

# Flexible Regression

Session 1 - Introduction and Nonparametric Regression  
(Chs 1 & 2)

Electronic notes available at:

[https://warwick.ac.uk/fac/sci/statistics/apts/  
students/resources/](https://warwick.ac.uk/fac/sci/statistics/apts/students/resources/)

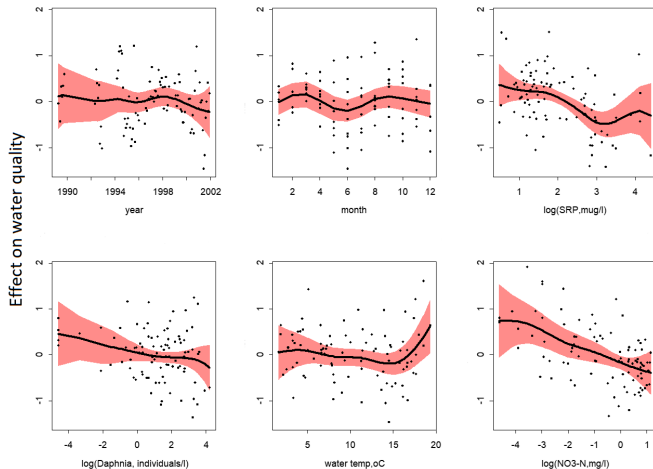
Slides available at:

[https://www.stats.gla.ac.uk/~claire/APTS\\_FR\\_  
session\\_1.pdf](https://www.stats.gla.ac.uk/~claire/APTS_FR_session_1.pdf)

Claire Miller & Tereza Neocleous

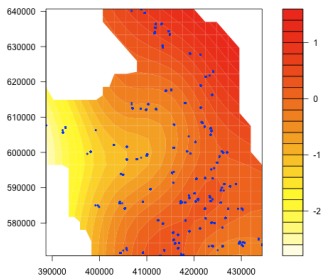
# What this course is about?

Fitting **smooth relationships** to describe drivers of water quality.



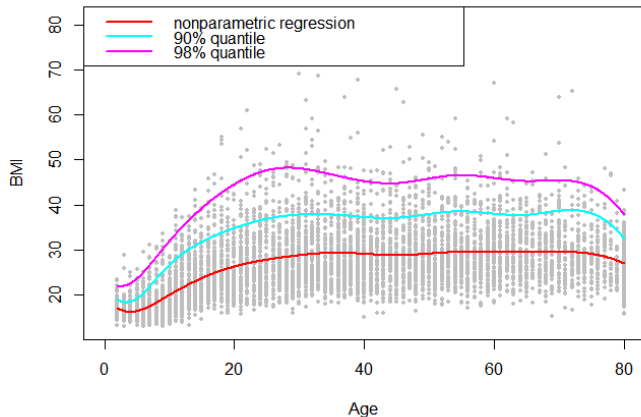
# What this course is about?

Fitting **smooth surfaces** to explain spatiotemporal variation in river nutrient levels.



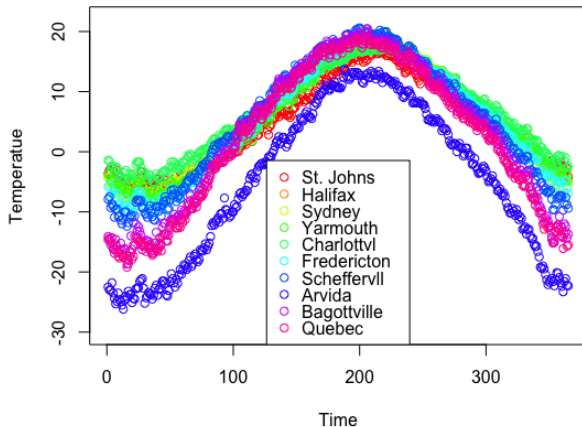
# What this course is about?

Fitting **relationships to appropriate quantiles** to explain BMI using age.



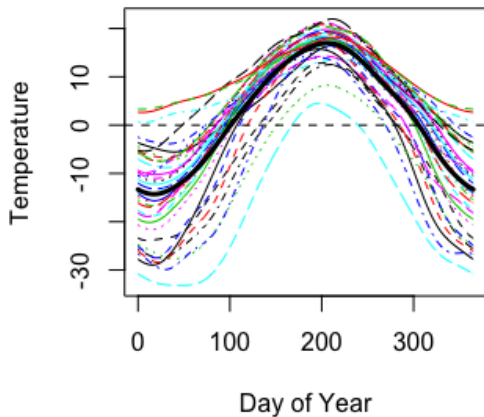
# What this course is about?

Extending the previous approaches to summarise temperature across multiple seasonal patterns.



## What this course is about?

Extending the previous approaches to **summarise temperature across multiple seasonal functions.**



# What this course is about? - Flexible Regression

- ▶ **flexibility in the mean:**

$$Y_i = f(x_i, \beta) + \varepsilon_i.$$

minimise

$$\sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2.$$

- ▶ **flexibility in the response quantile:**

e.g. median regression minimise

$$\sum_{i=1}^n |y_i - f(\mathbf{x}_i)|$$

# What this course is about? - Flexible Regression

- ▶ **flexibility in the mean:**

- ▶ Nonparametric regression - Chapter 2 (Session 1);
- ▶ (Generalised) Additive Models (GAMs) - Chapter 4 (Session 3)

- ▶ **flexibility in the response quantile:**

- ▶ Quantile regression - Chapter 3 (Session 2);
- ▶ (Generalised) additive quantile regression - Chapter 5 (Session 2 & 4)



# What this course is about? - Flexible Regression

To help illustrate the ideas we'll also have 2 labs:

- ▶ **Lab 1 (Tues)** - nonparametric regression/quantile regression with splines
- ▶ **Lab 2 (Thurs)** - (Generalised) Additive (quantile) Models (GAMs)

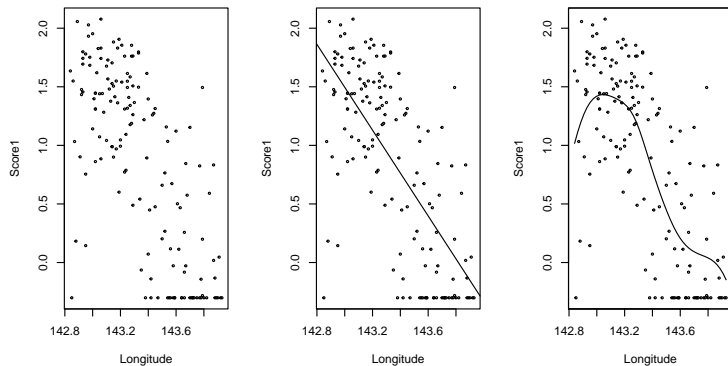
## Chapter 2 - Session motivation

### Example 2.1 Great Barrier Reef data

Zone	an indicator for the closed (1) and open (0) zones
Year	an indicator of 1992 (0) or 1993 (1)
Latitude	latitude of the sampling position
Longitude	longitude of the sampling position
Depth	bottom depth
Score1	catch score 1
Score2	catch score 2

## Chapter 2 - Session motivation

**Figure 2.1: Great Barrier Reef data**



# What we will be discussing in this session?

## Nonparametric regression

- ▶ Approaches for nonparametric regression
- ▶ Properties of smooth functions
- ▶ Why use splines?
- ▶ How to construct splines in 1D?
- ▶ Penalty-based approaches

## 2.1 Nonparametric regression

A simple **nonparametric regression model** has the form

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n$$

where the data  $(x_i, y_i)$  are described by a smooth curve  $f$  plus independent errors  $\varepsilon_j$ .

**Smoothing** is used to estimate  $f()$ .

## 2.1 Nonparametric regression

**Smoothers** have two main uses:

**Description** - to aid 'visually' in the exploration of a relationship or pattern.

**Estimation** - to estimate the dependence of the mean of  $Y$  on the predictor  $x$ .

## 2.1 Nonparametric regression

The two key questions that arise regarding the definition of a smoother are:

- ▶ Which **smoothing method** should be used?
- ▶ What **level of smoothing** is appropriate?

## 2.1 Nonparametric regression

Which **smoothing method** should be used?

- ▶ local fitting approaches;
- ▶ spline based methods.

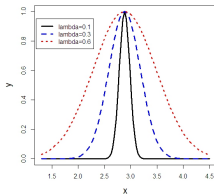
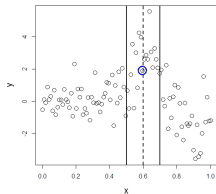


## 2.2 A local fitting approach

For example, **local linear regression**. Solve the least squares problem:

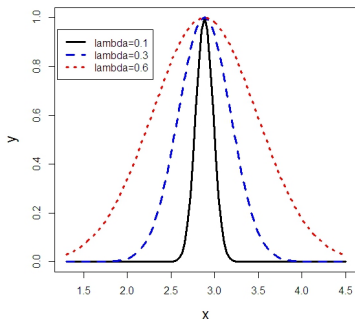
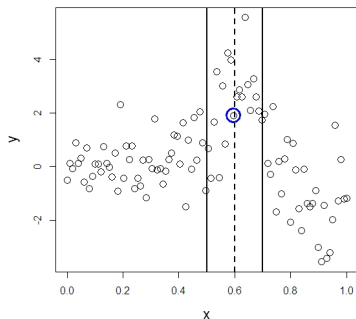
$$\min_{\alpha, \beta} \sum_{i=1}^n \{y_i - \alpha - \beta(x_i - x)\}^2 w(x_i - x; h)$$

and take as the estimate at  $x$  the value of  $\hat{\alpha}$ , as this defines the position of the local regression line at the point  $x$ . The weight function,  $w(x_i - x; h)$ , is a kernel function (see preliminary material).



## 2.2 A local fitting approach

$$\min_{\alpha, \beta} \sum_{i=1}^n \{y_i - \alpha - \beta(x_i - x)\}^2 w(x_i - x; h)$$

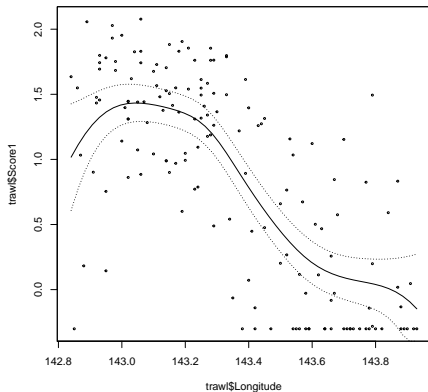


## 2.2.2 A local fitting approach - properties

- ▶ Sometimes computational or practical reasons can constrain our choice of smoother.
- ▶ Expressions for **bias and variance** (derived in section 2.2.1) can help us to choose between smoothing approaches.
- ▶ There is a trade-off between following the data closely (low bias, possibly large variance) and obtaining a smooth function (low variance, possibly large bias).

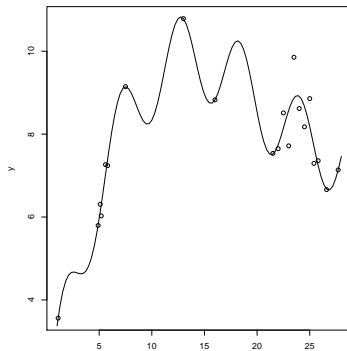
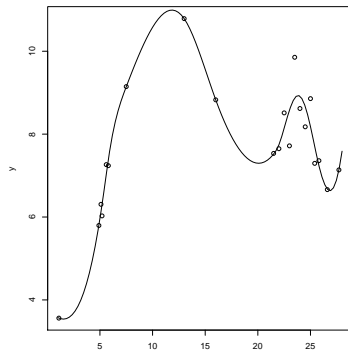
## 2.2.3 Local linear regression in R

A **local linear regression fit** for the Reef data can be obtained using the R library `sm`.



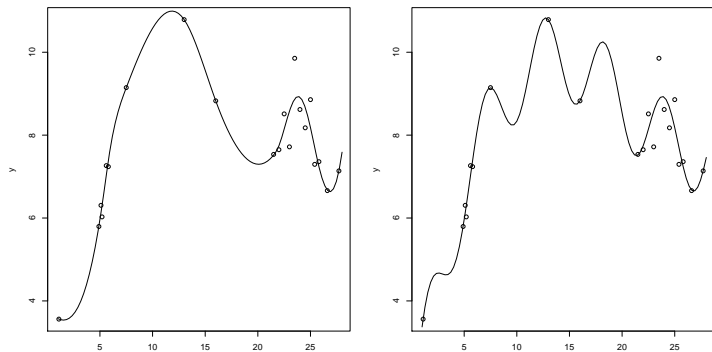
## 2.3 Regression splines

**Example 2.3 - Which function fits the data better?** - bias versus variance (Figure 2.4)



## 2.3 Regression splines

**Example 2.3 - Which function fits the data better?** - bias versus variance (Figure 2.4)



Both have the same fitted values  $\hat{y}_i = \hat{f}(x_i)$ !

## 2.3 Regression splines

### Example 2.3 - Which can we learn from this example?

- ▶ Difficult to do smoothing without knowing / understanding the context.
- ▶ Family of smooth functions too rich to be able to only rely on the data.
- ▶ Alternatives to local fitting approaches
  - ▶ Splines based on truncated power series and B-splines
  - ▶ Penalties or Bayesian approaches to penalise wiggleness

## 2.3.1 Regression splines - polynomial regression

### Linear regression

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i \quad \text{for } i = 1, \dots, n,$$

In matrix-vector notation:

$$\mathbb{E}(\mathbf{y}) = \mathbf{B}\boldsymbol{\beta} \quad \text{with } \mathbf{y} = (Y_1, \dots, Y_n)^\top \text{ and } \mathbf{B} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

$\mathbf{B}$  - matrix of basis functions

$\boldsymbol{\beta}$  - vector of basis coefficients

Basis functions:  $B_0(x) = 1, B_1(x) = x.$



## 2.3.1 Regression splines - polynomial regression

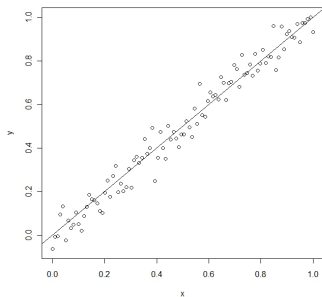


Figure: 2.5 A simple linear regression line with underlying simulated data

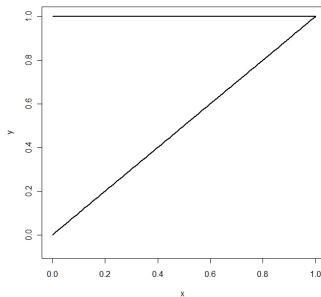


Figure: 2.6 The basis functions for simple linear regression  $1, x$

## 2.3.1 Regression splines - polynomial regression

### Polynomial regression

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i + \dots + \beta_r x_i^r \quad \text{for } i = 1, \dots, n,$$

just corresponds to

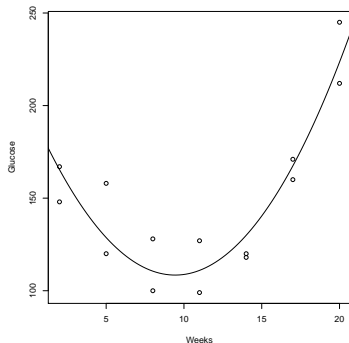
$$\mathbf{B} = \begin{pmatrix} 1 & x_1 & \dots & x_1^r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^r \end{pmatrix}.$$

Polynomial regression is a basis expansion technique.

$$\hat{\boldsymbol{\beta}} = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{y}$$

## 2.3.1 Regression splines - polynomial regression

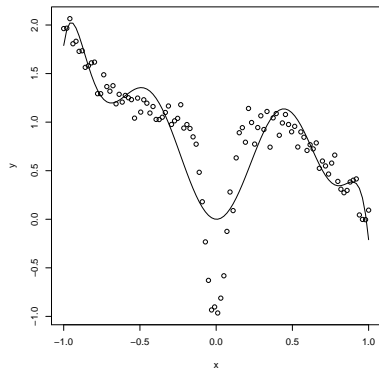
### Example 2.4: Glucose levels in potatoes



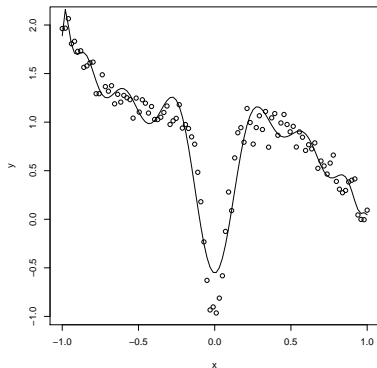
Polynomial regression can be a useful tool for small datasets.  
(if a small degree of polynomial is used)

## 2.3.1 Regression splines - polynomial regression

**Simulated example 2.5** :  $y_i = 1 - x_i^3 - 2 \exp(-100x_i^2) + \varepsilon_i$  with  $\mathbf{x} = (-1, -0.98, \dots, 0.98, 1)$  and  $\varepsilon_i \sim N(0, 0.1^2)$ .



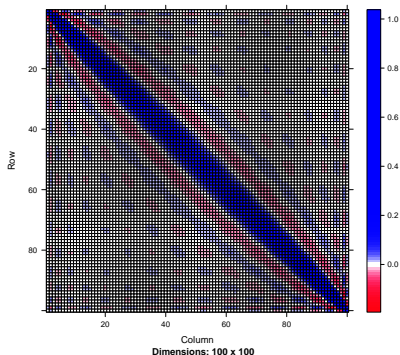
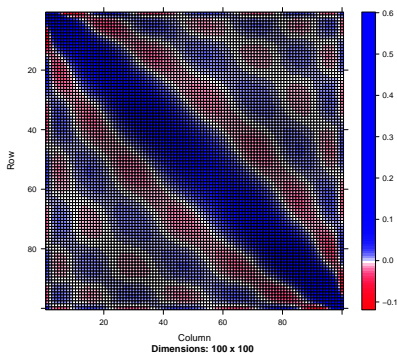
Polynomial regression (degree  $r = 10$ )



Polynomial regression (degree  $r = 17$ )

## 2.3.1 Regression splines - polynomial regression - what is going wrong?

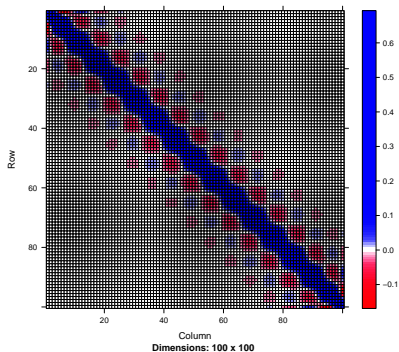
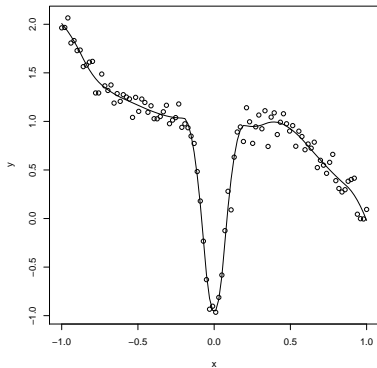
Let's look at the hat matrix (Figure 2.9):  $\mathbf{S} = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top$   
( $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$ )



Polynomial regression (degree  $r = 10$ )

Polynomial regression (degree  $r = 17$ )

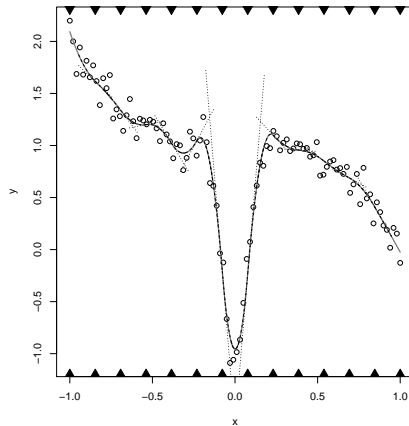
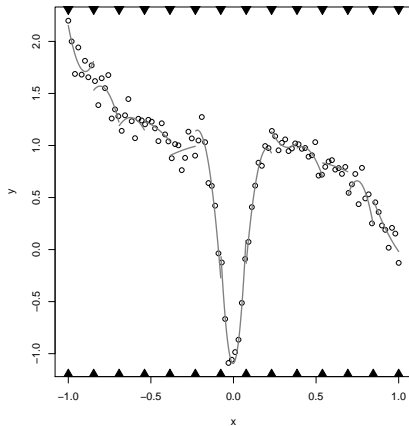
## 2.3.1 Regression splines: spline-based model



## 2.3.1 Regression splines - polynomial regression - problems

- ▶ Polynomials are *not* a local model.  
↪ Oscillations (Runge's phenomenon), "derivative propagation"
- ▶ Likely to produce spurious edge effects on both ends of range. (polynomial has to diverge to  $\pm\infty$  as  $x \rightarrow \pm\infty$ )
- ▶ Also very likely to be numerically unstable.  
In our example:
  - ▶ Condition number of  $\mathbf{B}^T \mathbf{B}$ :  $1.56 \times 10^{12}$
  - ▶ For a B-spline: Condition number of  $\mathbf{B}^T \mathbf{B}$ : 32.49
- ▶ Possible solution: Use piecewise polynomial functions.

## 2.3.1 Regression splines - piecewise polynomials



Maybe we need to “glue” the polynomials together a bit  $\rightsquigarrow$  splines (Figure 2.10).



## 2.3.2 Regression splines

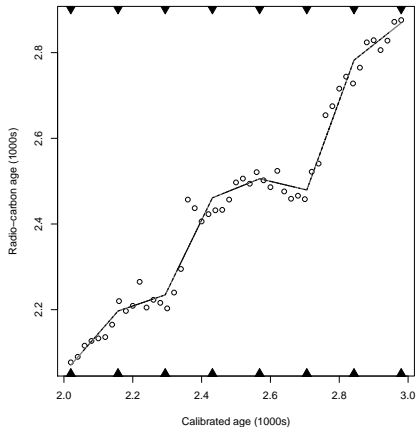
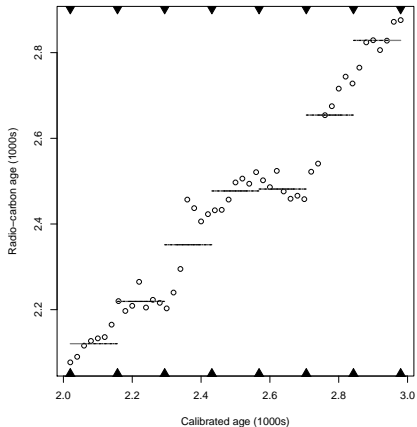
### Definition 2.1: (Polynomial) spline

Given a set of knots  $a = \kappa_1 < \kappa_2 < \dots < \kappa_l = b$ , a function  $f : [a, b] \rightarrow \mathbb{R}$  is called a (*polynomial*) *spline* of degree  $r$  if

- ▶  $f(\cdot)$  is a polynomial of degree  $r$  on each interval  $(\kappa_j, \kappa_{j+1})$  ( $j = 1, \dots, l - 1$ ).
- ▶  $f(\cdot)$  is  $r - 1$  times continuously differentiable.

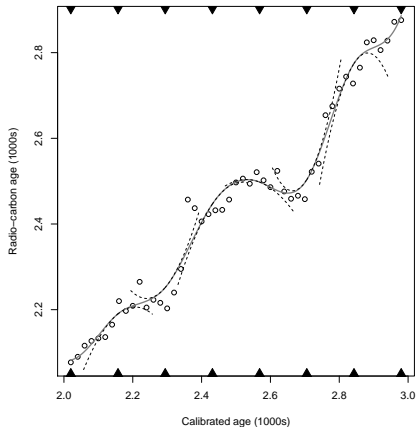
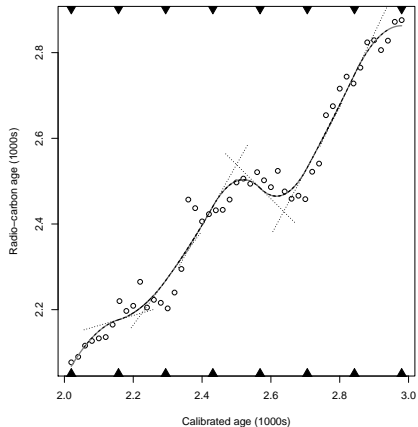
## 2.3.2 Regression splines: degrees $r = 0$ and $r = 1$

### Radiocarbon dating (Figure 2.12)



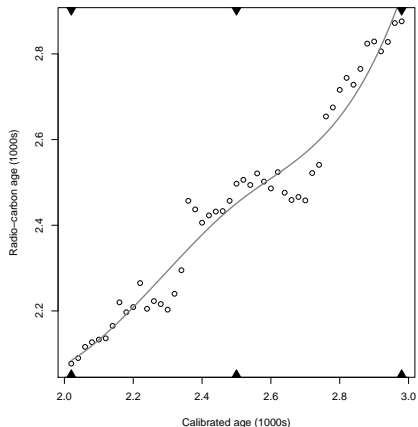
## 2.3.2 Regression splines: degrees $r = 2$ and $r = 3$

### Radiocarbon dating

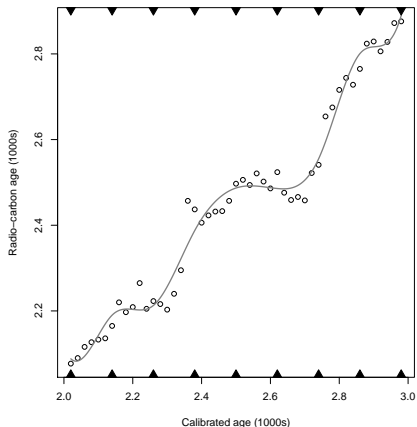


## 2.3.2 Regression splines: different numbers of knots (1)

### Radiocarbon dating (Figure 2.13)



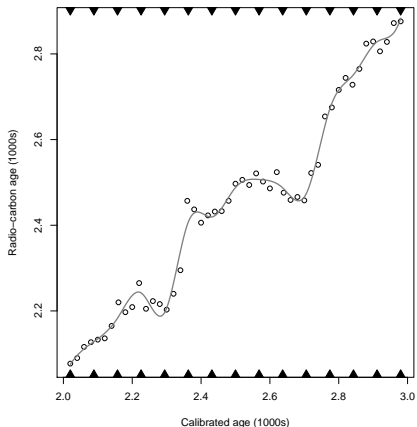
$l = 3$  knots



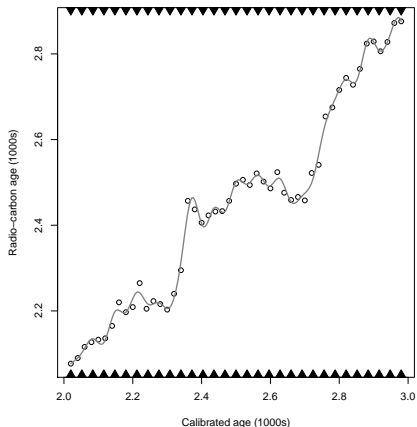
$l = 9$  knots

## 2.3.2 Regression splines: different numbers of knots (2)

### Radiocarbon dating



$l = 15$  knots



$l = 31$  knots

## 2.3.2 Regression splines

### Choice of degree and number of knots

#### Choice of degree $r$

- ▶ Degree  $r$  controls smoothness / differentiability
- ▶ The larger the degree  $r$  the more the spline behaves like a polynomial.
- ▶ Rarely necessary to go beyond  $r = 3$ .

#### Choice of number of knots $l$

- ▶ Number of knots  $l$  controls smoothness / flexibility
- ▶ Alternative: Use “too many” knots and control flexibility using roughness penalty (see later)

## 2.3.4 Regression splines - how to fit?

Minimise

$$\sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2.$$

Represent  $f(\mathbf{x}_i)$  as  $\mathbf{B}\beta$ .

**How to construct a basis?**

**B** is formulated through:

- ▶ Truncated power basis;
- ▶ B-splines.

## 2.3.4 Regression splines - truncated power series

### Definition 2.6: Truncated power basis

Given a set of knots  $\kappa_1 < \dots < \kappa_m$  the truncated power basis of degree  $r$  is given by

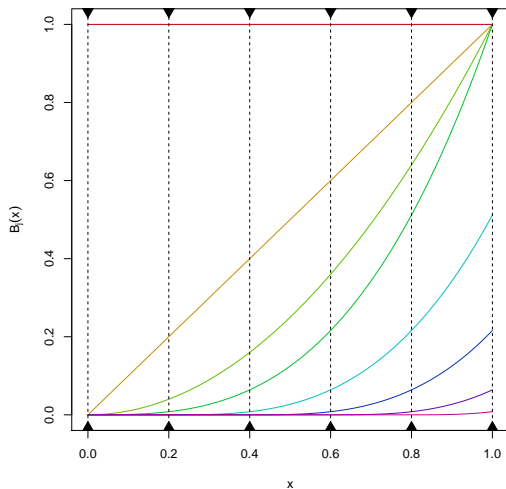
$$(1, x, \dots, x^r, (x - \kappa_1)_+^r, (x - \kappa_2)_+^r, \dots, (x - \kappa_m)_+^r),$$

where  $(z)_+^r = \begin{cases} z^r & \text{for } z > 0 \\ 0 & \text{otherwise.} \end{cases}$



## 2.3.4 Regression splines - truncated power series

### Truncated power series of degree 3 (Figure 2.15)



## 2.3.4 Regression splines - truncated power series

### How to fit a model using the truncated power basis?

- Use basis expansion

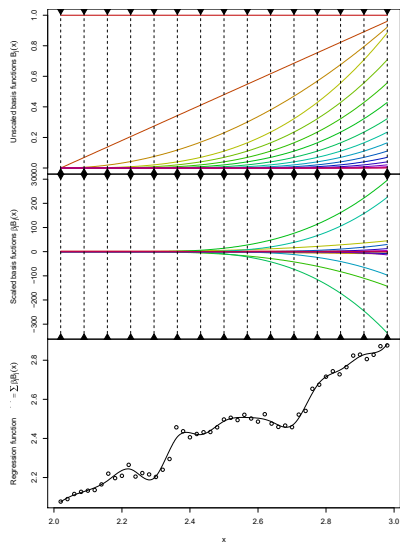
$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_r x^r \\ + \beta_{r+1} (x - \kappa_1)_+^r + \dots + \beta_{r+m} (x - \kappa_m)_+^r$$

- This is just a linear model with design matrix

$$\mathbf{B} = \begin{pmatrix} 1 & x_1 & \dots & x_1^r & (x_1 - \kappa_1)_+^r & \dots & (x_1 - \kappa_m)_+^r \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^r & (x_n - \kappa_1)_+^r & \dots & (x_n - \kappa_m)_+^r \end{pmatrix}$$

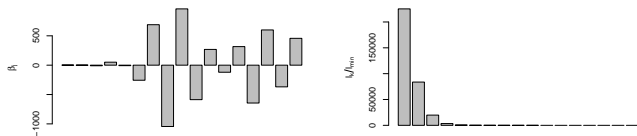
## 2.3.4 Regression splines - truncated power series

### Illustration: Radiocarbon dating (Figure 2.19)



## 2.3.4 Regression splines - truncated power series

### Numerical problems (Figures 2.17, 2.18)



Problem: Basis functions highly correlated (0.99921)  
 $\rightsquigarrow \mathbf{B}^T \mathbf{B}$  ill conditioned (condition number:  $5.85 \times 10^9$ )

## 2.3.4 Regression splines - B-splines

### Definition 2.7: B-spline basis

- (a) Given a set of  $l$  knots the B-spline basis of degree 0 is given by the functions  $(B_1^0(x), \dots, B_{l-1}^0(x))$  with

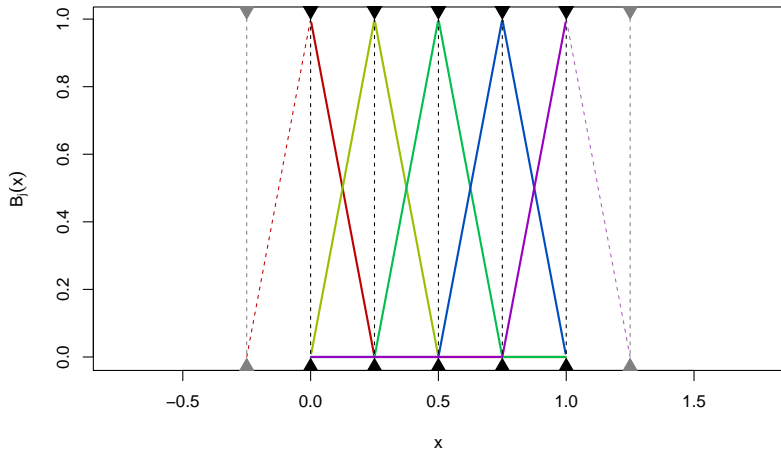
$$B_j^0(x) = \begin{cases} 1 & \text{for } \kappa_j \leq x < \kappa_{j+1} \\ 0 & \text{otherwise.} \end{cases}$$

- (b) Given a set of  $l$  knots the B-spline basis of degree  $r > 0$  is given by the functions  $(B_1^r(x), \dots, B_{l+r-1}^r(x))$  with

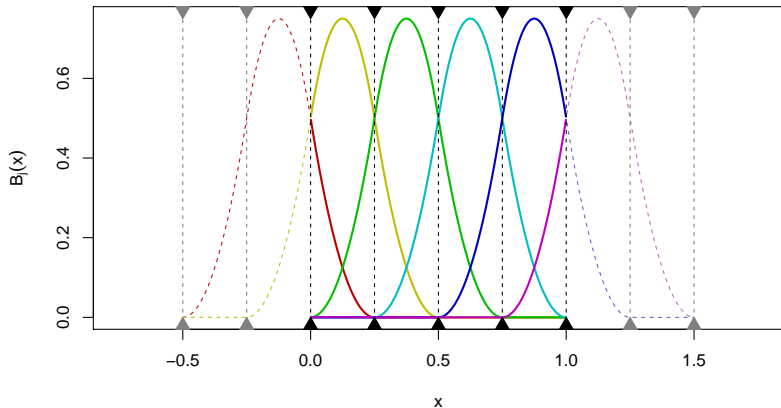
$$B_j^r(x) = \frac{x - \kappa_{j-r}}{\kappa_j - \kappa_{j-r}} B_{j-1}^{r-1}(x) + \frac{\kappa_{j+1} - x}{\kappa_{j+1} - \kappa_{j+1-r}} B_j^{r-1}(x).$$

## 2.3.4 Regression splines: B-spline basis of degree $r = 1$

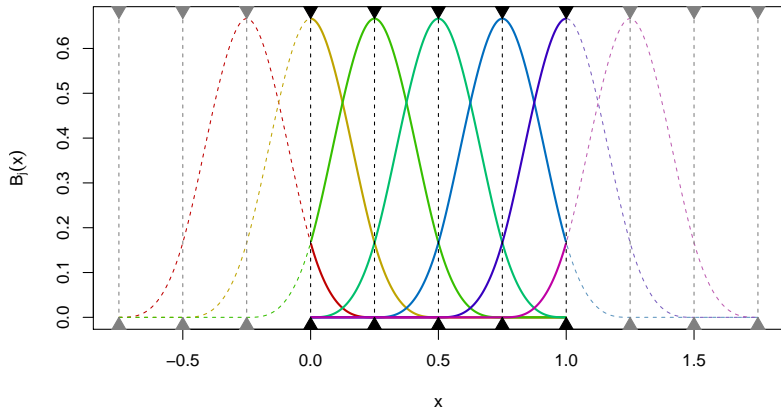
Figure 2.20:



## 2.3.4 Regression splines: B-spline basis of degree $r = 2$



## 2.3.4 Regression splines: B-spline basis of degree $r = 3$





## 2.3.4 Regression splines: B-splines

### Model fitting using B-splines

- Use basis expansion

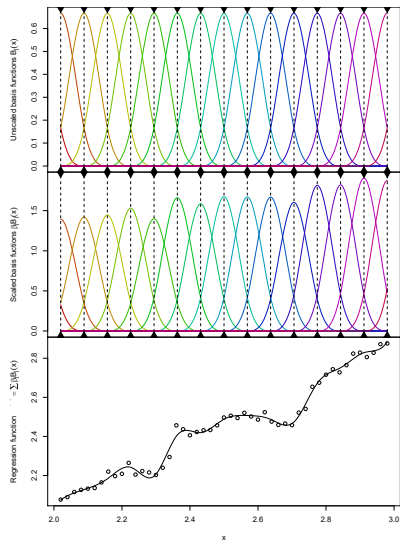
$$f(x) = \sum_{j=1}^{l+r-1} \beta_j B_j(x)$$

- This is just a linear model with design matrix

$$\mathbf{B} = \begin{pmatrix} B_1^r(x_1) & \cdots & B_{l+r-1}^r(x_1) \\ \vdots & \ddots & \vdots \\ B_1^r(x_n) & \cdots & B_{l+r-1}^r(x_n) \end{pmatrix}.$$

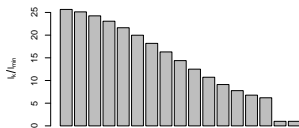
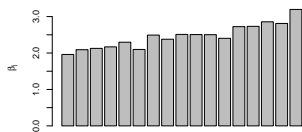
## 2.3.4 Regression splines: B-splines

### Illustration: Radiocarbon dating (Figure 2.24)



## 2.3.4 Regression splines: B-splines

No more numerical problems (Figures 2.22, 2.23)

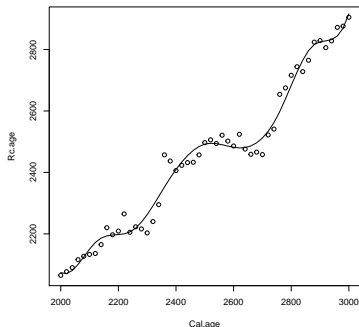


Problem solved: Basis functions not highly correlated (0.8309 at most)

$\rightsquigarrow \mathbf{B}^\top \mathbf{B}$  not ill-conditioned (condition number: 358.263)

## 2.3.4 Regression splines: B-splines in R(Figure 2.21)

```
library(splines)
model <- lm(Rc.age~bs(Cal.age, df=10), data=radiocarbon)
with(radiocarbon, {
  plot(Cal.age, Rc.age)
  lines(Cal.age, predict(model))
})
```



## 2.3.5 Penalised regression splines (P-splines) – Idea

- ▶ Positioning of knots can have large influence of fitted function (especially if number of knots is small);
- ▶ Solution: Use “too many” knots and control flexibility using roughness penalty;
- ▶ We will only consider quadratic roughness penalties of the form  $\|\mathbf{D}\beta\|^2$ .

## 2.3.5 Penalised regression splines (P-splines) – Idea

- ▶ Objective function (\*\*see ridge regression):

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \|\mathbf{D}\boldsymbol{\beta}\|^2$$

- ▶  $\lambda$  controls the trade-off between following the data ( $\lambda \downarrow$ ) and a strongly regularised curve ( $\lambda \uparrow$ ).

## 2.3.5 Penalised regression splines (P-splines) – Solution

- ▶ Solution for P-splines is

$$\hat{\beta} = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{B}^T \mathbf{y},$$

- ▶ Numerically more stable to use a QR decomposition to minimise augmented system

$$\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{B} \\ \sqrt{\lambda} \mathbf{D} \end{pmatrix} \beta \right\|^2$$

## 2.3.6 Penalised regression splines - how to choose **D**?

### Smoothing splines

One can show that

$$\int_a^b f''(x)^2 dx = \beta^\top \begin{pmatrix} \int_a^b B_1''(x)B_1''(x) dx & \dots & \int_a^b B_1''(x)B_{l+r-1}''(x) dx \\ \vdots & \ddots & \vdots \\ \int_a^b B_1''(x)B_{l+r-1}''(x) dx & \dots & \int_a^b B_{l+r-1}''(x)B_{l+r-1}''(x) dx \end{pmatrix} \beta$$

↪ Set  $\mathbf{D}^\top \mathbf{D}$  equal to this matrix of cross-products.

### 2.3.6 Penalised regression splines - difference penalties

For equally-spaced knots we can also use difference penalties (much simpler).

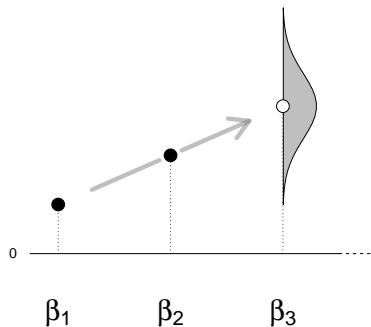


## 2.3.6 Penalised regression splines - second-order difference penalty

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 1 & -2 & 1 \end{pmatrix}.$$

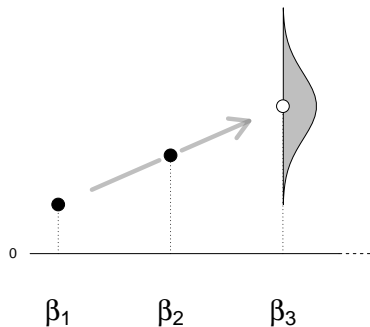
$$\|\mathbf{D}_2\boldsymbol{\beta}\|^2 = \sum(\beta_{j+2} - 2\beta_{j+1} + \beta_j)^2$$

(second-order differences)



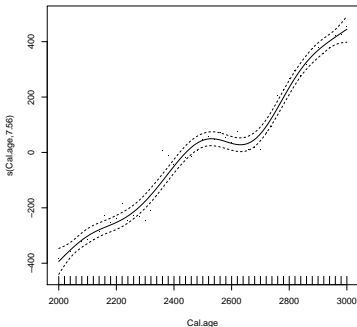
## 2.3.6 Penalised regression splines - second-order difference penalty

- ▶ Shrinks the coefficients towards a linear sequence.  
↪ Shrinks the regression function  $f(\cdot)$  towards linear function.
- ▶ Adding a linear function to  $f(\cdot)$  does not change the penalty.
- ▶ Natural choice for spline basis of degree  $r = 3$ .



## 2.3.7 Penalised regression splines in R

```
library(mgcv)
model <- gam(Rc.age~s(Cal.age), data=radiocarbon)
model
plot(model, residuals=TRUE)
```



# Summary

## Nonparametric regression

- ▶ Approaches for nonparametric regression
- ▶ Properties of smooth functions
- ▶ Why use splines?
- ▶ How to construct splines in 1D? (truncated power and B-splines)
- ▶ Penalty-based approaches (P-splines)