M7777 Applied Functional Data Analysis
10. Functional Response with Functional Covariate

Jan Koláček (kolacek@math.muni.cz)

Dept. of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno

Functional Response Models

1. The Concurrent Model

Let us recall model (1) with scalar covariates

$$y_i(t) = \beta_0(t) + \sum_{j=1}^{K} \beta_j(t) z_{ij} + \varepsilon_i(t). \tag{1}$$

We can extend (1) to allow for functional covariates as follows

$$y_i(t) = \beta_0(t) + \sum_{j=1}^{K} \beta_j(t) z_{ij}(t) + \varepsilon_i(t). \tag{2}$$

Model (2) is called **concurrent** because it only relates the value of $y_i(t)$ to the value of $z_{ij}(t)$ at the same time points t.

Functional Response Models

Example Canadian Weather

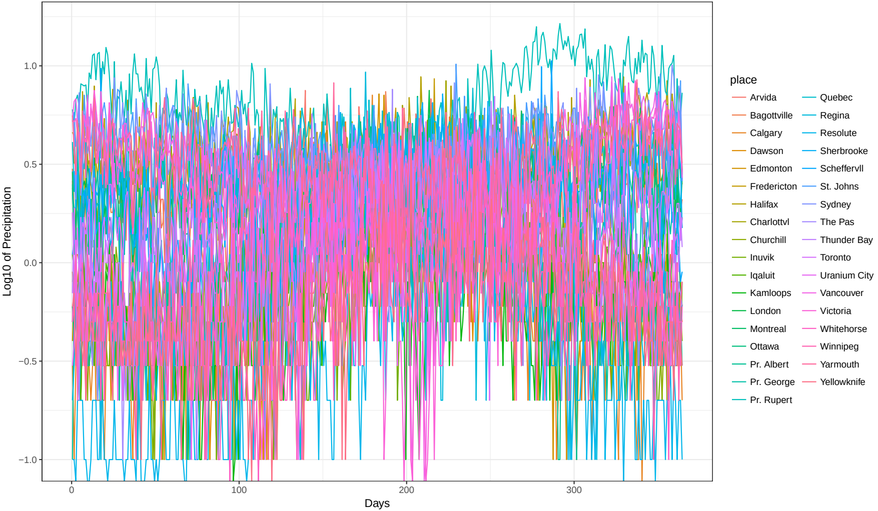
We would like to predict $\log_{10}$ of Precipitation in relation to Temperature. We consider a model

$$y_i(t) = \beta_0(t) + \beta_1(t) z_i(t) + \varepsilon_i(t), \qquad i = 1, \ldots, 35,$$

where

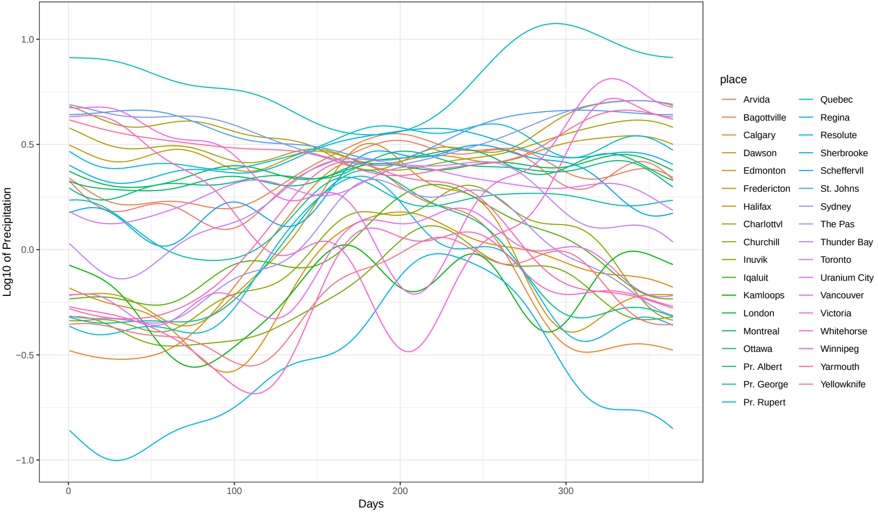
- $z_i(t)$ ... smoothed annual temperature at the i-th station
- $y_i(t)$ ... smoothed $\log_{10}$ of precipitation at the i-th station
- $\beta_0(t), \beta_1(t)$ ... regression parameters

Functional Response Models



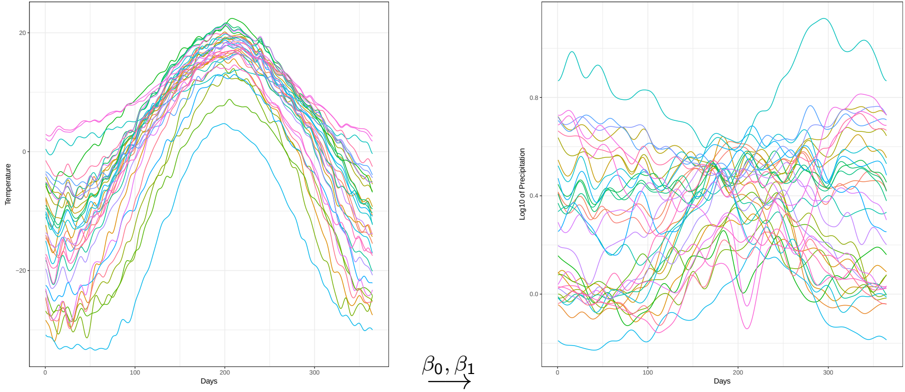
$\log_{10}$ of Precipitation

Functional Response Models



Smoothed data

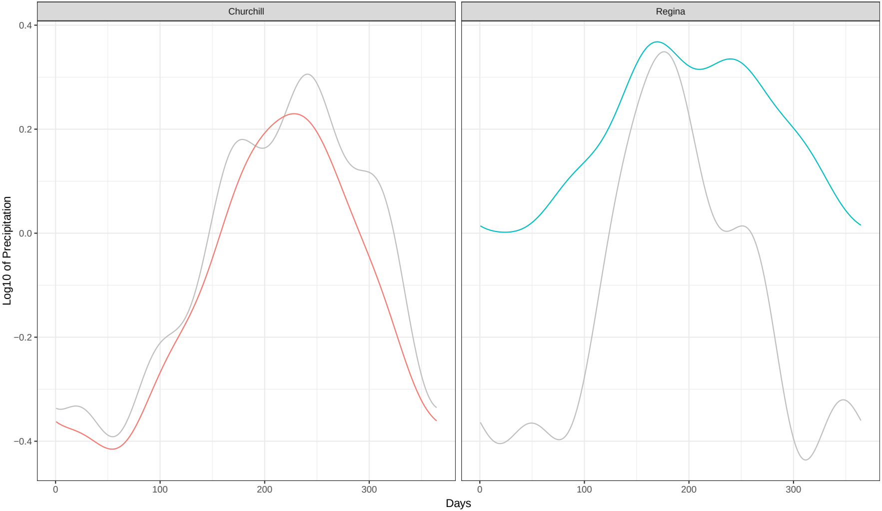
Functional Response Models



Regression model: $\log_{10}(\text{precipitation}) \sim \text{temperature}$

Functional Response Models

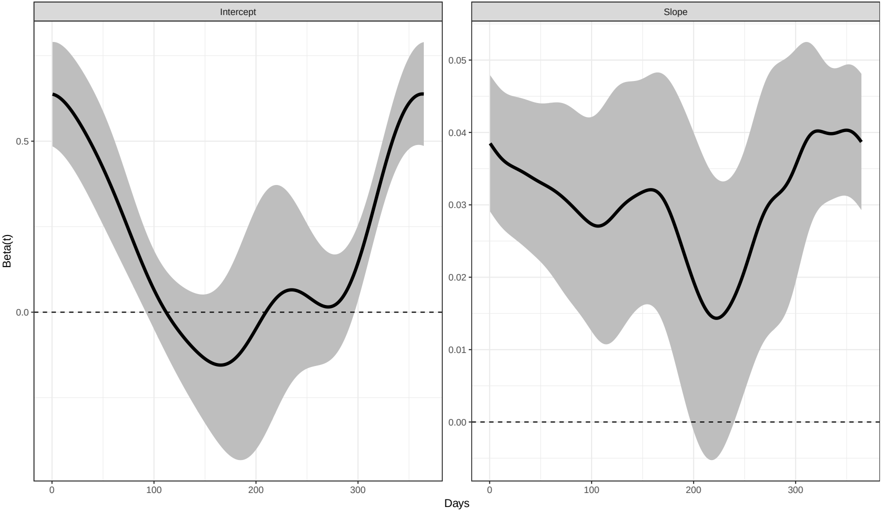
Fitted Model



Precipitation estimates

Functional Response Models

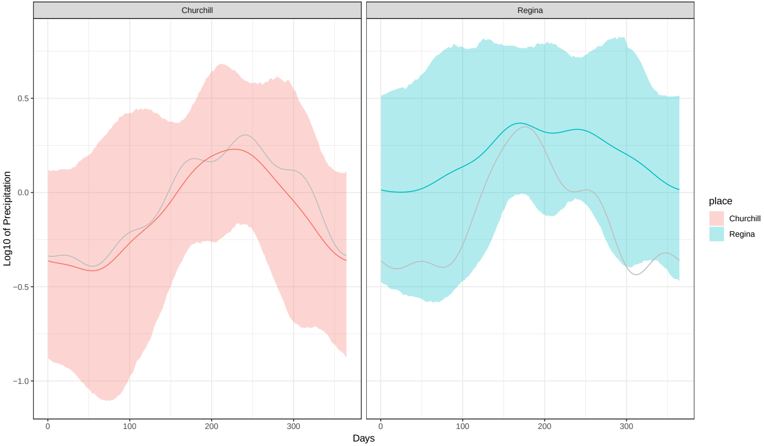
Confidence Intervals for parameters



Model parameters estimates, $\widehat{\beta}_0(t)$, $\widehat{\beta}_1(t)$

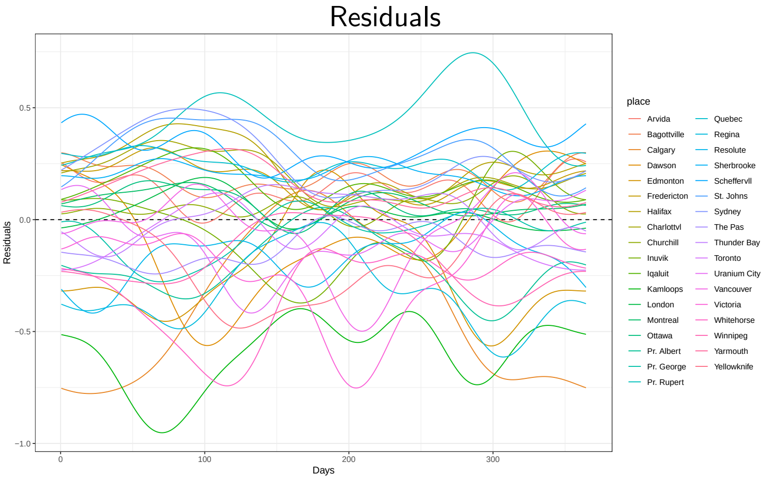
Confidence Intervals

?Main idea: Generate B times $y_i^b \sim N_N(\hat{\mathbf{y}}_i, \hat{\mathbf{\Sigma}}) \Rightarrow$ pointwise CI for $\hat{\mathbf{y}}_i$



Functional Response Models

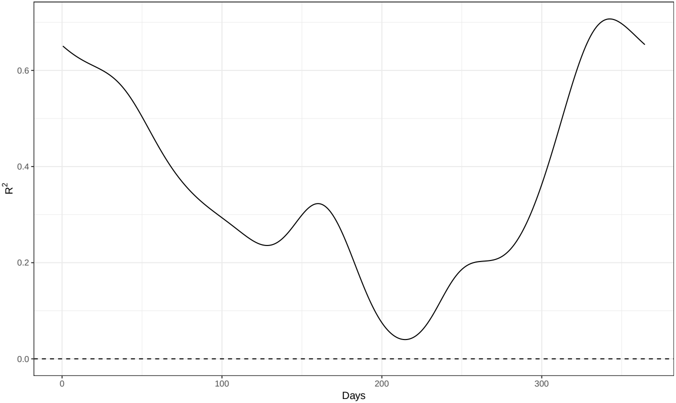
Assessing the fit



Functional Response Models

Functional R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2}$$



Functional Response Models

F-statistic

To test significance, we can define a point-wise F-statistic

$$F(t) = \frac{\text{Var}(\hat{\mathbf{y}}(t))}{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2 / n}$$

indicates where there is a large amount of signal relative to variance.

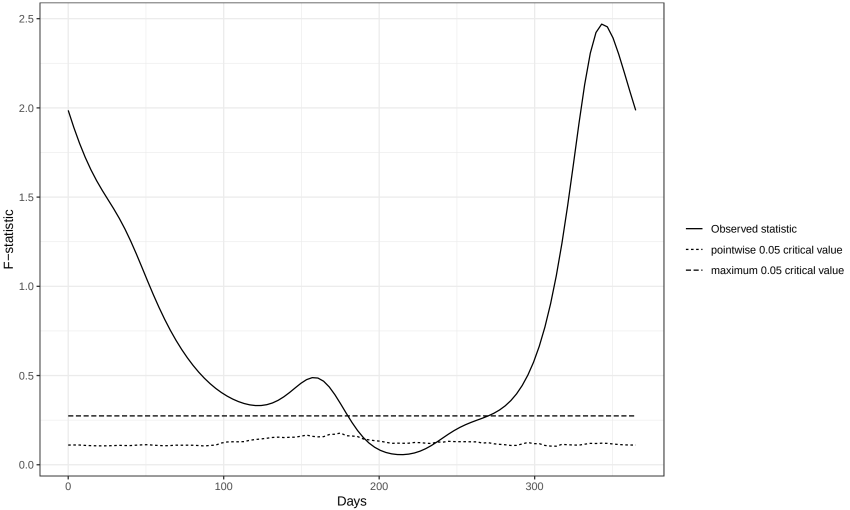
Test over-all regression significance based on

$$F^* = \max F(t).$$

Practical implementation is based on the **permutation test**.

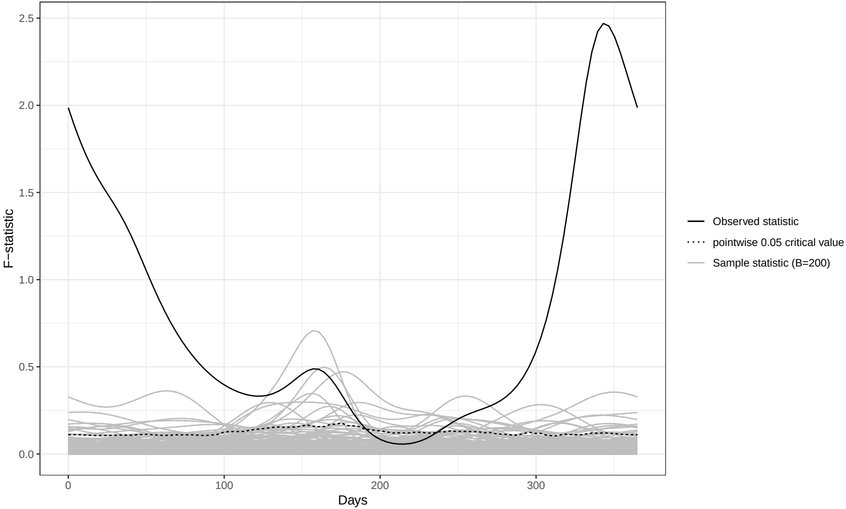
Canadian Weather

$B = 200, p_B = 0$



Canadian Weather

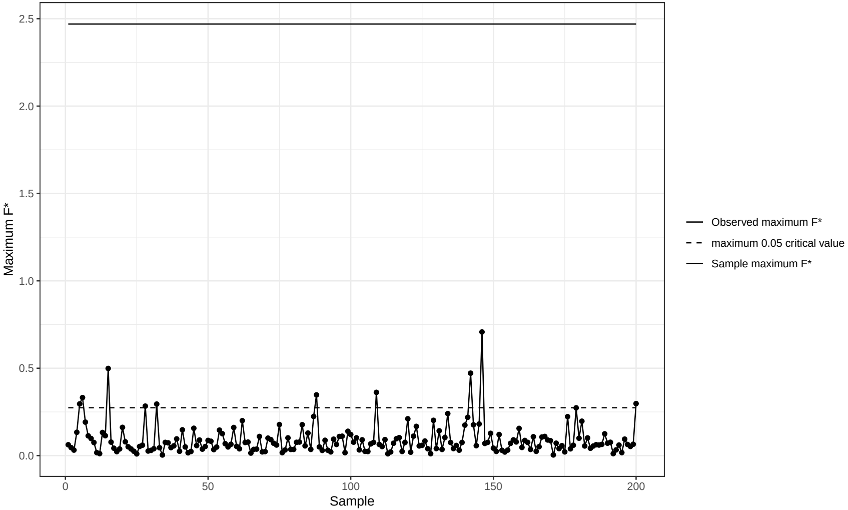
detailed test results



Functional Response Models

Canadian Weather

$F^* = 2.47, F^*_{0.95} = 0.274$



Functional Response Models

Another Approach for Canadian Weather (Ramsey et al. 2005)

They combined the original model

$$y_i(t) = \beta_0(t) + \beta_1(t)z_i(t) + \varepsilon_i(t), \qquad i = 1, \dots, 35,$$

where

- $z_i(t)$ ... smoothed annual temperature at the i-th station
- $y_i(t)$ ... smoothed $\log_{10}$ of precipitation at the i-th station

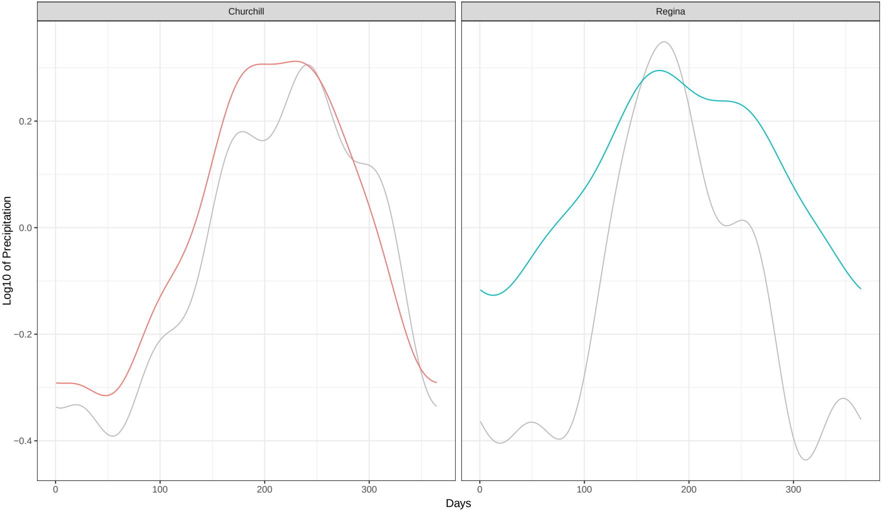
with **fANOVA** for regions, where $z_{ij}(t) = \mu(t) + \alpha_j(t) + e_{ij}(t)$.
Thus they considered a model

$$y_{ij}(t) = \mu(t) + \alpha_j(t) + \beta(t)e_{ij}(t) + \varepsilon_{ij}(t), \; j = 1, \dots, 4,$$

where

- $e_{ij}(t)$ ... the residual temperature at the i-th station after removing the temperature effect of climate zone j by using fANOVA.

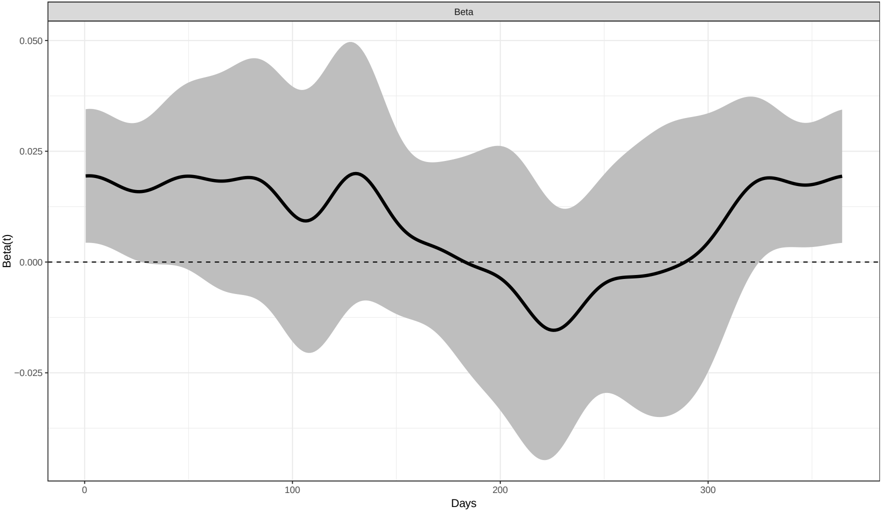
Fitted Model



Precipitation estimates

Functional Response Models

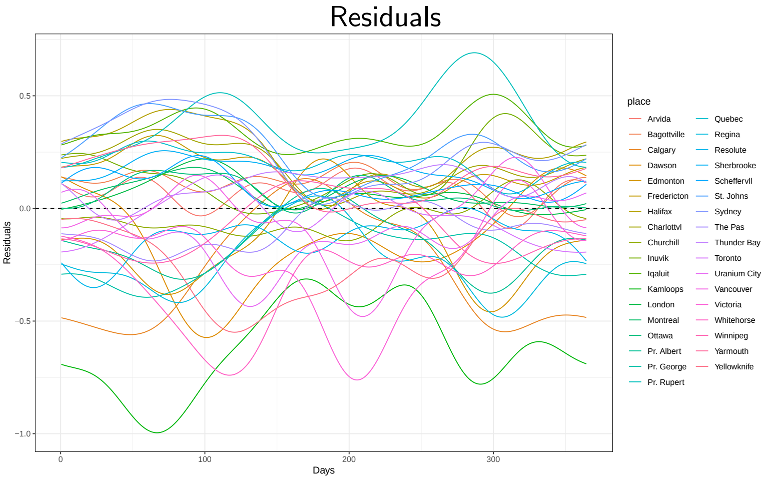
Confidence Interval for parameter β



Model parameters estimates, $\widehat{\beta}(t)$

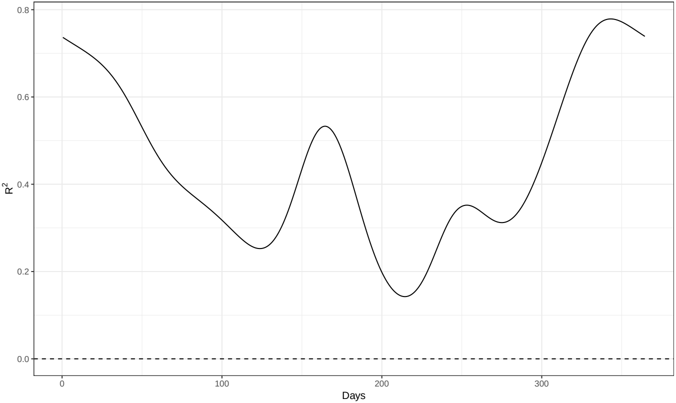
Functional Response Models

Assessing the fit



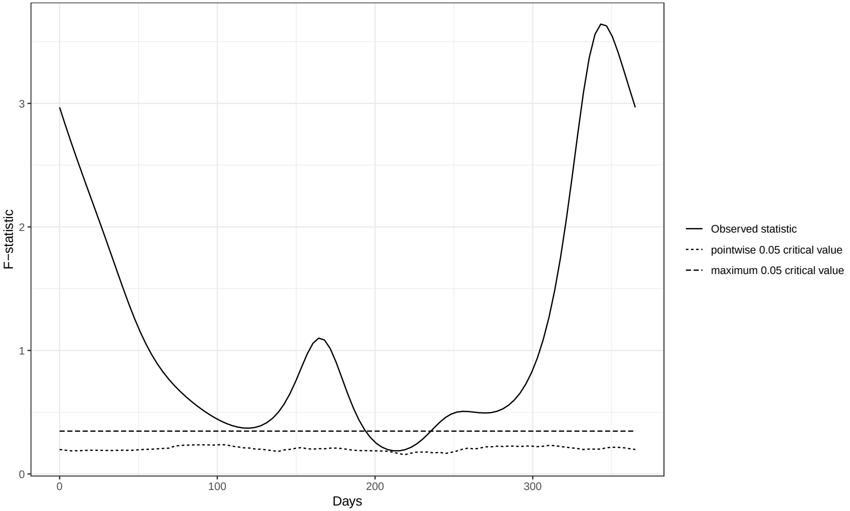
Functional R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2}$$



Canadian Weather

$B = 200, p_B = 0$



Functional Response Models

2. Fully Functional Regression Model

Let us recall the **concurrent** model (2)

$$y_i(t) = \beta_0(t) + \sum_{j=1}^{K} \beta_j(t) z_{ij}(t) + \varepsilon_i(t).$$

We can generalize (2) for $K \to \infty$

$$y_i(t) = \beta_0(t) + \int_{\Omega_t} \beta_1(t,s) z_i(s) ds + \varepsilon_i(t), \tag{3}$$

where

- $\Omega_t = \{s | s < t\}$ ... **historical linear model**
- $\Omega_t = \{s \text{ unconstrained}\}$ ... **full integration regression**
- $\beta_1(t,s)$ ... defines the dependence of $y_i(t)$ on covariate $z_i(s)$ at each time t ($z_i(s)$ need not be defined over the same range as $y_i(t)$)

Functional Response Models

Example Canadian Weather

We consider a model

$$y_i(t) = \beta_0(t) + \int_{\Omega_t} \beta_1(t,s) z_i(s) ds + \varepsilon_i(t), \qquad i = 1, \dots, 35,$$

where

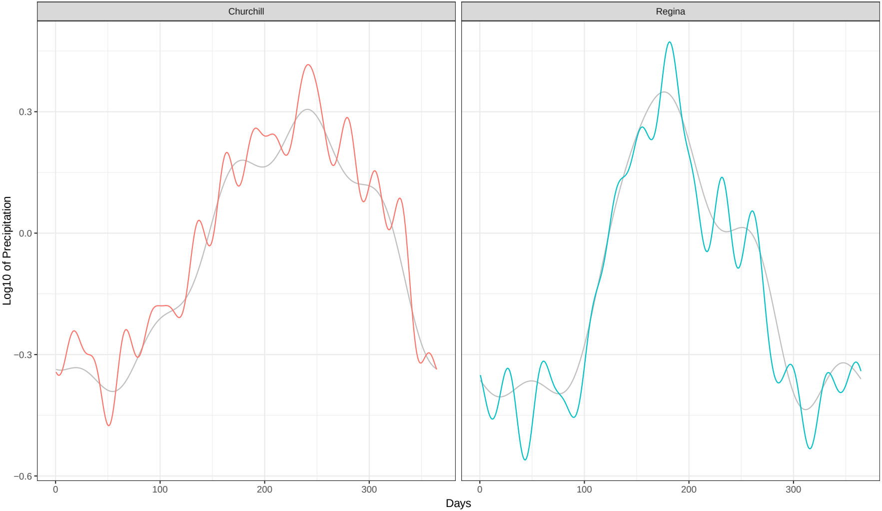
- $z_i(t)$ ...smoothed annual temperature at the i-th station
- $y_i(t)$ ...smoothed $\log_{10}$ of precipitation at the i-th station
- $\beta_0(t), \beta_1(t,s)$ ...regression parameters

Parameters of smoothing

- "Full" basis, i.e. 65 Fourier basis for $z(t), y(t)$ $\Rightarrow$ overfitted model
- "Restricted" basis, i.e. 21, 11 Fourier basis for $z(t), y(t)$, respectively $\Rightarrow$ quite satisfactory compromise

Functional Response Models

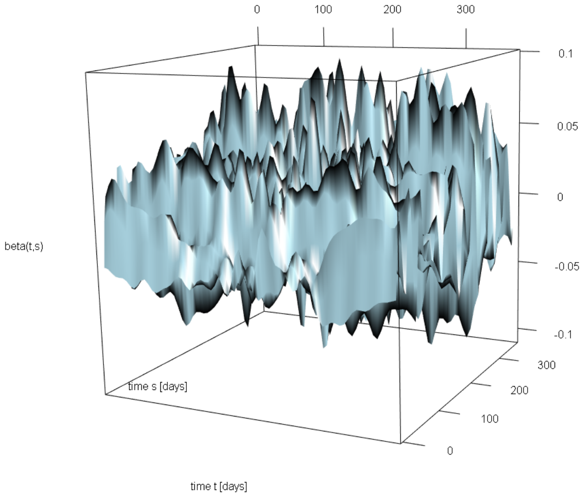
Fitted Model – full basis



Precipitation estimates

Functional Response Models

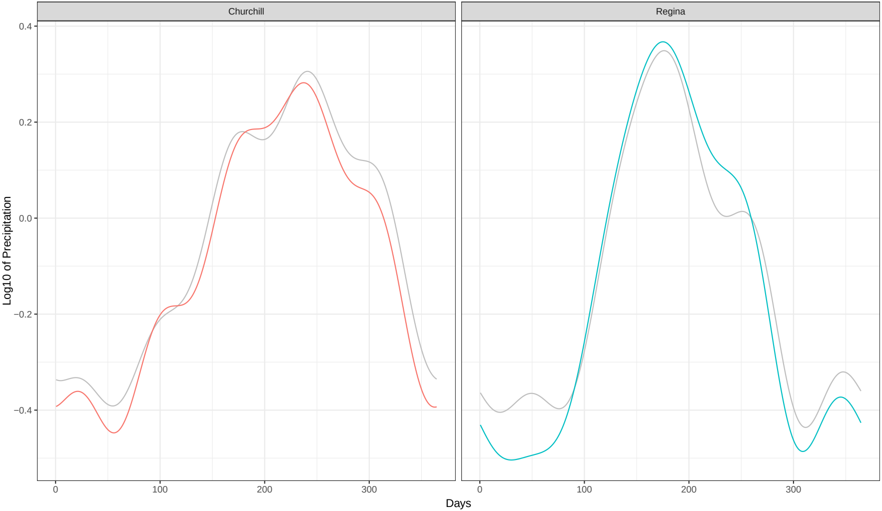
Parameters



Model parameter estimate, $\widehat{\beta}_1(t, s)$

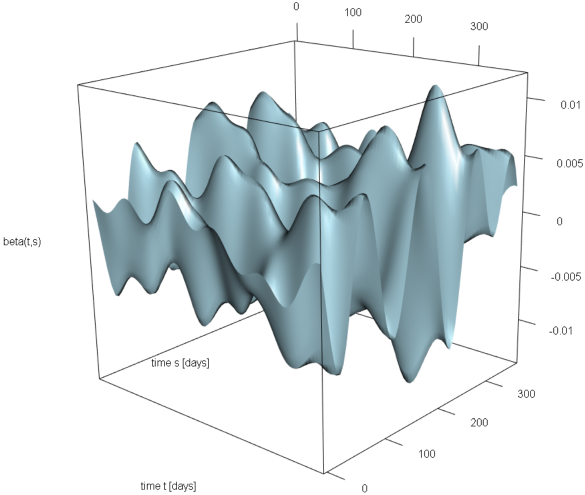
Functional Response Models

Fitted Model – restricted basis



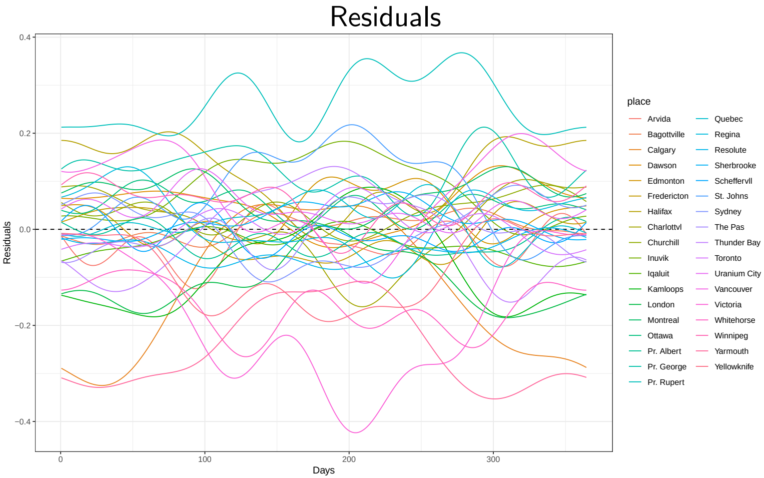
Precipitation estimates

Parameters


0
100
200
300
0.01
0.005
0
-0.005
-0.01
beta(t,s)
time s [days]
time t [days]
300
200
100
0

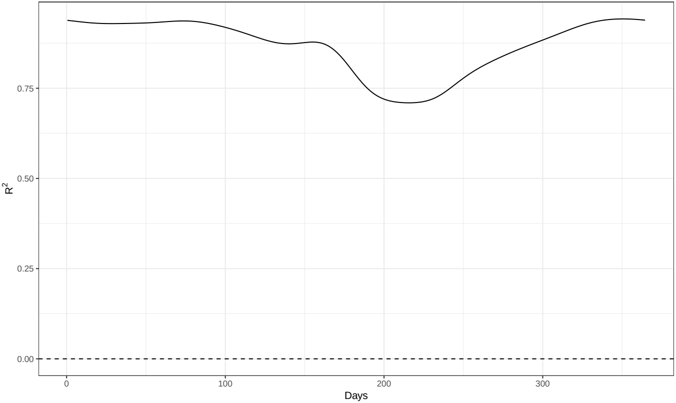
Functional Response Models

Assessing the fit



Functional R^2

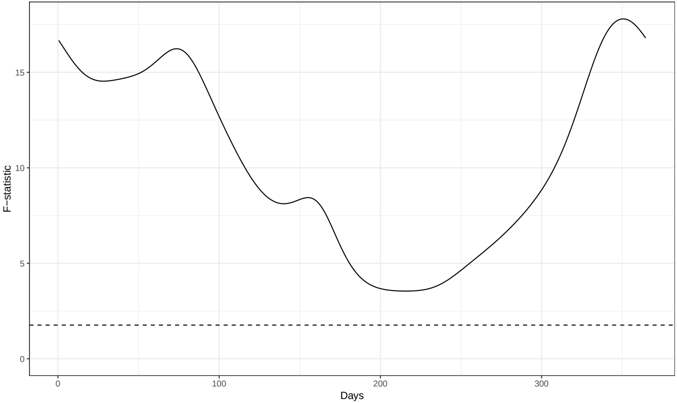
$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2}$$



Functional Response Models

Functional F-statistic

$$F(t) = \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2/(n-1)}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2/n}$$



Example Swedish Mortality Data

- Log hazard rates calculated from tables of mortality at ages 0 through 80 for Swedish women.
- Data available for birth years 1751 through 1894.
- Interest in looking at mortality trends.

Clear over-all reduction in mortality; but effects common to adjacent cohorts?

Functional Response Models

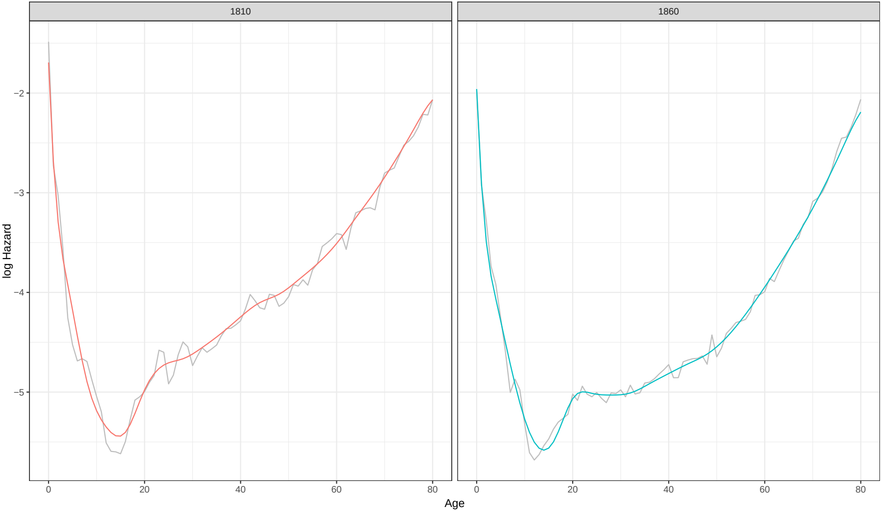
We consider a functional auto-regressive model

$$y_{i+1}(t) = \beta_0(t) + \int \beta_1(t,s) y_i(s) ds + \varepsilon_i(t), \qquad i = 1, \dots, 143,$$

Parameters of smoothing

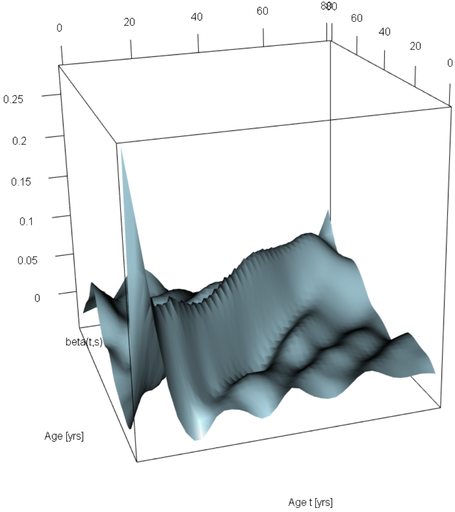
- 85 B-spline basis of order 6 for $y(t)$, data smoothed with parameter $\lambda = 10^{-7}$ and $J_4(y) = \int [D^4 y(t)]^2 dt$

Fitted Model



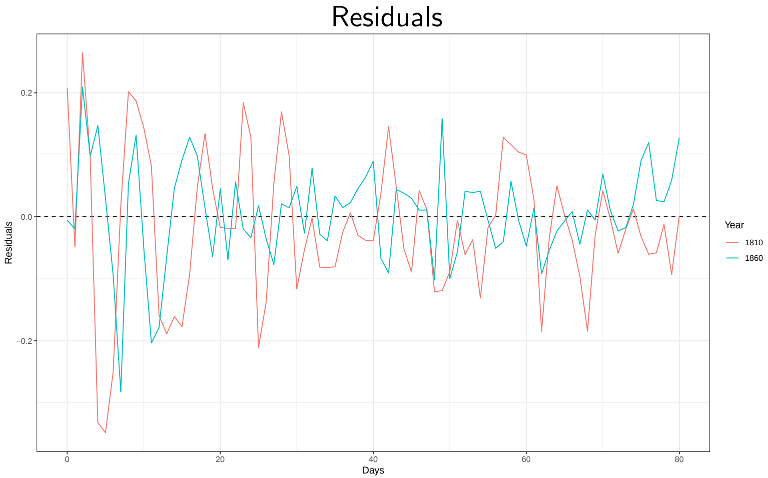
Log hazard estimates

Parameters



Model parameter estimate, $\widehat{\beta}_1(t,s)$

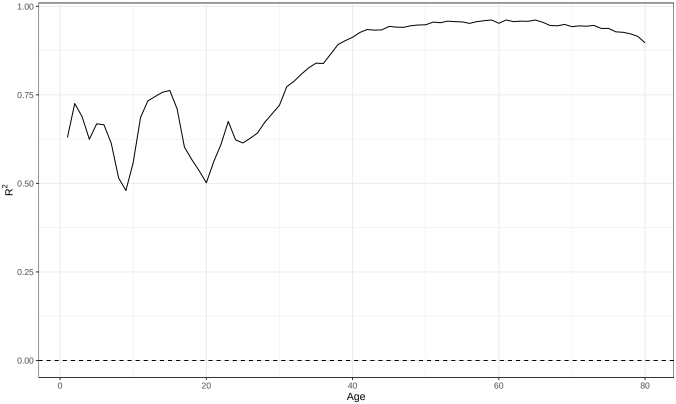
Assessing the fit



Functional Response Models

Functional R^2

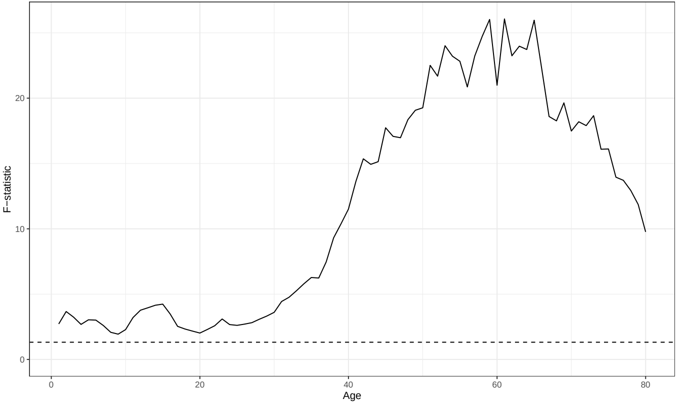
$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2}$$



Functional Response Models

Functional F-statistic

$$F(t) = \frac{\sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2/(n-1)}{\sum_{i=1}^{n} (y_i(t) - \bar{y}(t))^2/n}$$



Problems to solve

1. Gait Data
   Load the variable `gait` from the `fda` package. These data consist of the angles formed by the hip and knee of each of 39 children over each child's gait cycle.
   - Smooth the data by Fourier bases with harmonic acceleration penalties and plot all hip and knee curves (see Figure 1).
   - A question of interest is the extent to which the hip angle can explain the knee angle. Let us consider a model of the form

     $$Y_i(t) = \beta_0(t) + \beta_1(t)X_i(t) + \varepsilon_i(t).$$

     Estimate parameters of the model and plot them (see Figure 2).
   - Plot predictions for the first two boys with its bootstrap pointwise confidence bands (see Figure 3).
   - Asses the model by the permutation test for F-statistic and plot the result (see Figure 4).
   - Plot $\beta_0(t)$ together with mean curve for knee angle and $\beta_1(t)$ together with functional R^2 of the model (see Figure 5). Could we interpret it?

Problems to solve

2. Simulation
   - Using the attached script generate 30 curves $x_i(t)$ with its sample points.
   - Smooth them using cubic B-spline bases with GCV optimal λ.
   - Consider a fully functional regression model

     $$y_i(t) = \beta_0(t) + \int \beta_1(t,s)x_i(s)ds + \varepsilon_i(t).$$

     Generate sample points for 30 curves $y_i(t)$ with $\beta_0(t) = 1 + 2t/30 - (t/30)^2$, $\beta_1(t,s) = \sin(2pi(x-y)/365)$ and $\varepsilon_i(t) \sim N(0,10)$. For an example of generated curves see Figure 6.
   - Smooth y_i by the same way as x_i (see Figure 7).
   - Estimate parameters of the model and compare them with original (see Figure 8 for $\beta_0(t)$ comparison). Try several choices of `nbasis` for β's estimation.
   - Plot predictions for the first two curves together with originals (see Figure 9).
   - Plot functional R^2 of the model (see Figure 10) and interpret it.

Problems to solve

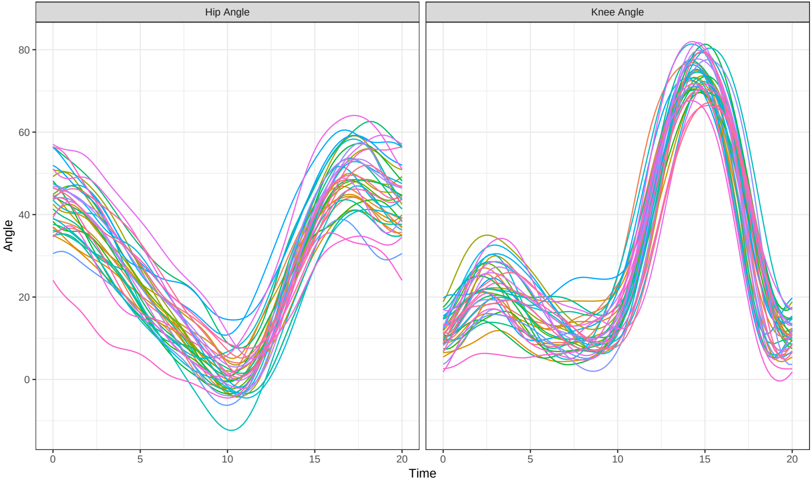


Figure 1.

Problems to solve

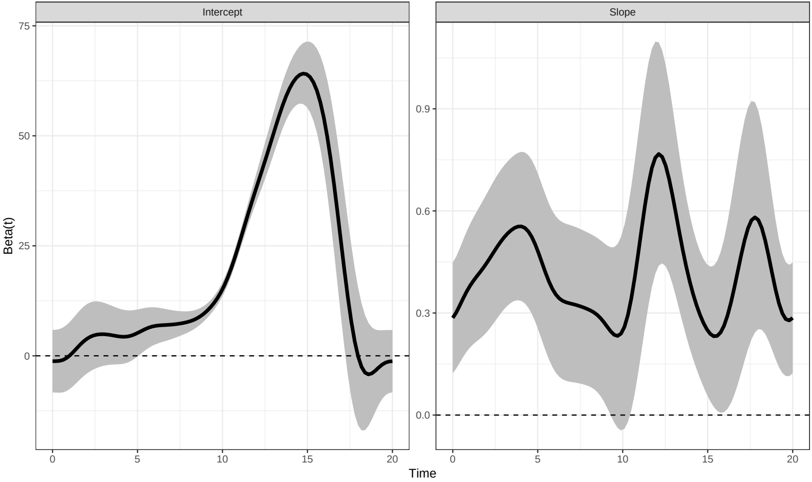


Figure 2.

Problems to solve

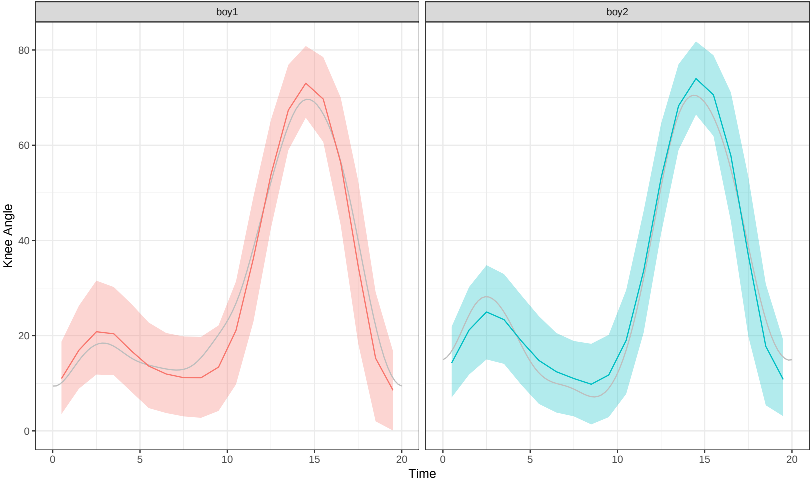


Figure 3.

Problems to solve

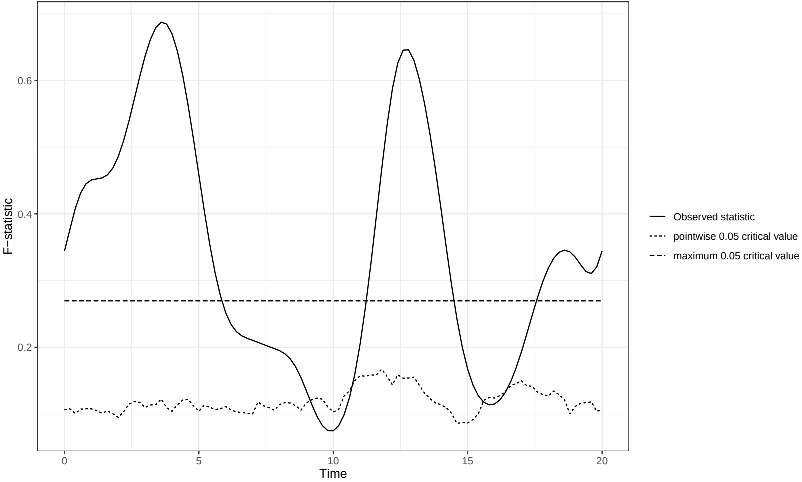


Figure 4.

Problems to solve

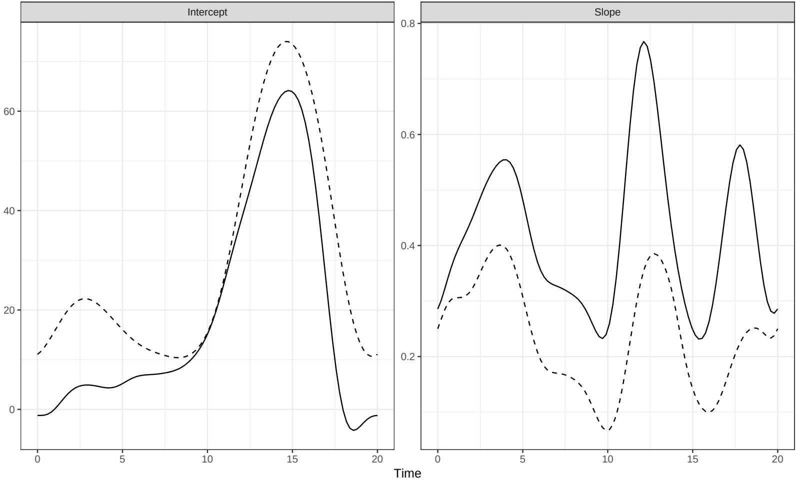


Figure 5.

Problems to solve

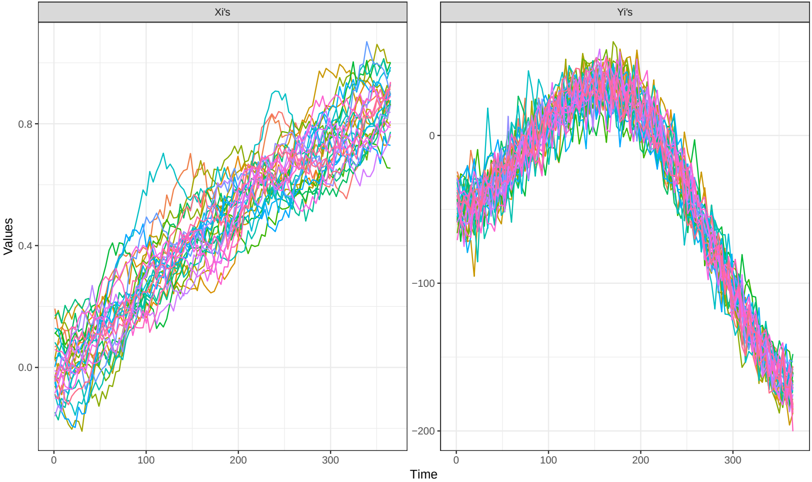


Figure 6.

Problems to solve

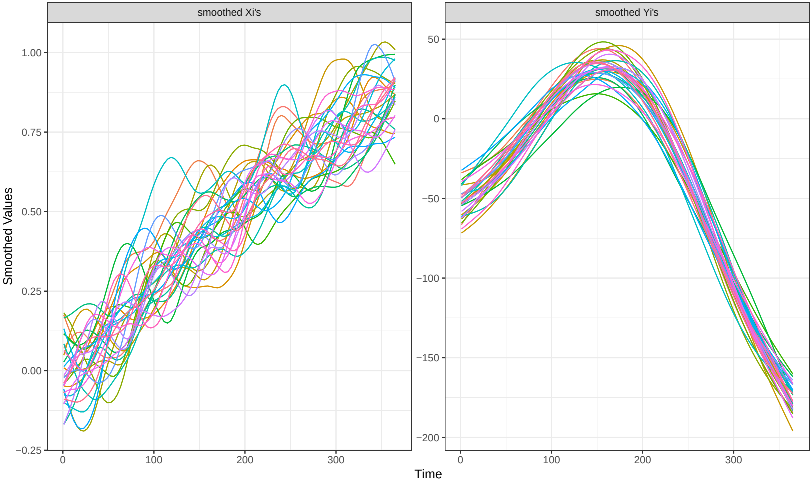


Figure 7.

Problems to solve

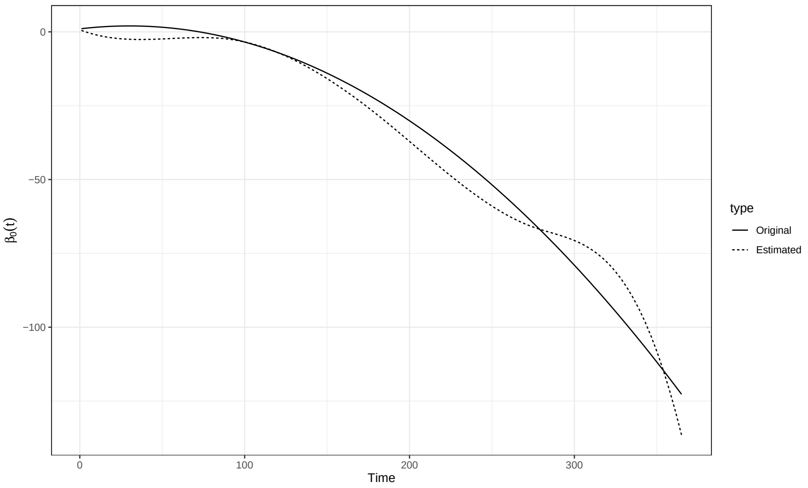


Figure 8.

Problems to solve

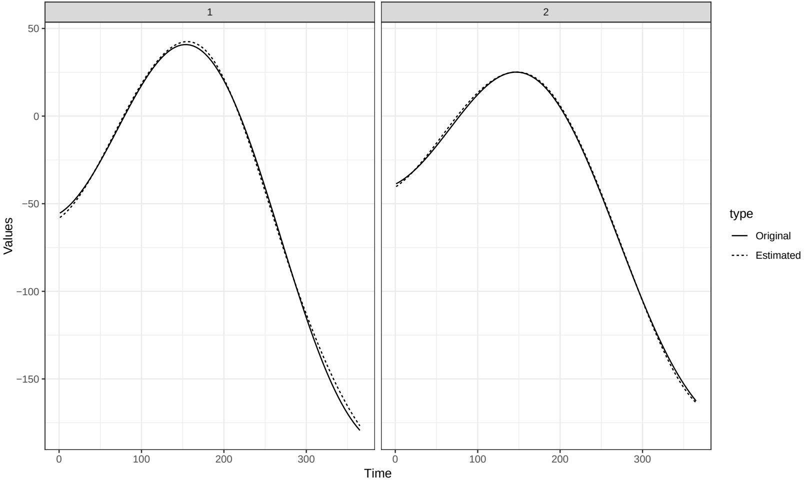


Figure 9.

Problems to solve

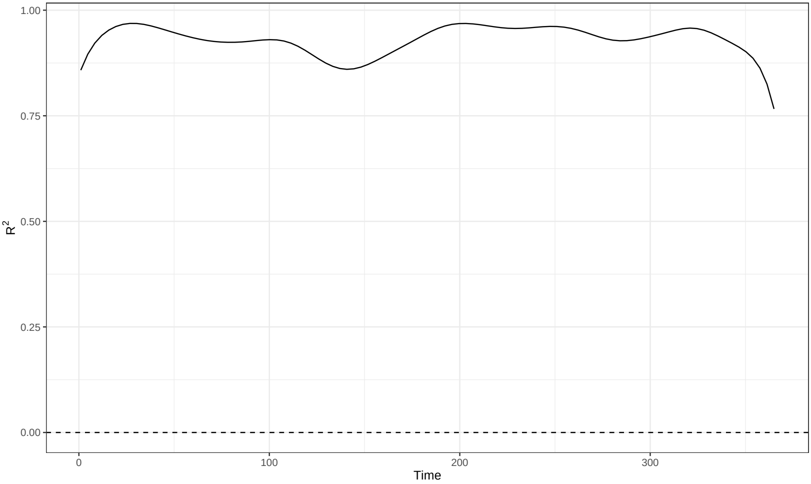


Figure 10.