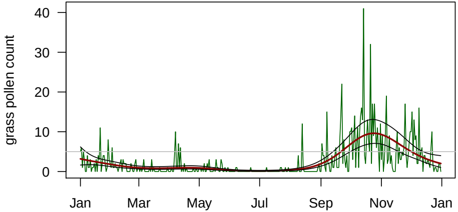
SEEC Toolbox Seminar

Generalized Additive Models

Birgit Erni
September 2018

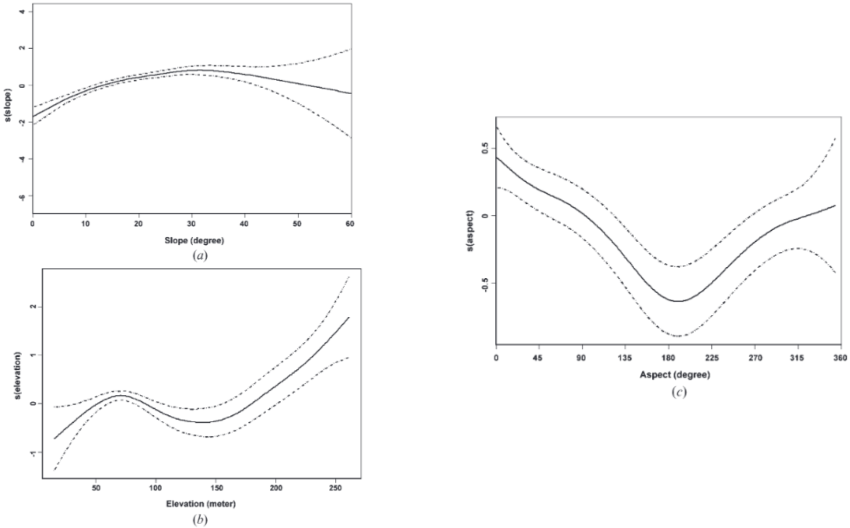


SEEC - Statistics in Ecology, Environment and Conservation

Overview: Introduction to Splines and GAMs

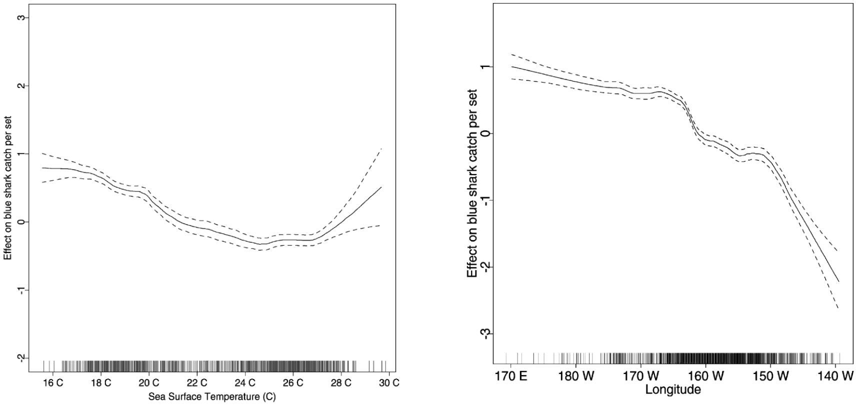
1. motivating examples
2. splines
3. GAMs
4. implementation in mgcv
5. knots, penalites, confidence bands
6. 2-dimensional splines

Landslide susceptibility



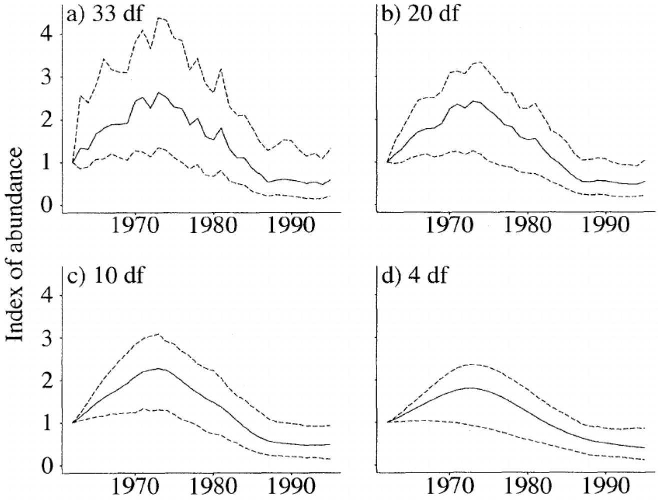
Park & Chi 2008 (International Journal of Remote Sensing)

Shark catch



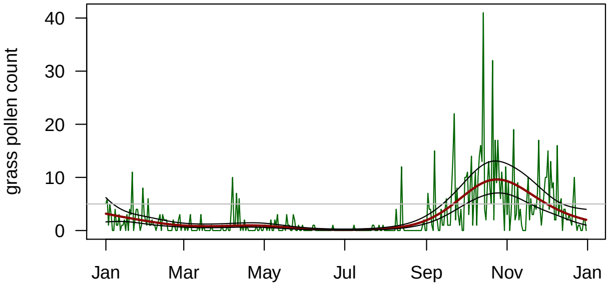
Walsh & Kleiber 2001 (Fisheries Research)

Long-term trends

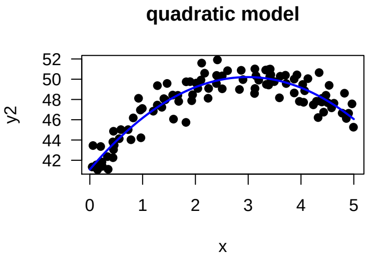


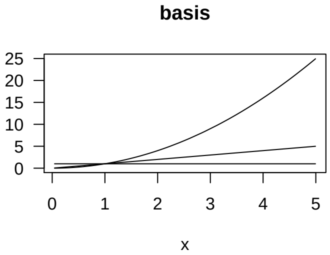
Fewster et al. 2000 (Ecology)

Seasonal effects in time series



Polynomial regression



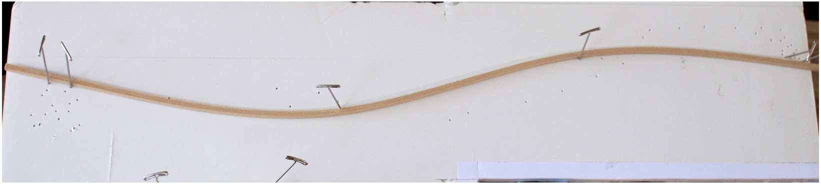


extension of linear model to handle non-linear relationships

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

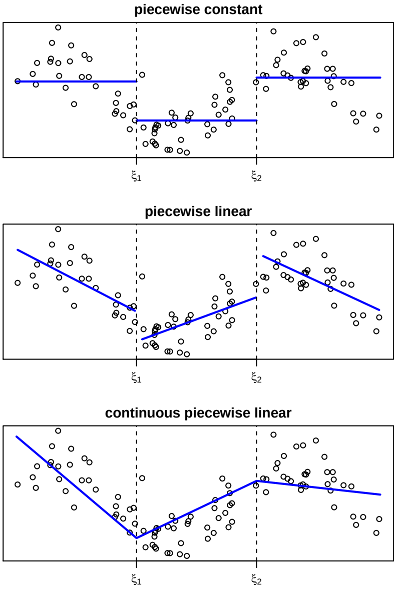
$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

Splines – before computer graphics

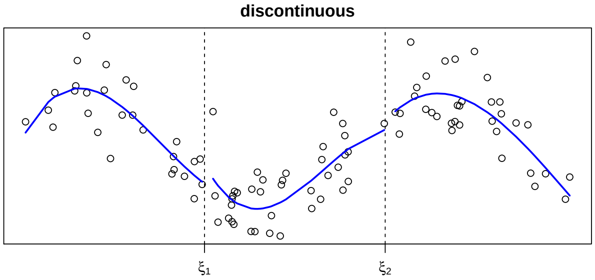


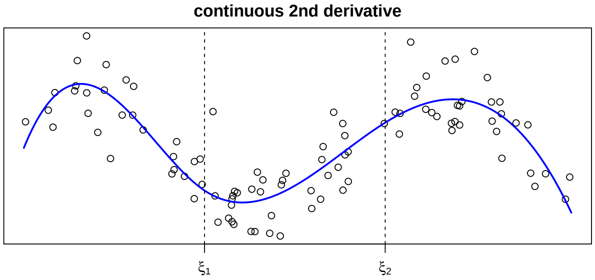
least energy – smoothest curve

Piecewise Polynomials

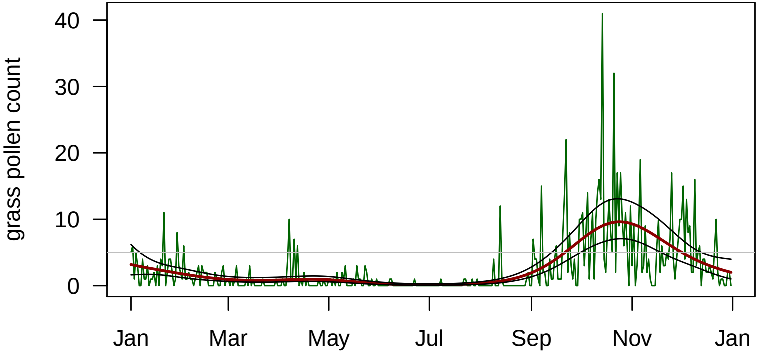


Cubic Polynomials



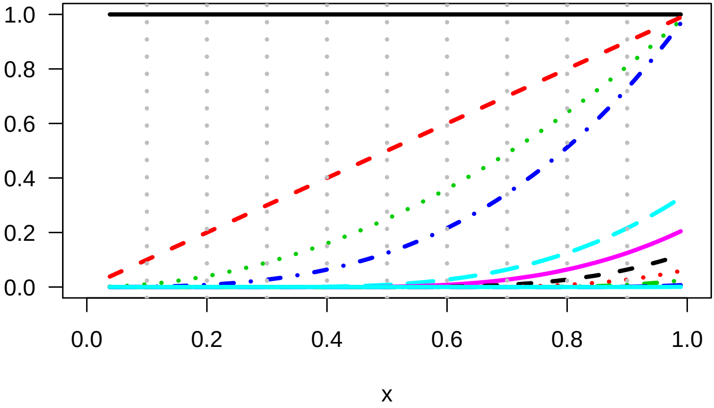


Splines



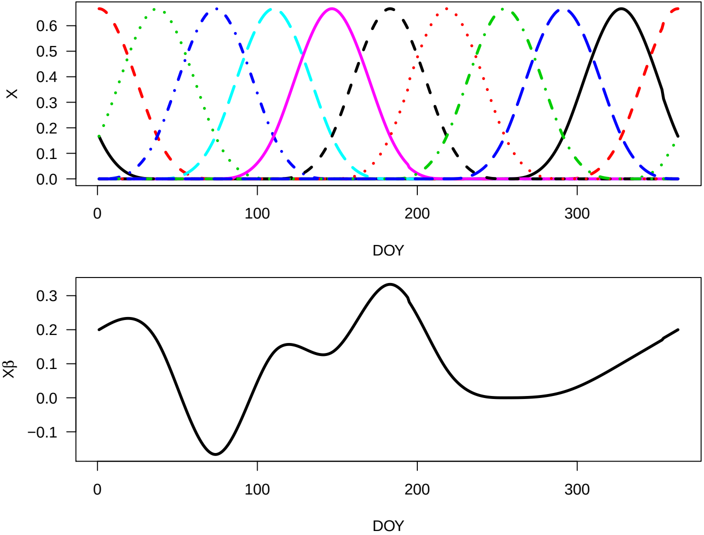
Different types of spline basis functions

cubic spline basis

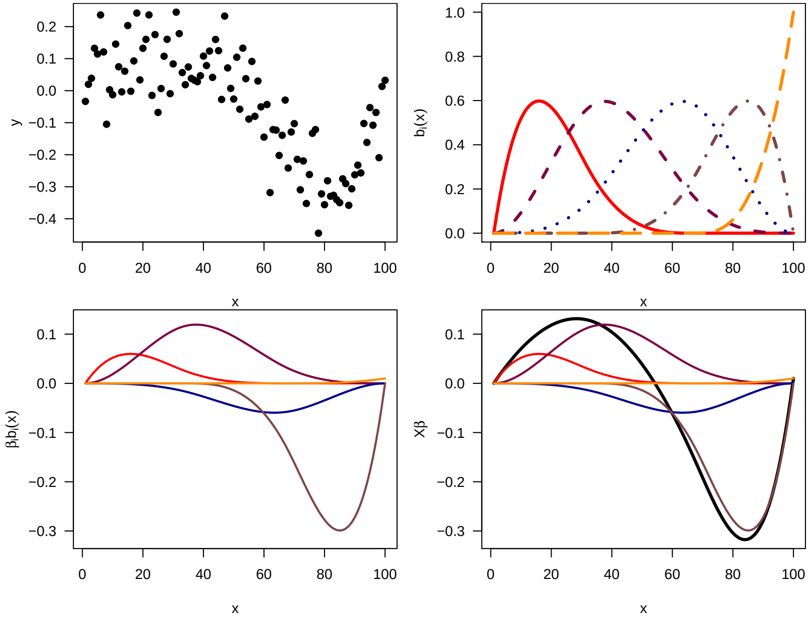


From basis to fitted curve

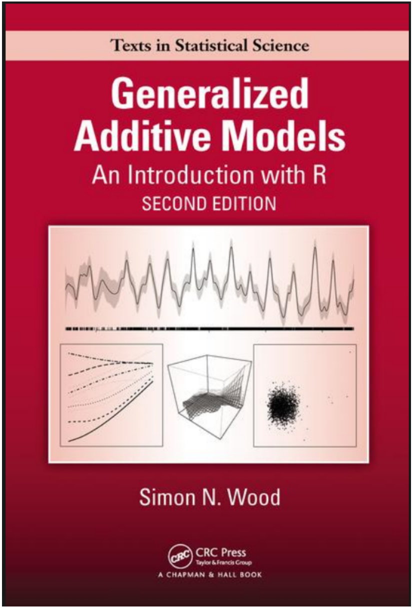
B-spline Basis



From basis to fitted curve



GAMs using mgcv



Wood, S. Generalized Additive Models.

R package mgcv
mixed gam computation vehicle with automatic smoothness estimation

Linear Model – GLM – GAM

1. linear models

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

2. non-normal response — GLM

$$Y_i \sim NegBin(\mu_i, \theta)$$

(or normal, Poisson, Binomial, etc....)

$$g(\mu_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$$

3. add smooth functions — GAM

$$Y_i \sim NegBin(\mu_i, \theta)$$

$$g(\mu_i) = \beta_0 + s_1(x_i) + s_2(z_i)$$

additive — no interactions

Specifying GAMs in mgcv

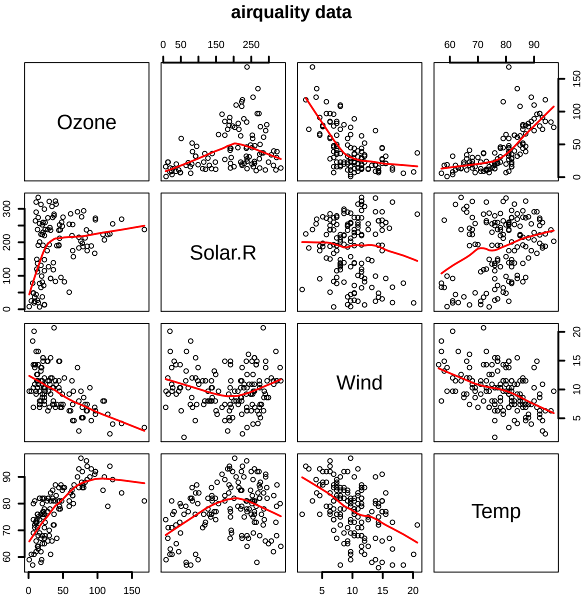
$$y_i = f(x_i) + e_i, \qquad \text{where } e_i \sim N(0, \sigma^2)$$

```
gam(y ~ s(x))
```

$$\log(\mu_i) = f1(xi) + f2(zi) + f3(vi) + wi, \qquad \text{where } yi \sim Poisson(\mu_i)$$

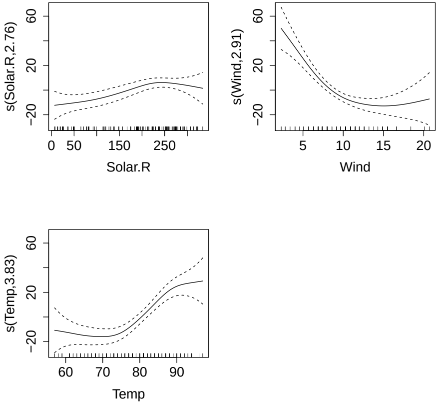
```
gam(y ~ s(x) + s(z) + s(v) + w, family = poisson)
```

Example: Airquality



Airquality Example

```
library(mgcv)
gam1 <- gam(ozone ~ s(rad) + s(wind) + s(temp),
            data = airquality)
plot(gam1)
```



ozone.R

Airquality Example

```
> summary(gam1)

   :

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   42.099      1.663   25.32   <2e-16

Approximate significance of smooth terms:
             edf Ref.df      F  p-value
s(Solar.R) 2.760  3.447  3.967  0.00849
s(Wind)    2.910  3.657 13.695 1.59e-08
s(Temp)    3.833  4.753 11.613 8.45e-09

R-sq.(adj) =  0.723   Deviance explained = 74.7%
GCV =  338.9  Scale est. = 306.83    n = 111
```

s term details

```
s(x, k = 10, bs = "tp", fx = FALSE)
```

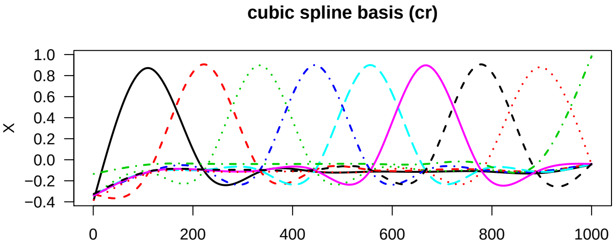
- x is the covariate
- bs defines the type of basis
- k is the basis dimension
- fx: fixed degrees of freedom or penalized
- by allows interactions with a factor

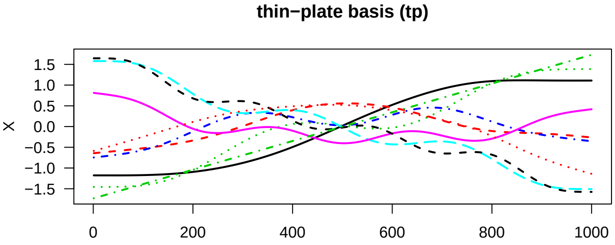
DEFAULT: thin-plate spline, with penalty (eigen-decomposition)

bs — different basis functions

- tp – default – thin-plate regression spline
- cr – cubic regression spline – cubic spline basis
- cc – cyclic cubic regression spline

Thin-plate Splines





Penalties (fx = FALSE)

Choose a large number of knots, but constrain their influence: e.g.

$$\sum \beta_j^2 \leq C$$

$\rightarrow$ choose $\boldsymbol{\beta}$ to minimize

$$||y - X\beta||^2 \qquad \text{subject to} \qquad \beta' D\beta \leq C$$

$\rightarrow$ minimize

$$||y - X\boldsymbol{\beta}||^2 + \lambda\boldsymbol{\beta}' D\boldsymbol{\beta}$$

$\lambda =$ smoothing parameter (sp)

$\rightarrow$ tradeoff between fit and smoothness

Estimation: generalized cross-validation (GCV), or Maximum Likelihood (or REML)

$\rightarrow$ **penalized regression splines**

2-dimensional smooths

1. **Tensor Product:**

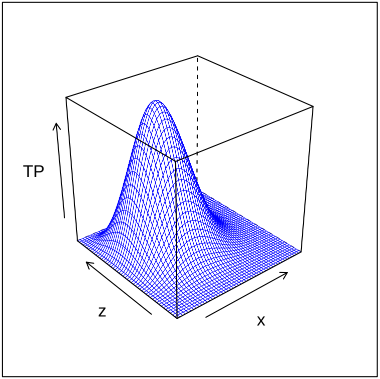
$$\mu_i = s(\mathsf{Solar.R}_i, \mathsf{Temp}_i)$$

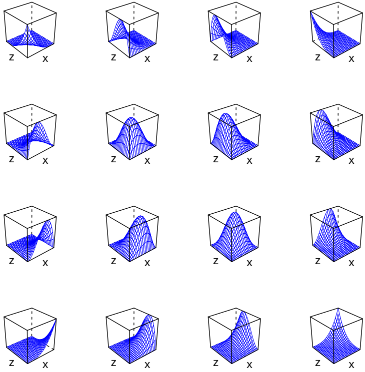
```
gam(y ~ te(Solar.R, Temp), data = airquality)
```

2. **Thin-plate Spline:**

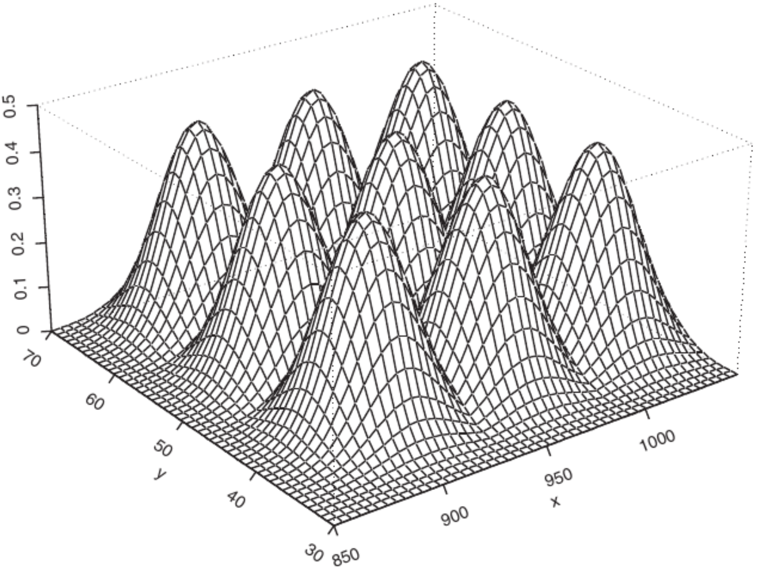
```
gam(y ~ s(longitude, latitude), data = ... )
```

Tensor Product





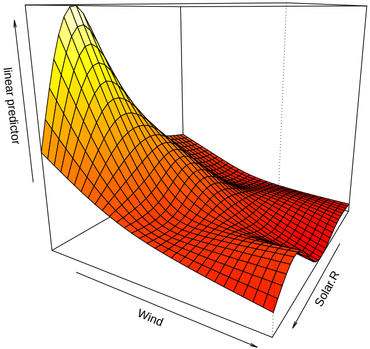
Tensor Product Basis



Tensor Product

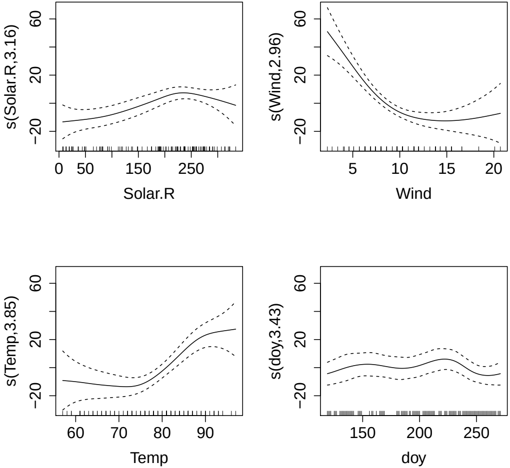
```
gam3 <- gam(Ozone ~
            te(Solar.R, Wind),
            data = airquality)

vis.gam(gam3, theta = 120)
```



Seasonal Spline

```
gam.seas <- gam(Ozone ~ s(Solar.R) + s(Wind) + s(Temp) +
                s(doy, bs = "cc"), data = airquality)
plot(gam.seas)
```



Seasonal Spline

```
visreg(gam.seas)
```