Density surface models: spatial modelling of distance sampling data

David L Miller

Spatial models of distance sampling data

- Collect spatially referenced data
- Why not make spatially-explicit models?
- Go beyond stratified estimates
- Relate environmental covariates to counts

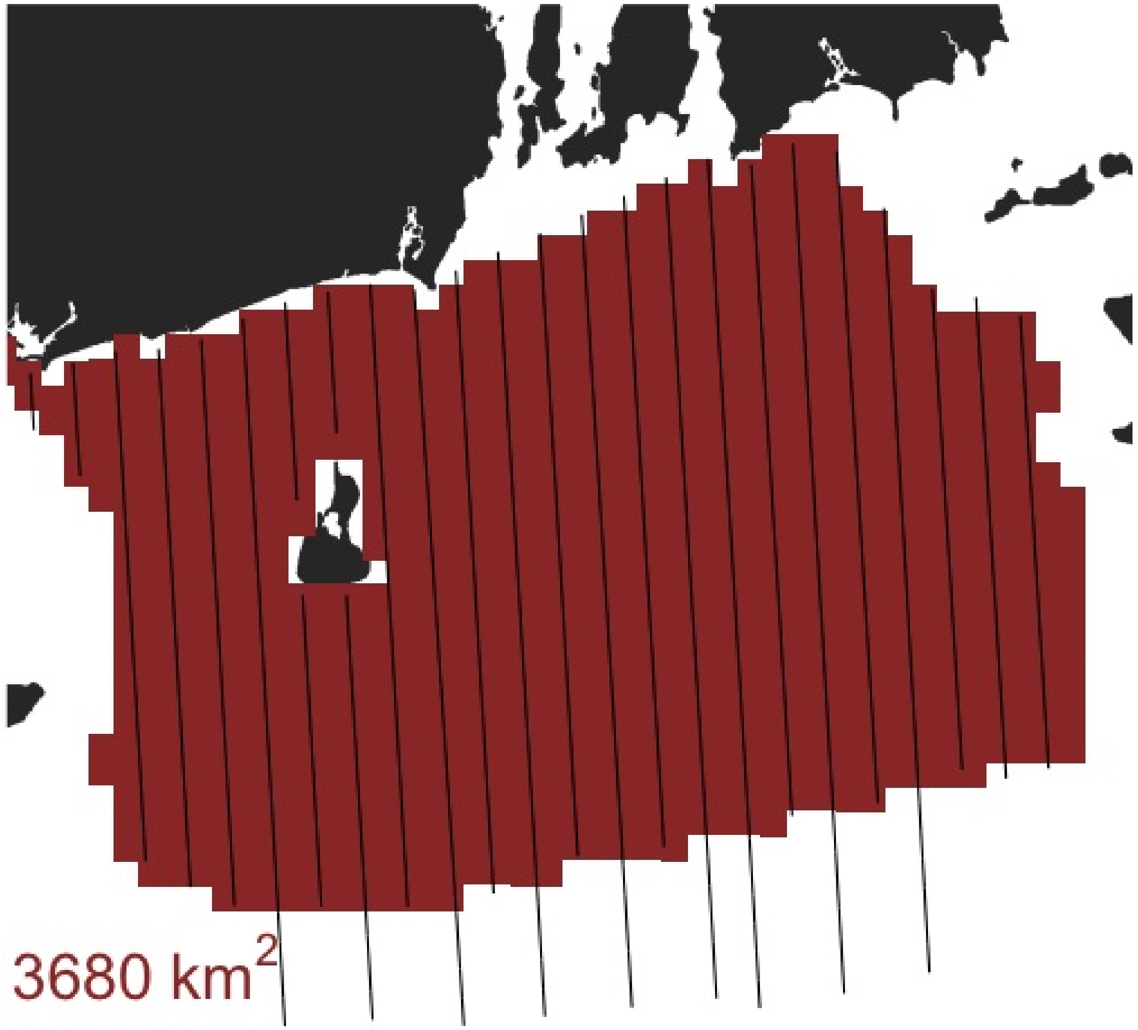
Motivating example

Rhode Island loons

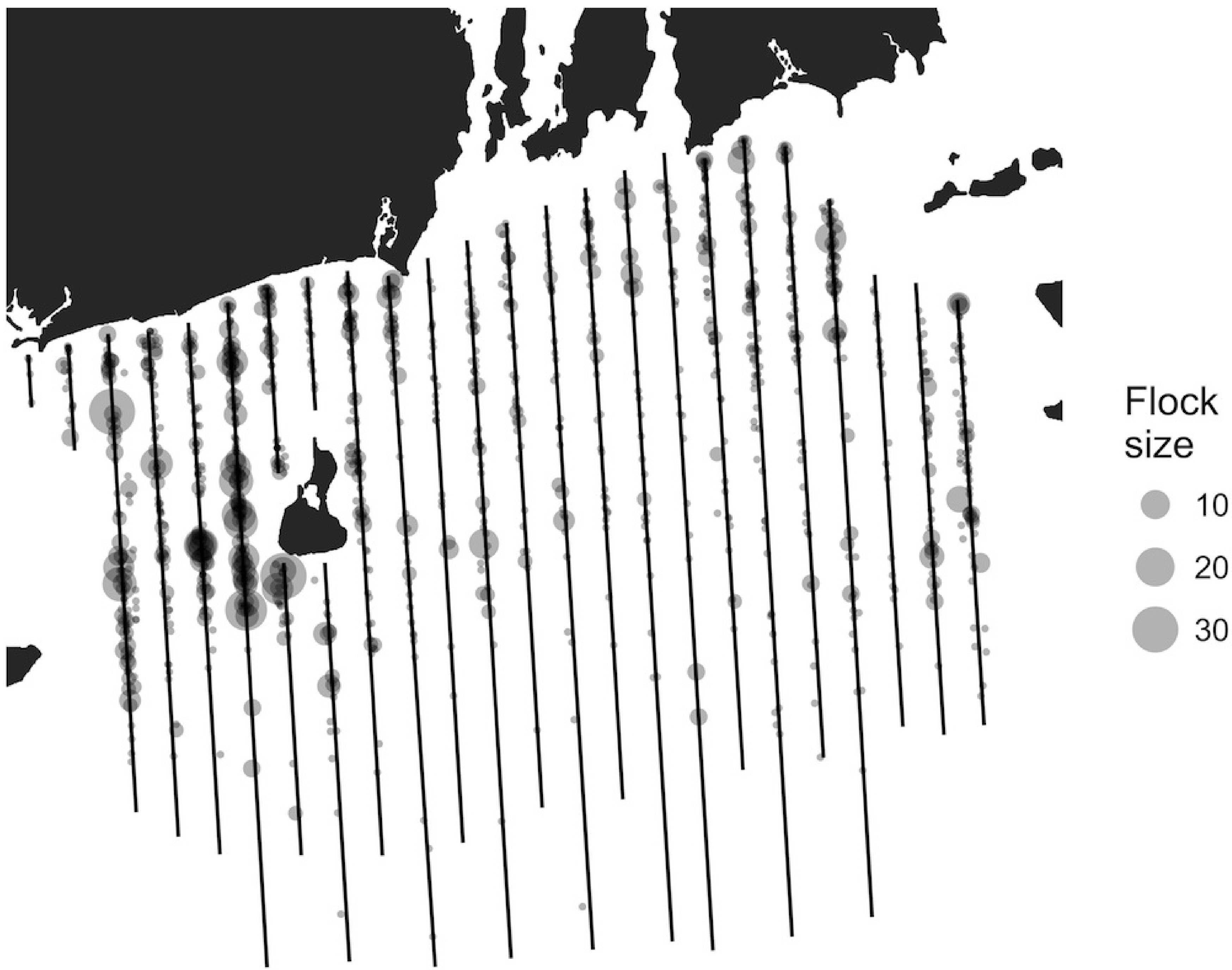
- Wind development off RI/MA
- Map usage
- Estimate abundance
- Estimate uncertainty

Aerial surveys

- Small plane
- 2 observers (1 each side)
- Record on dictaphone
- Transcribe later
- Glare an issue (roughly N-S transects)

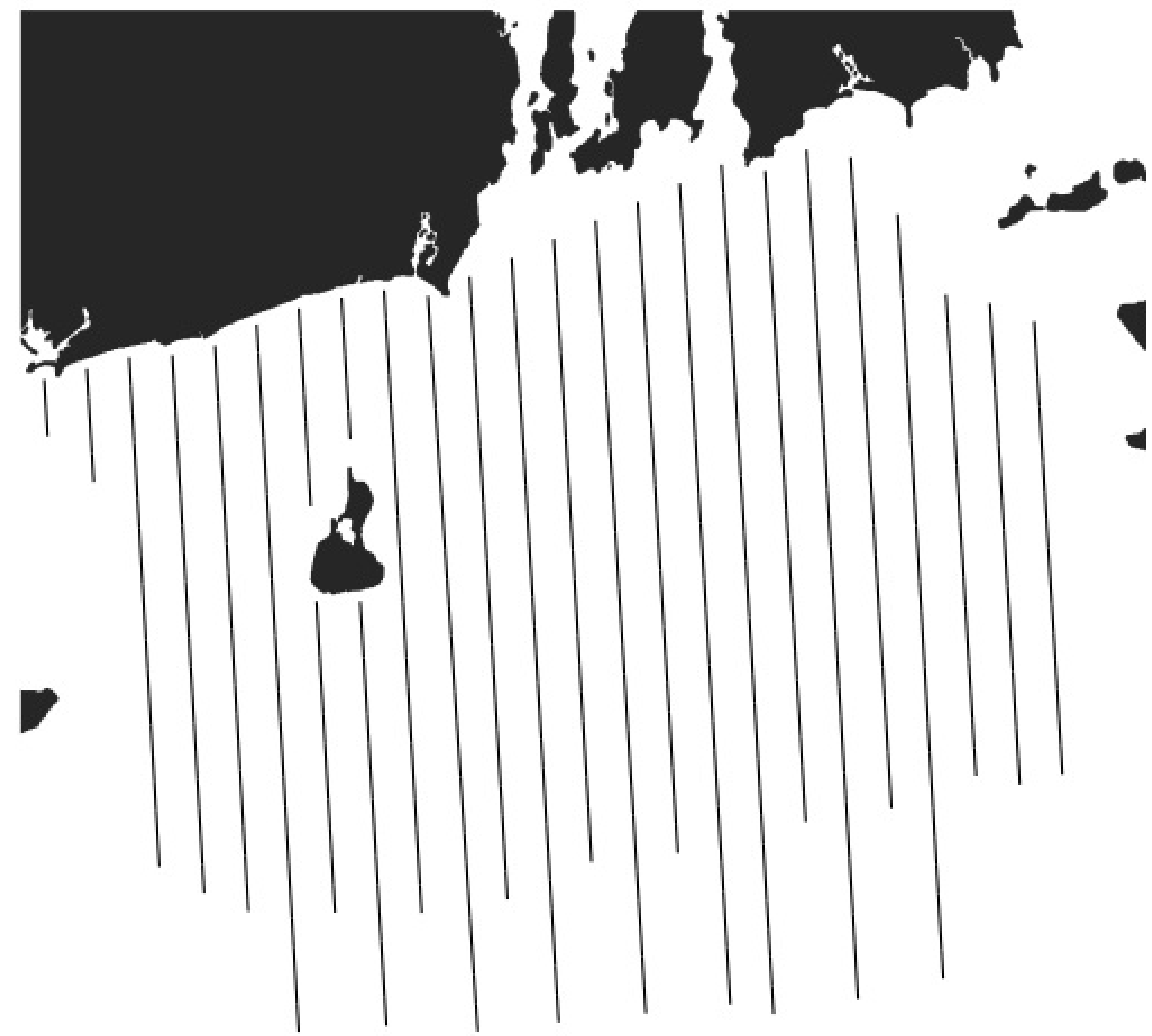


Non-uniform spatial distribution



What can we do?

- Stratification is not the solution
  - Arbitrary
  - Discard information



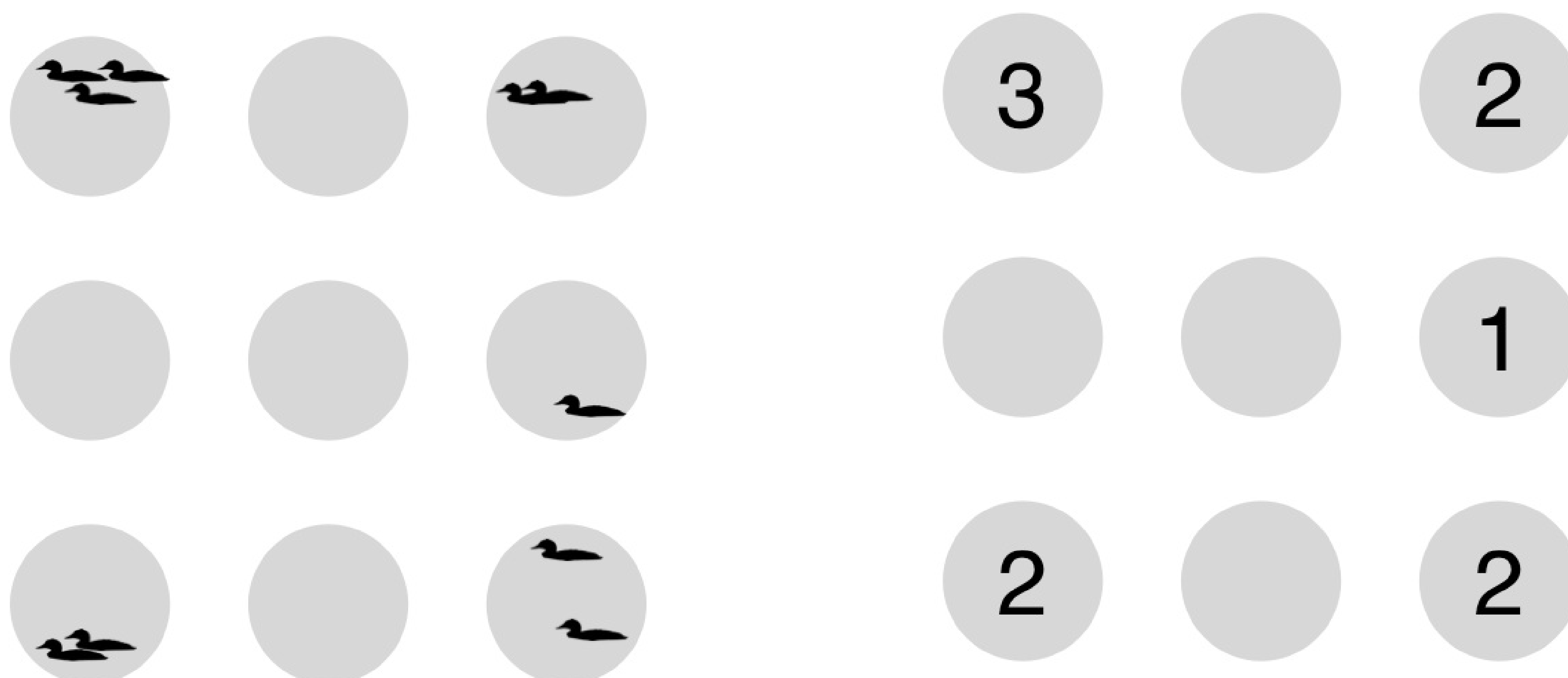
Aims

- Want to estimate abundance
- Distribution in space can be important
    - (Don't always have enough information to estimate this)
- What *correlates* with distribution?
    - Environmental covariates
- Still doing *distance sampling*!
    - Need to account for detectability

Putting together distance sampling and spatial modelling

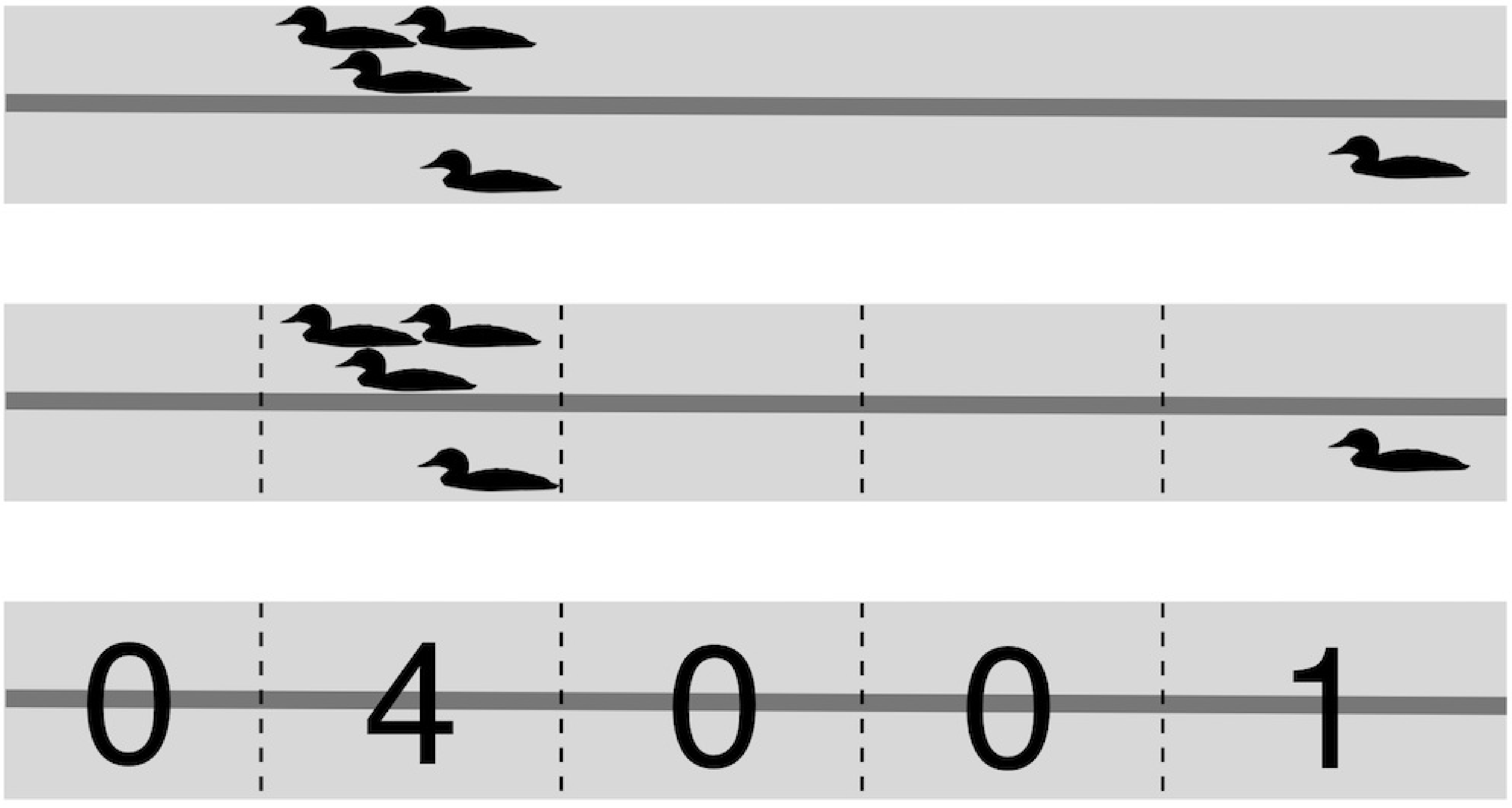
Making our data spatial -- points

- Points
  - Know point location
  - Know number of animals seen per point



Making our data spatial -- lines

- Lines
  - Lines are **too long**
  - Can't use as sample units
  - *Segment* the lines



Including detectability

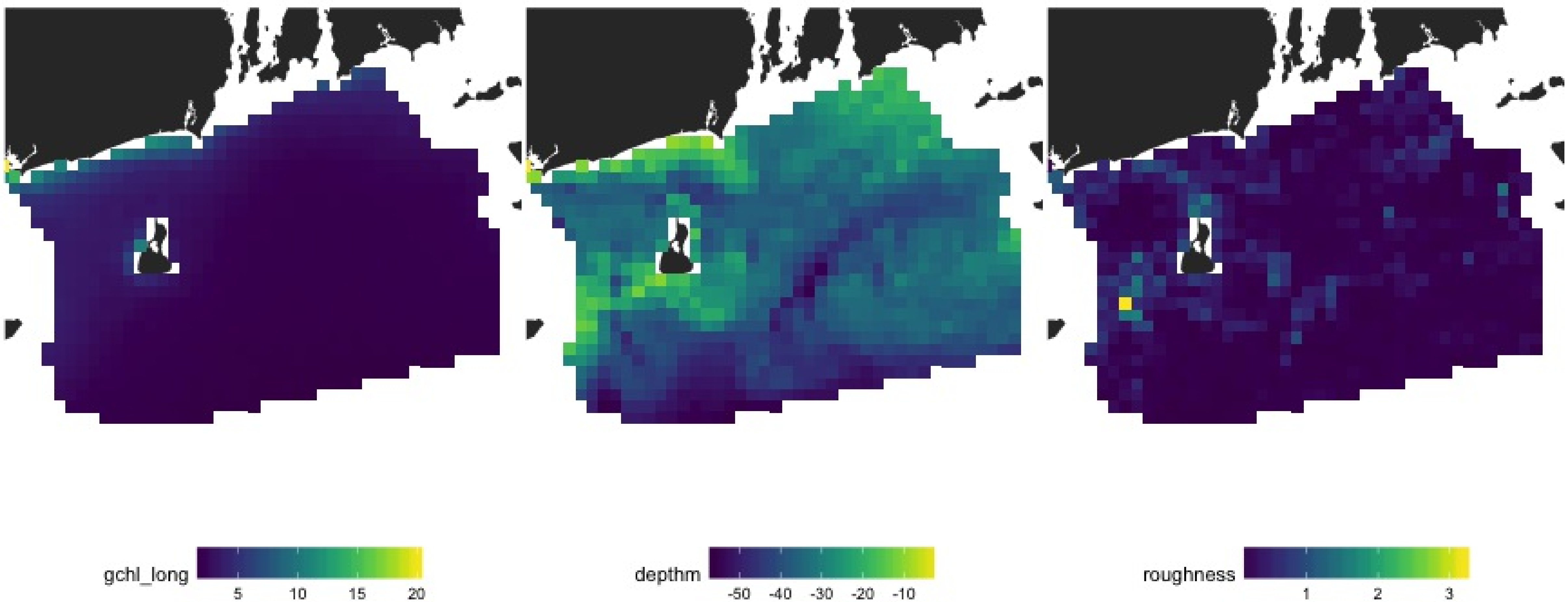
- Going back to *effective strip width*
    - think: *effective area*
- Effective areas are then:
    - $2wl_j\hat{p}_j$ for lines
    - $\pi w^2\hat{p}_j$ for points

How do we include this in our models?

Building spatial models

- Want to relate *counts* per *sample unit* to covariates
- Covariates like: space, also environmental covariates?
- Spatial modelling using Generalized Additive Models (GAMs)
- 2 stage approach
  - get the detection function right
  - then model spatial distribution

Spatially-referenced data



Spatial modelling

Generalised additive models

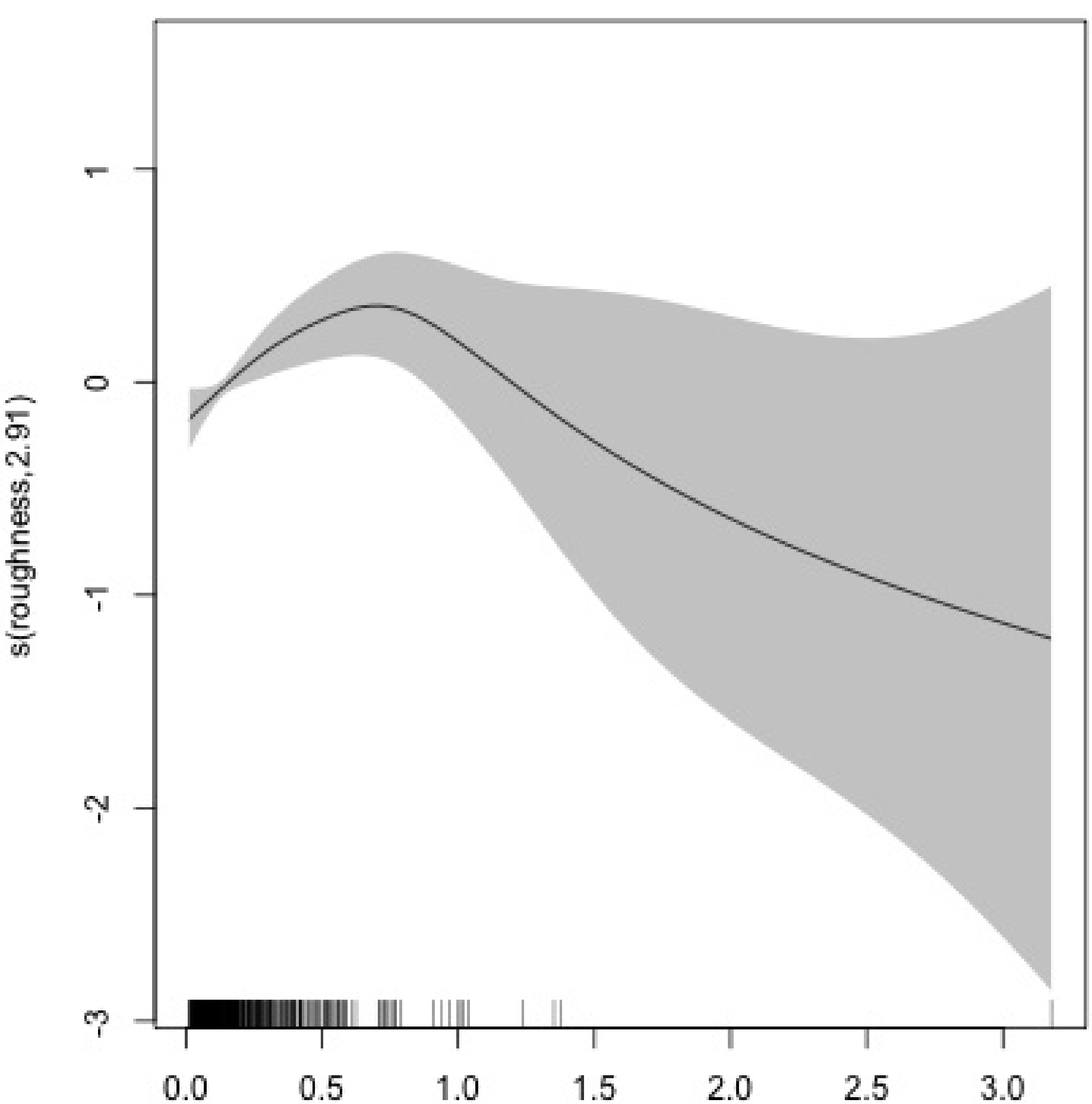
$$\mathbb{E}\left(n_j\right) = A_j \hat{p}_j \exp\left[\beta_0 + \sum_k s_k(z_{kj})\right]$$

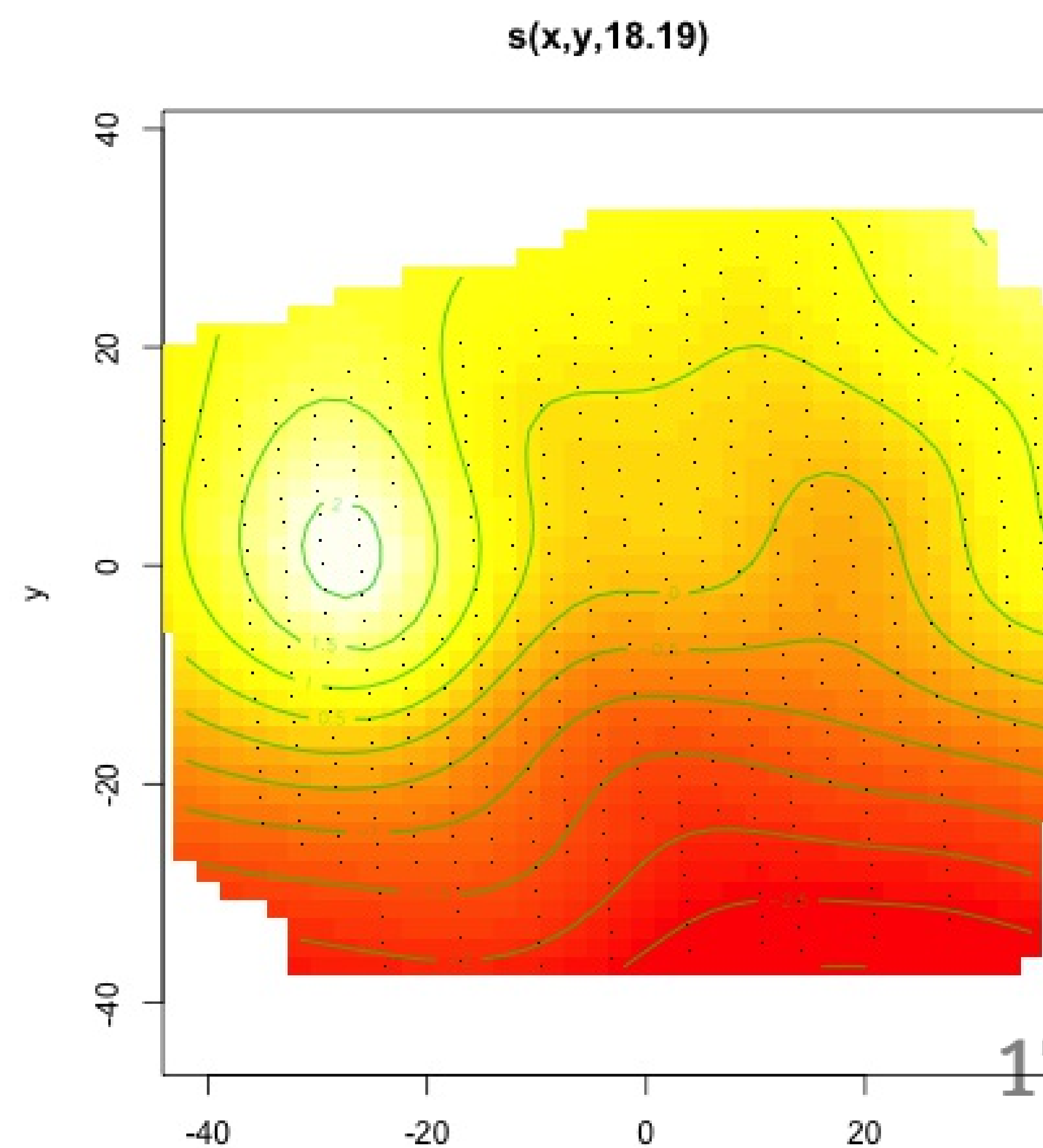
$n_j \sim$ some count distribution

- area of segment or circle
- probability of detection in segment
- (inverse) link function
- model terms

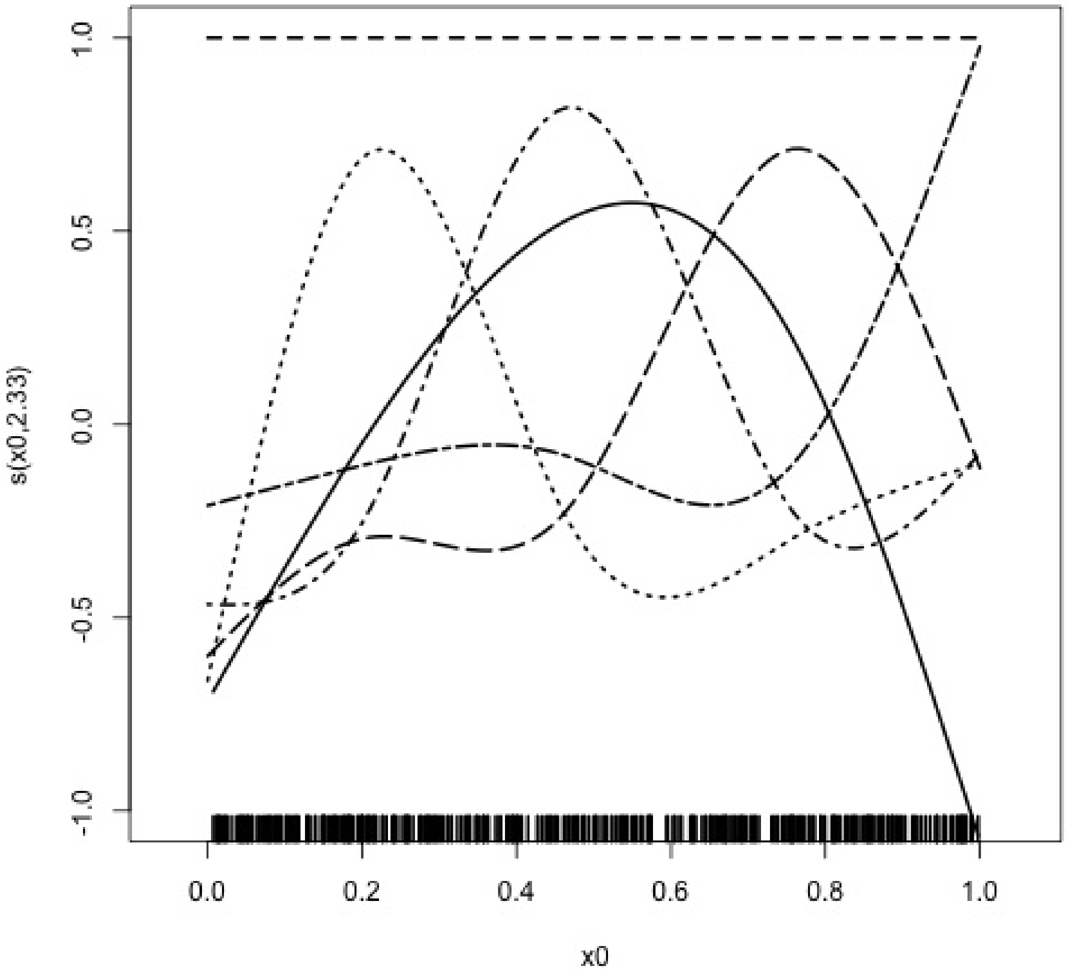
What are those s thingos?

- "Smooth" functions of (spatial?) covariates
- Could be univariate, could be multivariate





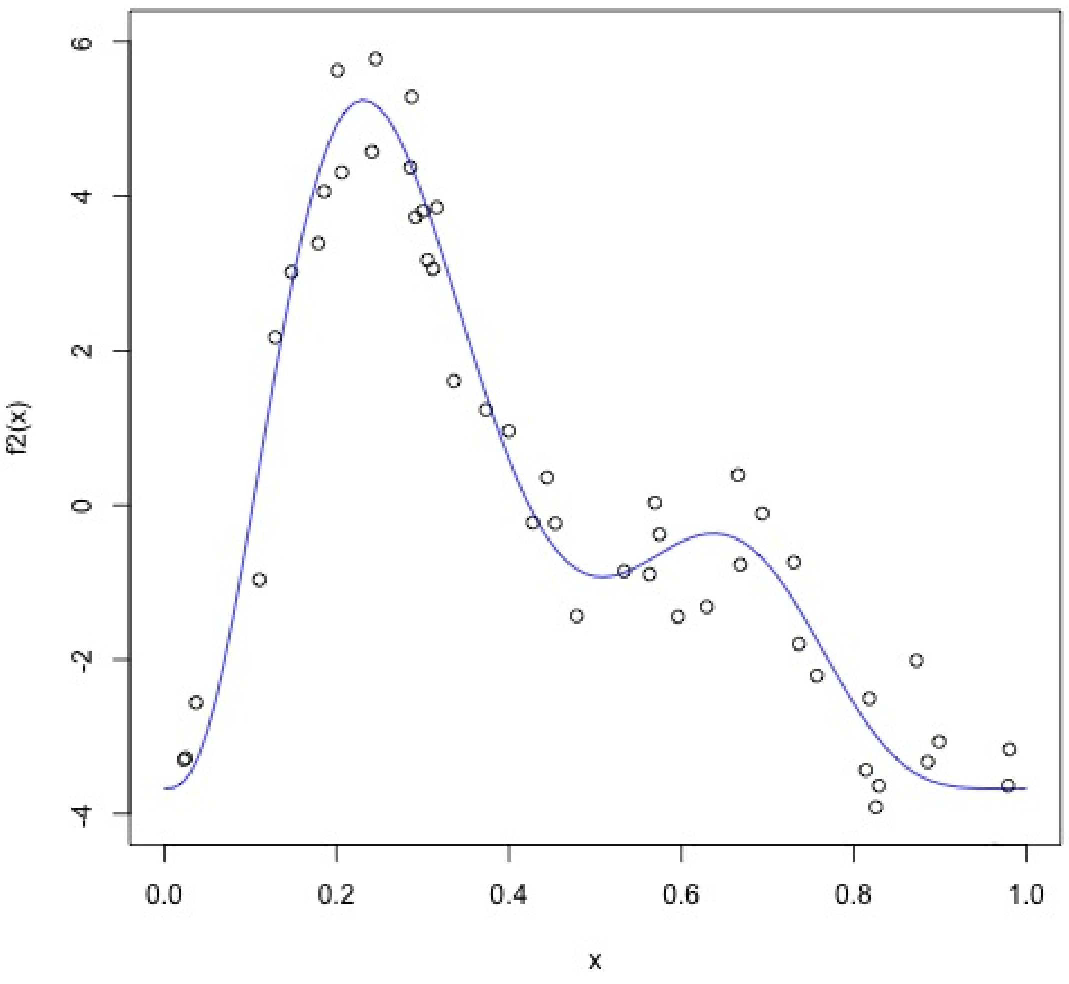
How do we build them?



- Functions made of other, simpler functions
- **Basis functions** b_k**,** estimate β_k

$$s(x) = \sum_{k=1}^{K} \beta_k b_k(x)$$

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 **generality**

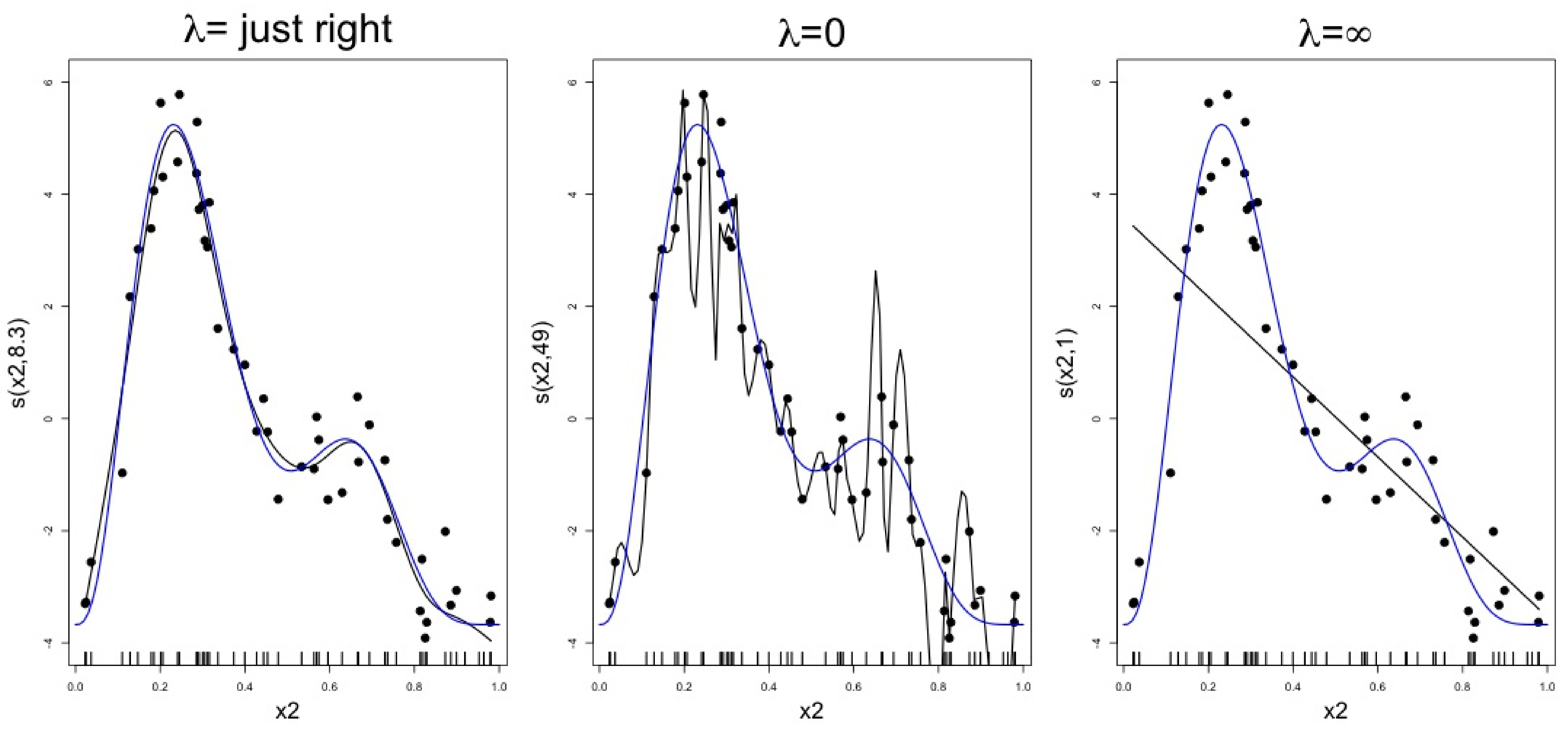
Making wigglyness matter

- Fit needs to be **penalised**
- What should we penalize? How wiggly the function is!
- *Something* like:

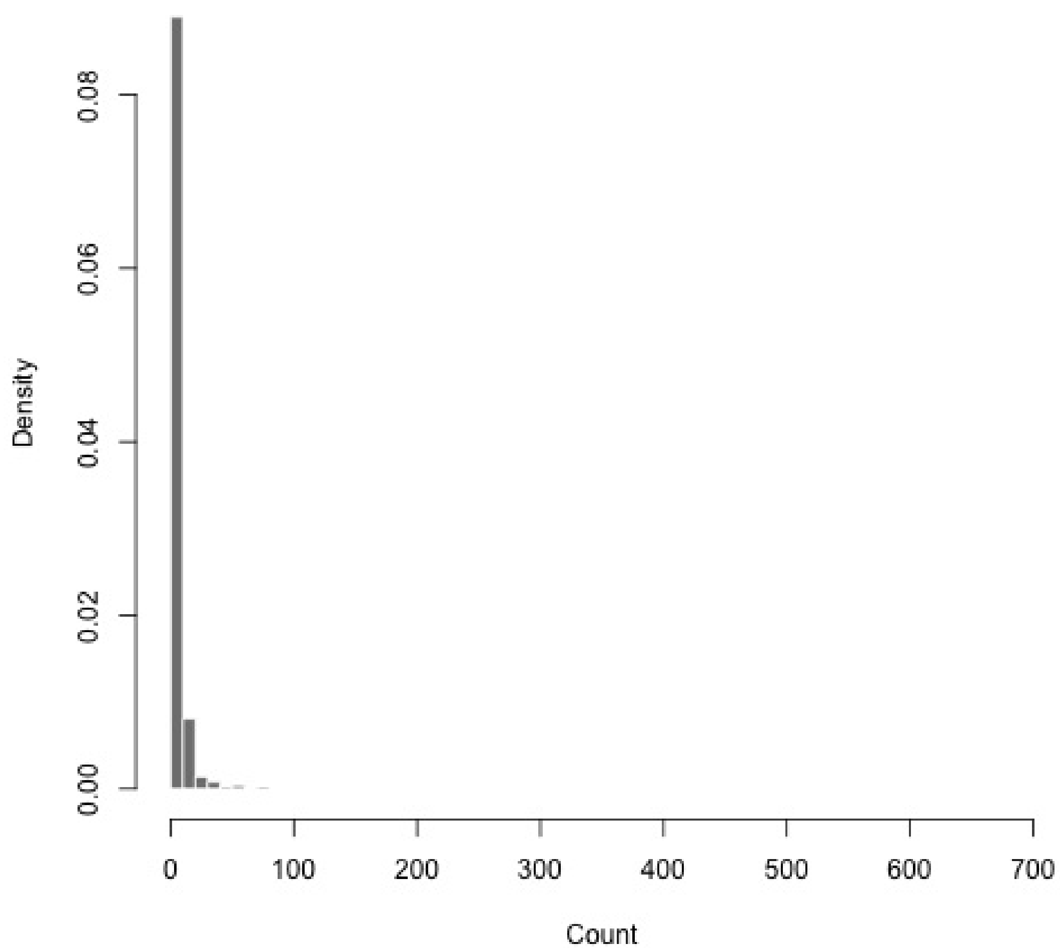
$$\int_{\mathbb{R}} \left(\frac{\partial^2 s(x)}{\partial x^2} \right)^2 \mathrm{d}x$$

- Estimate the β_k terms but penalise objective
  - "closeness to data" + penalty (REML?)

Smoothing parameter

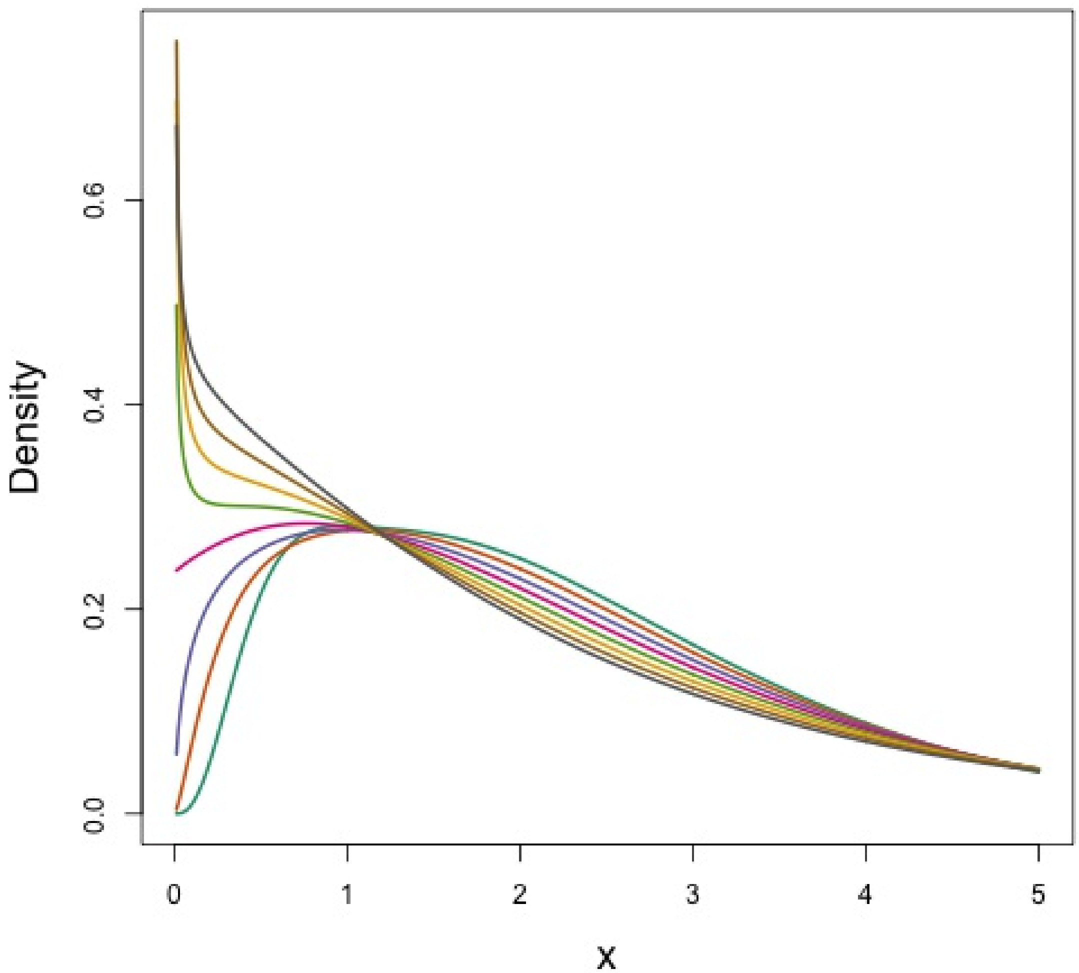


Count distributions



- Response is a count (not not always integer)
- Often, there are a lot of zeros
- Want response distribution that deals with that
- Flexible mean-variance relationship

Tweedie distribution



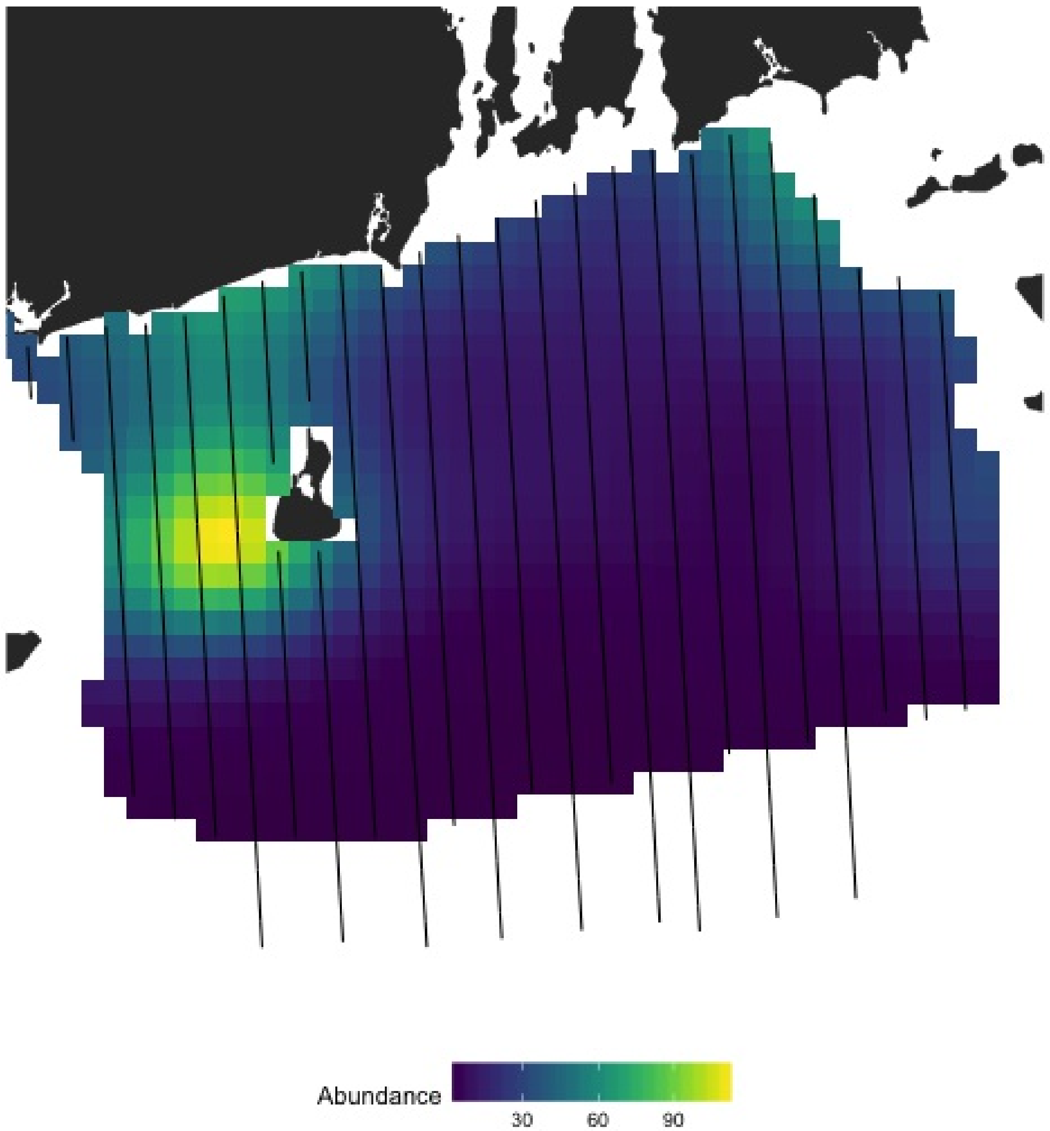
- $\text{Var}\,(\text{count}) = \phi\mathbb{E}(\text{count})^q$
- Common distributions are sub-cases:
  - $q = 1 \Rightarrow$ Poisson
  - $q = 2 \Rightarrow$ Gamma
  - $q = 3 \Rightarrow$ Normal
- We are interested in $1 < q < 2$
- (here $q = 1.2, 1.3, \ldots, 1.9$)

Abundance estimation

- If we have covariates available over a grid evaluate equation

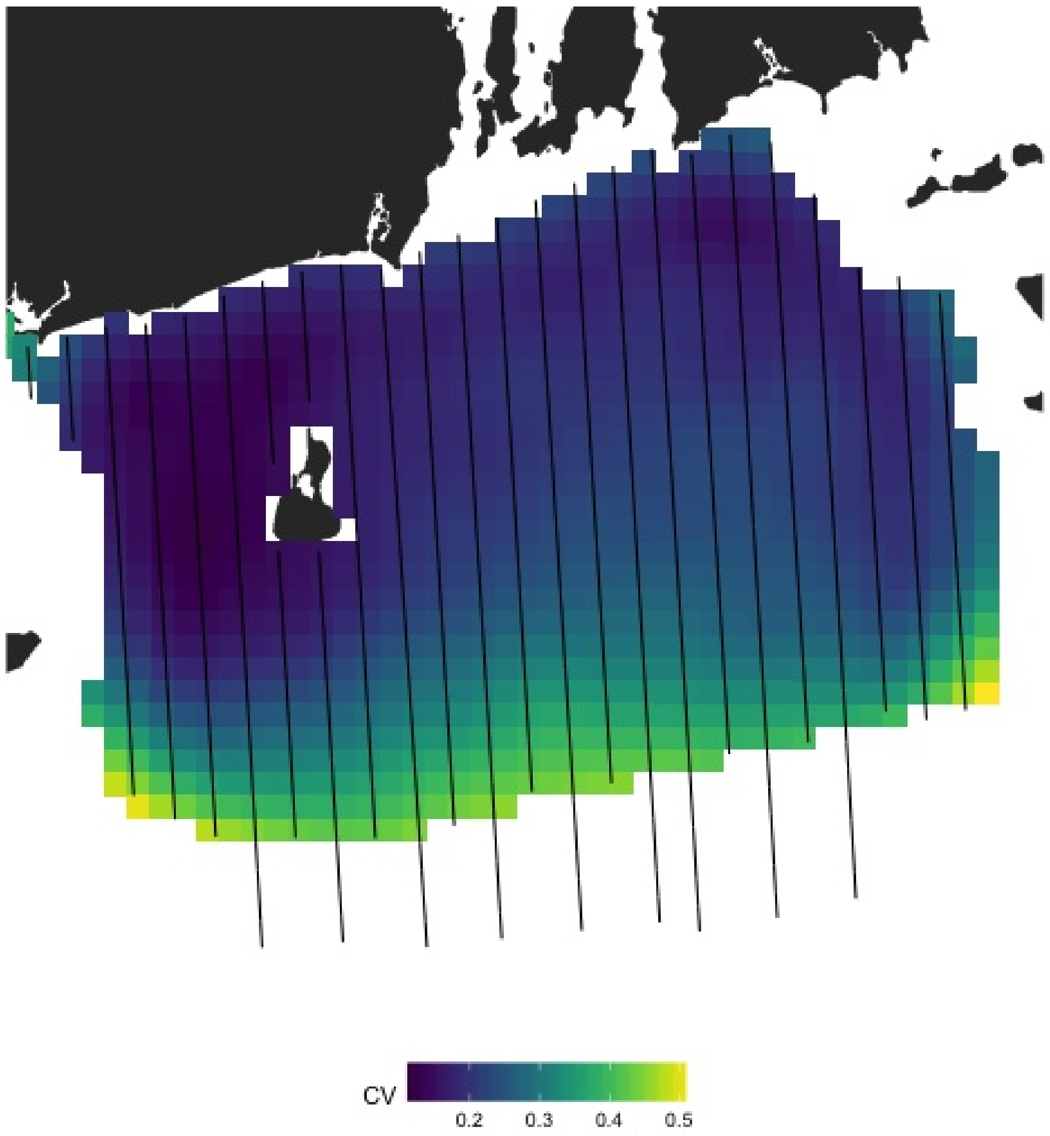
$$n_p = A_p \exp\left[\beta_0 + s_{x,y}(x_p, y_p) + s_{\text{depth}}(\text{depth}_p)\right]$$

- Sum grid cells for total abundance estimate



Uncertainty estimation

- Point estimates (even spatial ones) are not the end of the story
- Want to know how certain we are about our map
- Mapping uncertainty can help

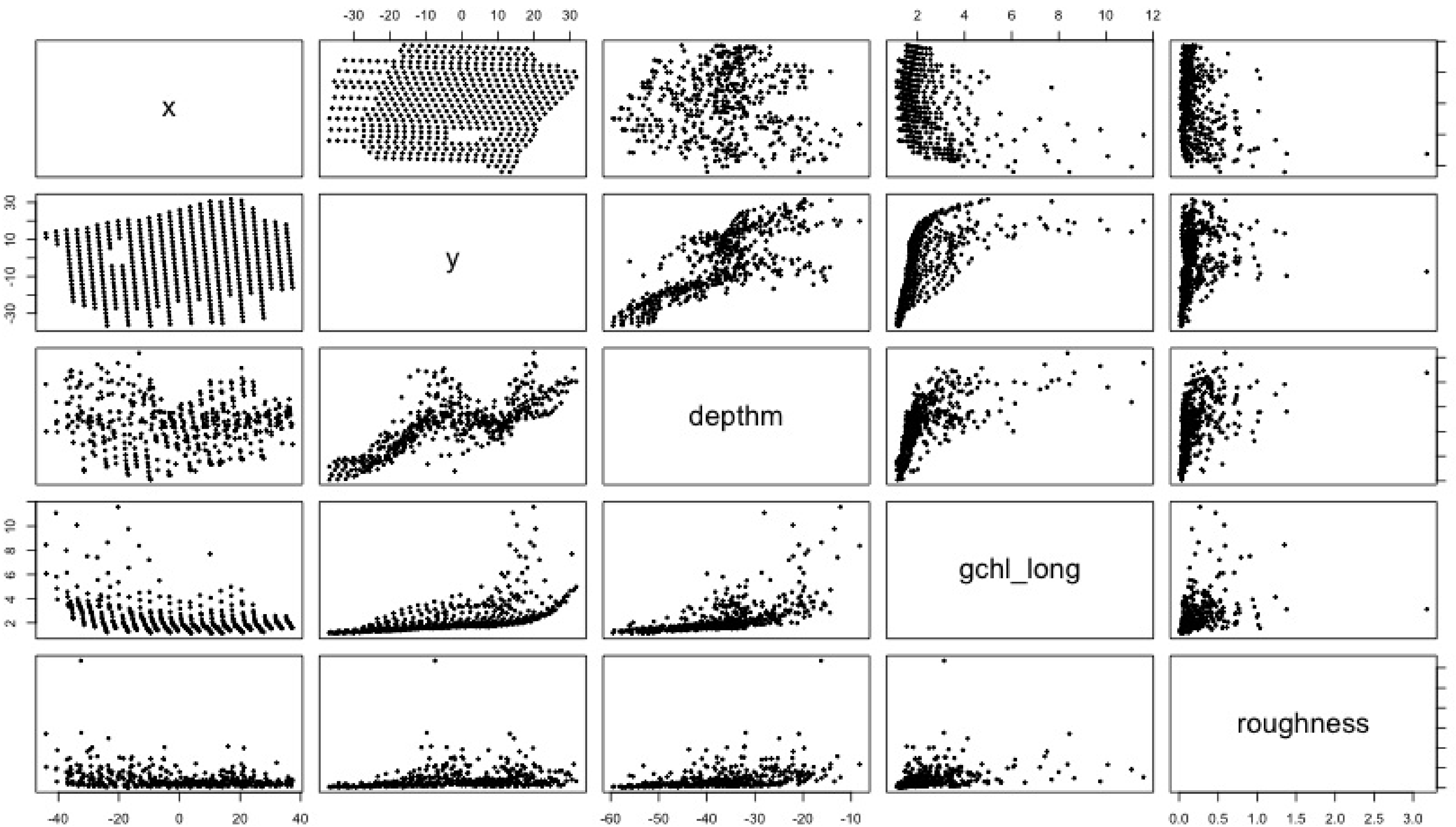


Potential pitfalls

Tobler's first law of geography

"Everything is related to everything else,
but near things are more related than distant things"

Implications of Tobler's law

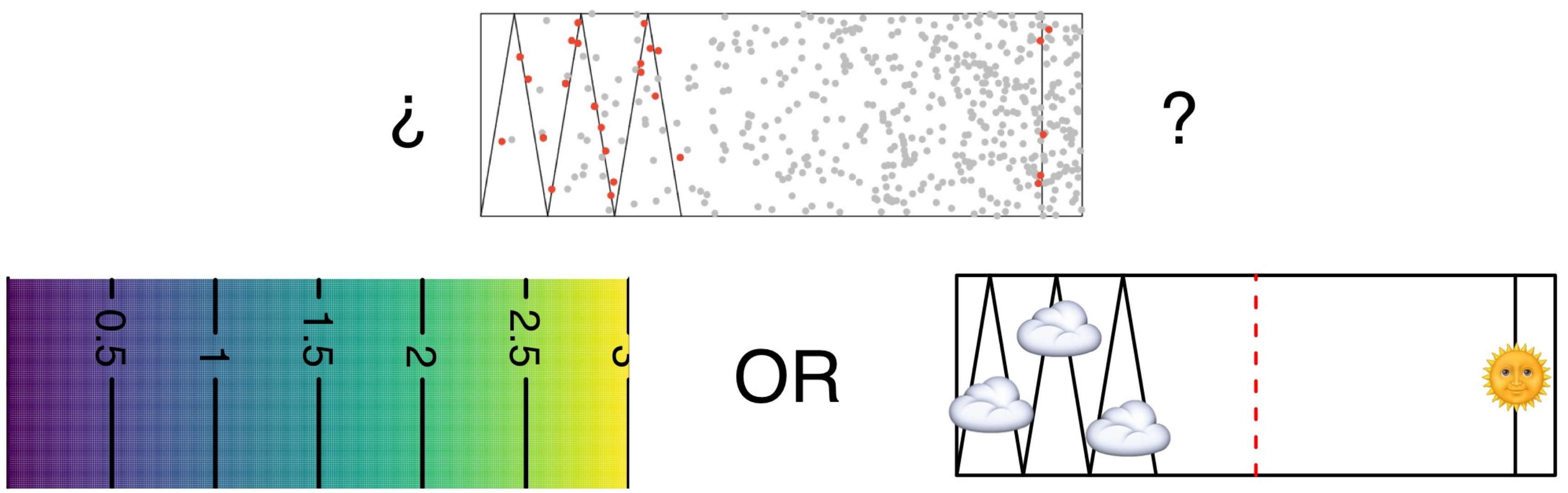


Selecting spatial covariates

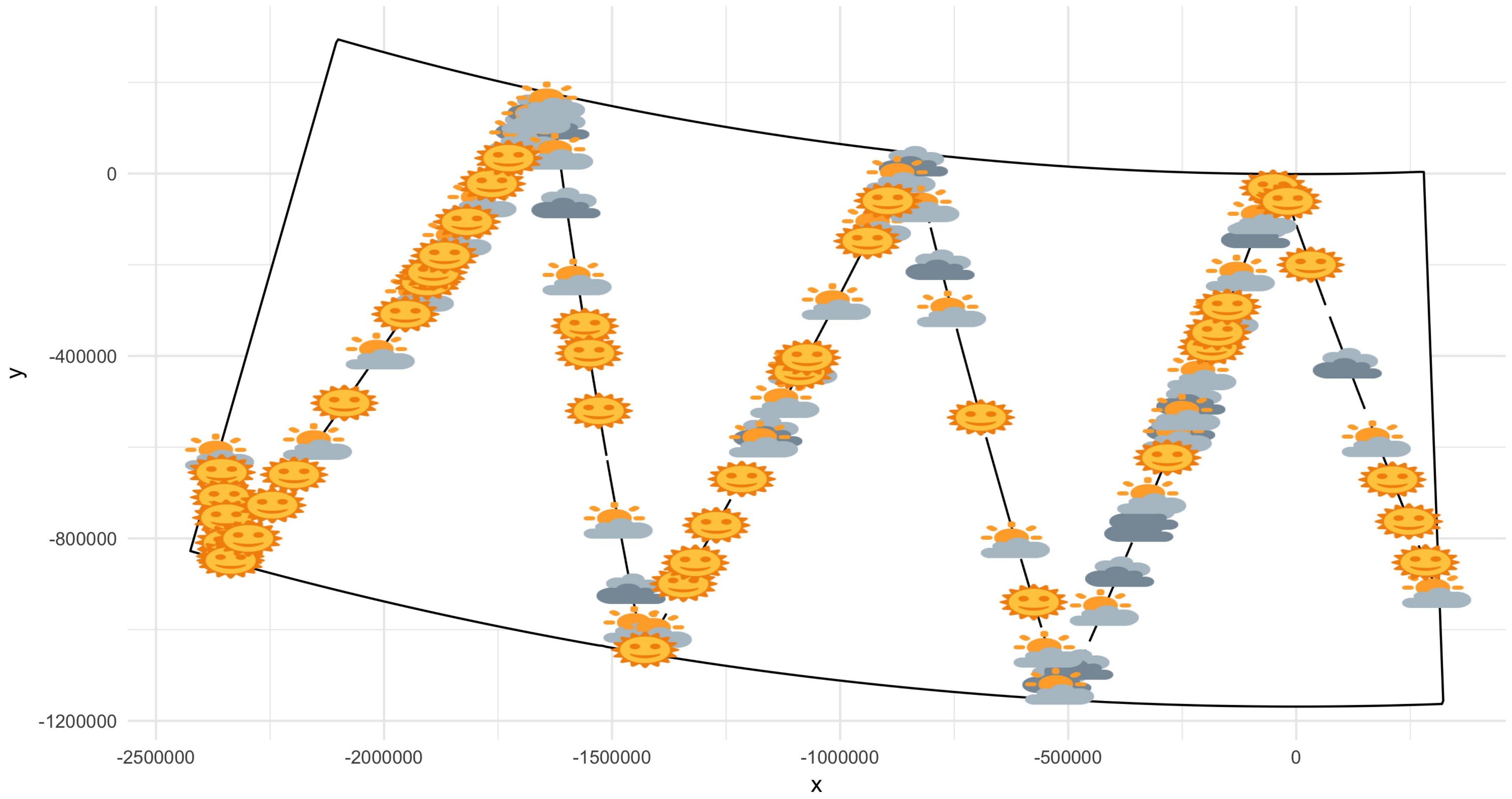
- Complicated topic
  - What are covariates really *doing*? See later
- Several strategies
  - Stepwise AIC selection (path dependence)
  - "Shrinkage" type approaches
  - (Emprical Bayes)
- Often smooths of location are adequate
  - And keep you honest (extrapolation)

Detection-level covariates

- Confounding between e.g., sea state and space important
- Pattern the same but abundance can change a LOT

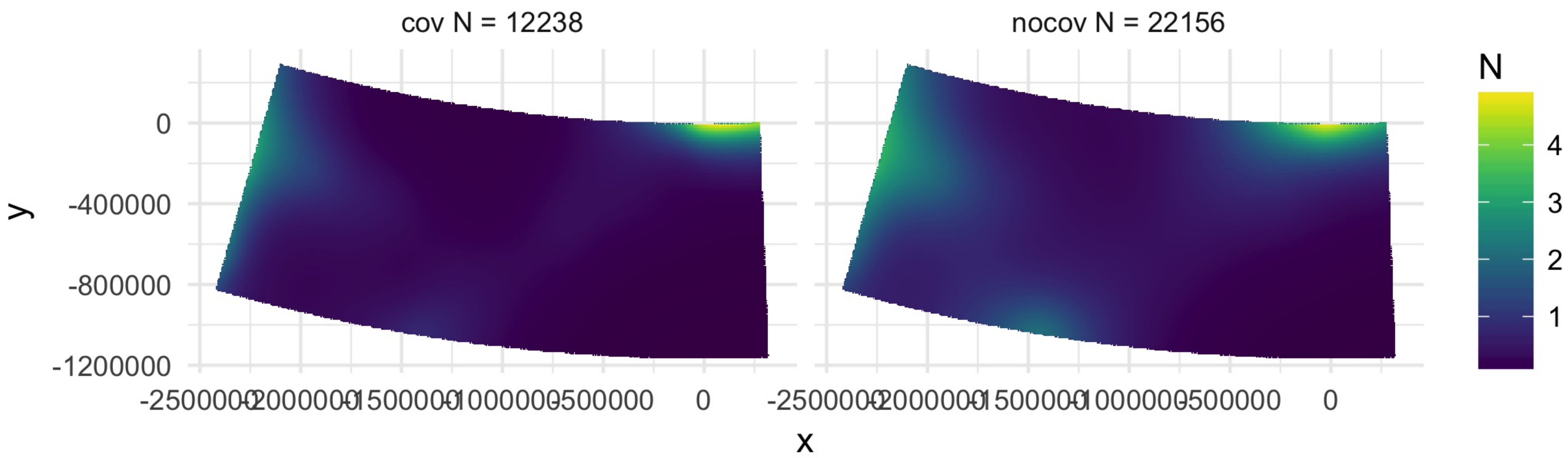


Visibility during POWER 2014

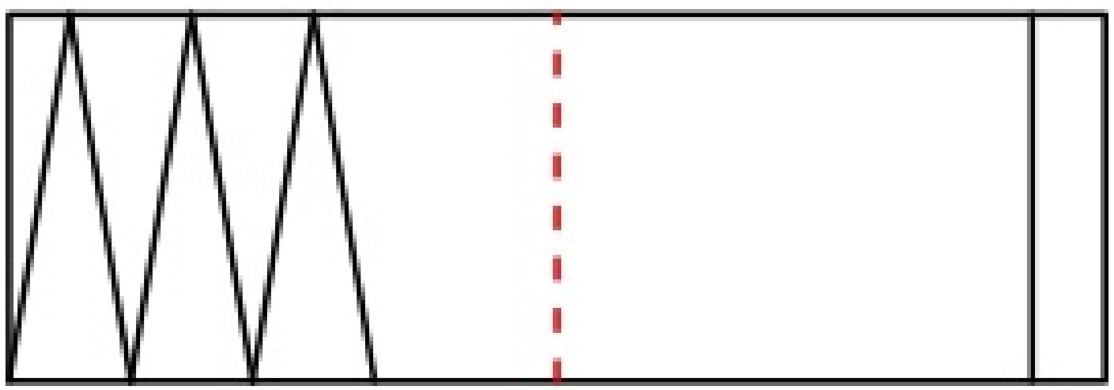


Thanks to Hiroto Murase and co for this data!

Covariates can make a big difference!



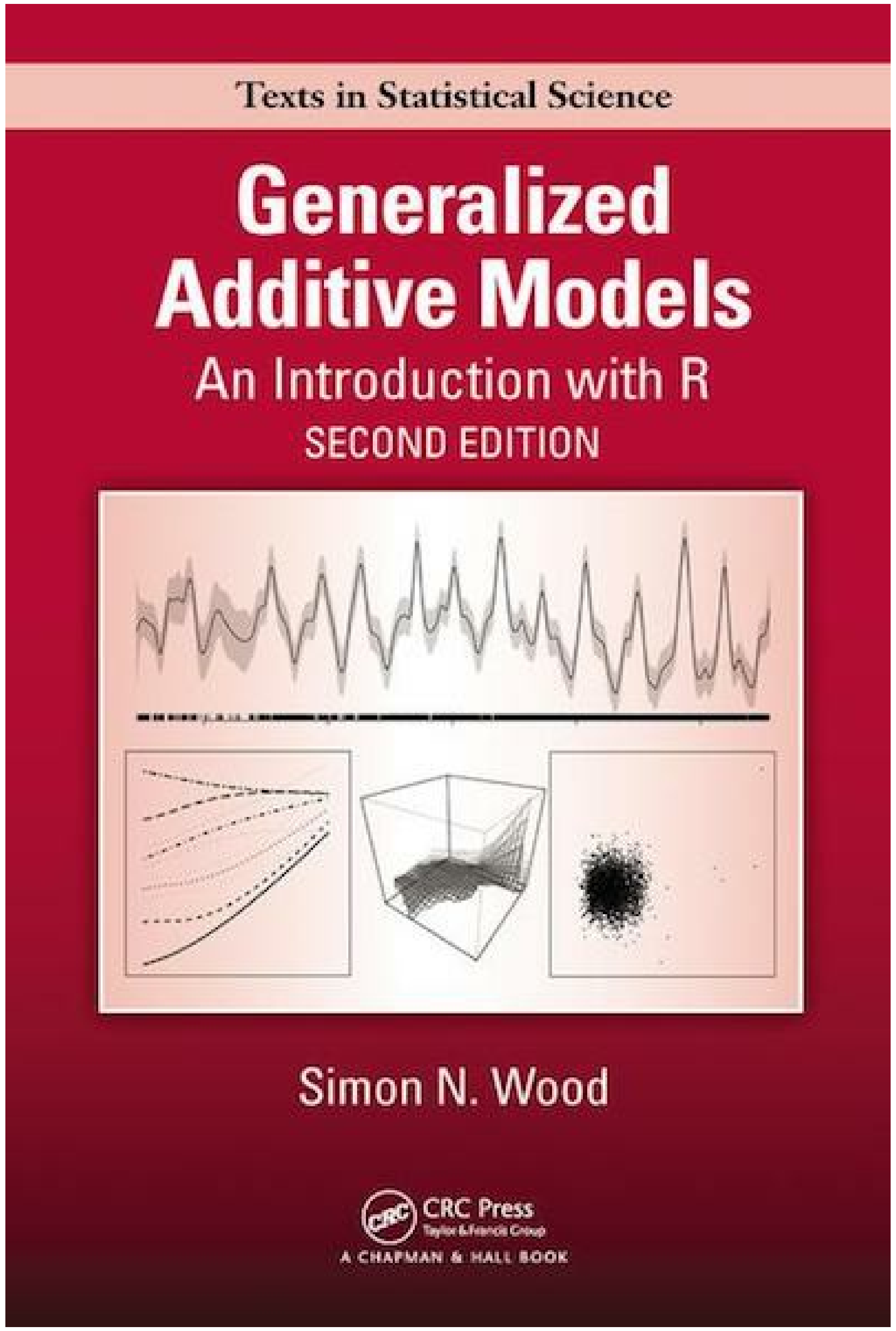
Spatial modelling won't solve all yr problems



- Design issues
  - Ludicrous extrapolation
  - Survey plan not robust to weather issues
  - Non-uniform distribution wrt sampler
  - Migration

Spatial models alone can't solve these issues

Resources



- Loon analysis:
    - *Distance Sampling* chapter in *Quantitative Analyses in Wildlife Science*, Buckland, Miller & Rexstad
    - Winiarski, Miller, Paton, and McWilliams (2013) *Spatially explicit model of wintering common loons: conservation implications*. Marine Ecology Progress Series.
- Miller, Burt, Rexstad and Thomas (2013) *Spatial models for distance sampling data: recent developments and future directions*. Methods in Ecology and Evolution.

I am going to stop talking very soon...

Summary

- We often collect spatially-explicit data
- There is information there to harness

- This *won't* solve all the problems you have

- Use 2 stage approach

    - First get your detection function right
    - *Then* do spatial modelling
- Spatial modelling is a multi-headed hydra
- Lots of things to think about