

# Beyond mean modelling: GAMLSS models

Matteo Fasiolo (University of Bristol, UK)

*matteo.fasiolo@bristol.ac.uk*

June 27, 2018

# Beyond mean modelling: GAMLSS models

## Structure:

- 1 Intro to GAMs for Location Scale and Shape
- 2 GAM modelling using `mgcv` and `mgcViz`

# Beyond mean modelling: GAMLSS models

## Structure:

- 1 **Intro to GAMs for Location Scale and Shape**
- 2 GAM modelling using `mgcv` and `mgcViz`

# Intro to GAMLSS models

Recall GAM model structure:

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

where

$$\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = g^{-1}\left\{\sum_{j=1}^m f_j(\mathbf{x})\right\},$$

and  $g$  is the link function.

Example, Scaled Student-t distribution:

- location  $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$
- scale  $\theta_2 = \sigma$
- shape  $\theta_3 = \nu$

# Intro to GAMLSS models

In Generalized Additive Models for Location Scale and Shape (GAMLSS) we let scale and shape change with the covariates  $\mathbf{x}$ .

GAMLSS model structure:

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu_1(\mathbf{x}), \theta_2 = \mu_2(\mathbf{x}), \dots, \theta_p = \mu_p(\mathbf{x})\},$$

where

$$\begin{aligned}\mu_1(\mathbf{x}) &= g_1^{-1}\left\{\sum_{j=1}^m f_j^1(\mathbf{x})\right\}, \\ &\dots \\ \mu_p(\mathbf{x}) &= g_p^{-1}\left\{\sum_{j=1}^m f_j^p(\mathbf{x})\right\},\end{aligned}$$

and  $g_1, \dots, g_p$  are link function.

Example: **Gaussian model for location and scale**

Model is

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

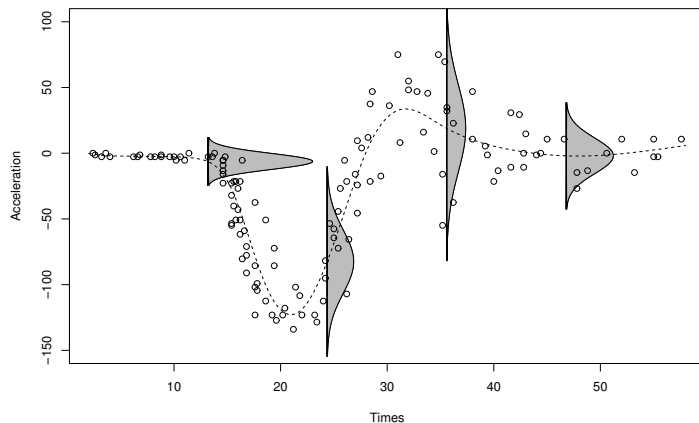
where

$$\mathbb{E}(y|\mathbf{x}) = \mu(\mathbf{x}) = \sum_{j=1}^m f_j^1(\mathbf{x})$$

$$\text{var}(y|\mathbf{x})^{1/2} = \sigma(\mathbf{x}) = \exp \left\{ \sum_{j=1}^m f_j^2(\mathbf{x}) \right\}$$

that is  $g_2 = \log$  to guarantee  $\sigma > 0$ .

# Intro to GAMLSS models



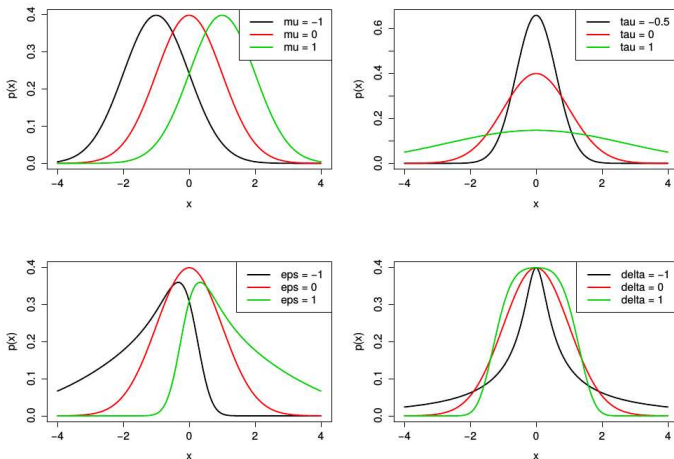
**Figure:** Gaussian model with variable mean and variance.

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

# Intro to GAMLSS models

## Example: **Sinh-arcsinh (shash) distribution**

Four parameter distribution where location, scale, skewness (asymmetry) and kurtosis (tail behaviour) can depend on  $x$ .





# Intro to GAMLSS models

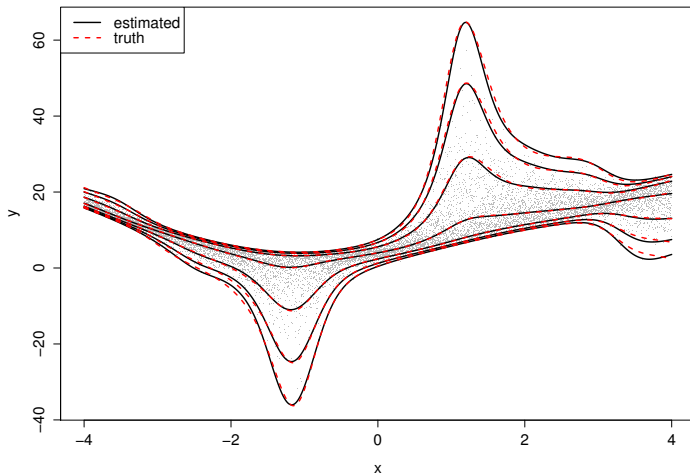
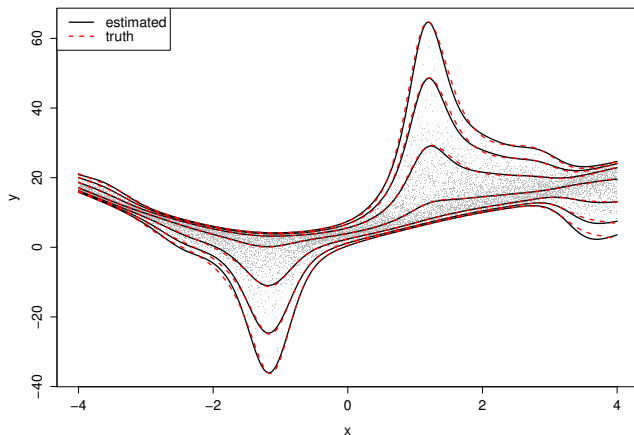


Figure: `gam(list(y~s(x), ~s(x), ~s(x), ~s(x)), family=shash).`

# Intro to GAMLSS models

## Why is this useful?

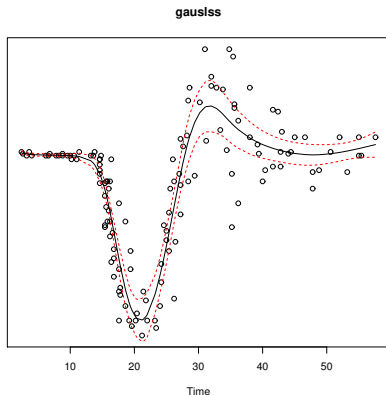
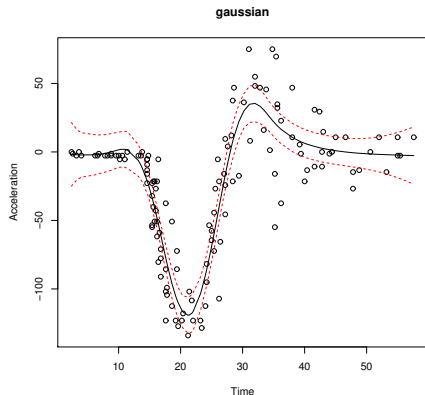
R1: you might be interested in whole distribution  $y|\mathbf{x}$  not just  $\mathbb{E}(y|\mathbf{x})$ .



# Intro to GAMLSS models

## Why is this useful?

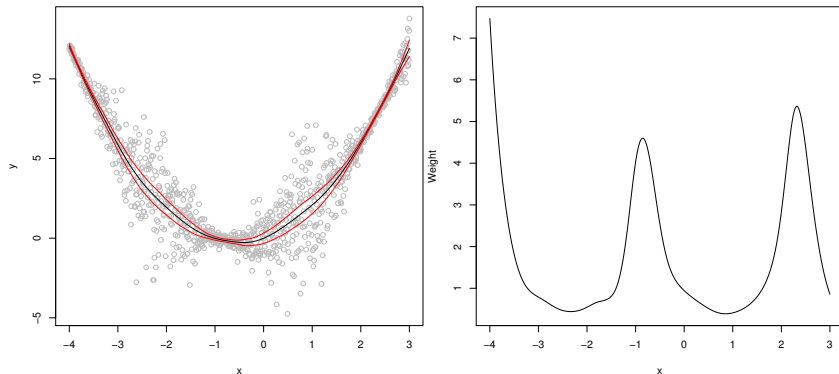
R2: standard GAM inference (e.g. p-value & confidence interval) is valid if the model for  $y|x$  is correct



# Intro to GAMLSS models

## Why is this useful?

R3: the accuracy of the fit is improved if the weight of each observation is inversely proportional to  $\text{Var}(y|\mathbf{x})$ .

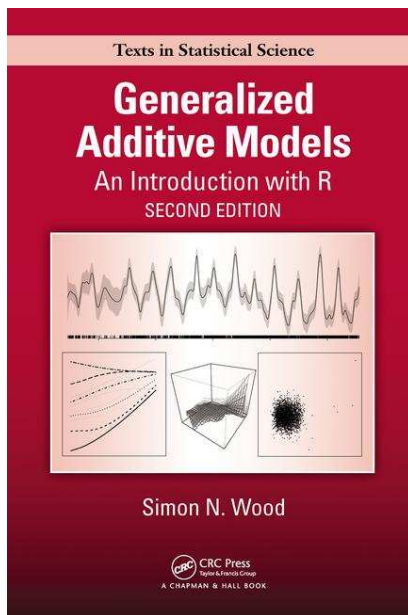


# Beyond mean modelling: GAMLSS models

## Structure:

- 1 Intro to GAMs for Location Scale and Shape
- 2 **GAM modelling using mgcv and mgcViz**

# Further reading



# References I

Fasiolo, M., Y. Goude, R. Nedellec, and S. N. Wood (2017). Fast calibrated additive quantile regression. *arXiv preprint arXiv:1707.03307*.