Reproducing Kernel Hilbert Spaces and Smoothing Spline Regression

Alexej Gossmann

Tulane University

agossman@tulane.edu

March 25, 2014

Overview

Regression

Reproducing Kernel Hilbert Spaces

Smoothing Spline Models

Penalized Least Squares Estimation

Examples

Regression Models

Given observations (x_i, y_i) for $i \in \{1, \ldots, n\}$ a *regression model* relates the dependent variable y and the independent variable x as follows,

$$y_i = f(x_i) + \varepsilon_i,$$

where f is the regression function, ε_i are zero-mean independent random errors with a common variance.

Parametric and Non-Parametric Regression

$y_i = f(x_i) + \varepsilon_i$, for $i \in \{1, \ldots, n\}$.

Parametric Regression The form of f is known except for finitely many unknown parameters, e.g. $f(x) = \beta_0 + \beta_1 x + \ldots + \beta_{m-1} x^{m-1}$ a polynomial of order m.

Non-Parametric Regression No specific form is imposed on f. We let the data speak for themselves in order to decide which function fits best. For example, one merely assumes that f lies in the space of "smooth" functions.

Hilbert Spaces

Definition (Hilbert Space)

A complete inner product space is called a *Hilbert Space*.

Definition (Evaluational Functional)

Let $\mathcal{H}$ be a Hilbert space of real-valued functions from $\mathcal{X}$ to $\mathbb{R}$, and let $x \in \mathcal{X}$. The functional $\mathcal{L}_x : \mathcal{H} \to \mathbb{R} : f \mapsto f(x)$ is called an *evaluational functional*.

Reproducing Kernel Hilbert Spaces

Definition (RKHS)

A Hilbert space of real-valued functions is a *reproducing kernel Hilbert space* (RKHS) if every evaluational functional is continuous.

Definition (RK)

Let $\mathcal{H}$ be an RKHS. By the Riesz representation theorem, $\exists R_x \in \mathcal{H} : \mathcal{L}_x f = (R_x, f)$. Denote, $R(x, z) := R_x(z)$. Then $R(x, z)$ is called the *reproducing kernel* (RK) of the RKHS $\mathcal{H}$.

Let $\mathcal{H}$ be an RKHS with an RK R. In particular, R is non-negative definite, i.e.

- Symmetry, $R(x, z) = R(z, x)$,
- $(\forall \alpha_1, \ldots, \alpha_n \in \mathbb{R})(\forall x_1, \ldots, x_n \in \mathcal{X}) : \sum_{i,j=1}^{n} \alpha_i \alpha_j R(x_i, x_j) \geq 0$.

Theorem (Moore-Aronszajn theorem)

For every non-negative definite function R on $\mathcal{X} \times \mathcal{X}$, there exists a unique RKHS $\mathcal{H}$ on $\mathcal{X}$ with R as its RK.

That is, there is a one-to-one correspondence between RKHS's and non-negative definite functions.

Properties

- For a finite dimensional RKHS $\mathcal{H}$ with orthonormal basis $\phi_1(x), \ldots, \phi_p(x)$, $R(x,z) = \sum_{i=1}^{p} \phi_i(x)\phi_i(z)$ is the RK of $\mathcal{H}$.
- All closed subspaces of RKHS's are RKHS's.
- If $\mathcal{H}$ is an RKHS, and $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, and R, R_0, R_1 are the RK's respectively, then $R = R_0 + R_1$.
- If $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ is a Hilbert space, and if $\mathcal{H}_0$ and $\mathcal{H}_1$ are RKHS's with RK's R_0 and R_1 respectively, then $\mathcal{H}$ is an RKHS with RK $R = R_0 + R_1$.

Construction of a Smoothing Spline Model

1. An RKHS $\mathcal{H}$ as the model space.
2. A decomposition of the model space into two subspaces $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0$ is a finite dimensional space of functions which are not penalized.
3. A penalty $\|P_1 f\|^2$, where P_1 is the projection onto $\mathcal{H}_1$.

Example: Model Space for Polynomial Splines

Definition (Polynomial Spline)

Given a set of fixed points $t_0 = a < t_1 < \ldots < t_k < b = t_{k+1}$ (called knots), a *polynomial spline of order r* is a function $f : [a, b] \to \mathbb{R}$ such that

- f is a piecewise polynomial of order r on $[t_i, t_{i+1})$ for all $i \in \{0, \ldots, k\}$,
- f has $r - 2$ continuous derivatives and the $r - 1$st derivative is a step function with jumps at knots.

f is a *natural polynomial spline of order* $r = 2m$ if additionally $f^{(j)}(a) = f^{(j)}(b) = 0$ for $j = m, \ldots, 2m - 1$.

For instance, for a cubic natural spline we have $r = 4$ and $m = 2$.

Example: Model Space for Polynomial Splines

Given $y_i = f(x_i) + \varepsilon_i$, where ε_i are zero-mean independent random errors with a common variance, assume that

$$f \in W_2^m[a,b] := \left\{ f : [a,b] \to \mathbb{R} \middle| f, f', \ldots, f^{(m-1)} \text{ abs. cont.}, \right. \\ \left. \text{and } \int_a^b (f^{(m)})^2 \mathrm{d}x < \infty \right\}.$$

$W_2^m[a,b]$ is an RKHS with the inner product

$$(f,g) = \sum_{\nu=0}^{m-1} f^{(\nu)}(a) g^{(\nu)}(a) + \int_a^b f^{(m)} g^{(m)} \mathrm{d}x.$$

Example: Model Space for Polynomial Splines

- $W_2^m[a,b] = \mathcal{H}_0 \oplus \mathcal{H}_1$,
- $\mathcal{H}_0 := \mathrm{span}\left\{1, (x-a), \ldots, \frac{(x-a)^{m-1}}{(m-1)!}\right\}$,
- $\mathcal{H}_1 := \left\{f : f^{(\nu)} = 0, \nu = 0, \ldots, m-1, \int_a^b (f^{(m)})^2 \mathrm{d}x < \infty\right\}$.

$\rightsquigarrow$ Minimize the penalized least squares (PLS),

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \|P_1 f\|^2,$$

where P_1 is the projection onto $\mathcal{H}_1$, and $\|P_1 f\|^2 = \int_a^b (f^{(m)})^2 \mathrm{d}x$.

Example: Model Space for Polynomial Splines

For a fixed smoothing parameter λ, the PLS

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_a^b (f^{(m)})^2 \mathrm{d}x$$

has a unique minimizer $\hat{f}$, and $\hat{f}$ is a natural polynomial spline of order $2m$ with knots at distinct design points x_i ($i \in \{1, \ldots, n\}$).

General Smoothing Spline Regression Models

- $y_i = f(x_i) + \varepsilon_i$, $i \in \{1, \ldots, n\}$, where ε_i are zero-mean independent random errors with a common variance,
- $f \in \mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, i.e.
  $f = \underbrace{f_0}_{\text{parametric component}} + \underbrace{f_1}_{\textit{smooth component}}$,
- The *null space* $\mathcal{H}_0$ is a finite dimensional RKHS with basis $\phi_1(x), \ldots, \phi_p(x)$, and with RK $R_0(x, z)$,
- $\mathcal{H}_1$ is an RKHS with RK $R_1(x, z)$,
- For greater generality, consider $y_i = \mathcal{L}_i f + \varepsilon_i$, where $\mathcal{L}_i$ are bounded linear functionals on $\mathcal{H}$. For simplicity, here, we only consider the case that $\mathcal{L}_i$ is the evaluational functional at x_i, i.e. $\mathcal{L}_i f = f(x_i)$.

Penalized Least Squares Estimation

The smoothing spline estimate of f, $\hat{f}$, is the minimizer of the PLS

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \|P_1 f\|^2,$$

where λ is a smoothing parameter controlling the balance between the goodness-of-fit measured by the least squares and the departure from the null space $\mathcal{H}_0$ measured by $\|P_1 f\|^2$.

Kimeldorf-Wahba representer theorem

Theorem

The PLS has a unique minimizer given by,

$$\hat{f}(x) = \sum_{\nu=1}^{p} d_\nu \phi_\nu(x) + \sum_{i=1}^{n} c_i \xi_i(x),$$

where $\xi_i = P_1 R_{x_i}$ *(in particular,* $\xi_i(x_j) = (P_1 R_{x_i}, R_{x_j}) = R_1(x_i, x_j)$*), and where the coefficients* $\mathbf{d} = (d_1, \ldots, d_p)^T$ *and* $\mathbf{c} = (c_1, \ldots, c_n)^T$ *are determined by,*

$$\begin{aligned}
\mathbf{d} &= (T^T M^{-1} T)^{-1} T^T M^{-1} y, \\
\mathbf{c} &= M^{-1}(I - T(T^T M^{-1} T)^{-1} T^T M^{-1}) y, \\
M &= \Sigma + n\lambda I, \\
\Sigma &= \{R_1(x_i, x_j)\}_{i,j=1}^{n}.
\end{aligned}$$

Smoothing Parameter Estimation

Ideally, we want to select λ that minimizes the MSE,

$$\mathrm{MSE} = \mathrm{E}\left(\frac{1}{n}\|\hat{f} - f\|^2\right),$$

where $f = (f(x_1), \ldots, f(x_n))^T$ and $\hat{f} = (\hat{f}(x_1), \ldots, \hat{f}(x_n))^T$.
Bias-Variance Trade-Off:

$$\mathrm{MSE} = \frac{1}{n}\|\mathrm{E}\hat{f} - f\|^2 + \frac{1}{n}\mathrm{E}\|\hat{f} - \mathrm{E}\hat{f}\|^2 =: b^2(\lambda) + v(\lambda).$$

If λ increases from 0 to ∞, then $b^2(\lambda)$ increases from $b^2(0) = 0$ (interpolation), and $v(\lambda)$ decreases.

Smoothing Parameter Estimation

- MSE depends on the unknown true function f. $\rightsquigarrow$ Try to estimate MSE.
- Cross-validation and generalized cross validation (CV, GCV).
- Unbiased risk (UBR).
- Generalized maximum likelihood (GML).

Example: Model Space for Polynomial Splines

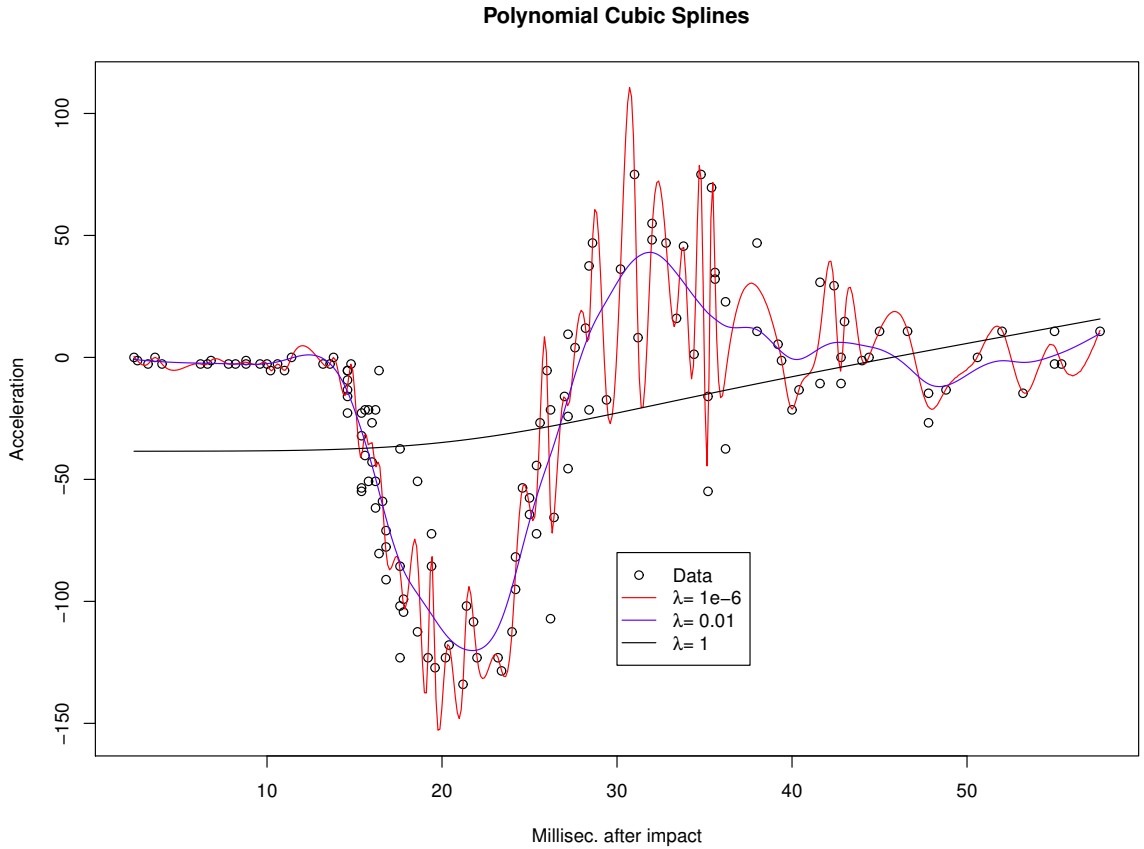


Figure: Polynomial cubic splines obtained by smoothing spline regression. The data (mcycle from the R-package MASS) gives a series of measurements of head acceleration in a simulated motorcycle accident (used to test crash helmets).

Example: Thin-Plate Splines

- $y_i = f(x_i) + \varepsilon_i$, $f : \mathbb{R}^d \to \mathbb{R}$, $x_i \in \mathbb{R}^d$, $i \in \{1, \ldots, n\}$,
- $f \in W_2^m(\mathbb{R}^d) := \left\{ f : \mathcal{J}_m^d(f) < \infty \right\}$,

$$\mathcal{J}_m^d(f) := \sum_{\alpha_1 + \ldots + \alpha_d = m} \frac{m!}{\alpha_1! \cdots \alpha_d!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)^2 \prod_{j=1}^{d} dx_j,$$

- $W_2^m(\mathbb{R}^d) = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0$ is the space spanned by polynomials in d variables of total degree up to $m - 1$,
- A thin-plate spline estimate is the minimizer to the PLS

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \mathcal{J}_m^d(f).$$

Example: Thin-Plate Splines

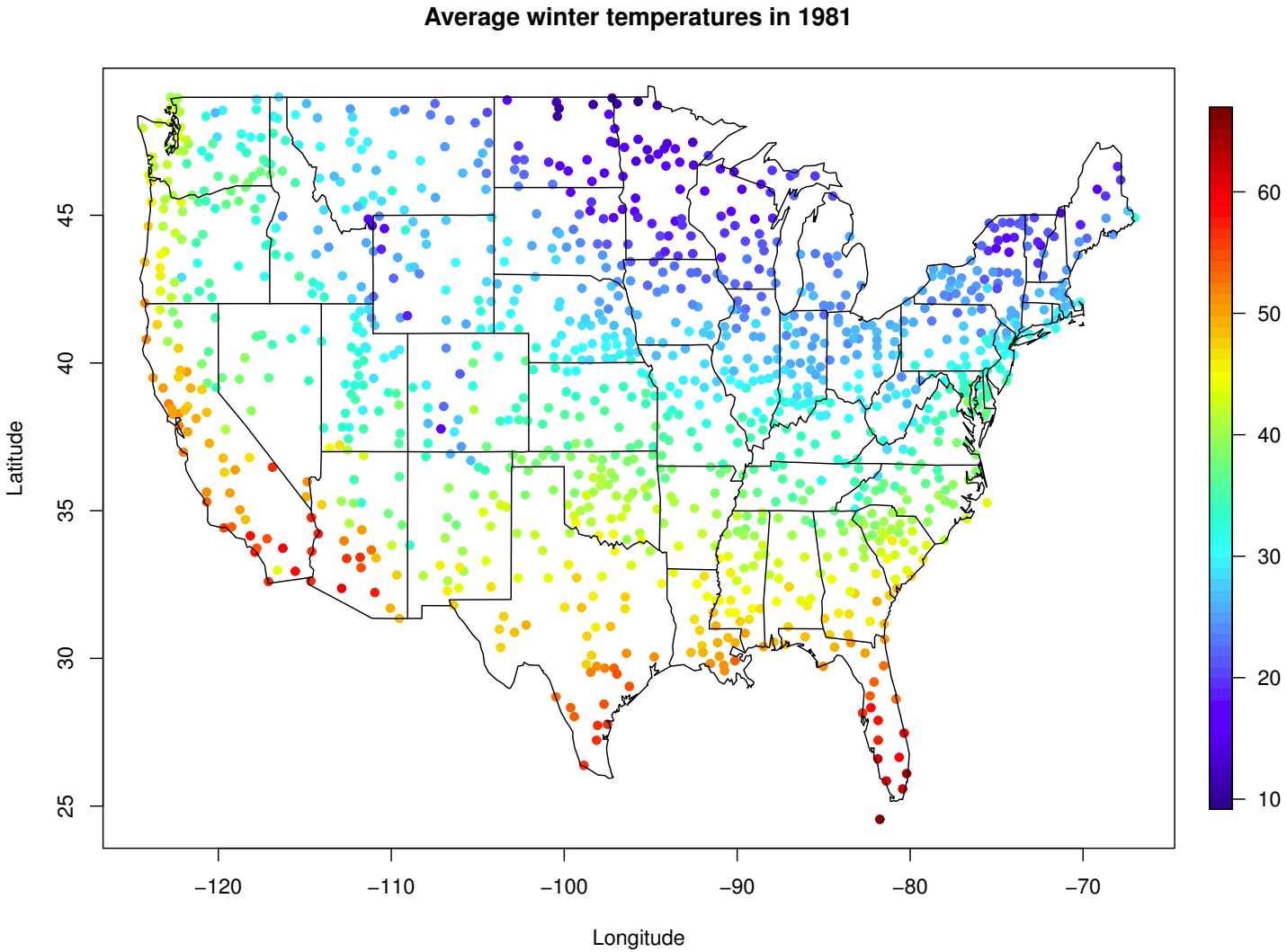
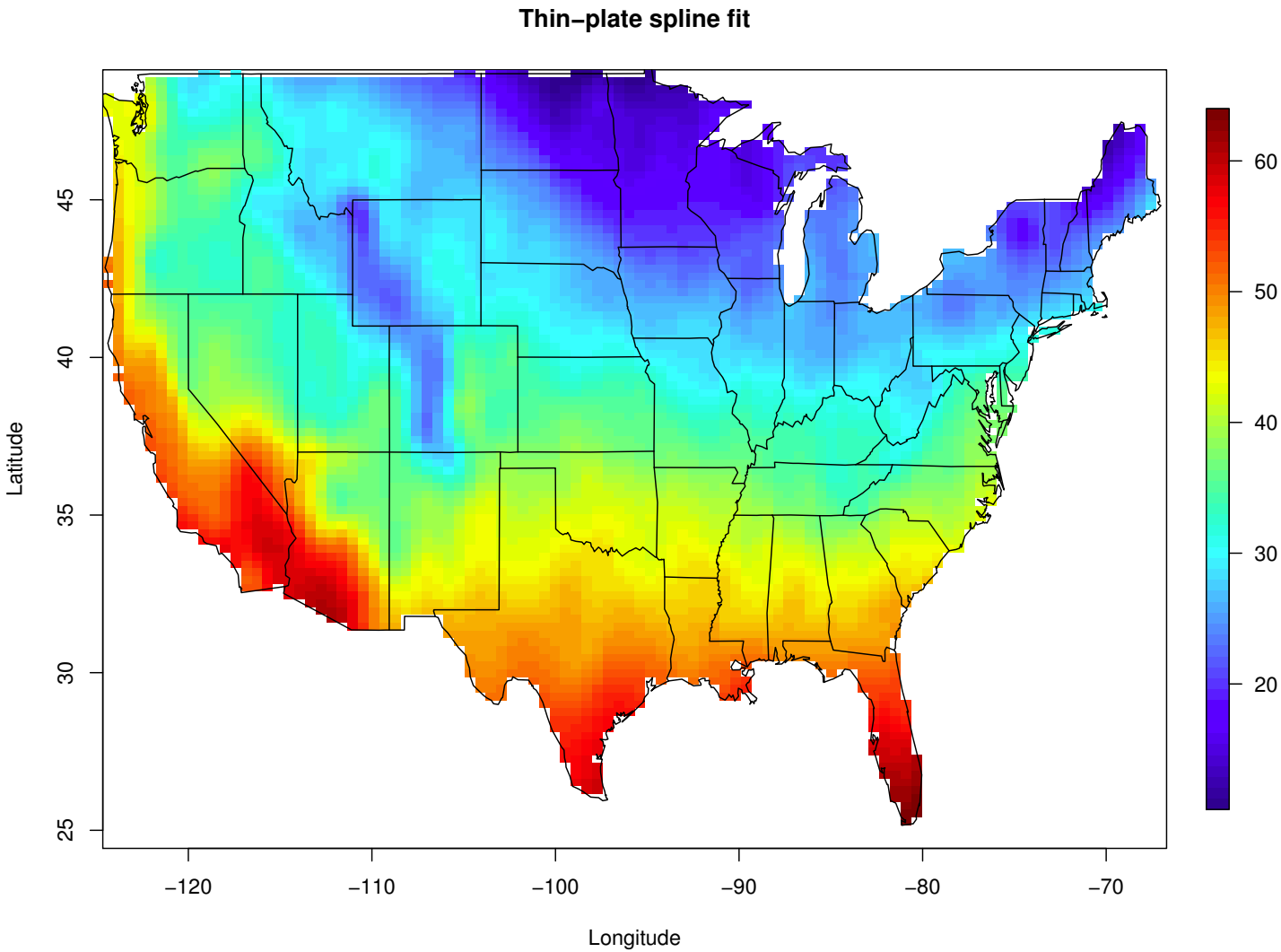


Figure: Average winter temperatures in 1981 from 1205 stations in USA.

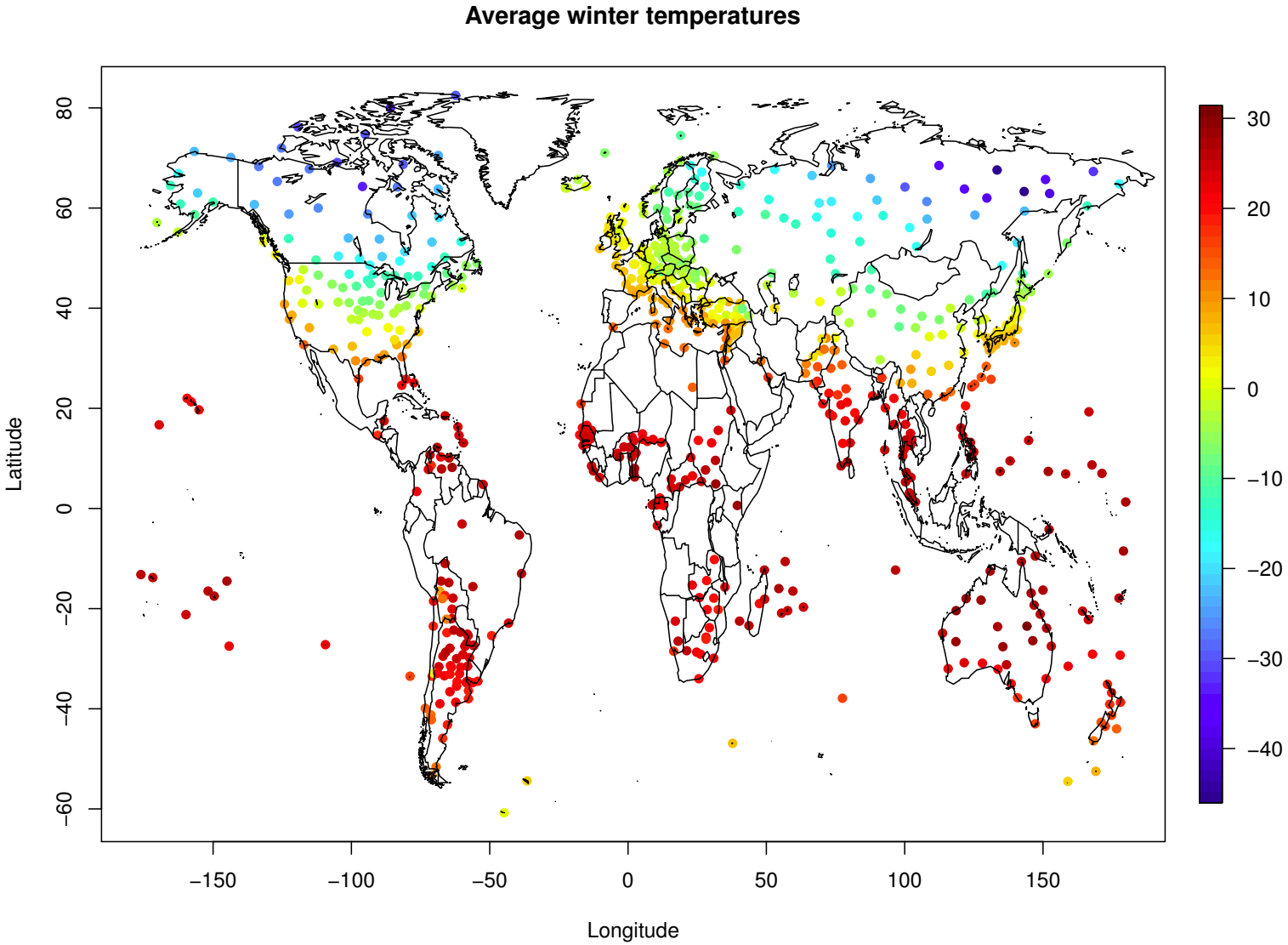
Example: Thin-Plate Splines



Example: Spherical Splines

- The *spherical spline* is an extension of both, the *periodic spline* defined on the unit circle and the *thin-plate spline* defined on R^2.
- $y_i = f(x_i) + \varepsilon_i$.
- $f : \mathcal{S} \to \mathbb{R}$, where $\mathcal{S}$ is the unit sphere.
- $x_i = (\theta_i, \phi_i) \in (0, 2\pi) \times (-\pi/2, \pi/2)$, where θ_i is the *longitude* and ϕ_i is the *latitude*.
- $f \in W_2^m(\mathcal{S}) := \left\{ f : \left| \int_{\mathcal{S}} f \, \mathrm{d}x \right| < \infty, \mathcal{J}(f) < \infty \right\}$, where $\mathcal{J}(f) =$"a very complicated expression".

Example: Spherical Splines

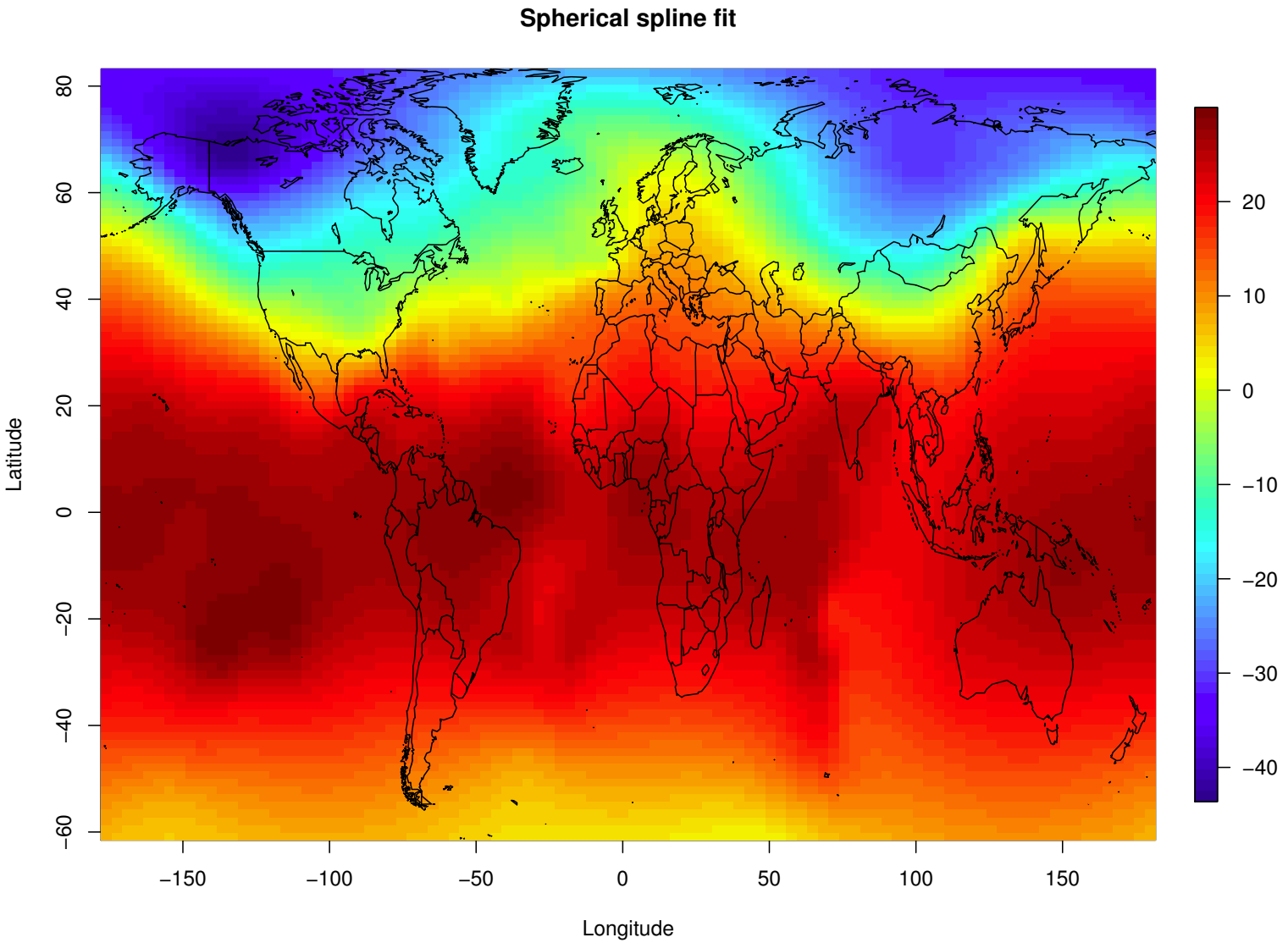


Example: Spherical Splines

Fit a spherical spline model in *R*:

```
> library(assist)
> data(climate)
> climate.ssr <- ssr(temp~1,
+ rk=sphere(cbind(long,lat)), data=climate)
```

Example: Spherical Splines



A remark concerning the thin-plate and the spherical splines examples

Both weather models are *not optimal* because the models assume the observations to be uncorrelated which is not a realistic assumption.

More Examples...

- Periodic spline
- Partial spline
- L-splines
- Exponential spline
- Logistic spline
- Linear-periodic spline
- Trigonometric spline

Bibliography...

Everything up to here comes from **Smoothing Splines: Methods and Applications** by *Yuedong Wang* (Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 2011, Taylor & Francis).

Analysis of Bone Growth Data by Mixed-Effects Smoothing Spline ANOVA Methods

- The data describes bone growth after amputation in digits of mice. The data was acquired by the Muneoka Lab at Tulane University department of cell and molecular biology.
- We have five sets of data corresponding to different treatments of the bone, and one data set corresponding to the control group.
- There is a phase of bone loss followed by a phase of bone growth. The change point occurs approximately between days 10 and 14.

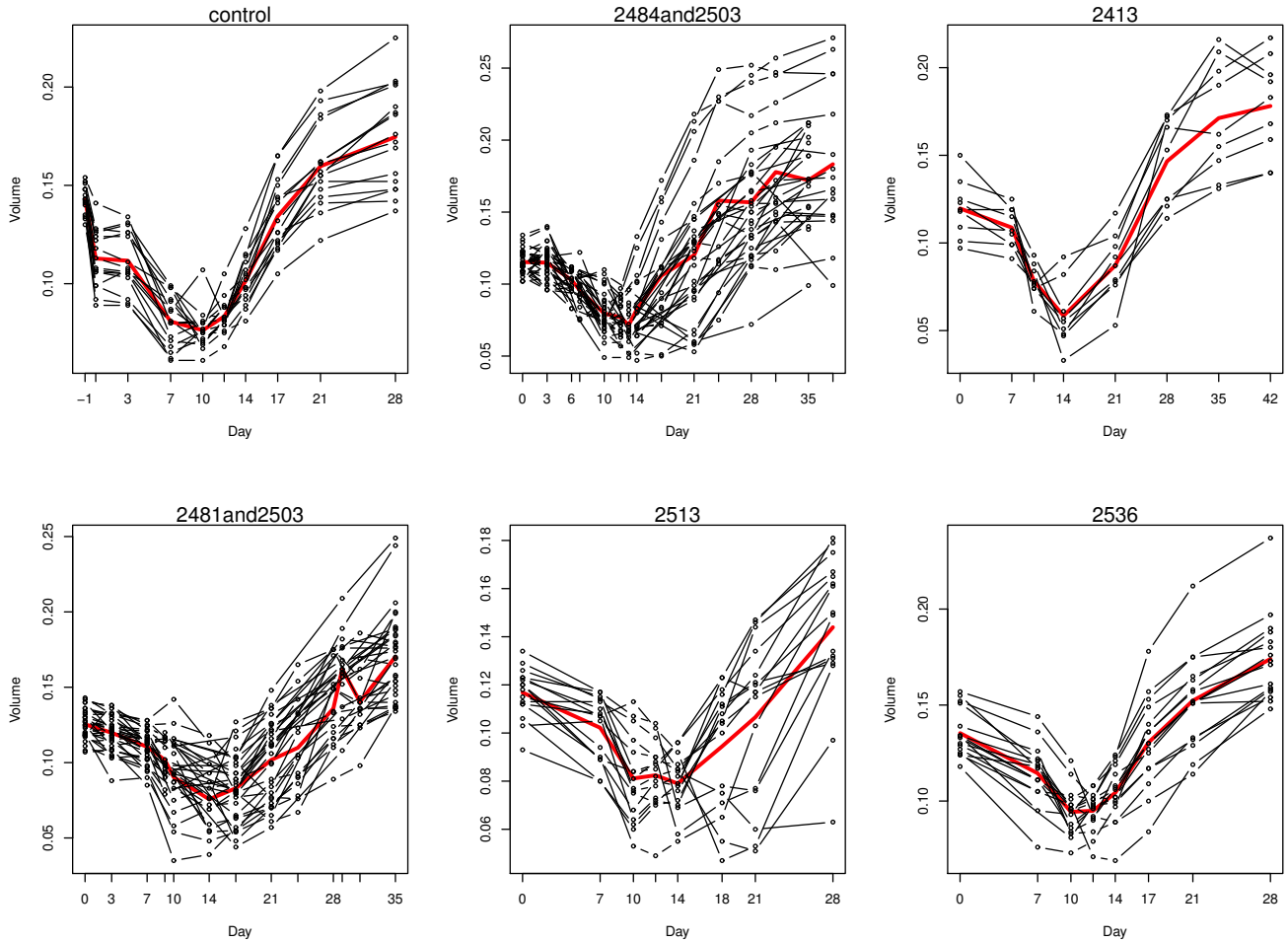


Figure: Bone growth after amputation in digits of mice. Each plot corresponds to a different treatment group. Thick red lines represent the pointwise mean values. Thin lines connect the observations corresponding to the same digit of the same mouse.

The Model

We build a separate model for each treatment group.

- The treatment group is designated by $j = 1$, the control group is designated by $j = 0$.
- Mice $m \in \{1, \ldots, n_m\}$.
- Digits $d \in \{1, \ldots, n_d\}$.
- Observation taken at times $t_k \in [0, 1]$ where $k \in \{1, \ldots, n_{md}\}$.

The Model

- Bone volume y_{mdjk} corresponding to treatment j, mouse m, digit d, at time t_k,

$$\begin{aligned} y_{mdjk} &= f_0(t_k) \cdot I_{[j=0]}(j) + f_1(t_k) \cdot I_{[j=1]}(j) + z_{jk}^T b_{md} + \varepsilon_{mdjk} \\ &= f_0(t_k) + (f_1(t_k) - f_0(t_k)) \cdot I_{[j=1]}(j) + z_{jk}^T b_{md} + \varepsilon_{mdjk}, \end{aligned}$$

- f_0 and f_1 are cubic smoothing splines defined on $[0, 1]$.
- b_{md} is a vector of random effects for digit d of mouse m, which are uncorrelated, homoscedastic and normally distributed (random intercept and slope terms for each mouse and for each digit).
- $\varepsilon_{mdjk} \sim \mathrm{N}(0, \sigma^2)$ (i.i.d.) are the random error terms, which are independent of the random effect terms.

Results: Treatment Group "2413"

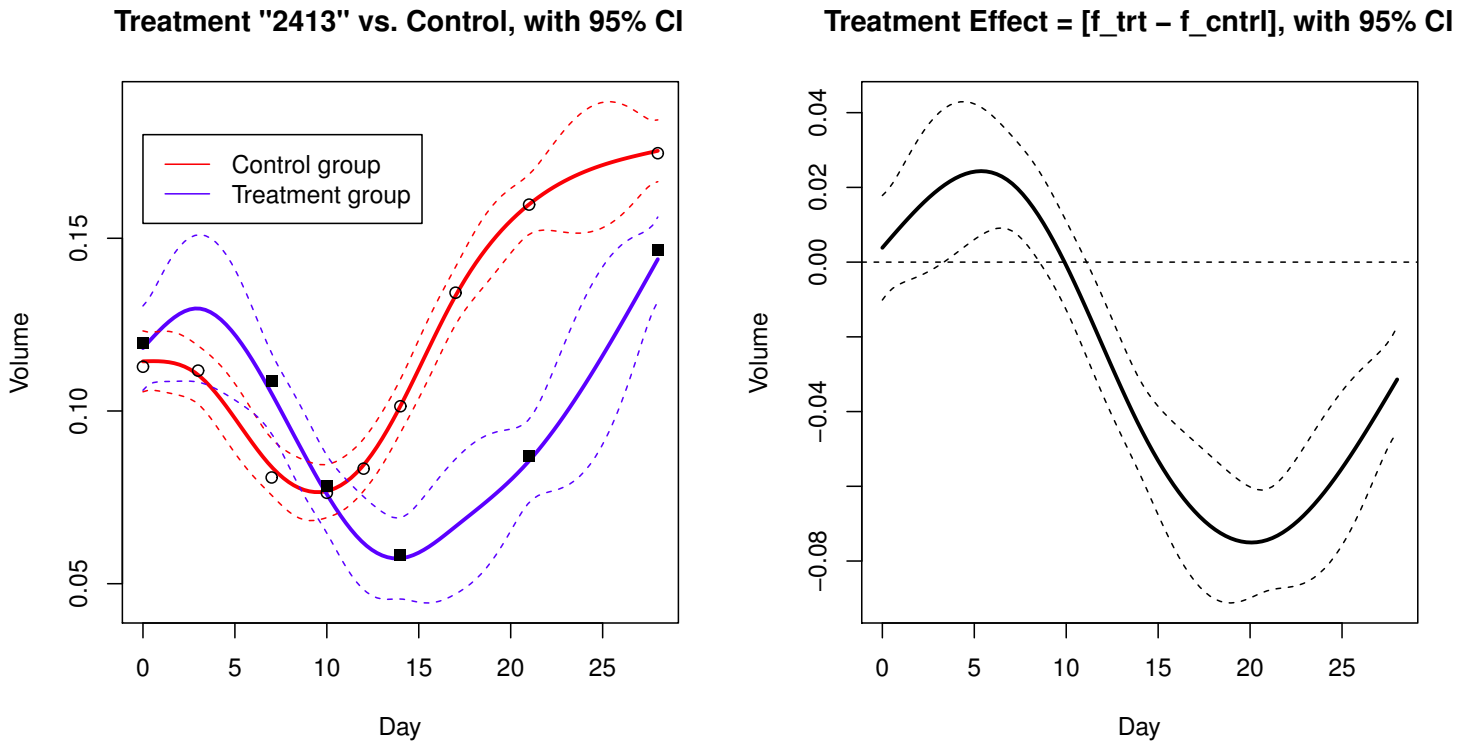


Figure: Bone recovery is significantly slower for the treatment group. The bone volume on the last day of observation is significantly lower for the treatment group.

Results: Treatment Group "2481 & 2503"

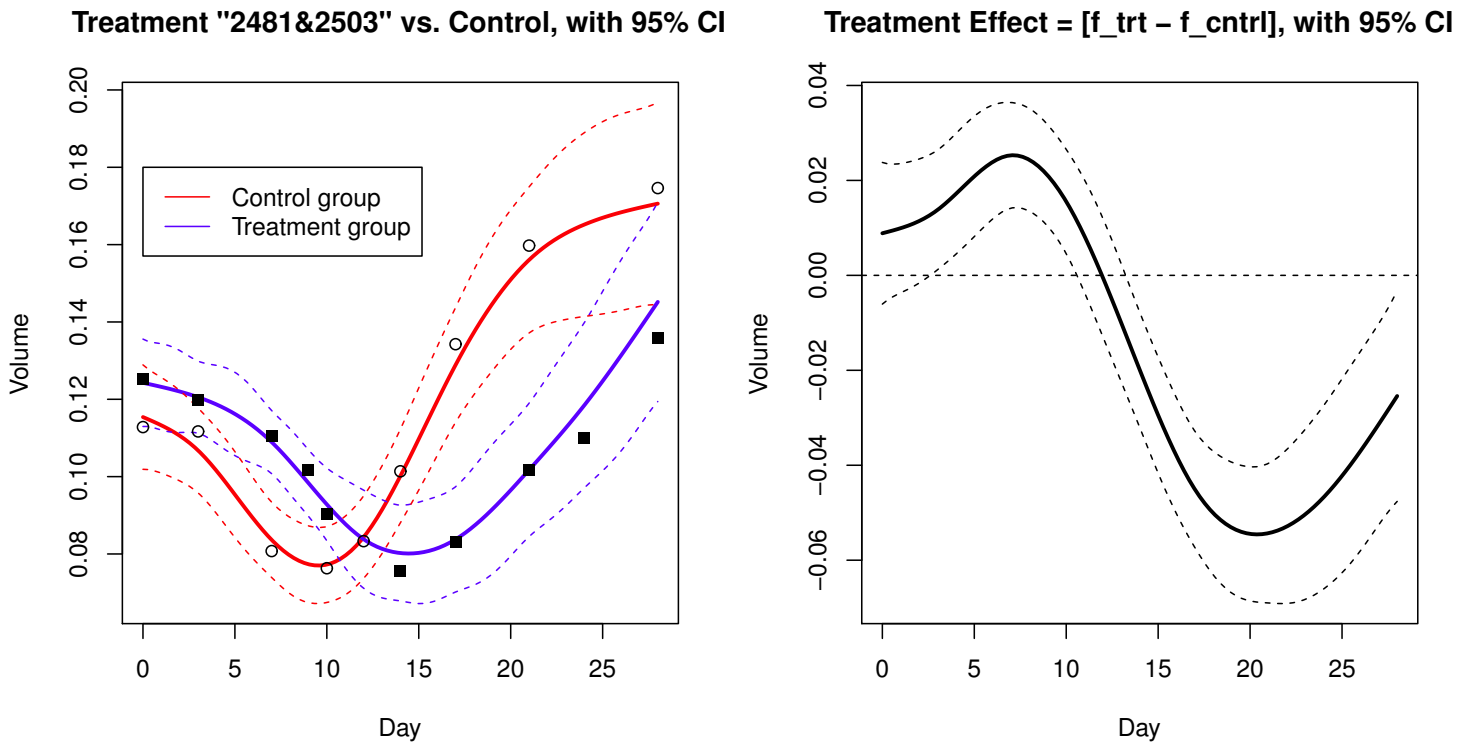


Figure: Bone decay is significantly slower for the treatment group. Bone recovery is significantly slower for the treatment group. The bone volume on the final day of observation is significantly lower for the treatment group.

Results: Treatment Group "2484 & 2503"

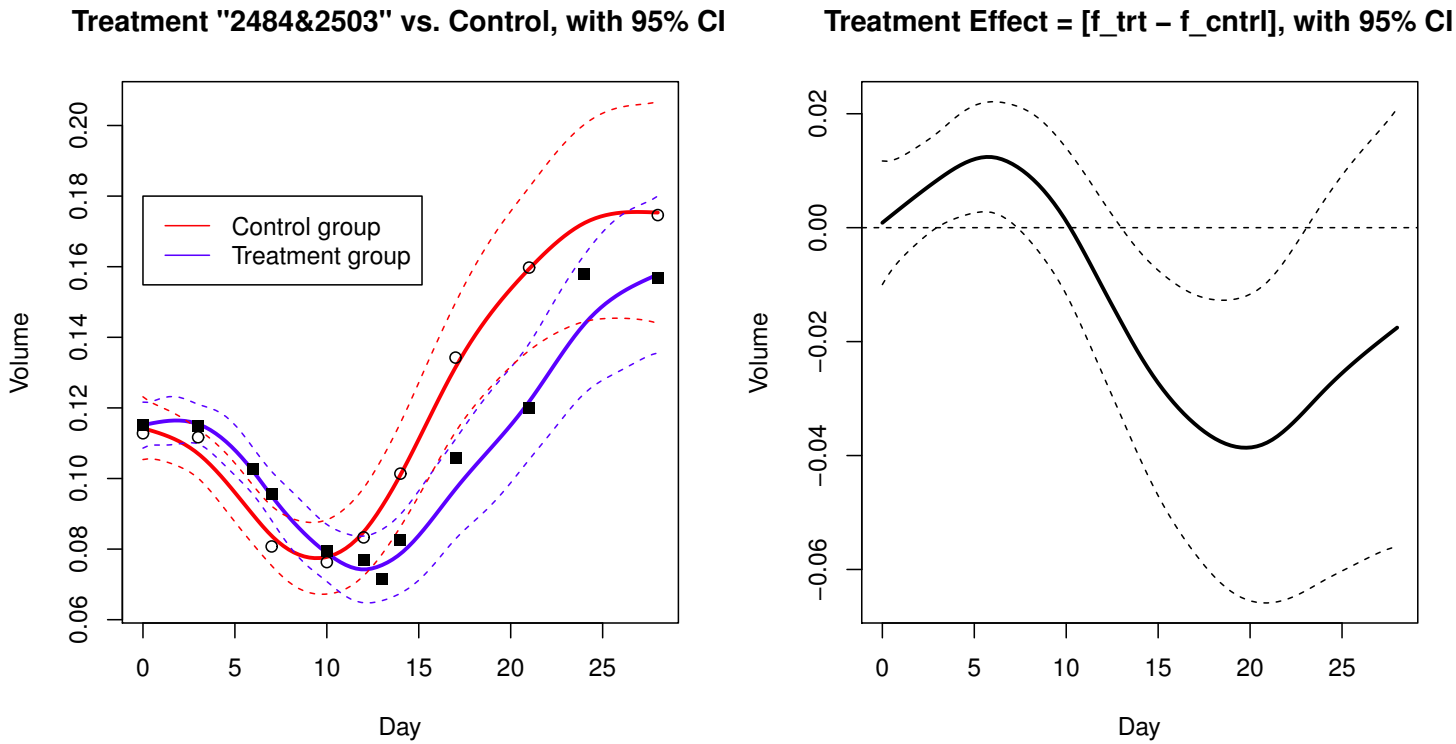


Figure: The bone decay process, as well as the bone recovery process, is significantly slower for the treatment group than it is for the control group.

Results: Treatment Group "2513"

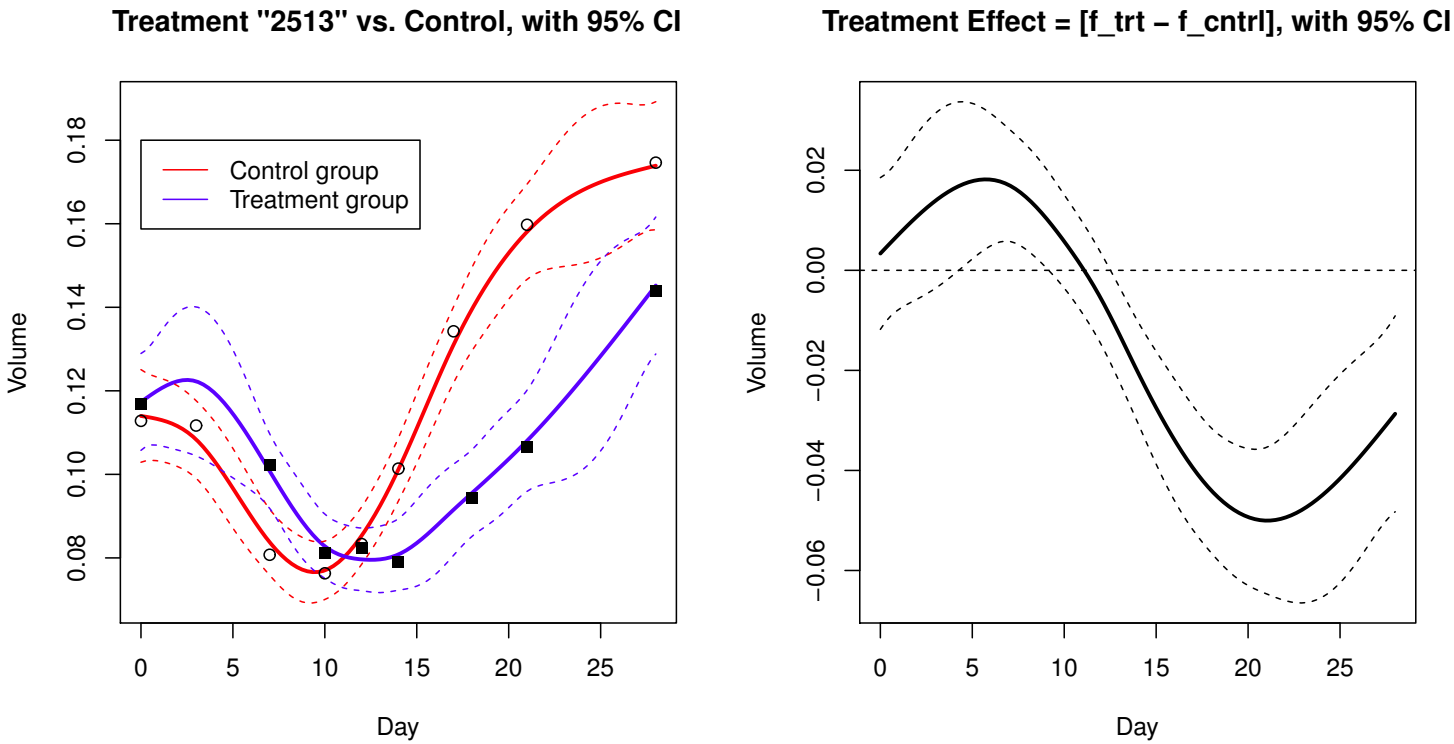


Figure: Bone recovery process is significantly slower for the treatment group. Bone volume on the final day of observation is significantly lower for the treatment group.

Results: Treatment Group "2536"

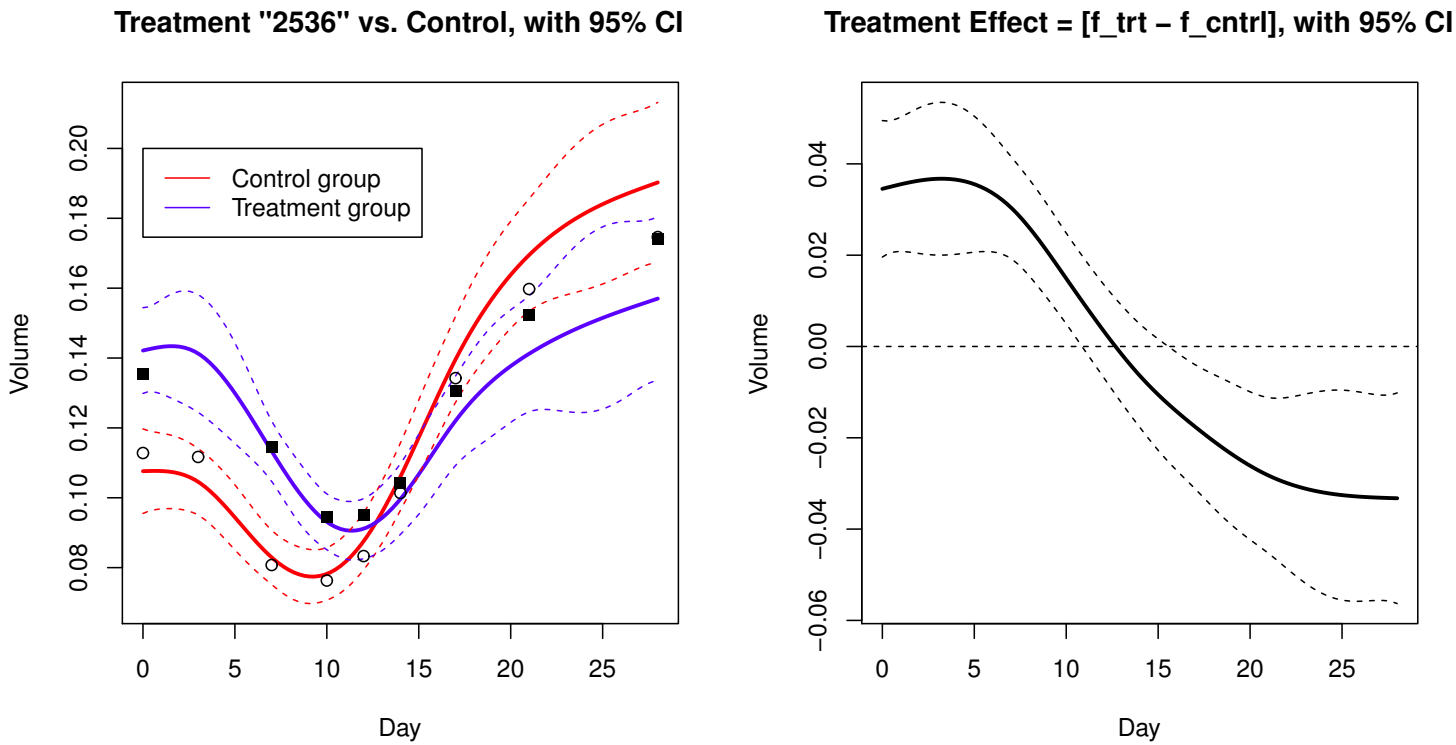


Figure: Bone recovery is significantly slower for the treatment group. The bone volume of the treatment group is significantly higher than that of the control group on the first day of observation, but significantly lower on the last day of observation.

Normality Assumptions

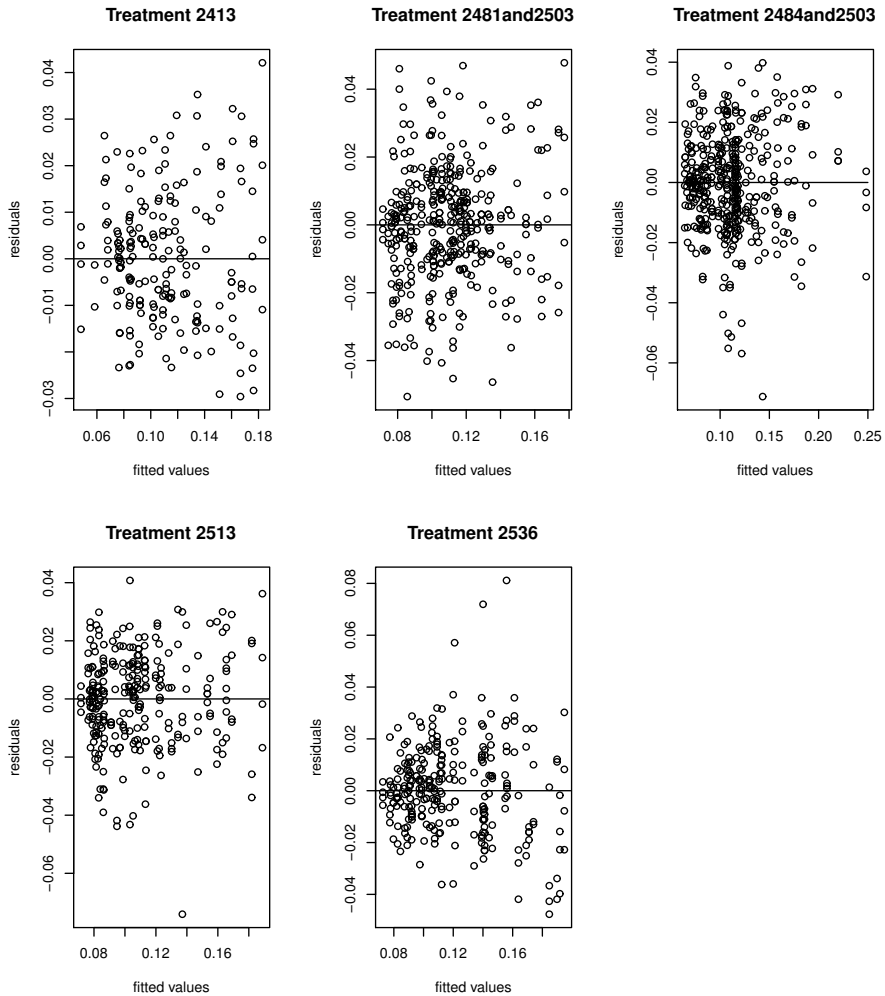


Figure: The fitted values plotted against the residuals.

Normality Assumptions

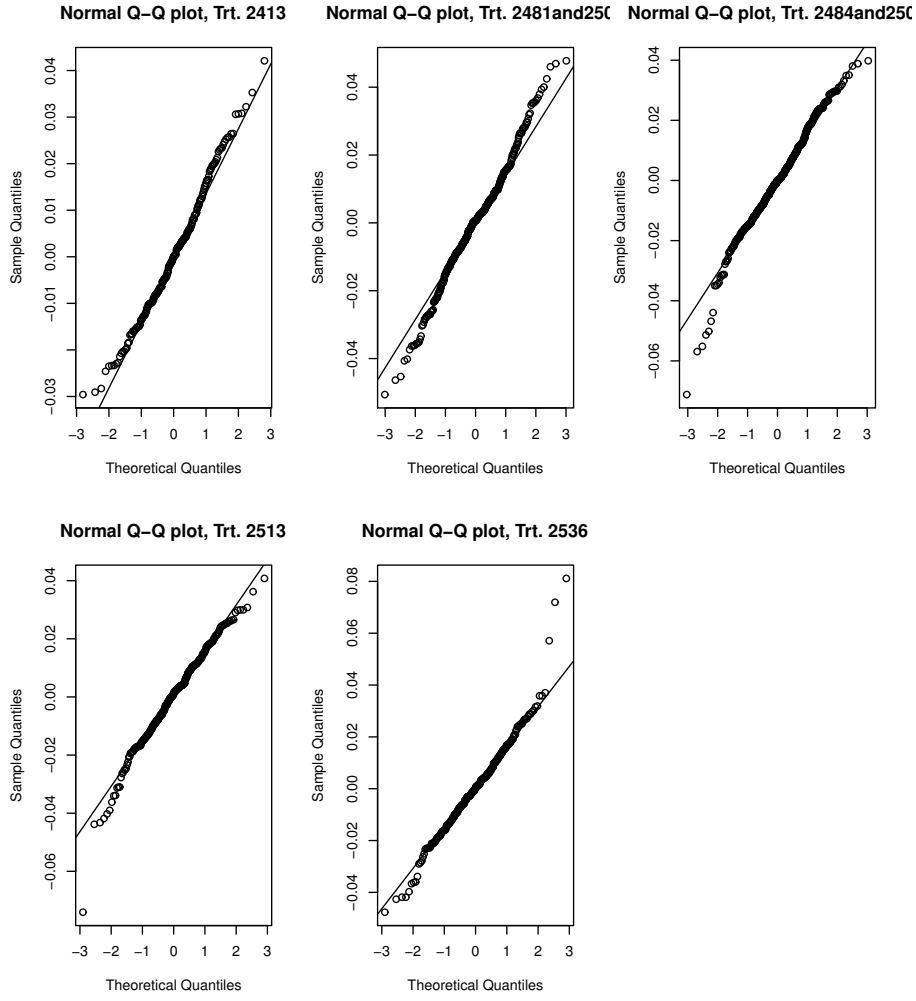


Figure: Normal QQ plots of the residuals.

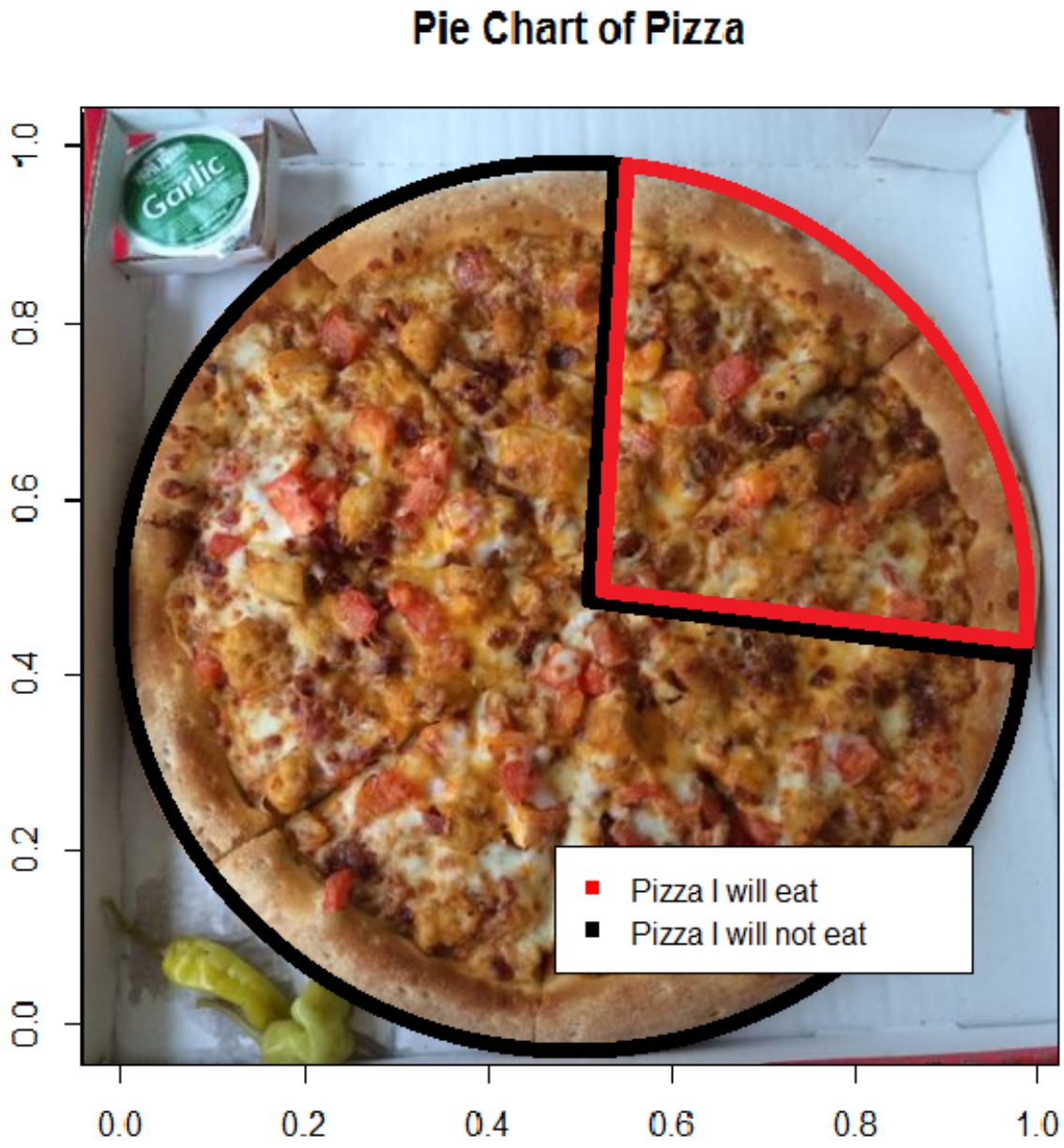


Figure: A joke about statistics and pizza.