

Intro to generalized additive models in R (with mgcv)

Matteo Fasiolo (University of Bristol, UK)

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/GAM_Workshop_Enbis_EDF_23

We will firstly cover:

- 1 What is an additive model?
- 2 Introducing smooth effects
- 3 GAM modelling with mgcv and mgcViz

What is an additive model

Regression setting:

- y is our response or dependent variable
- \mathbf{x} is a vector of covariates or independent variables

In **distributional regression** we want a good model for $\text{Dist}(y|\mathbf{x})$.

Model is $\text{Dist}_m\{y|\theta_1(\mathbf{x}), \theta_2, \dots, \theta_q\}$, where $\theta_1, \dots, \theta_q$ are param.

We assume that $\theta_2, \dots, \theta_q$ do not depend on \mathbf{x} .

Gaussian additive model:

$$y|\mathbf{x} \sim N\{y|\mu(\mathbf{x}), \sigma^2\},$$

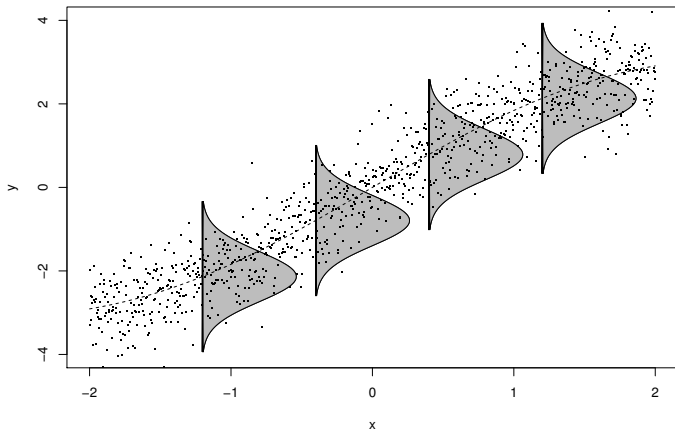
where

$$\mu(\mathbf{x}) = \sum_{j=1}^m f_j(\mathbf{x}),$$

and

$$\sigma^2 = \text{Var}(y).$$

f_j 's can be fixed, random or smooth effects.



Gaussian model with variable mean.

In mgcv: `gam(y~s(x), family=gaussian)`.

Generalized additive model (GAM) (Hastie and Tibshirani, 1990):

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu(\mathbf{x}), \theta_2, \dots, \theta_p\},$$

where

$$g\{\mu(\mathbf{x})\} = \sum_{j=1}^m f_j(\mathbf{x}),$$

and g is the link function.

Poisson GAM:

- $y|\mathbf{x} \sim \text{Pois}\{y|\mu(\mathbf{x})\}$
- $\mathbb{E}(y|\mathbf{x}) = \text{Var}(y|\mathbf{x}) = \exp\left\{\sum_{j=1}^m f_j(\mathbf{x})\right\}$
- $g = \log$ assures $\mu(\mathbf{x}) > 0$

These slides cover:

- 1 What is an additive model?
- 2 Introducing smooth effects
- 3 GAM modelling with mgcv and mgcViz

Introducing smooth effects

Consider additive model

$$g\{\mu(\mathbf{x})\} = f_1(\mathbf{x}) + f_2(\mathbf{x}) + f_3(\mathbf{x}),$$

where

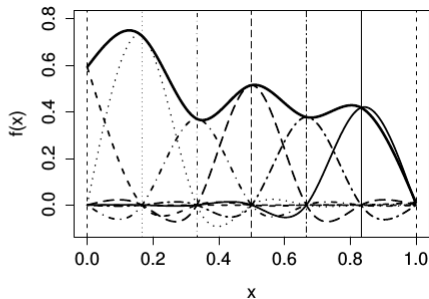
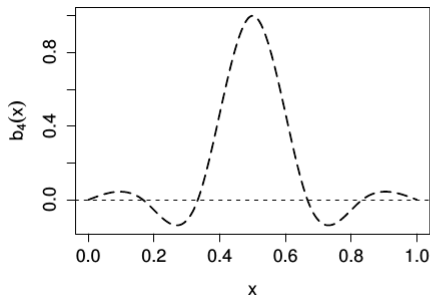
- $f_1(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$
- $f_2(\mathbf{x}) = \begin{cases} 0 & \text{if } x_2 = \text{FALSE} \\ \beta_4 & \text{if } x_2 = \text{TRUE} \end{cases}$
- $f_3(\mathbf{x}) = f_3(x_3)$ is a non-linear smooth function.

Smooth effects built using spline bases

$$f_3(x_3) = \sum_{k=1}^r \beta_k b_k(x_3)$$

where β_k are unknown coeff and $b_k(x_3)$ are known spline basis functions.

`s(x, bs = "cr", k = 20)`



Cubic regression splines are related to the optimal solution to

$$\sum_{i=1}^n \{y_i - f(x_i)\}^2 + \gamma \int f''(x)^2 dx.$$

mgcv offers a wide variety of smooths (see `?smooth.terms`).

Univariate types:

- $s(x) = s(x, bs = "tp")$ thin-plate-splines
- $s(x, bs = "cr")$ cubic regression spline
- $s(x, bs = "ad")$ adaptive smooth

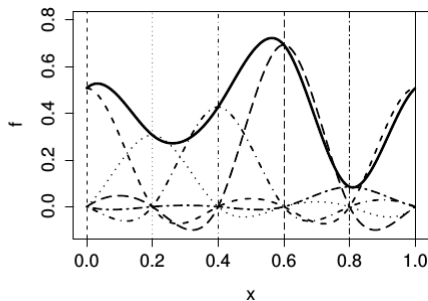
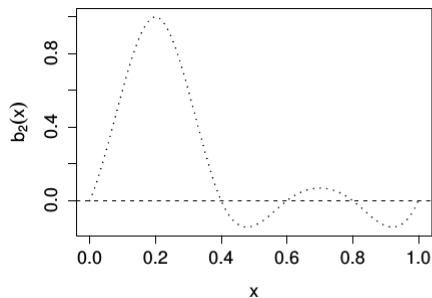
Multivariate type:

- $s(x_1, x_2) = s(x_1, x_2, bs = "tp")$ thin-plate-splines (isotropic)
- $te(x_1, x_2)$ tensor-product-smooth (anisotropic)
- $s(x, y, bs = "sos")$ smooth on sphere

They can depends on factors:

- $s(x, by = Subject)$
- $s(x, Subject, bs = "fs")$

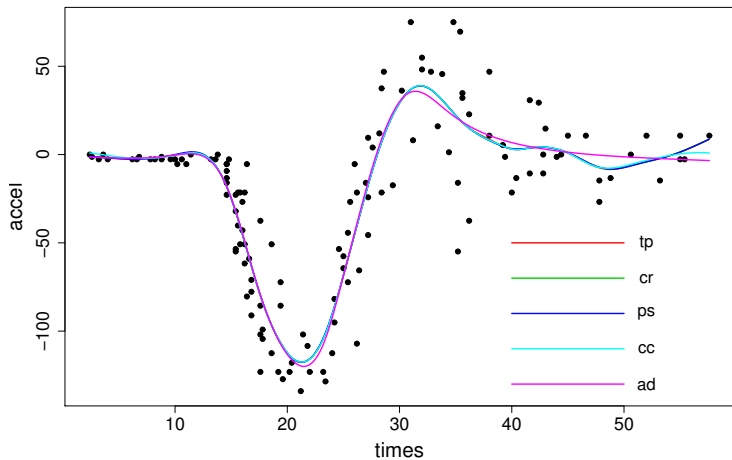
$s(x, \text{bs} = "cc")$



Cyclic cubic regression splines make so that

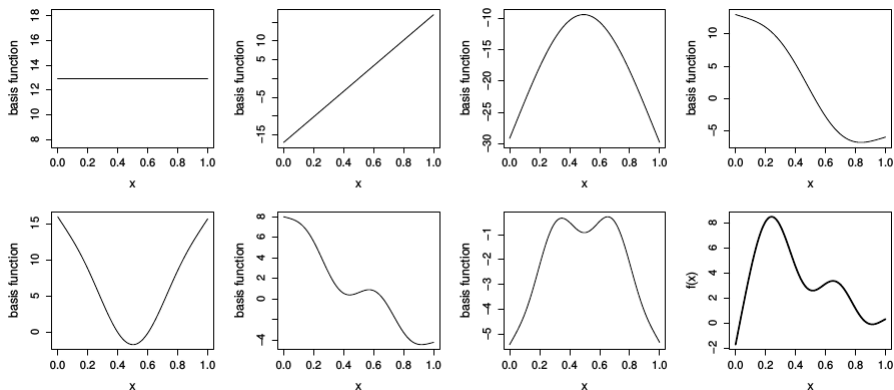
- $f(x_{min}) = f(x_{max})$
- $f'(x_{min}) = f'(x_{max})$
- $f''(x_{min}) = f''(x_{max})$

$s(x, \text{bs} = \text{"ad"})$

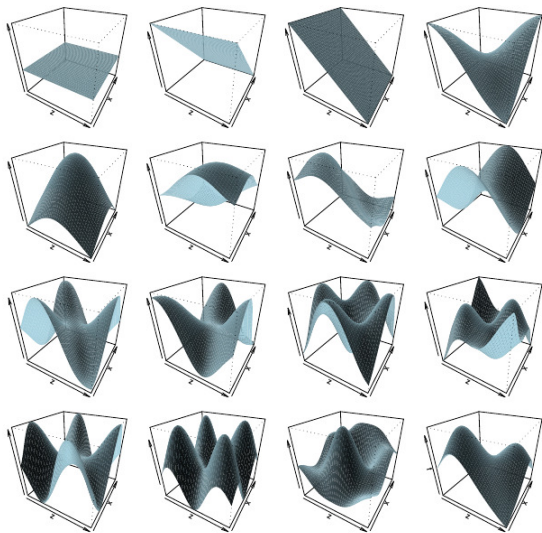


The wiggleness or smoothness of $f(x)$ depends on x .

$s(x, \text{bs} = \text{"tp"})$ or $s(x)$: Thin plate regression splines (TPRS)



Rank 7 TPRS basis. Image from Wood (2017).



Rank 17 2D TPRS basis. Courtesy of Simon Wood.

$s(x_1, x_2), s(x_1, x_2, x_3), \dots$

Based on thin plate regression splines basis.

Related to optimal solution to:

$$\sum_i \{y_i - f(x_i, z_i)\}^2 + \gamma \int f_{xx}^2 + 2f_{xz}^2 + f_{zz}^2 dx dz$$

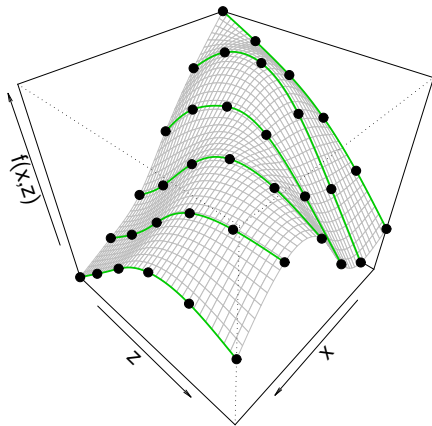
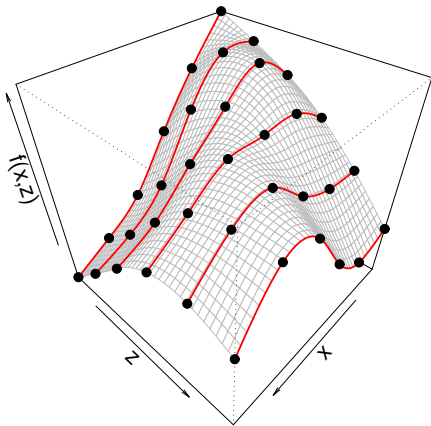
A single smoothing parameter γ .

Isotropic: same smoothness along x_1, x_2, \dots

Isotropic effect of x_1 , x_2 are in same unit (e.g. Km).

If different units better use tensor product smooths $\text{te}(x_1, x_2)$.

- x-penalty: average wiggleness of red curves
- z-penalty: average wiggleness of green curves



Can use (almost) any kind of marginal:

- `te(x1, x2, x3)` product of 3 cubic regression splines bases
- `te(x1, x2, bs = c("cc", "cr"), k = c(10, 6))`
- `te(L0, LA, t, d=c(2,1), bs=c("tp","cc"))`

Basis of `te` contains functions of the form $f(x_1)$ and $f(x_2)$.

To fit $f(x_1) + f(x_2) + f(x_1, x_2)$ separately use:

```
y ~ ti(x1) + ti(x2) + ti(x1, x2)
```

These slides cover:

- 1 What is an additive model?
- 2 Introducing smooth effects
- 3 GAM modelling with mgcv and mgcViz

mgcv is a recommended R package, included in R by default.

There are alternatives to mgcv, such as:

- mboost (Hothorn et al., 2010)
- gamlss (Rigby and Stasinopoulos, 2005)
- brms (Bürkner et al., 2017)
- BayesX (Brezger et al., 2003)
- INLA (Rue et al., 2009)

The mgcv ecosystem:

- mgcViz visualising GAMs
- qgam quantile GAMs
- SCM multivariate Gaussian GAMs
- gamFactory aggregation of experts with GAMs
- and many others gamm4, refund, scam, vagam, GJRM, itsadug, ...

GAM modelling with mgcv and mgcViz

Recall structure of smooth effects:

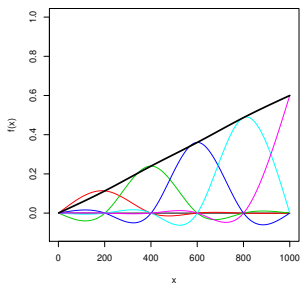
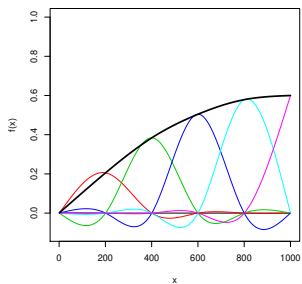
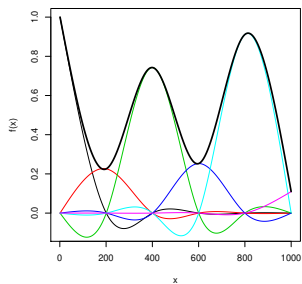
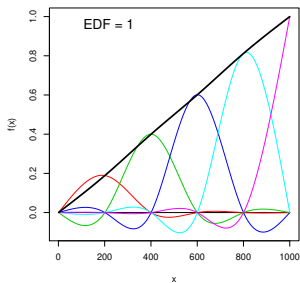
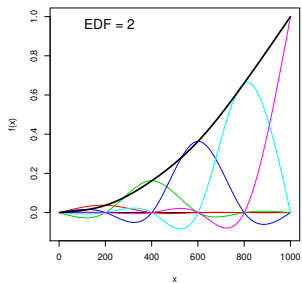
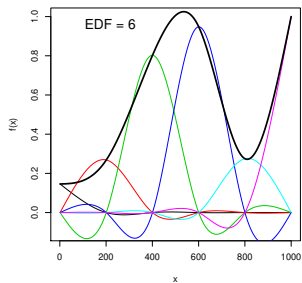
$$f(\mathbf{x}) = \sum_{j=1}^k \beta_j b_j(\mathbf{x}).$$

where β shrunk toward smoothness by penalty.

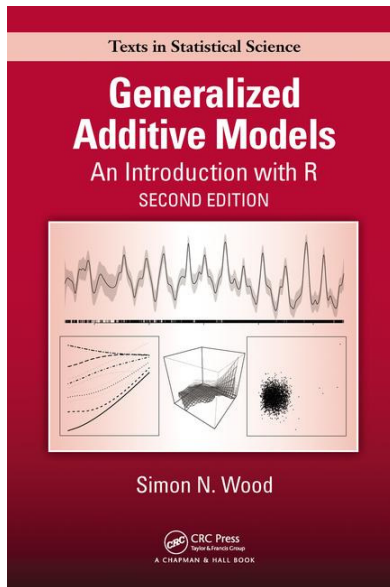
Exact k is unimportant, we choose it large enough and let penalty work.

Effective number of parameters we are using is $\leq k$.

Approximation is **Effective Degrees of Freedom** (EDF) $\leq k$.



Further reading



References I

- Brezger, A., T. Kneib, and S. Lang (2003). Bayesx: Analysing bayesian structured additive regression models. Technical report, Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians .
- Bürkner, P. C. et al. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software* 80(1), 1–28.
- Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*, Volume 43. CRC Press.
- Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner (2010). Model-based boosting 2.0. *The Journal of Machine Learning Research* 11, 2109–2113.
- Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54(3), 507–554.
- Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 71(2), 319–392.
- Wood, S. (2017). *Generalized additive models: an introduction with R*. CRC press.