Generalized Additive Models

University of
St Andrews

Overview

- What is a GAM?
- What is smoothing?
- How do GAMs work?
- Fitting GAMs using

What is a GAM?

"gam"

1.

2.

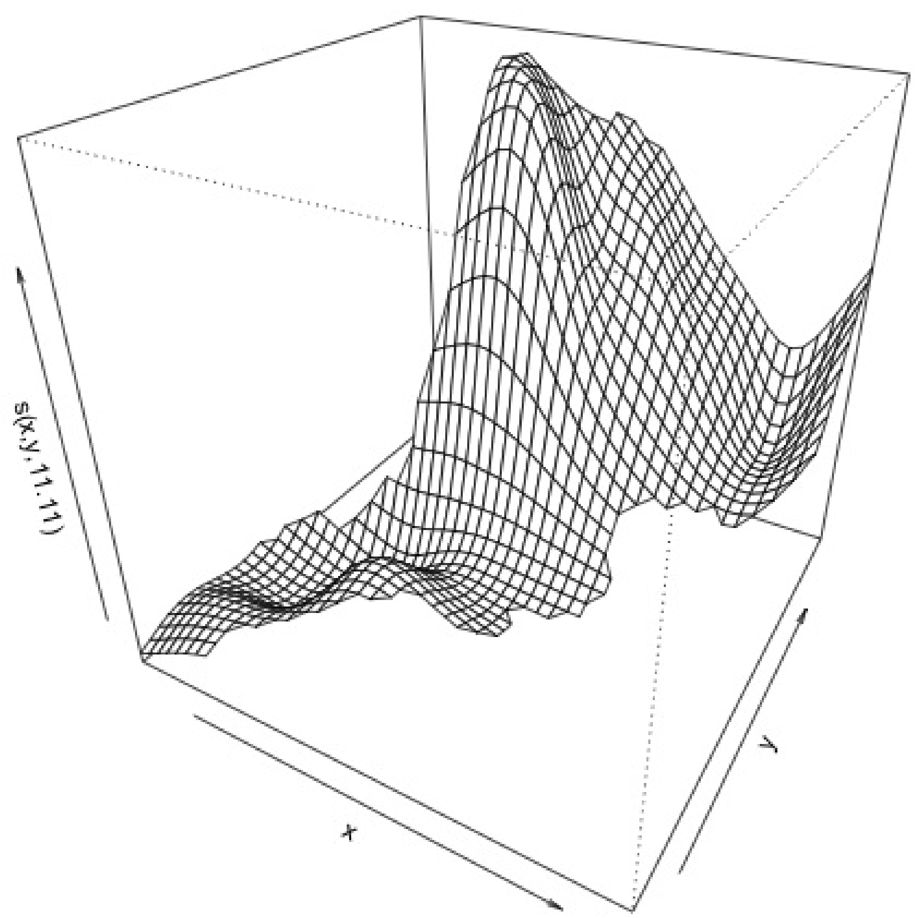
(via Natalie Kelly, AAD. Seen in Moby Dick.)

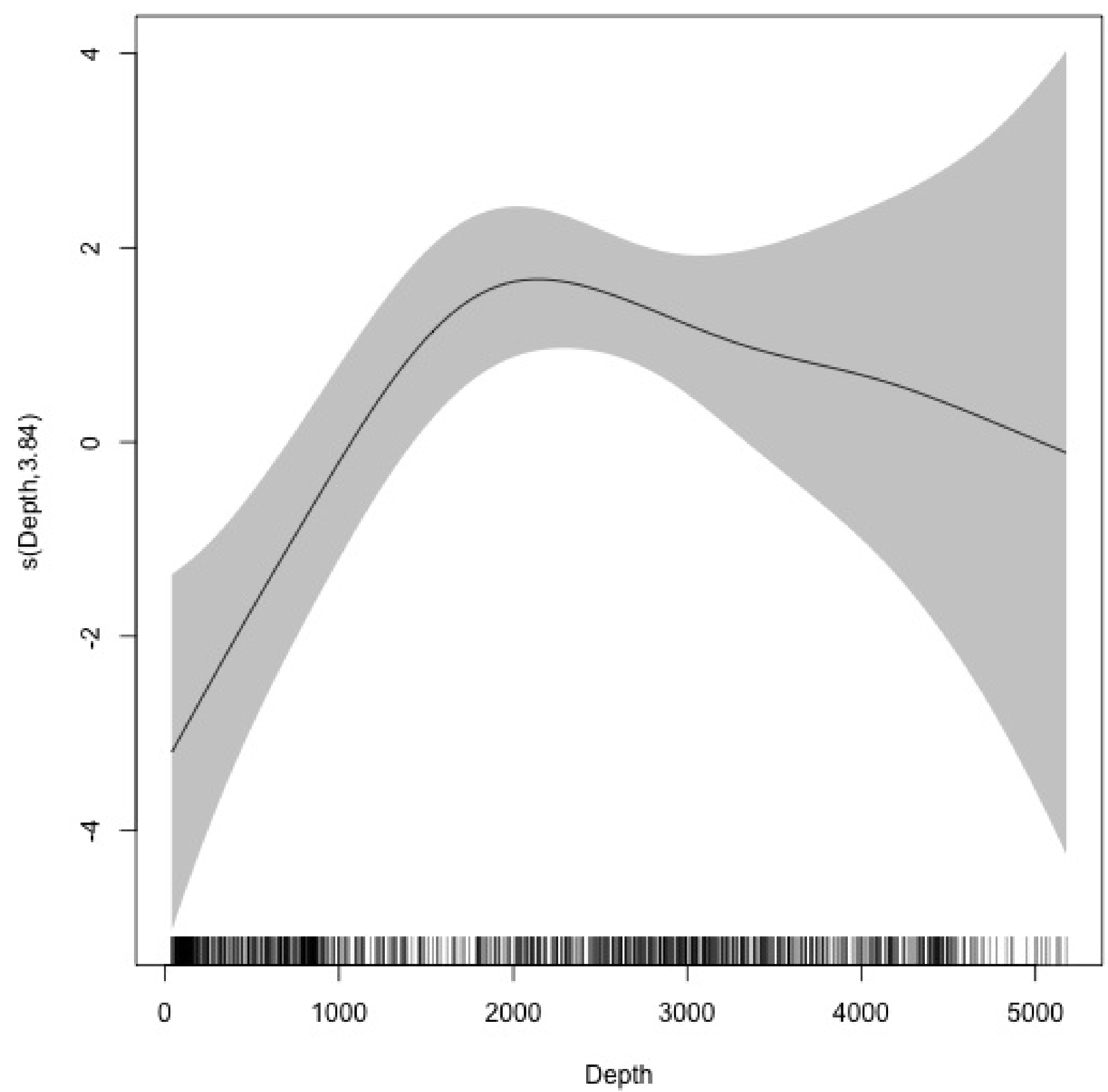
Generalized Additive Models

- Generalized: many response distributions
- Additive: terms **add** together
- Models: well, it's a model…

What does a model look like?

- Count n_j distributed according to some count distribution
- Model as sum of terms





Mathematically...

Taking the previous example...

$$n_j = A_j \hat{p}_j \exp\left[\beta_0 + s(y_j) + s(\text{Depth}_j)\right] + \epsilon_j$$

where $\epsilon_j \sim N(0, \sigma^2)$, $\quad n_j \sim$ count distribution

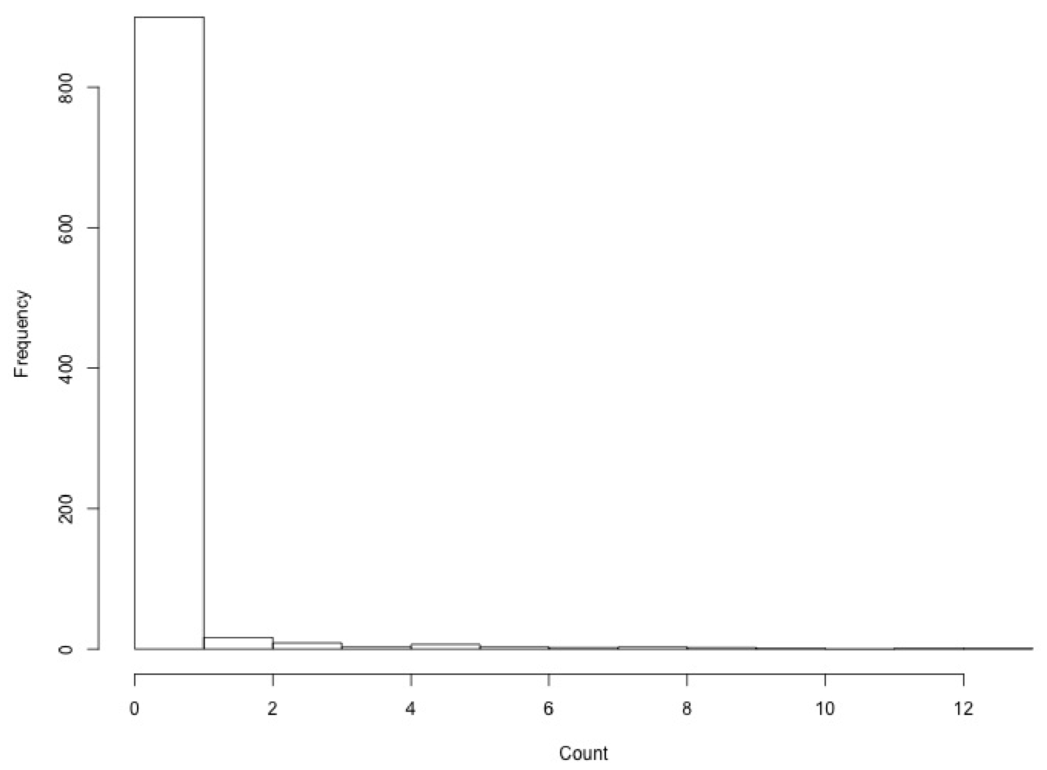
- area of segment - offset
- probability of detection in segment
- link function
- model terms

Response

$$n_j = A_j \hat{p}_j \exp\left[\beta_0 + s(y_j) + s(\text{Depth}_j)\right] + \epsilon_j$$

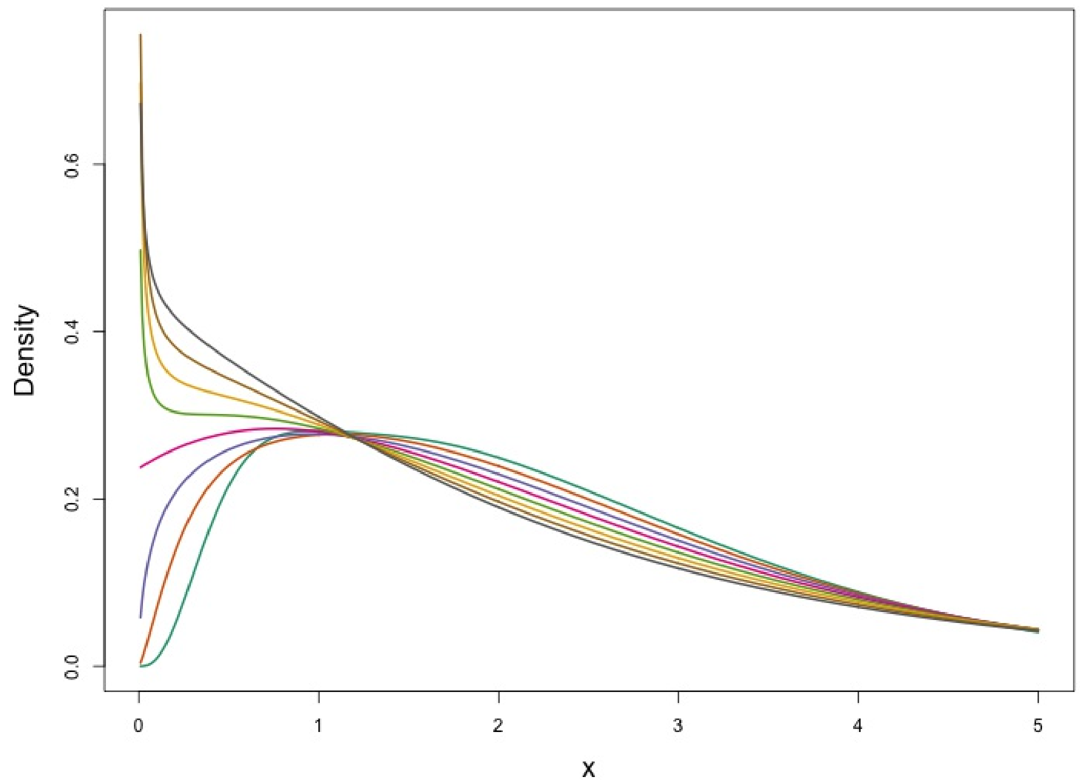
where $\epsilon_j \sim N(0, \sigma^2)$, $n_j \sim$ count distribution

Count distributions



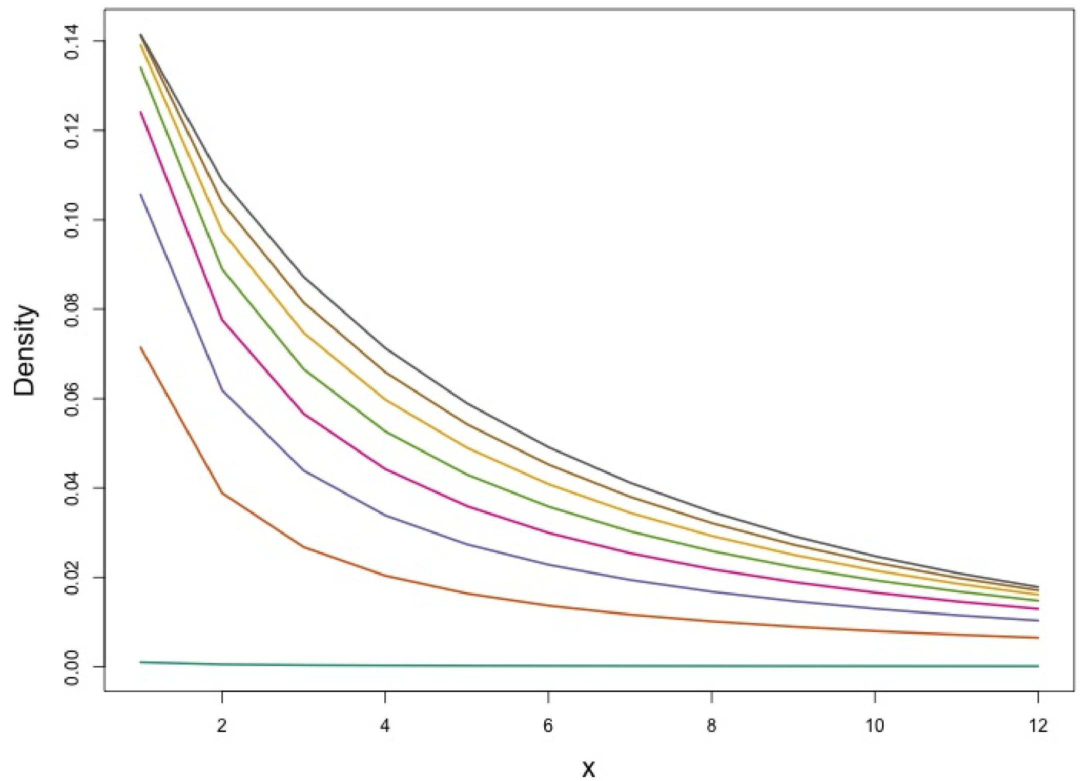
- Response is a count (not not always integer)
- Often, it's mostly zero (that's complicated)
- Want response distribution that deals with that
- Flexible mean-variance relationship

Tweedie distribution



- $\mathrm{Var}(\text{count}) = \varphi(\text{count})^q$
- Common distributions are sub-cases:
  - $q = 1 \Rightarrow$ Poisson
  - $q = 2 \Rightarrow$ Gamma
  - $q = 3 \Rightarrow$ Normal
- We are interested in $1 < q < 2$
- (here $q = 1.2, 1.3, \ldots, 1.9$)

Negative binomial distribution



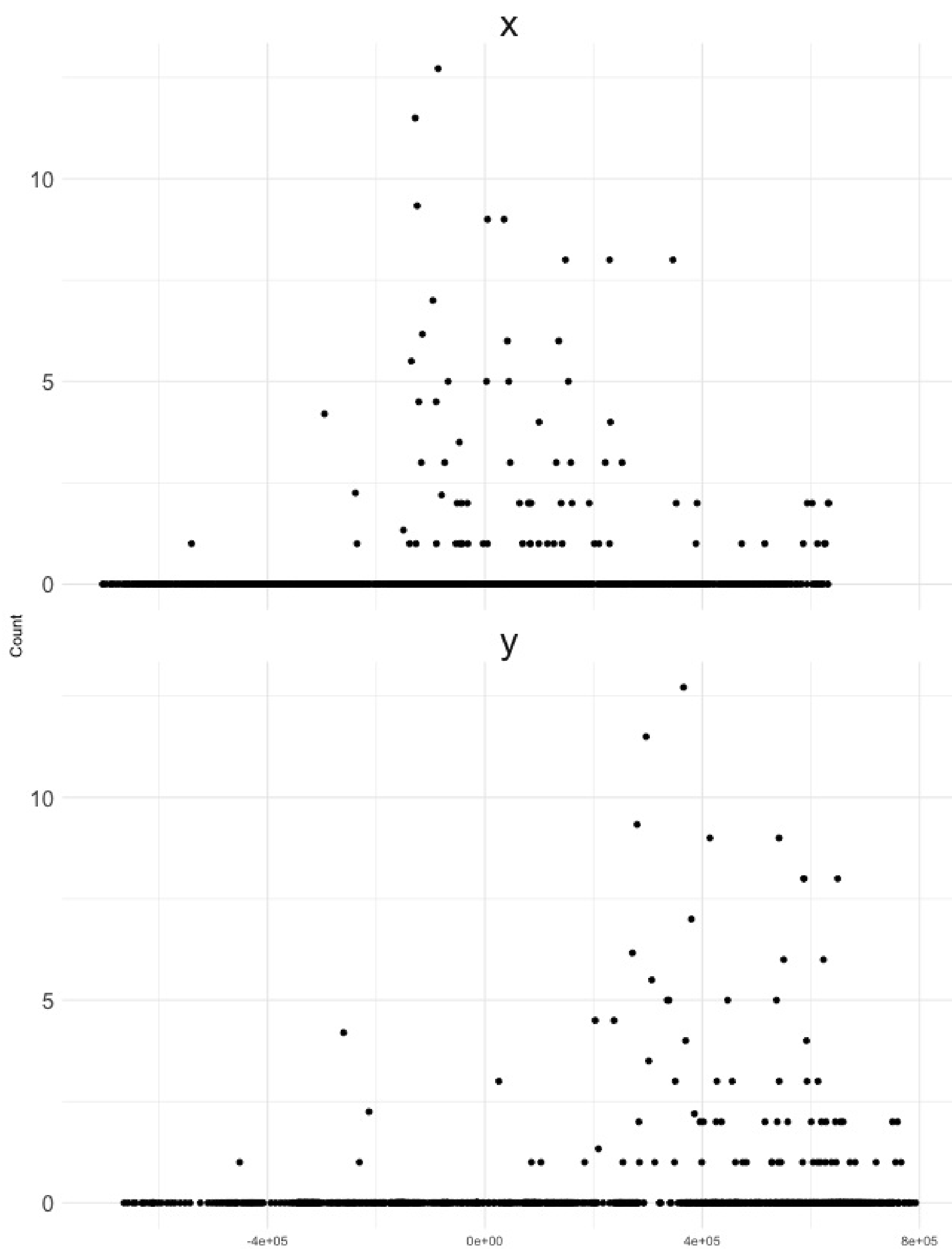
- Var (count) = (count) + $\varkappa$(count)^2
- Estimate $\varkappa$
- Is quadratic relationship a "strong" assumption?
- Similar to Poisson: Var (count) = (count)

Smooth terms

$$n_j = A_j \hat{p_j} \exp\left[\beta_0 + s(y_j) + s(\mathrm{Depth}_j)\right] + \epsilon_j$$

where $\epsilon_j \sim N(0, \sigma^2)$, $\quad n_j \sim$ count distribution

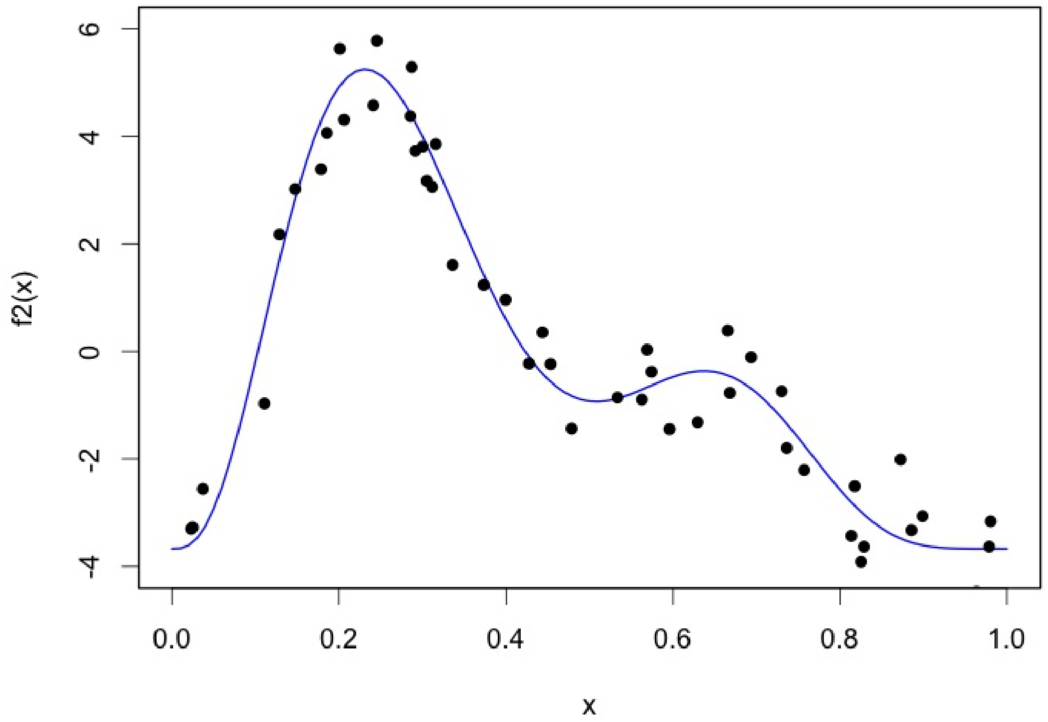
Okay, but what about these "s" things?



- Think s=**smooth**
- Want to model the covariates flexibly
- Covariates and response not necessarily linearly related!
- Want some wiggles

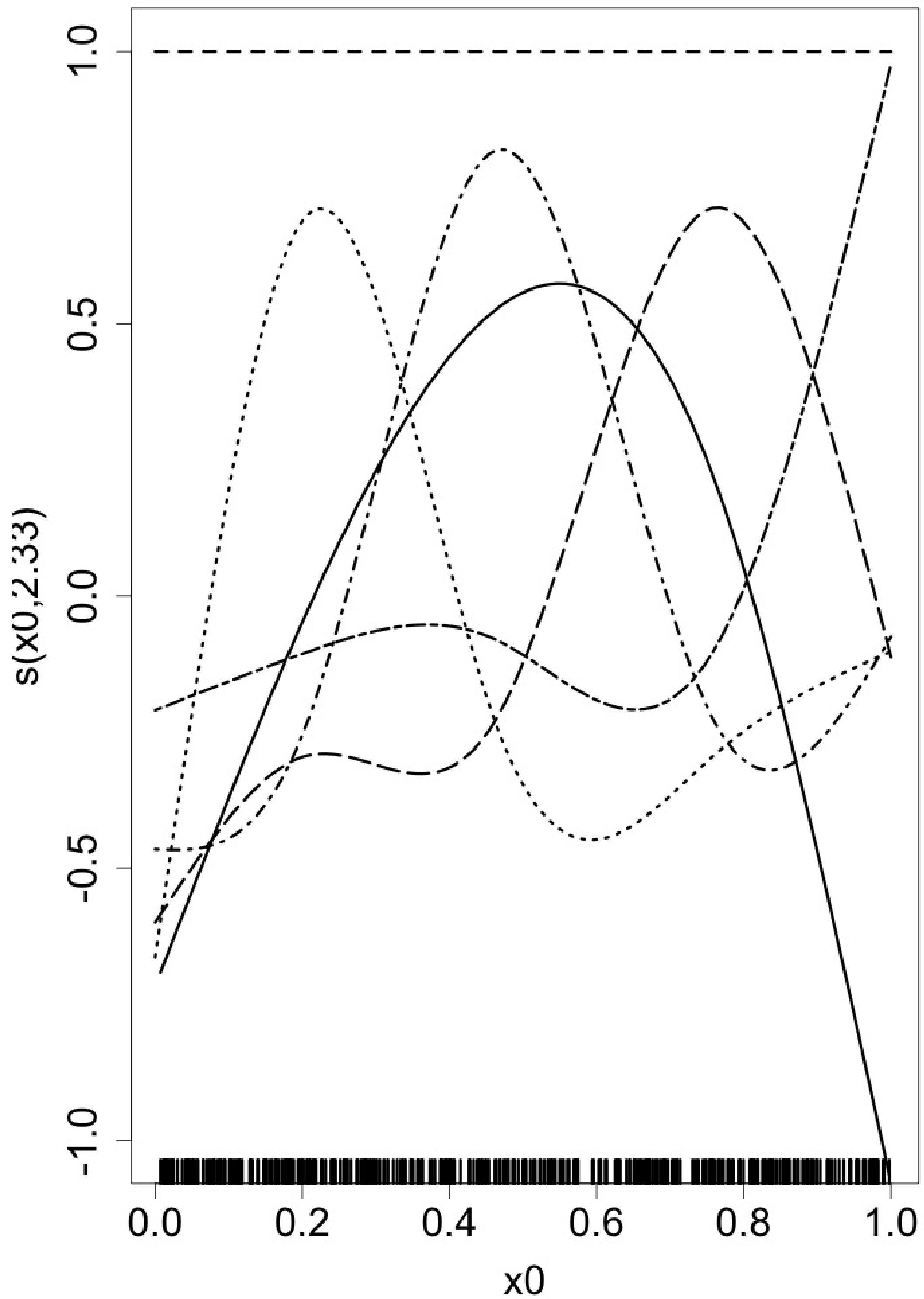
What is smoothing?

Straight lines vs. interpolation



- Want a line that is "close" to all the data
- Don't want interpolation -- we know there is "error"
- Balance between interpolation and "fit"

Splines

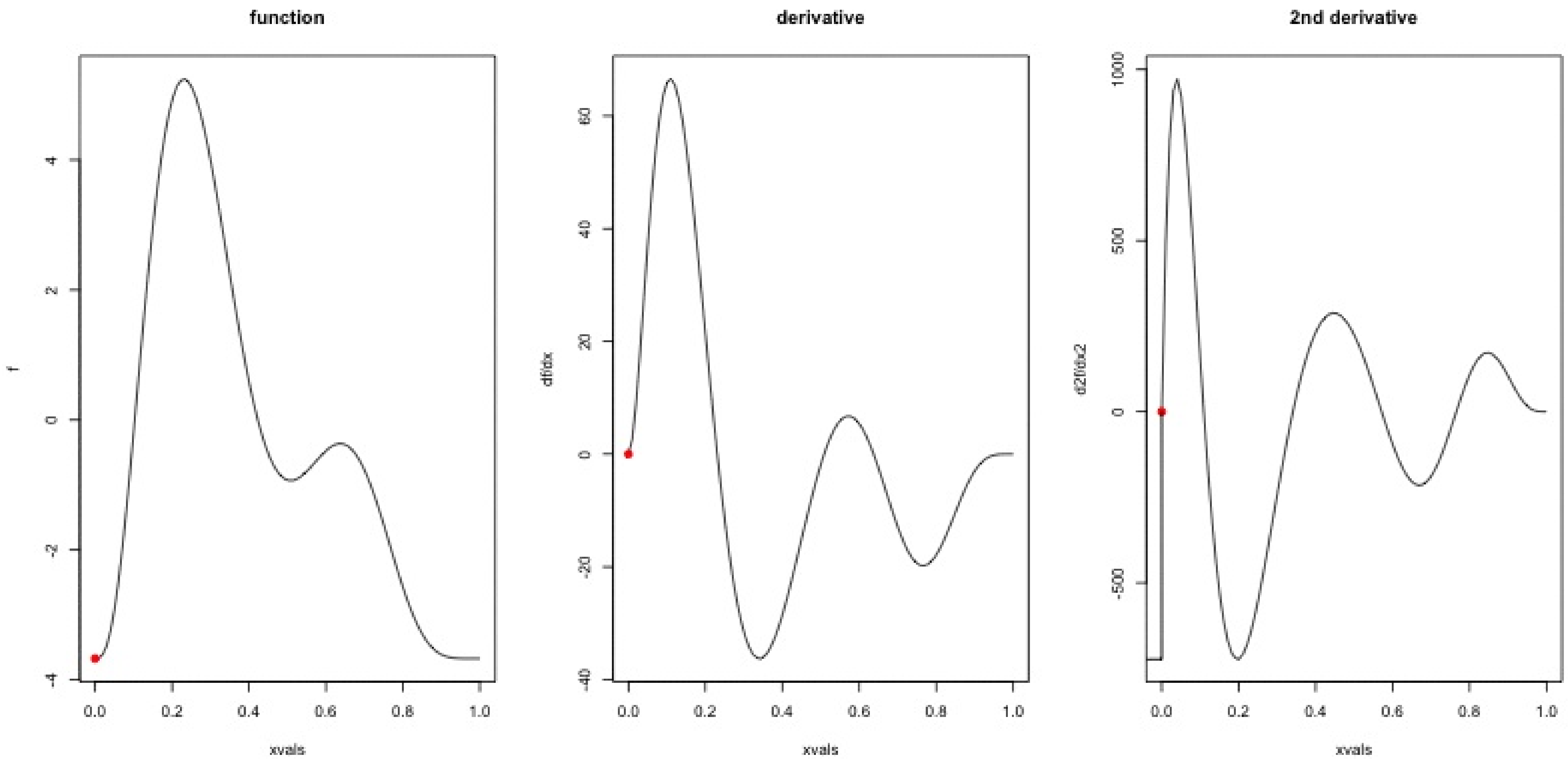


- Functions made of other, simpler functions
- **Basis functions** b_k, estimate β_k
- $s(x) = \sum_{k=1}^{K} \beta_k b_k(x)$
- Makes the maths much easier

Measuring wigglyness

- Visually:
  - Lots of wiggles == NOT SMOOTH
  - Straight line == VERY SMOOTH
- How do we do this mathematically?
  - Derivatives!
  - (Calculus a useful class after all)

Wigglyness by derivatives



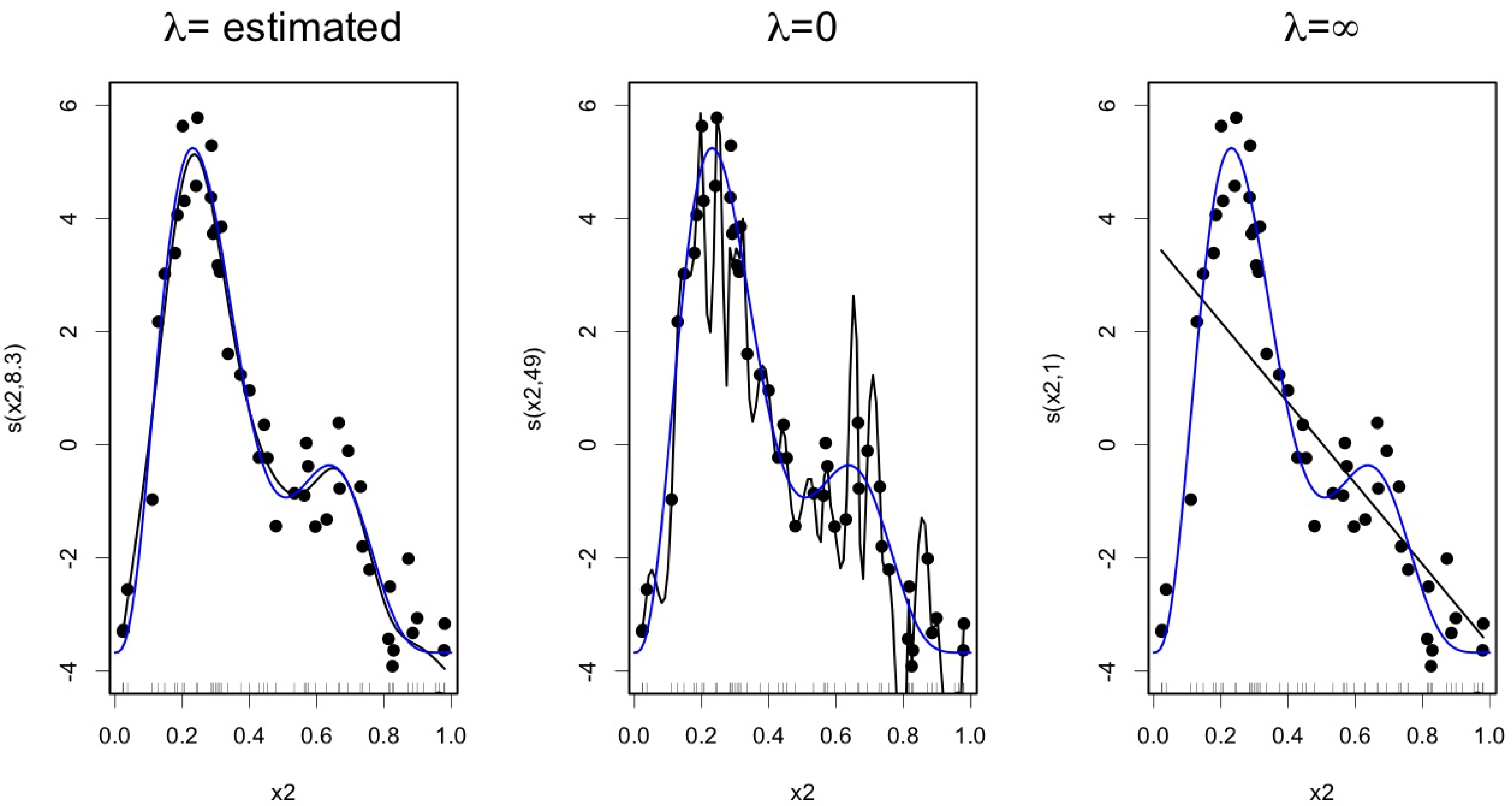
Making wigglyness matter

- Integration of derivative (squared) gives wigglyness
- Fit needs to be **penalised**
- **Penalty matrix** gives the wigglyness
- Estimate the β_k terms but penalise objective
  - "closeness to data" + penalty

Penalty matrix

- For each b_k calculate the penalty
- Penalty is a function of β
  - $\lambda\beta^T S\beta$
- S ( ) calculated once
- , λ dictates influence

Smoothing parameter



How wiggly are things?

- We can set **basis complexity** or "size" k
  - Maximum wigglyness
- Smooths have **effective degrees of freedom** (EDF)
- EDF < k
- Set k "large enough"

Why GAMs are cool...

Okay, that was a lot of theory…

Example data

Example data

Example data

Sperm whales off the US east coast

Model formulation

- Pure spatial, pure environmental, mixed?
- May have some prior knowledge
  - Biology/ecology
- What are drivers of distribution?
- Inferential aim
  - Abundance
  - Ecology

Fitting GAMs using dsm

Translating maths into R

$$n_j = A_j \hat{p_j} \exp\left[\beta_0 + s(y_j)\right] + \epsilon_j$$

where $\epsilon_j \sim N(0, \sigma^2)$, $\quad n_j \sim$ count distribution

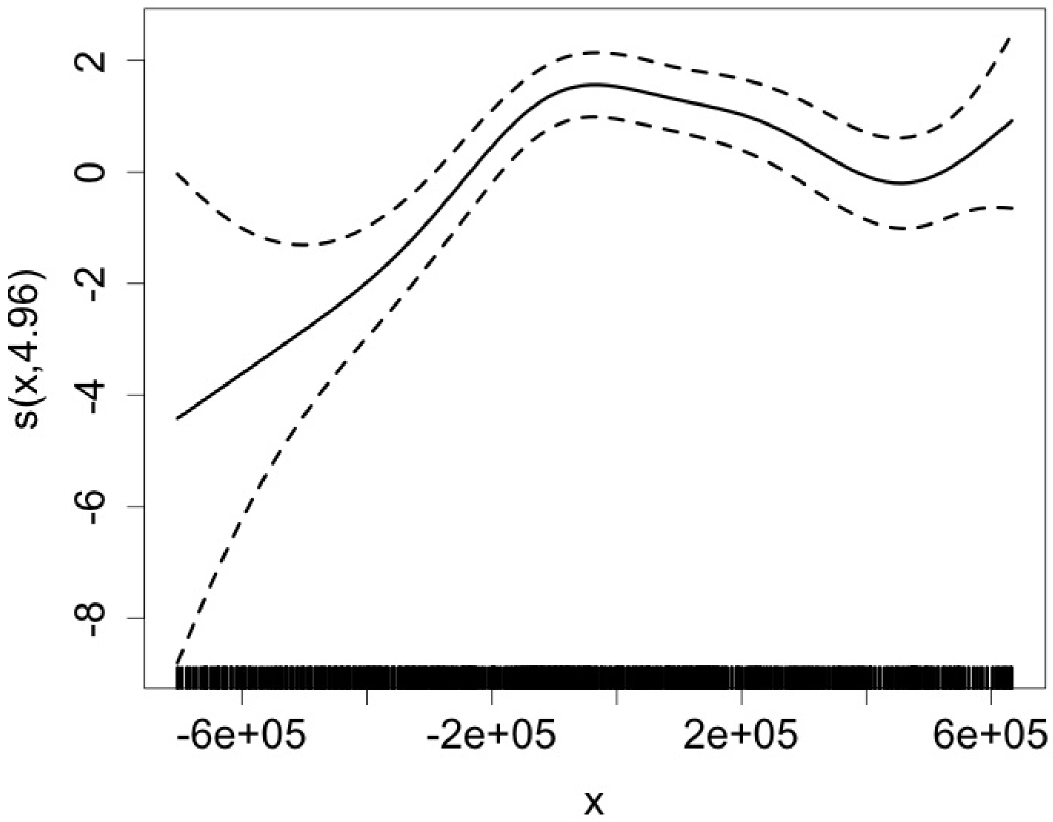
- inside the link:
- response distribution: or
- detectability:
- offset, data:

Your first DSM

is based on by Simon Wood

What did that do?

Plotting



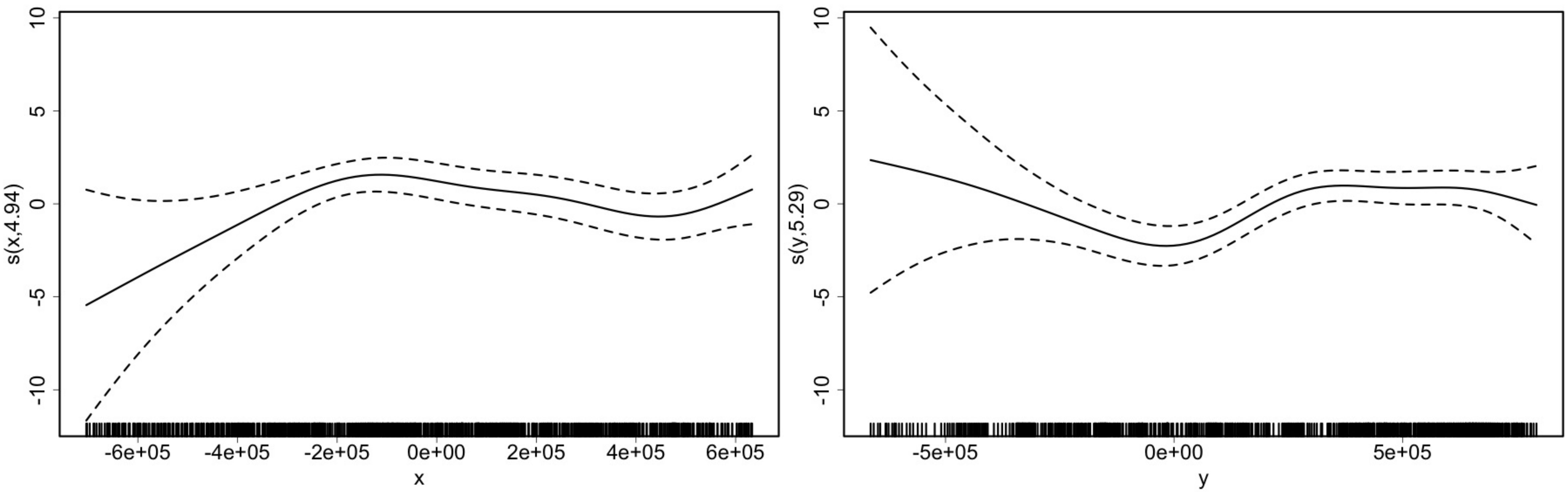
-
- Dashed lines indicate +/- 2 standard errors
- Rug plot
- On the link scale
- EDF on y axis

Adding a term

- Just use

Summary

Plotting



- : each plot on different scale
- : plot together

Bivariate terms

- Assumed an additive structure
- No interaction
- We can specify (and)

Thin plate regression splines

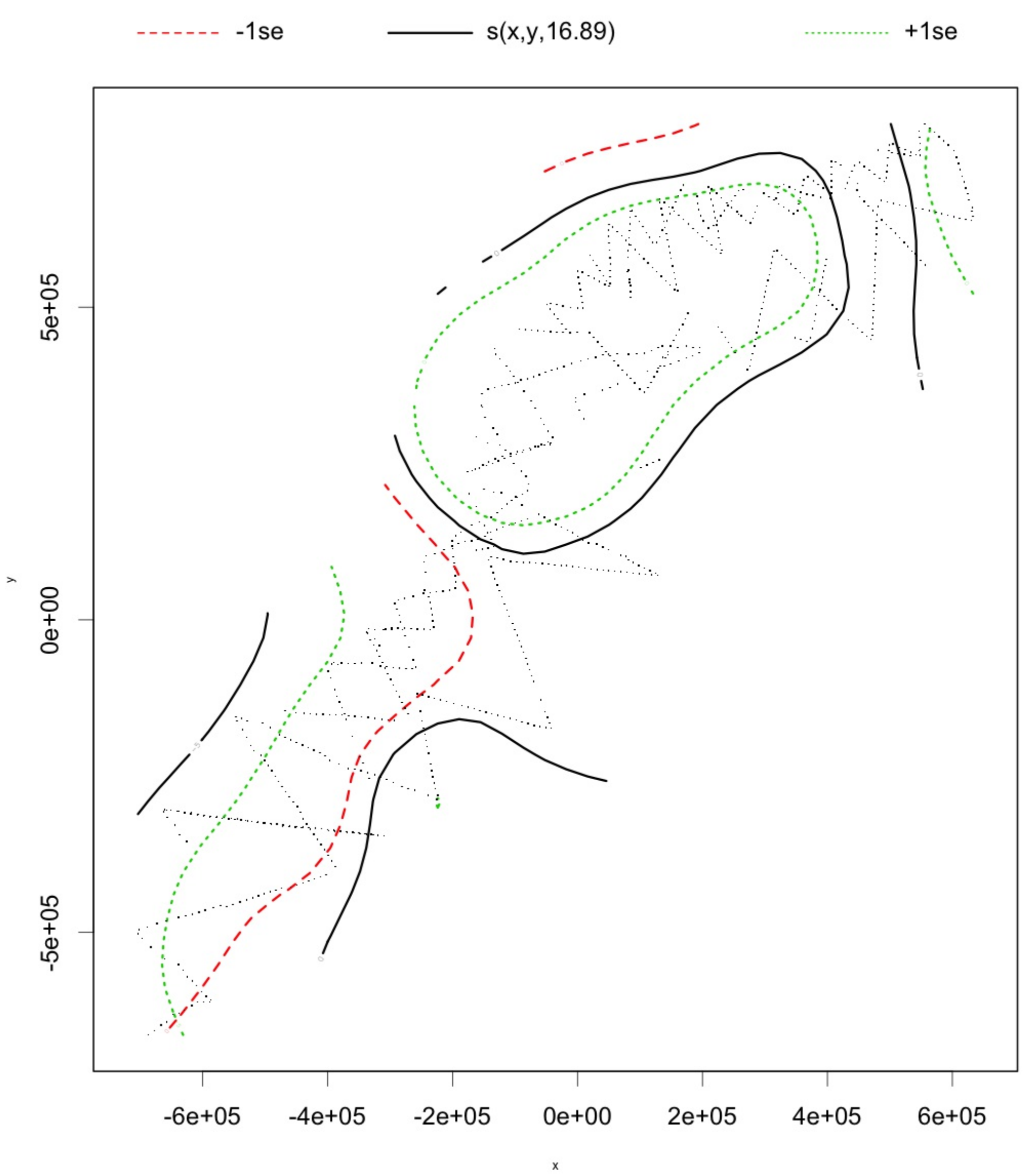
- Default basis
- One basis function per data point
- Reduce # basis functions (eigendecomposition)
- Fitting on reduced problem
- Multidimensional

Thin plate splines (2-D)

Bivariate spatial term

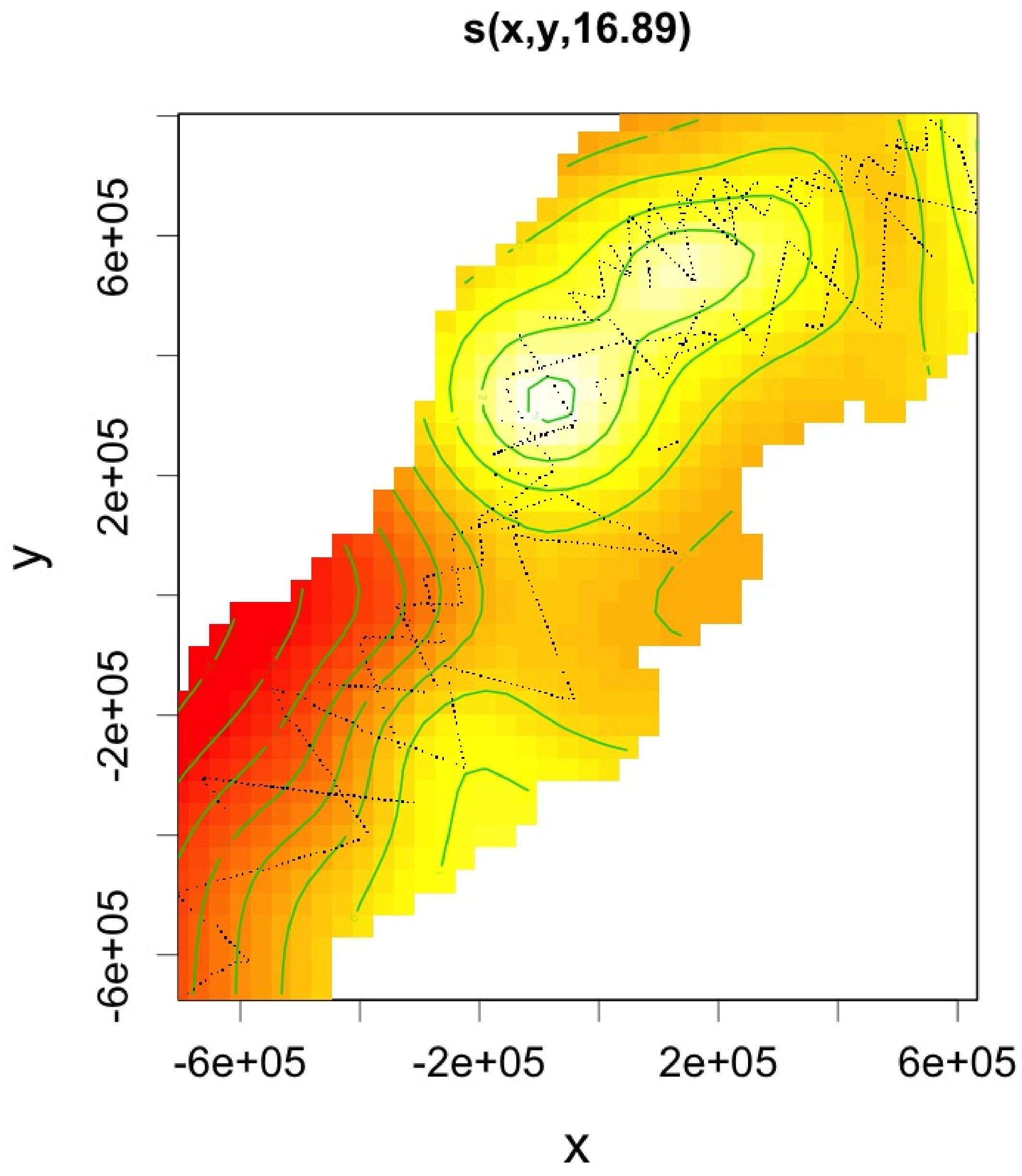
Summary

Plotting... erm...

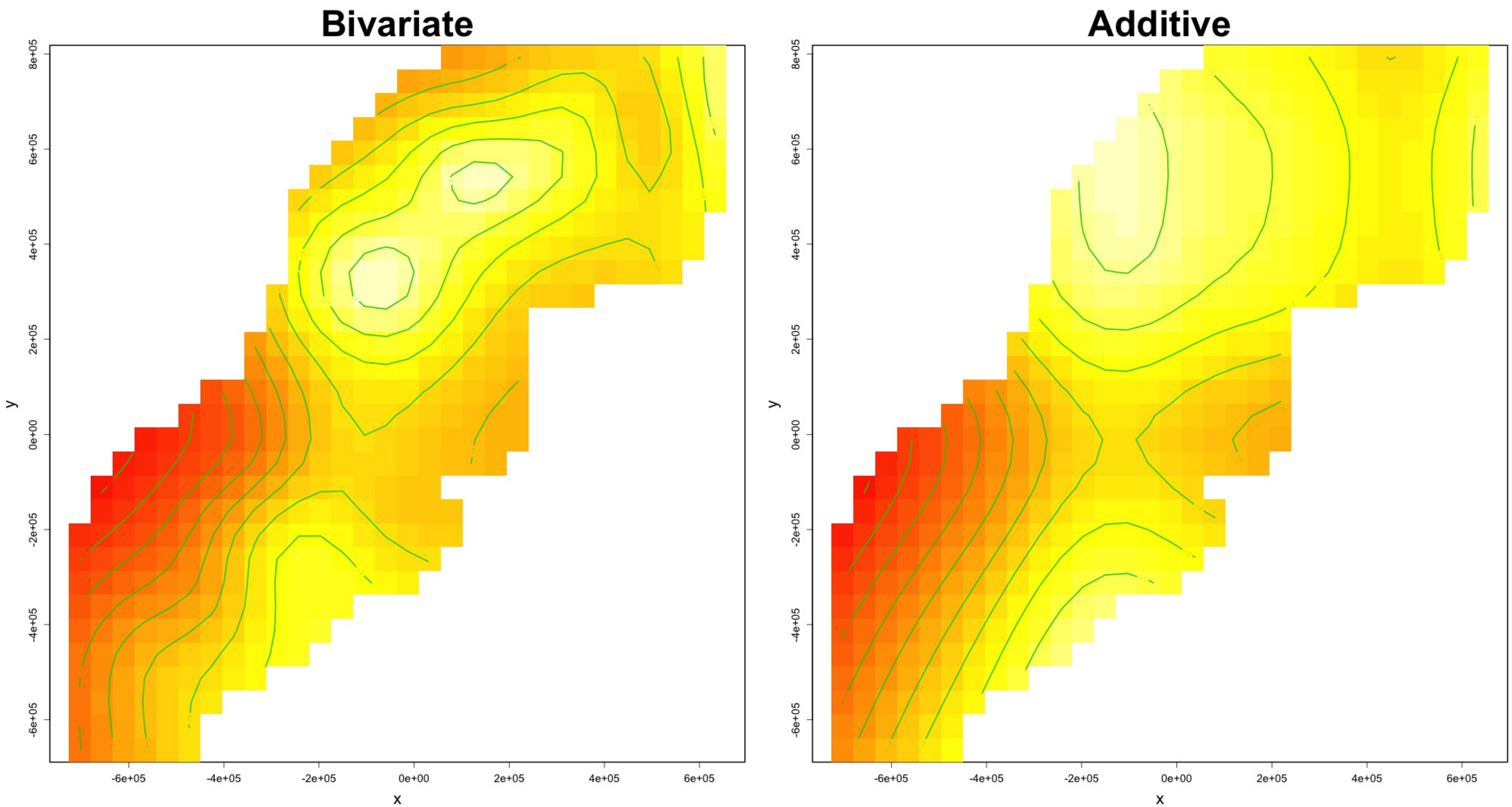


Let's try something different

- Still on link scale
- excludes points far from data



Comparing bivariate and additive models



Let's have a go...