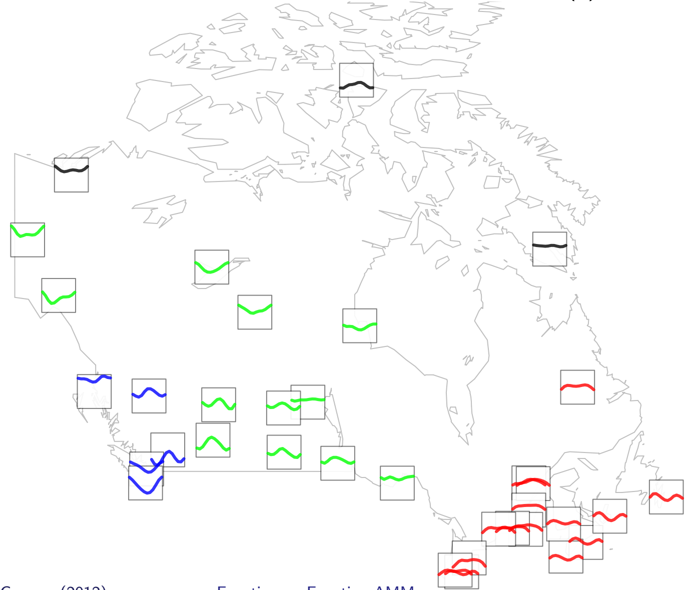
- region-specific functional intercepts $\alpha_{\texttt{region}}(t)$
- linear function-on-function effect of temperature profiles
- spatially correlated smooth residuals $E(t)$ with Matèrn-correlation function over station locations to capture deviations from regional, temperature-adjusted means

Canadian Weather Data: Effect Estimates Model



Canadian Weather Data: Effect Estimates

Smooth, correlated residual functions $\hat{E}_i(t)$



Simulation Study Results

- estimation accuracy robust against small number of observations
- estimation accuracy for $E(Y(t))$ good and very stable:
  - relative IMSE stays $< 3\%$ even for $SNR = 1$, 30 observations in 10 groups.
- computation time increases quickly, but still doable for large data
  - 30 observations in 10 groups, 30 gridpoints, two functional random effects: $\approx$ 5sec
  - 2000 observations in 100 groups, 100 gridpoints, two functional random effects: $\approx$ 12h
- REML much more stable than GCV
- not feasible (yet): smooth residuals for large data sets

Open questions & further directions:

- identifiability conditions for function-on-function terms: compare FPC- and spline-based approaches
  (Müller and Yao, 2008; Ivanescu et al., 2012; Scheipl et al., 2012)
- more efficient algorithms:
  - array models (Currie et al., 2006)
  - exploit sparse designs & penalties
- non-i.i.d. errors:
  - GLS-type approaches: $\epsilon_{ij}(t) \overset{iid}{\sim} \mathrm{GP}(0, K(t, t'))$ with known/fixed covariance $K(t, t')$ (Reiss et al., 2010)
  - bootstrap-based inference
- Bayes inference: better uncertainty measures, variable selection, robustification (Morris and Carroll, 2006; Zhu et al., 2011; Goldsmith et al., 2011)
- non-Gaussian data: binary data, count data, hazard rate models

Summary

In terms of **flexible, multiple** regression models for **correlated** functional responses and predictors, **refund**'s `pffr()` can already do (almost) everything that GAMMs can do for scalar responses.

But: Implementation more advanced than theoretical results and empirical evaluation

- framework pushes limits of underlying inference algorithms
  $\Rightarrow$ validate with more case studies
- do asymptotics & robustness results carry over from scalar case?

Find the papers & supplementary material with code examples on my homepage:

`http://www.statistik.lmu.de/~scheipl/research.html`

References

H. Cardot, F. Ferraty, and P. Sarda. Functional linear model. *Statistics and Probability Letters*, 45:11–22, 1999.

I.D. Currie, M. Durban, and P.H.C. Eilers. Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):259–280, 2006.

J. Goldsmith, M.P. Wand, and C. Crainiceanu. Functional regression via variational bayes. *Electronic Journal of Statistics*, 5: 572, 2011.

G. He, H.G. Müller, and J.L. Wang. Extending correlation and regression from multivariate to functional data. In M.L. Puri, editor, *Asymptotics in Statistics and Probability*, pages 301 – 315. VSP International Science Publishers, 2003.

A.E. Ivanescu, A.-M. Staicu, S. Greven, F. Scheipl, and C.M. Crainiceanu. **Penalized function-on-function regression**. 2012. URL http://biostats.bepress.com/jhubiostat/paper240/. *submitted*.

J.S. Morris and R.J. Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199, 2006.

H.G. Müller and F. Yao. Functional additive models. *Journal of the American Statistical Association*, 103(484):1534–1544, 2008.

P. Reiss, L. Huang, and M. Mennes. Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics*, 6(1):28, 2010.

F. Scheipl and S. Greven. Identifiability in penalized function-on-function regression models. Technical Report 125, LMU München, 2012. URL http://epub.ub.uni-muenchen.de/13060/.

F. Scheipl, A.-M. Staicu, and S. Greven. **Additive Mixed Models for Correlated Functional Data**. 2012. URL http://arxiv.org/abs/1207.5947. *submitted*.

S.N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, 2006.

H. Zhu, P.J. Brown, and J.S. Morris. Robust, adaptive functional regression in functional mixed model framework. *Journal of the American Statistical Association*, 106(495):1167–1179, 2011.

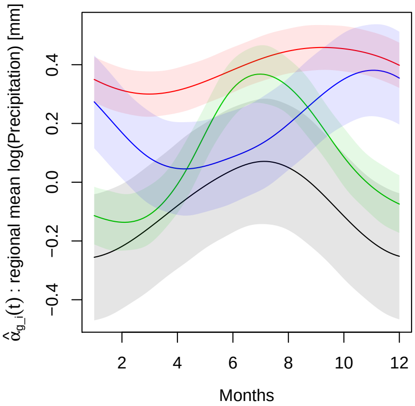
$\mathbf{P} \otimes \mathbf{I}$, $\mathbf{I} \otimes \mathbf{P}$: repeated penalties that apply to each subvector of $\boldsymbol{\theta}$ associated with a specific marginal basis function (Wood, 2006):

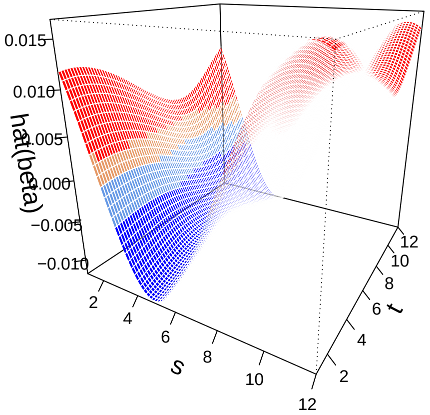
$$
\mathbf{B}_c \otimes \mathbf{B}_t = \begin{pmatrix} \perp_1 & \nabla_1 & \ltimes_1 \\ \perp_2 & \nabla_2 & \ltimes_2 \\ \perp_3 & \nabla_3 & \ltimes_3 \\ \perp_4 & \nabla_4 & \ltimes_4 \end{pmatrix} \otimes \begin{pmatrix} \maltese_1 & \bigstar_1 \\ \maltese_2 & \bigstar_2 \\ \maltese_3 & \bigstar_3 \end{pmatrix} = \begin{pmatrix}
\perp_1 \cdot \maltese_1 & \perp_1 \cdot \bigstar_1 & \nabla_1 \cdot \maltese_1 & \nabla_1 \cdot \bigstar_1 & \ltimes_1 \cdot \maltese_1 & \ltimes_1 \cdot \bigstar_1 \\
\perp_1 \cdot \maltese_2 & \perp_1 \cdot \bigstar_2 & \nabla_1 \cdot \maltese_2 & \nabla_1 \cdot \bigstar_2 & \ltimes_1 \cdot \maltese_2 & \ltimes_1 \cdot \bigstar_2 \\
\perp_1 \cdot \maltese_3 & \perp_1 \cdot \bigstar_3 & \nabla_1 \cdot \maltese_3 & \nabla_1 \cdot \bigstar_3 & \ltimes_1 \cdot \maltese_3 & \ltimes_1 \cdot \bigstar_3 \\
\perp_2 \cdot \maltese_1 & \perp_2 \cdot \bigstar_1 & \nabla_2 \cdot \maltese_1 & \nabla_2 \cdot \bigstar_1 & \ltimes_2 \cdot \maltese_1 & \ltimes_2 \cdot \bigstar_1 \\
\perp_2 \cdot \maltese_2 & \perp_2 \cdot \bigstar_2 & \nabla_2 \cdot \maltese_2 & \nabla_2 \cdot \bigstar_2 & \ltimes_2 \cdot \maltese_2 & \ltimes_2 \cdot \bigstar_2 \\
\perp_2 \cdot \maltese_3 & \perp_2 \cdot \bigstar_3 & \nabla_2 \cdot \maltese_3 & \nabla_2 \cdot \bigstar_3 & \ltimes_2 \cdot \maltese_3 & \ltimes_2 \cdot \bigstar_3 \\
\perp_3 \cdot \maltese_1 & \perp_3 \cdot \bigstar_1 & \nabla_3 \cdot \maltese_1 & \nabla_3 \cdot \bigstar_1 & \ltimes_3 \cdot \maltese_1 & \ltimes_3 \cdot \bigstar_1 \\
\perp_3 \cdot \maltese_2 & \perp_3 \cdot \bigstar_2 & \nabla_3 \cdot \maltese_2 & \nabla_3 \cdot \bigstar_2 & \ltimes_3 \cdot \maltese_2 & \ltimes_3 \cdot \bigstar_2 \\
\perp_3 \cdot \maltese_3 & \perp_3 \cdot \bigstar_3 & \nabla_3 \cdot \maltese_3 & \nabla_3 \cdot \bigstar_3 & \ltimes_3 \cdot \maltese_3 & \ltimes_3 \cdot \bigstar_3 \\
\perp_4 \cdot \maltese_1 & \perp_4 \cdot \bigstar_1 & \nabla_4 \cdot \maltese_1 & \nabla_4 \cdot \bigstar_1 & \ltimes_4 \cdot \maltese_1 & \ltimes_4 \cdot \bigstar_1 \\
\perp_4 \cdot \maltese_2 & \perp_4 \cdot \bigstar_2 & \nabla_4 \cdot \maltese_2 & \nabla_4 \cdot \bigstar_2 & \ltimes_4 \cdot \maltese_2 & \ltimes_4 \cdot \bigstar_2 \\
\perp_4 \cdot \maltese_3 & \perp_4 \cdot \bigstar_3 & \nabla_4 \cdot \maltese_3 & \nabla_4 \cdot \bigstar_3 & \ltimes_4 \cdot \maltese_3 & \ltimes_4 \cdot \bigstar_3
\end{pmatrix}
$$

$$
\mathbf{I}_{K_c} \otimes \mathbf{P}_t = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \otimes \begin{pmatrix} P_{\maltese} & P_{\bigstar\maltese} \\ P_{\bigstar\maltese} & P_{\bigstar} \end{pmatrix} = \begin{pmatrix}
P_{\maltese} & P_{\bigstar\maltese} & & & & \\
P_{\bigstar\maltese} & P_{\bigstar} & & & & \\
& & P_{\maltese} & P_{\bigstar\maltese} & & \\
& & P_{\bigstar\maltese} & P_{\bigstar} & & \\
& & & & P_{\maltese} & P_{\bigstar\maltese} \\
& & & & P_{\bigstar\maltese} & P_{\bigstar}
\end{pmatrix}
$$

$$
\mathbf{P}_c \otimes \mathbf{I}_{K_t} = \begin{pmatrix} P_{\perp} & P_{\nabla\perp} & P_{\ltimes\perp} \\ P_{\nabla\perp} & P_{\nabla} & P_{\nabla\ltimes} \\ P_{\ltimes\perp} & P_{\ltimes\nabla} & P_{\ltimes} \end{pmatrix} \otimes \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} = \begin{pmatrix}
P_{\perp} & & P_{\nabla\perp} & & P_{\ltimes\perp} & \\
& P_{\perp} & & P_{\nabla\perp} & & P_{\ltimes\perp} \\
P_{\nabla\perp} & & P_{\nabla} & & P_{\nabla\ltimes} & \\
& P_{\nabla\perp} & & P_{\nabla} & & P_{\nabla\ltimes} \\
P_{\ltimes\perp} & & P_{\ltimes\nabla} & & P_{\ltimes} & \\
& P_{\ltimes\perp} & & P_{\ltimes\nabla} & & P_{\ltimes}
\end{pmatrix}
$$

Canadian Weather Data: Effect Estimates without Smooth Residuals





Canadian Weather Data: Spatio-temporal smooth effect + uncorrelated scalar random effects

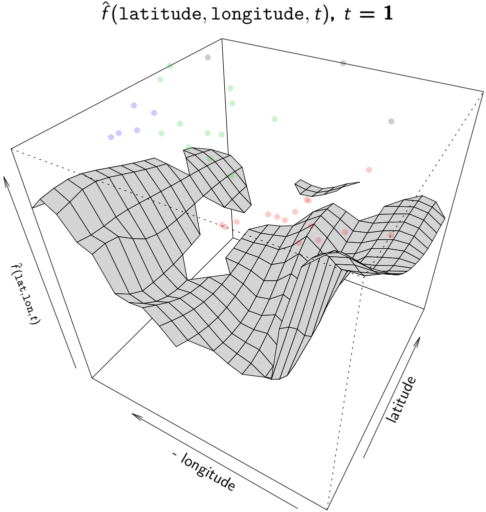
$$\log_{10}(\texttt{precipitation}_i(t)) = \alpha(t) + b_{0i} + \int \texttt{temperature}_i(s)\beta(t,s)ds + f(\texttt{latitude}_i, \texttt{longitude}_i, t) + \epsilon_i(t),$$

- uncorrelated scalar random effects $\mathbf{b}_0$ to capture small-scale spatial variation in precipitation levels
- linear function-on-function effect of temperature profiles
- smooth spatio-temporal effect $f(\texttt{latitude}, \texttt{longitude}, t)$ to capture large-scale spatial variation over time

R-Code:

```
m2 <- pffr(l10precip ~  c(s(place, bs="re")) +
             ff(temp, yind=month.t, xind=month.s, splinepars=list(bs=c("cc", "cc")) +
             s(lat,lon),
          bs.int = list(bs="cc", k=10), bs.yindex = list(bs="cc"),  data=dataM, yind = month.t)
```

Canadian Weather Data: Effect Estimates



Flexibility - Overfitting - Concurvity

very flexible models $\rightarrow$ danger of overfitting
$\rightarrow$ overparameterization? identifiability?

functional concurvity $\rightarrow$ useful analogies to scalar concurvity?

no good metrics/diagnostics for these issues yet

Identifiability for function-on-function terms

Functional Analysis:
"Coefficient surfaces $\beta(s,t)$ are not identifiable outside the span of the eigenfunctions of the cross-covariance of $Y(t)$ and $X(s)$."

- **FPCA approach:**
  "Restrict $\hat{\beta}(s,t)$ to span(eigenfunctions)."
- **penalized splines approach:**
  "Not actually a problem for most finite-sample data, get smoothest $\hat{\beta}(s,t)$ that yields a good fit. Most problematic settings can be identified and circumvented." (Scheipl and Greven, 2012)

But:
Can we really interpret $\hat{\beta}(s,t)$ meaningfully?

Identifiability of function-on-function terms

In theory, coefficient surfaces $\beta(s,t)$ are not identifiable except in the span of the covariate's eigenfunctions — for $X_i(s) = \sum_{m=1}^{M} \phi_m(s)\xi_{m,i}$:

$$\int X_i(s)\beta(s,t)ds = \sum_{m=1}^{M} \xi_{im} \int \phi_m(s)\beta(s,t)ds$$
$$= \sum_{m=1}^{M} \xi_{im} \int \phi_m(s)(\beta(s,t) + o(s))ds$$

$$\forall\ o(s) \notin \text{span}(\{\phi_m(s) : m = 1, \dots, M\}) \text{ as } \int \phi_m(s)o(s)ds = 0.$$

$\Rightarrow$ adding any $o(s)$ from kernel of X's covariance does not change fit (Cardot et al., 1999; He et al., 2003)

Identifiability of function-on-function terms

In theory, coefficient surfaces $\beta(s, t)$ are not identifiable except in the span of the covariate's eigenfunctions — for $X_i(s) = \sum_{m=1}^{M} \phi_m(s)\xi_{m,i}$:

$$\int X_i(s)\beta(s,t)ds = \sum_{m=1}^{M} \xi_{im} \int \phi_m(s)\beta(s,t)ds$$
$$= \sum_{m=1}^{M} \xi_{im} \int \phi_m(s)(\beta(s,t) + o(s))ds$$

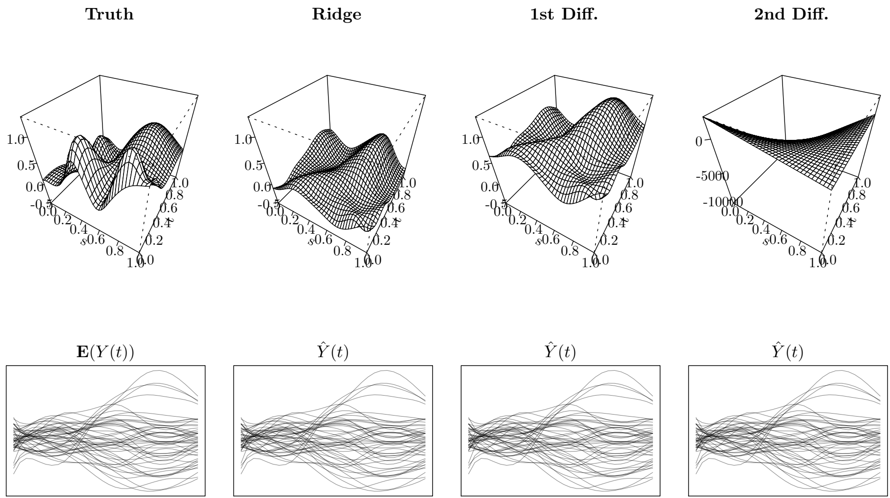
$$\forall\ o(s) \notin \text{span}(\{\phi_m(s) : m = 1, \dots, M\}) \text{ as } \int \phi_m(s)o(s)ds = 0.$$

⇒ adding any $o(s)$ from kernel of X's covariance does not change fit (Cardot et al., 1999; He et al., 2003)

- in practice: if
  - $K_s > M$ and
  - kernel of $X(s)$-covariance overlaps penalty nullspace

  ⇒ adding $o(s)$ does not change penalty term either ⇒ **not identifiable** (Scheipl and Greven, 2012)

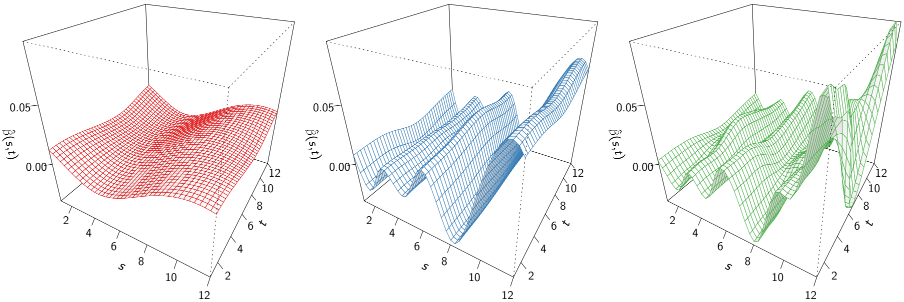
Identifiability: Example I



Identifiability: Example II

$\hat{\beta}(s, t)$ for Canadian Weather:

$(K_s, K_t) = (3, 3)$ $(K_s, K_t) = (8, 8)$ $(K_s, K_t) = (11, 11)$



Fitted values & remaining term estimates are (almost) exactly the same!