Overview of density surface modelling

David L Miller & Mark V Bravington

International Whaling Commission Scientific Committee
2017

Why are we interested in spatially-explicit estimation?

Horvitz-Thompson estimation: the good, the bad and the ugly

Horvitz-Thompson-like estimators

- Rescale the (flat) density and extrapolate

$$\hat{N} = \frac{\text{study area}}{\text{covered area}} \sum_{i=1}^{n} \frac{s_i}{\hat{p}_i}$$

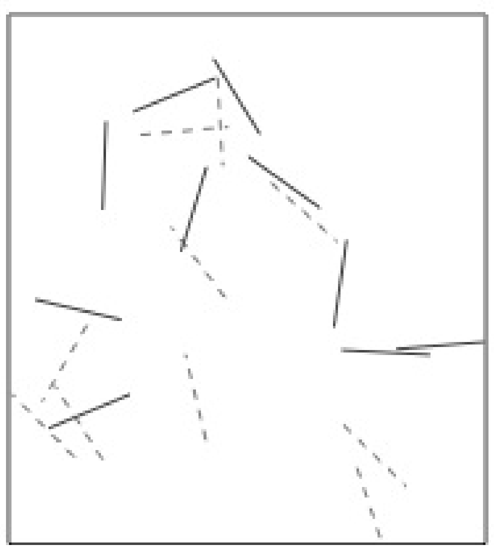
- s_i are group/cluster sizes
- $\hat{p}_i$ is the detection probability (from distance sampling)

Hidden in this formula is a simple assumption

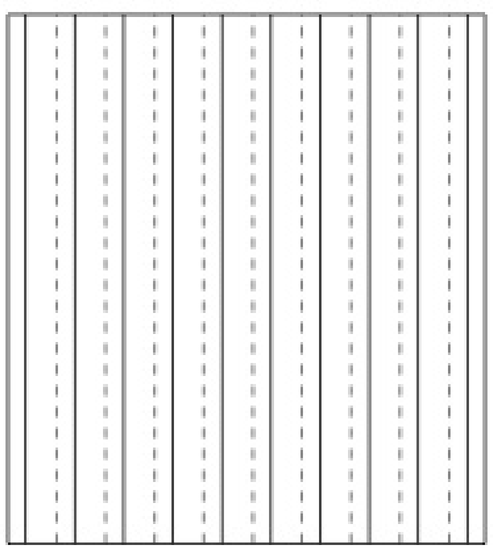
- Probability of sampling every point in the study area is equal
- Is this true? Sometimes.
- If (and only if) the design is randomised

Many faces of randomisation

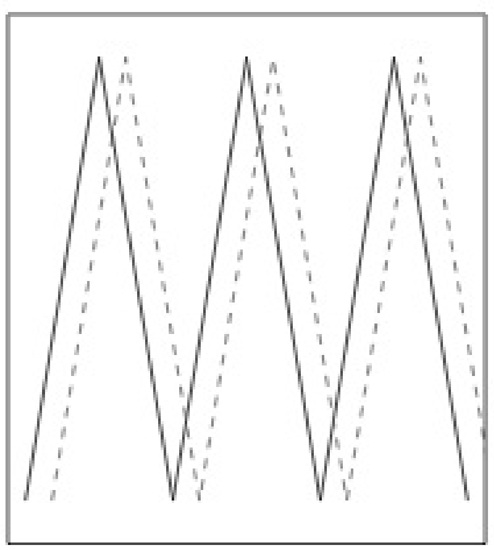
random placement



random offset parallel lines



random offset zigzag



What does this randomisation give us?

- Coverage probability
- H-T estimator assumes even coverage
- (or you can estimate)
- Otherwise not really valid

Estimating coverage

- We can estimate coverage of a non-uniform design!
- In Distance!

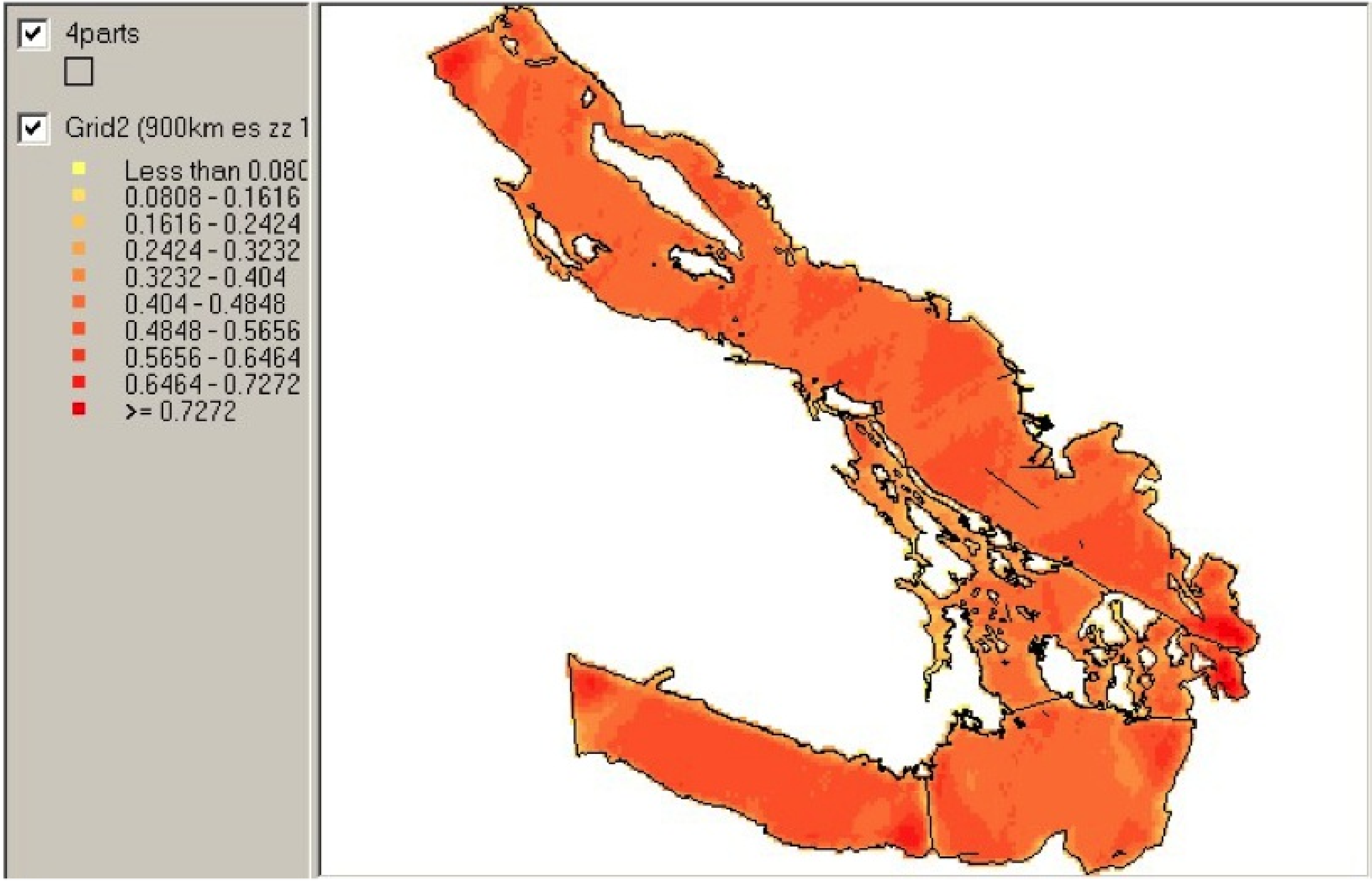
J. CETACEAN RES. MANAGE. 9(1):1–13, 2007 1

Designing line transect surveys for complex survey regions

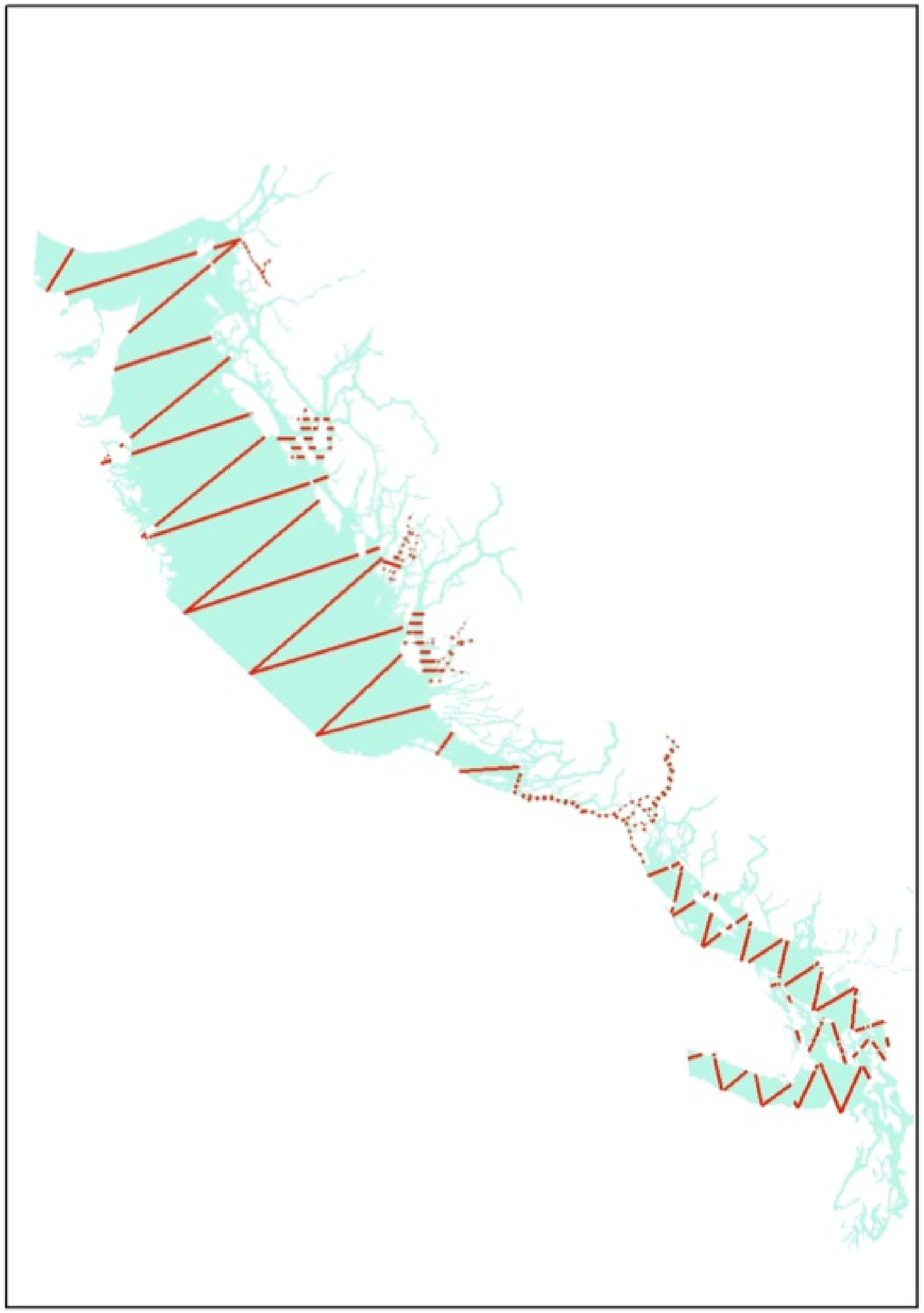
LEN THOMAS*, ROB WILLIAMS+# AND DOUG SANDILANDS++

Contact e-mail: len@mcs.st-and.ac.uk

Estimating coverage



A complex survey plan



- Thomas, Williams and Sandilands (2007)
- Different areas require different strategies
- Zig-zags, parallel lines, census
- Analysis in Distance

Sideline: alternative terminology

“A design is an algorithm for laying down samplers in the survey area”

“A realization (from that algorithm) is called a survey plan”

Len Thomas (Talk @CREEM 2004)

H-T estimation again

- Can't estimate w/ H-T w/o coverage
- “Fixed” “designs” violate assumptions
  - Some animals have $\mathbb{P}(\text{included}) = 0$
- “Deteriorate” pooling robustness property
- What can we do?

Spatial models

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

This is the rosey picture talk

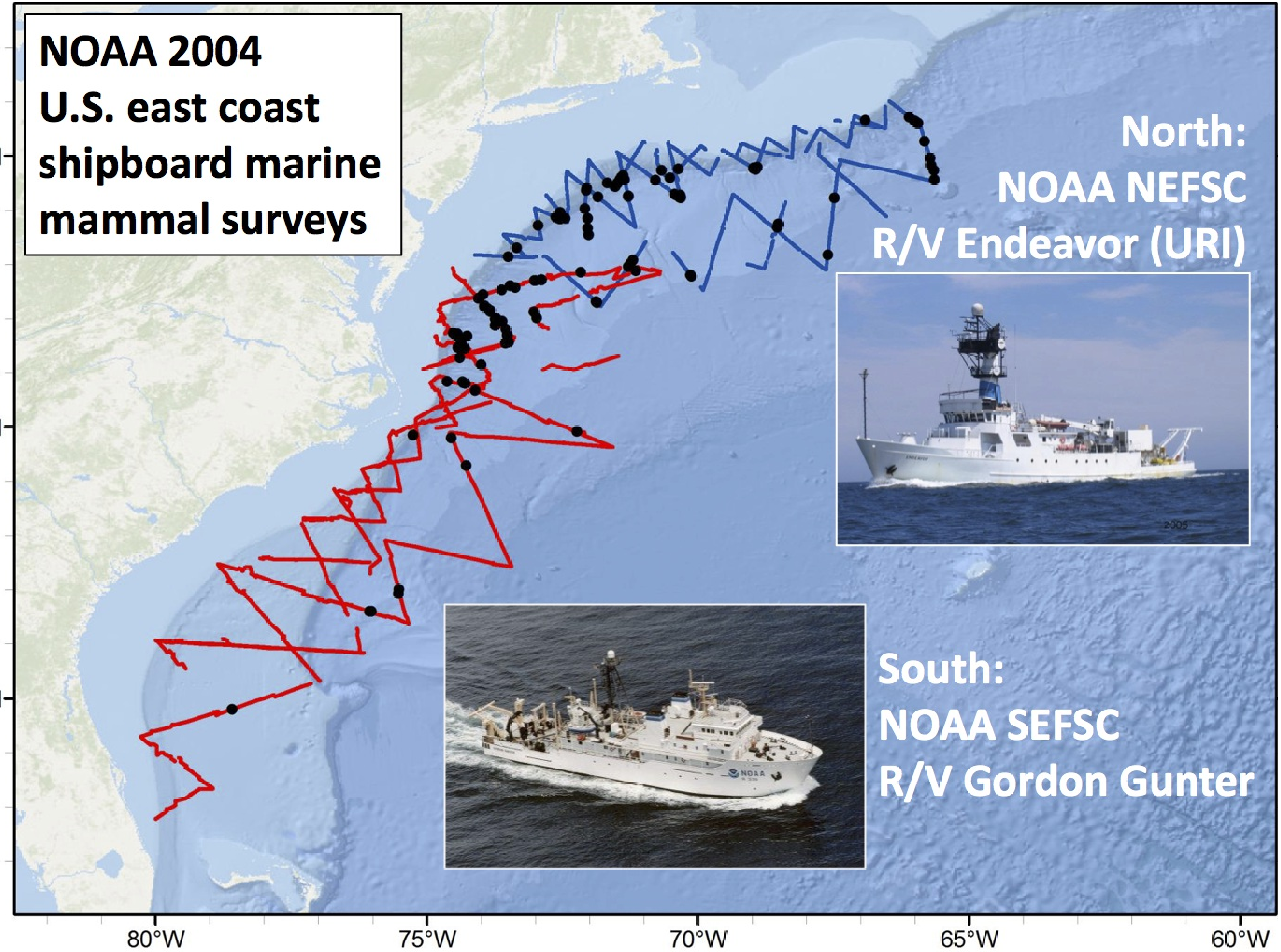
We’ll talk about the grim reality later

Example data in this talk

Sperm whales off the US east coast

- Hang out near canyons, eat squid
- Surveys in 2004, US east coast
- Combination of data from 2 NOAA cruises
- Thanks to Debi Palka, Lance Garrison for data. Jason Roberts for data prep.

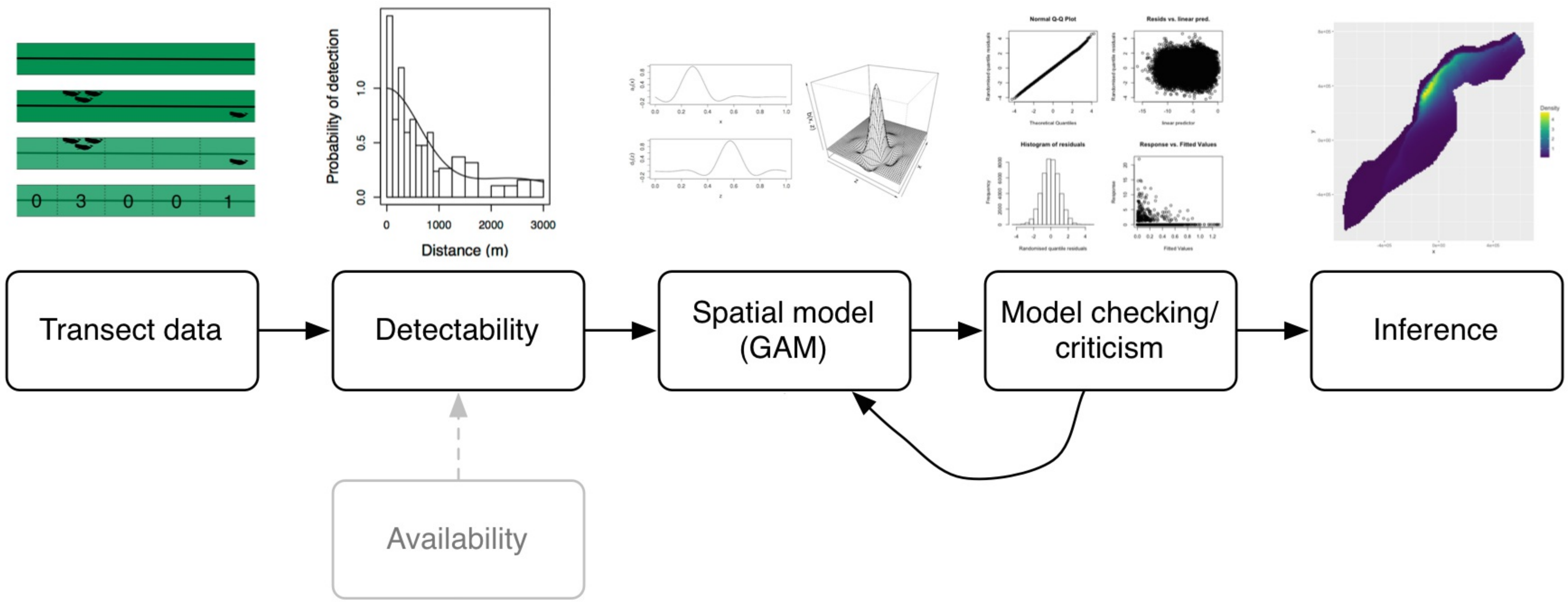
Example data



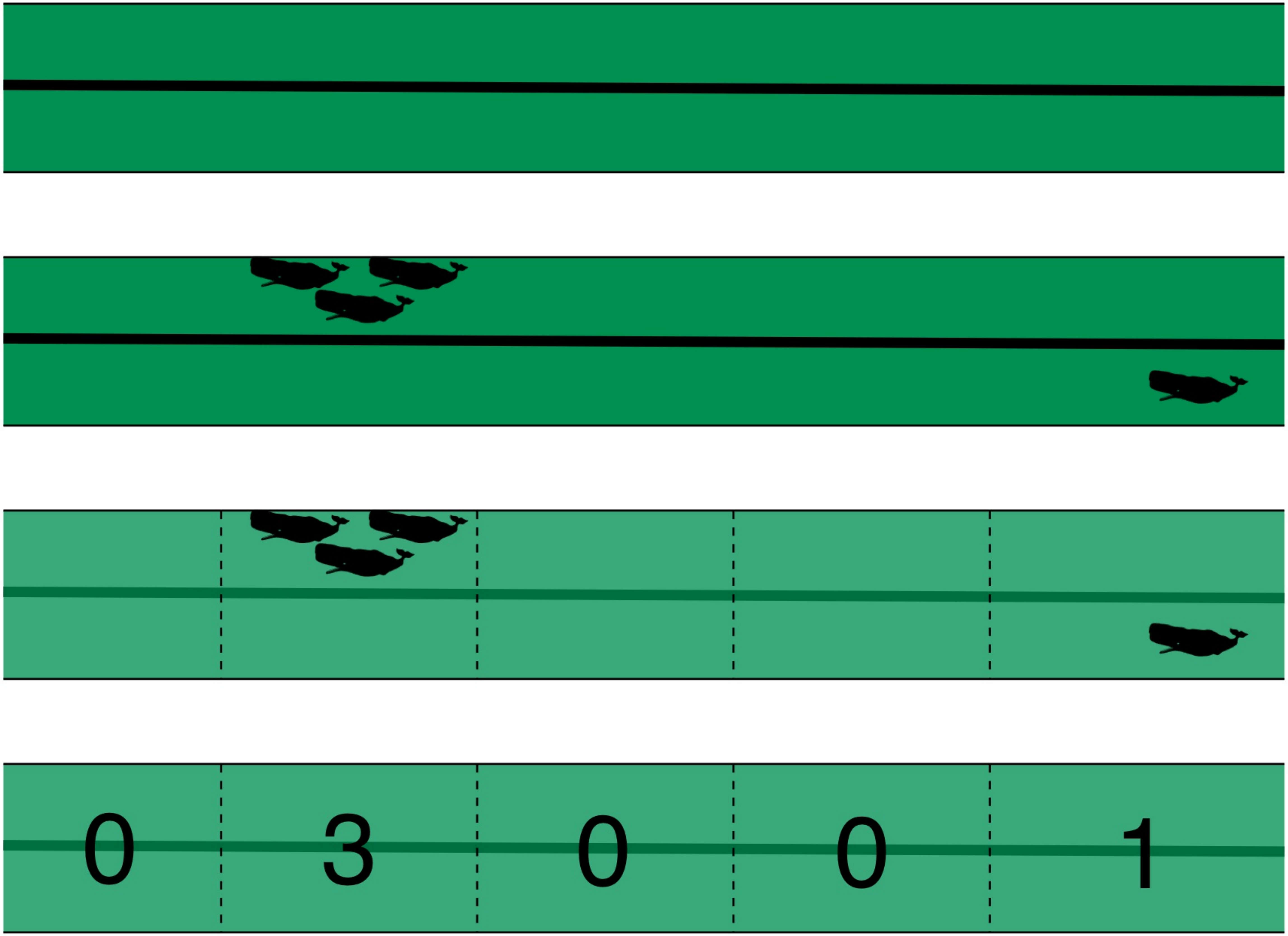
Density surface models

Hedley and Buckland (2004)

Miller et al. (2013)



Probability of detection
Distance (m)
Normal Q-Q Plot
Resids vs. linear pred.
Histogram of residuals
Response vs. Fitted Values
Density
Transect data
Detectability
Spatial model (GAM)
Model checking/ criticism
Inference
Availability



Physeter catodon by Noah Schlottman

How do we model that?

SPOILER ALERT: your model is probably just a very fancy GLM

Generalised additive models (in 1 slide)

Taking the previous example...

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

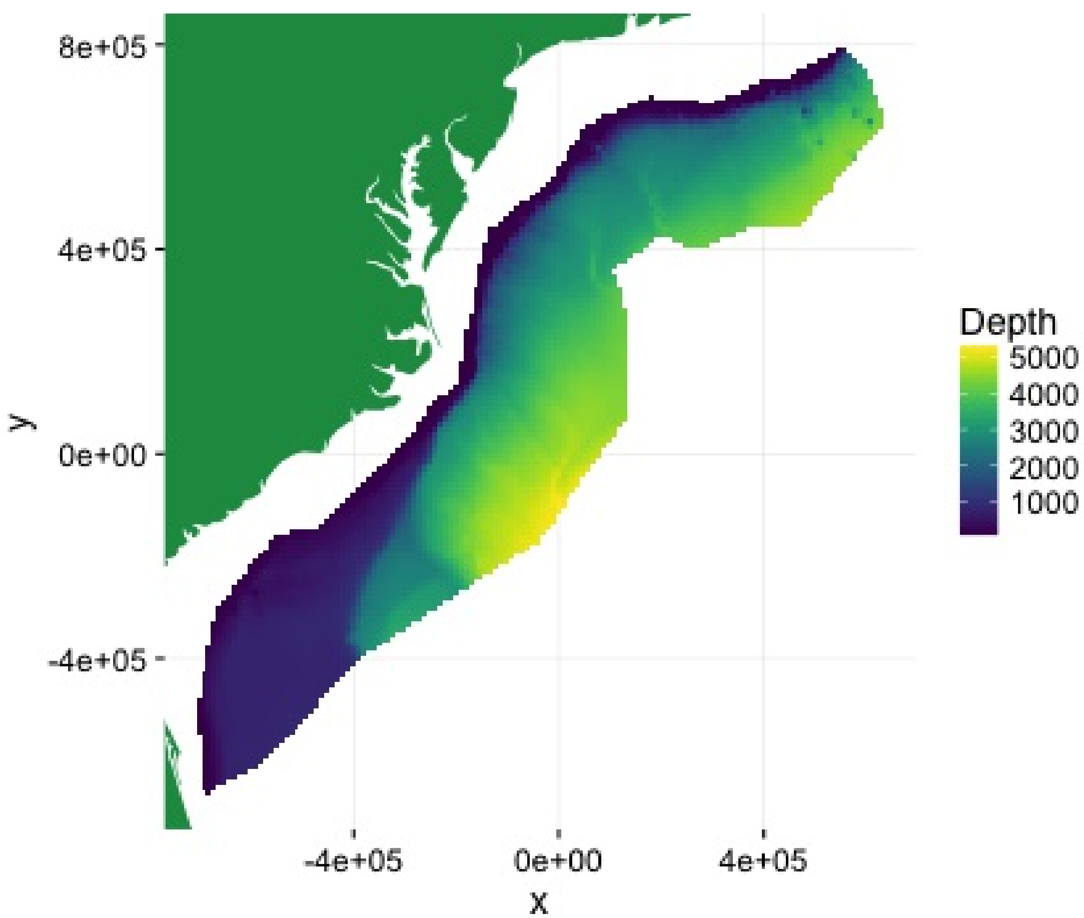
$n_j \sim$ some count distribution

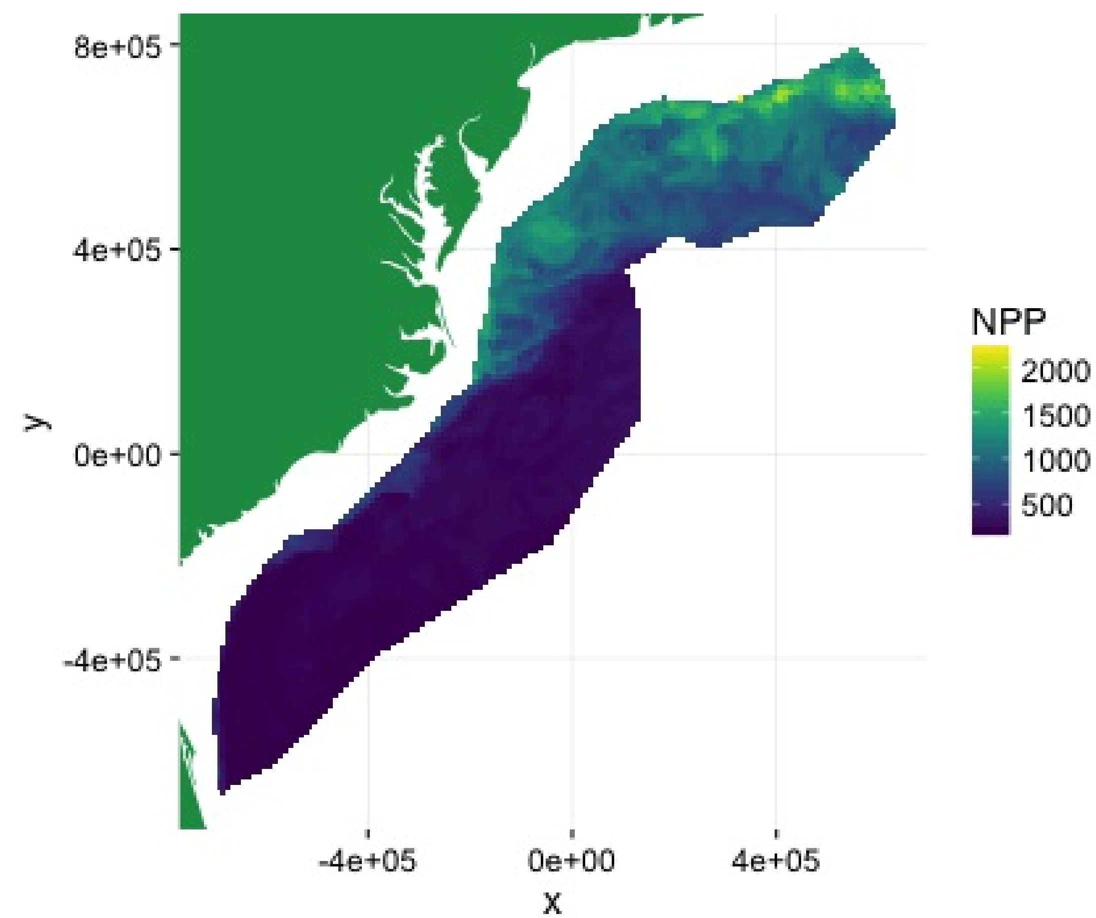
- area of segment
- probability of detection in segment
- (inverse) link function
- model terms

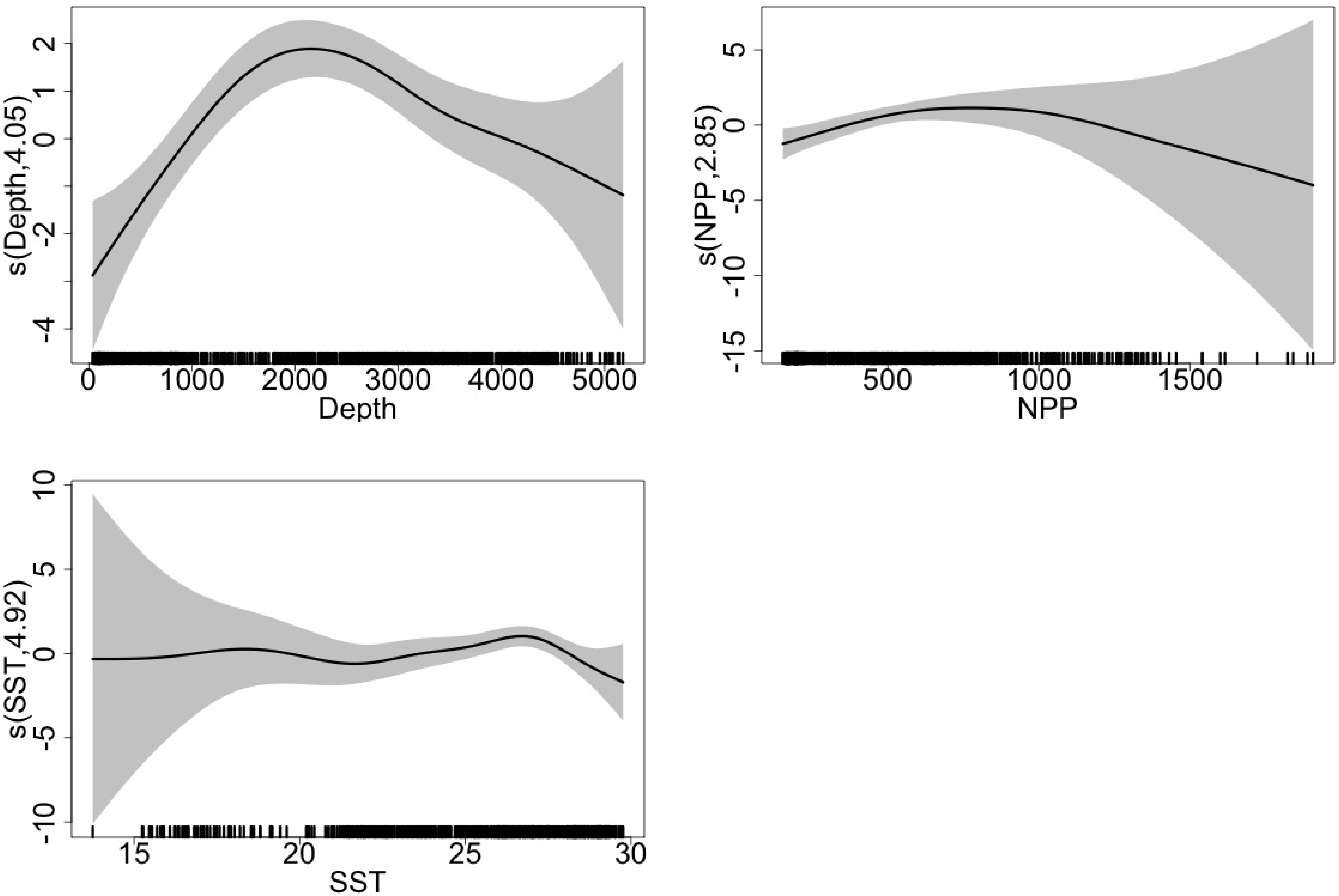
What about those s thingys?

Covariates

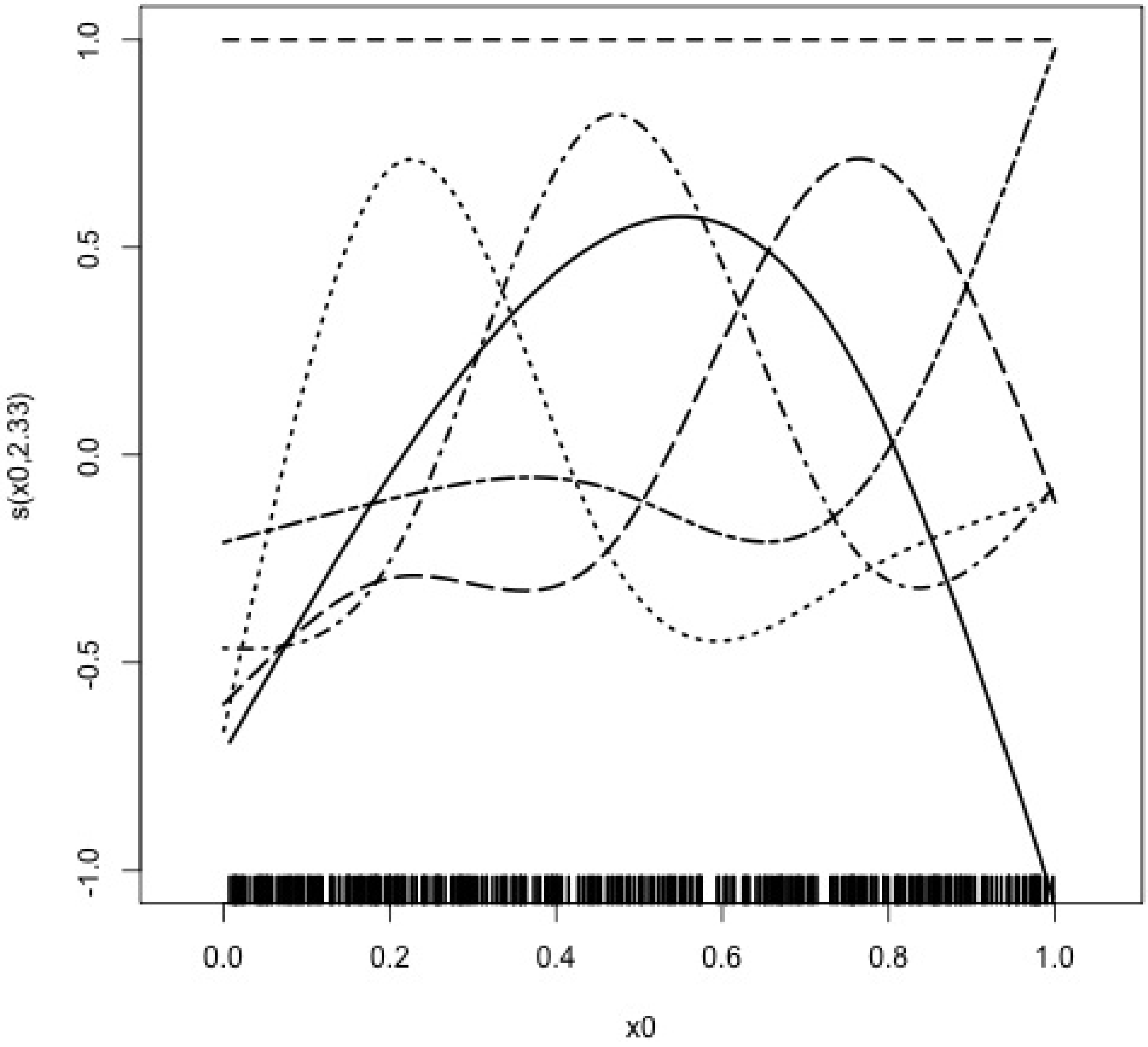
- space, time, environmental (remotely sensed?) data





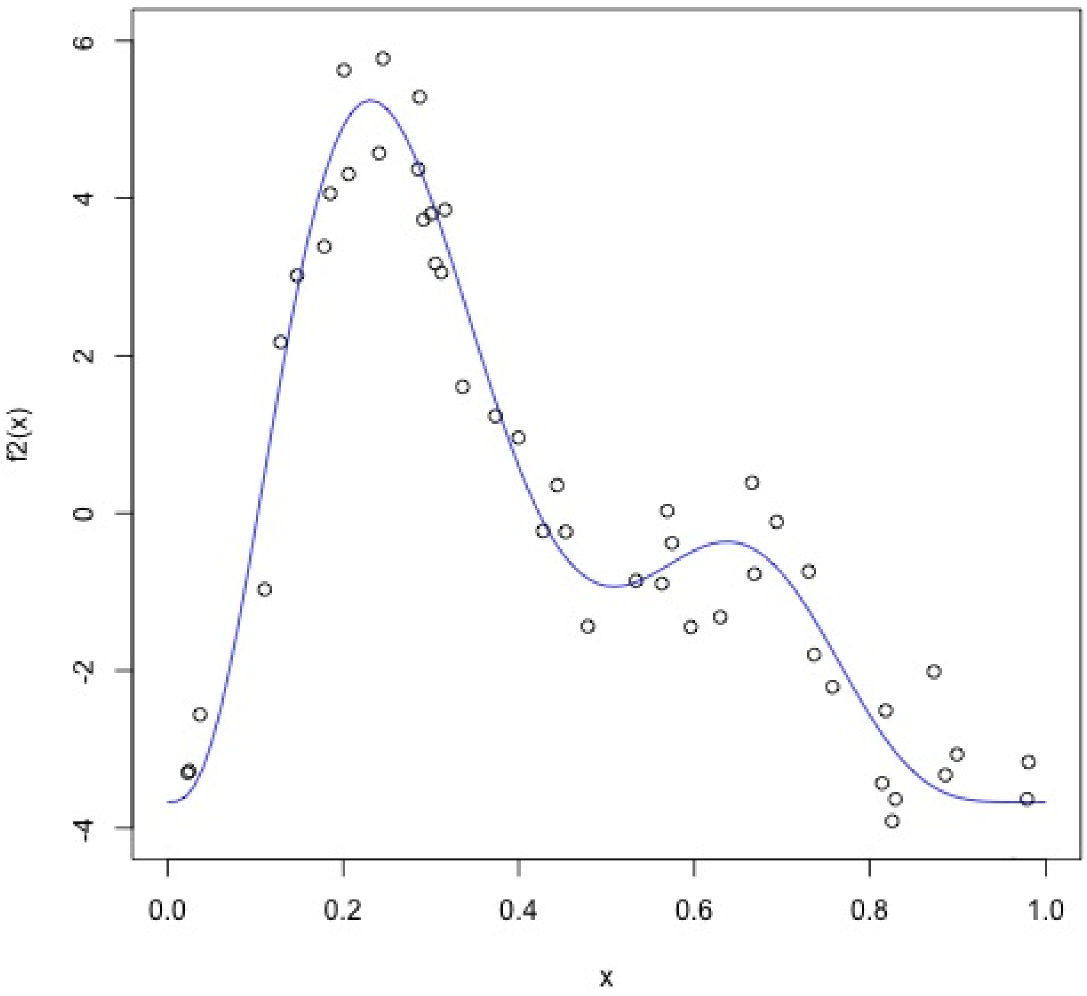

s(Depth,4.05)
2
1
0
-2
-4
0
1000
2000
3000
4000
5000
Depth
s(NPP,2.85)
5
0
-5
-10
-15
500
1000
1500
NPP
s(SST,4.92)
10
5
0
-5
-10
15
20
25
30
SST

How do we build them?



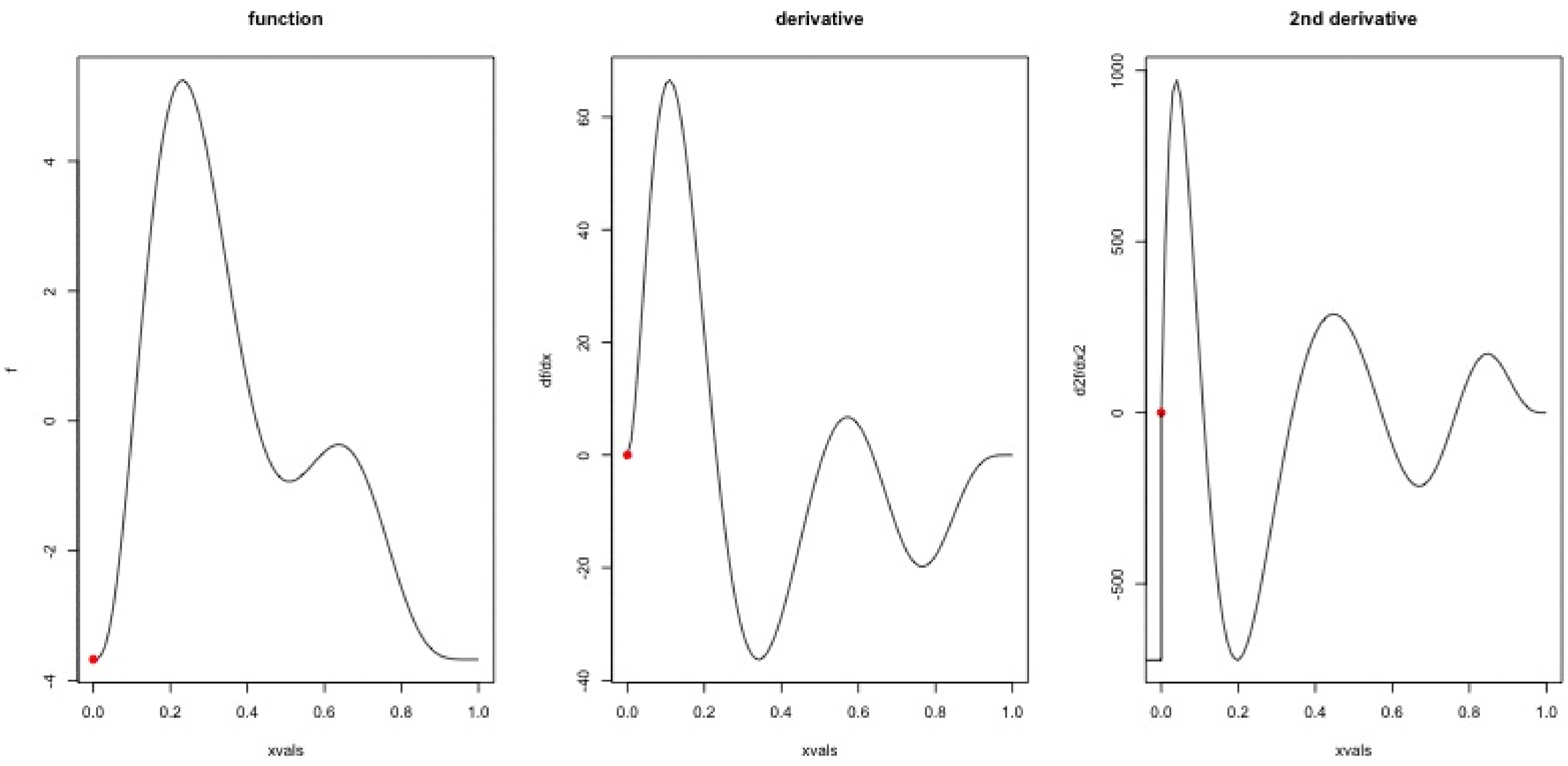
- Functions made of other, simpler functions
- **Basis functions**, b_k
- Estimate β_k
- $s(x) = \sum_{k=1}^{K} \beta_k b_k(x)$

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**

How wiggly is a function?



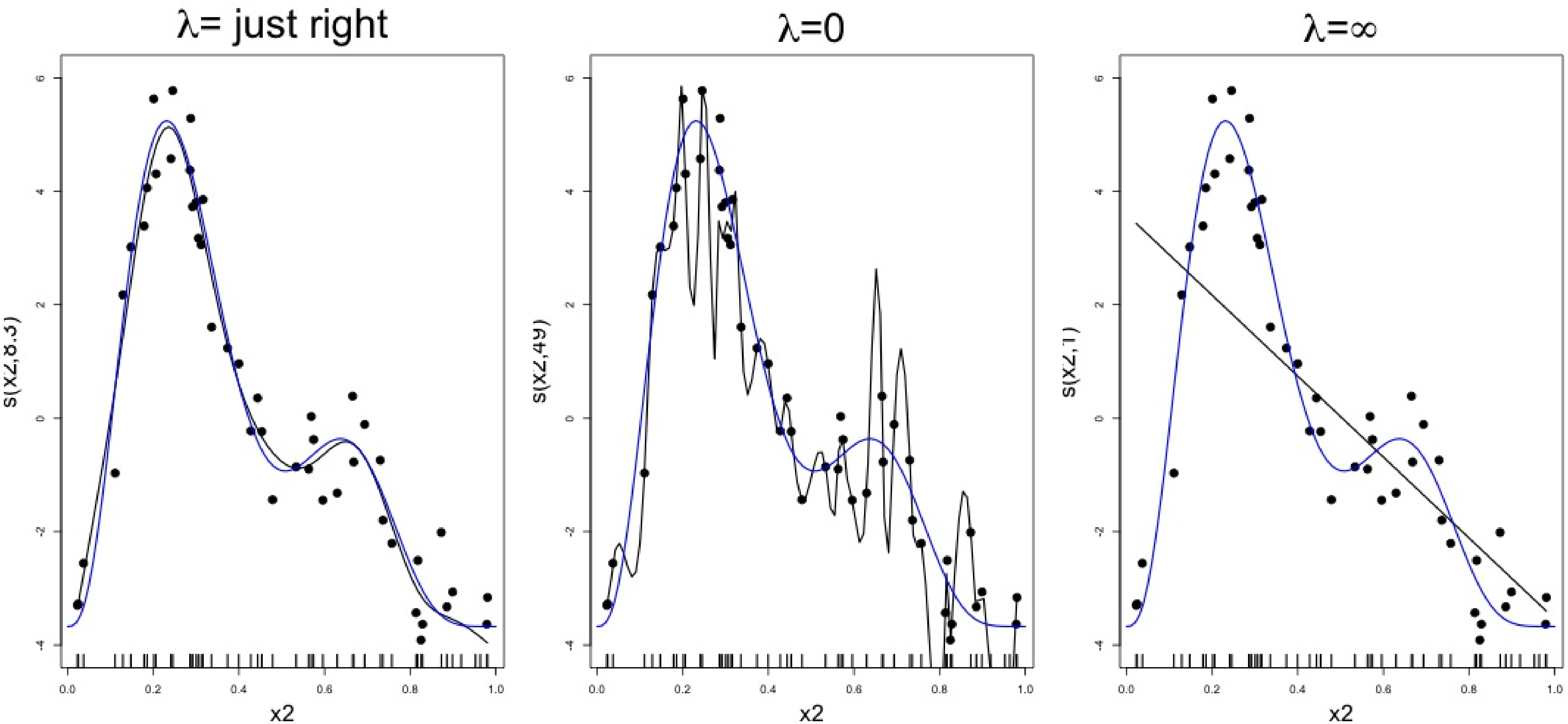
Making wigglyness matter

- Fit needs to be **penalised**
- *Something* like:

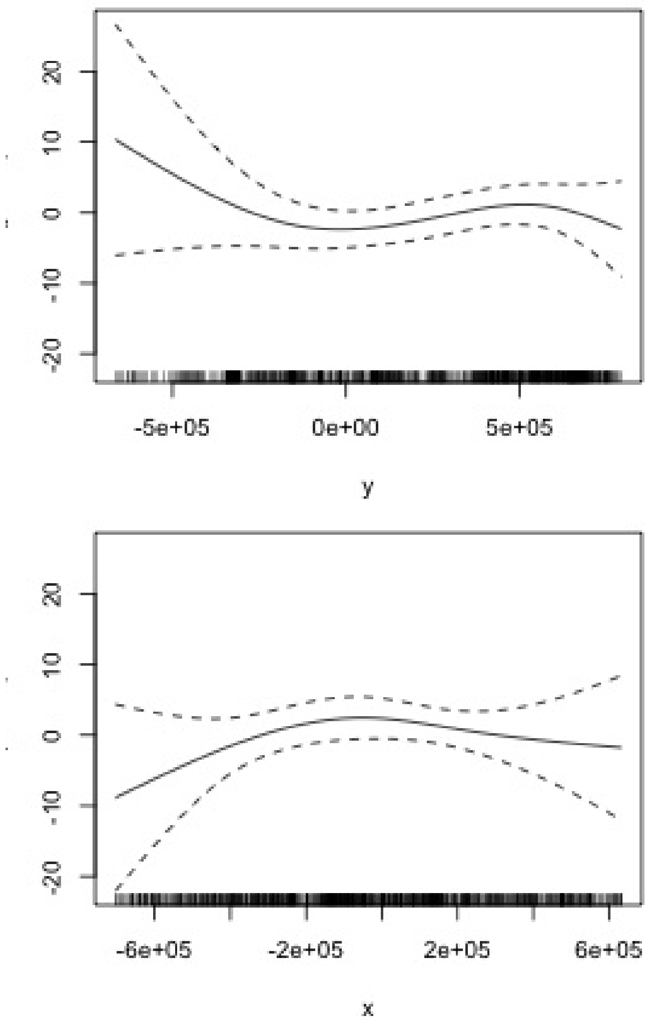
$$\int_{\mathbb{R}} \left(\frac{\partial^2 s(x)}{\partial x^2} \right)^2 dx$$

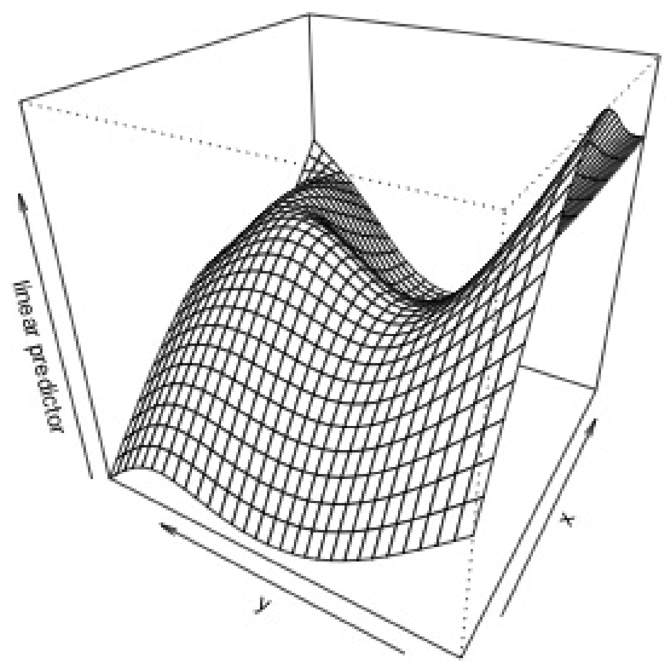
- (Can always re-write this in the form $\beta^T \mathbf{S} \beta$)
- Estimate the β_k terms but penalise objective
  - "closeness to data" + penalty (REML/ML)

Smoothing parameter



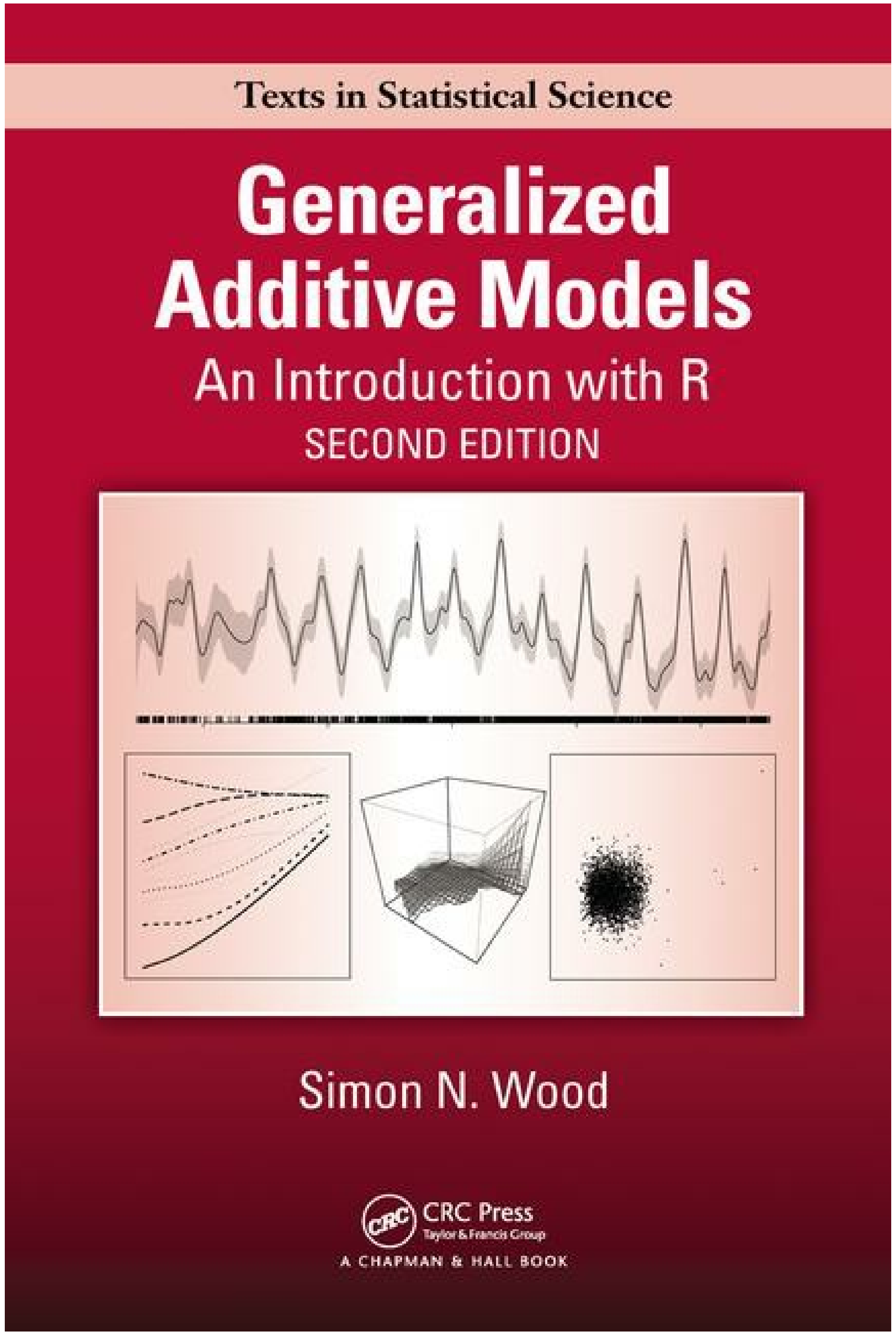
Beyond univariate smooths?





- Can build **tensor product** terms
- Take 2 or more univariate terms
- Thin plate regression splines allow multivariate terms (isotropic)

Why GAMs are cool...



- Fancy smooths (cyclic, boundaries, …)
- Fancy responses (exp family and beyond!)
- Random effects (by equivalence)
- Markov random fields
- Correlation structures
- See Wood (2006/2017) for a handy intro

Let's fit a model

```
library(dsm)
dsm_env_tw <- dsm(count~s(Depth) + s(NPP) + s(SST),
                  ddf.obj=df_hr,
                  segment.data=segs, observation.data=obs,
                  family=tw())
```

`dsm` is based on `mgcv` by Simon Wood

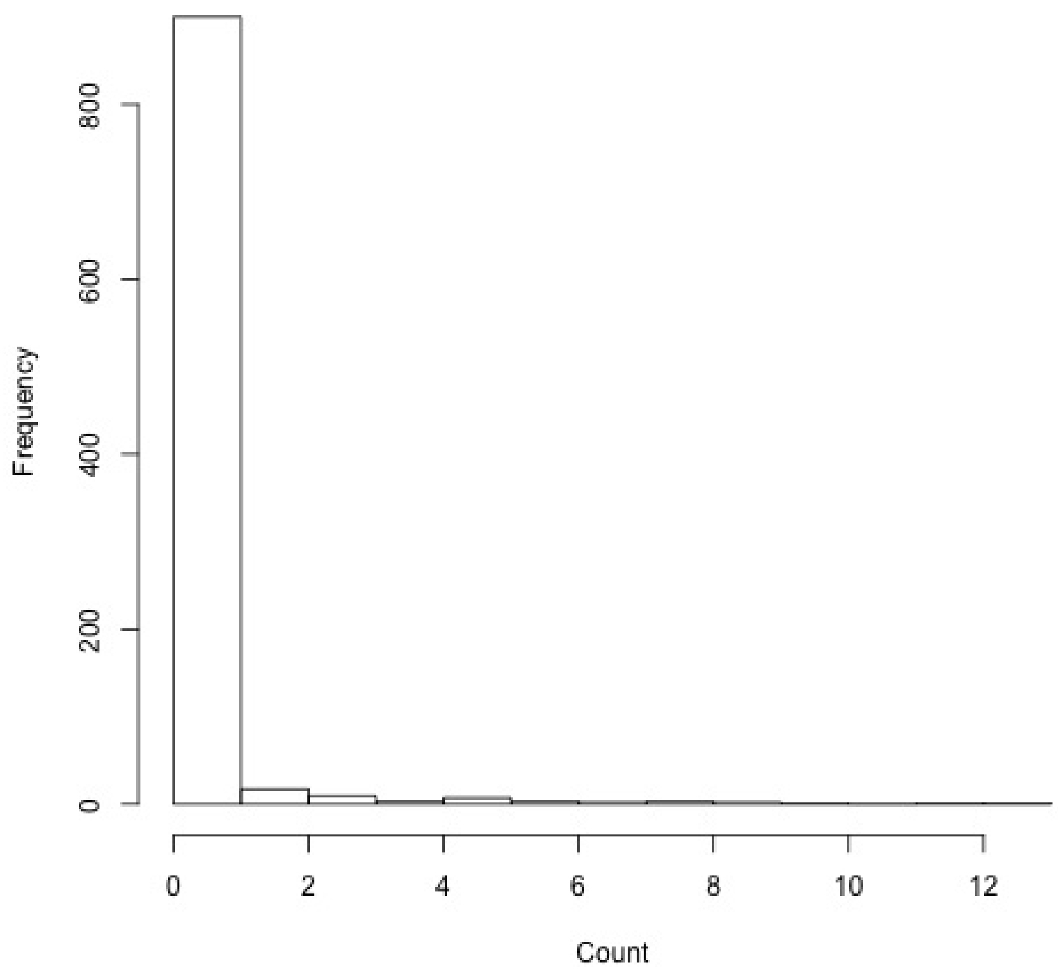
Simple! Done?

NO

Model checking

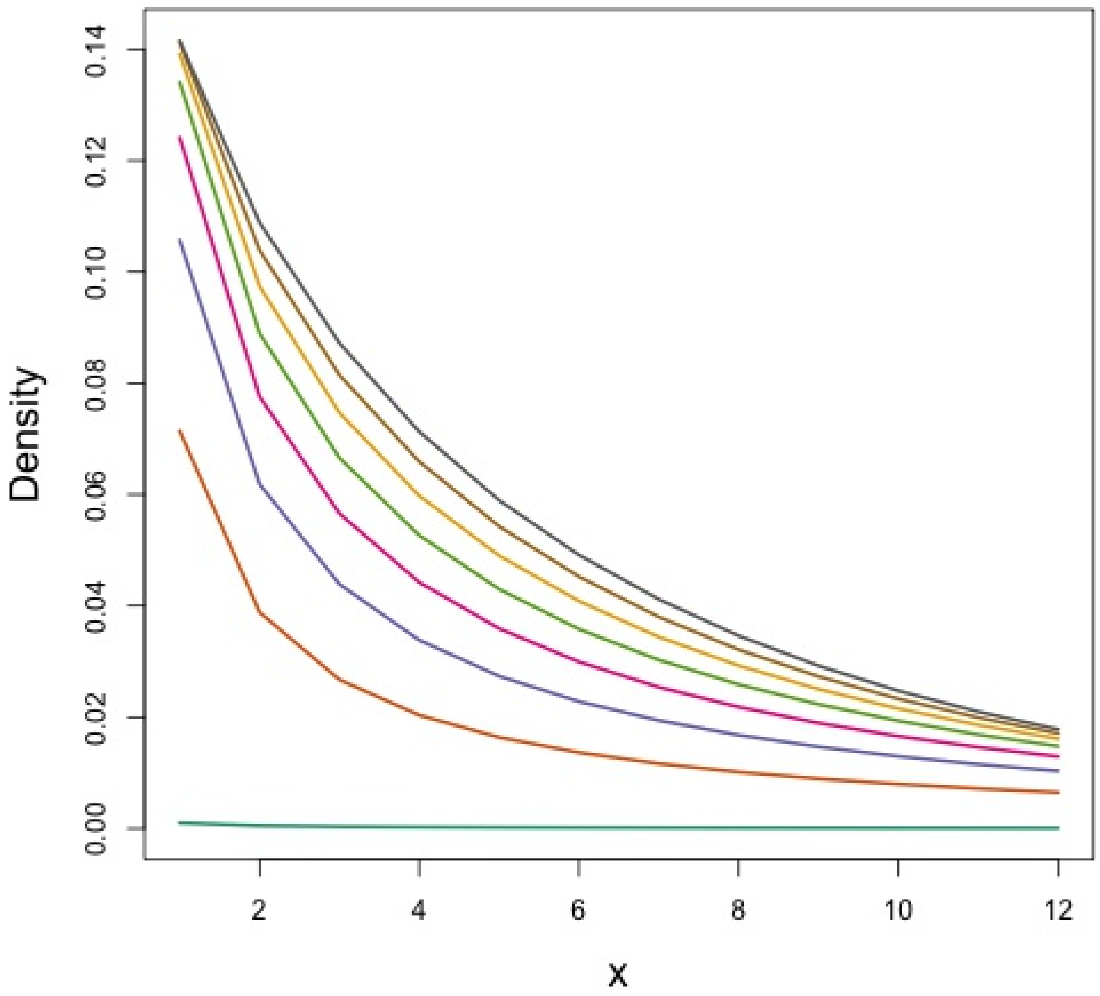
- Response distribution
- Model (term) selection
- Sensitivity
- Cross-validation (replicability)

Count distributions



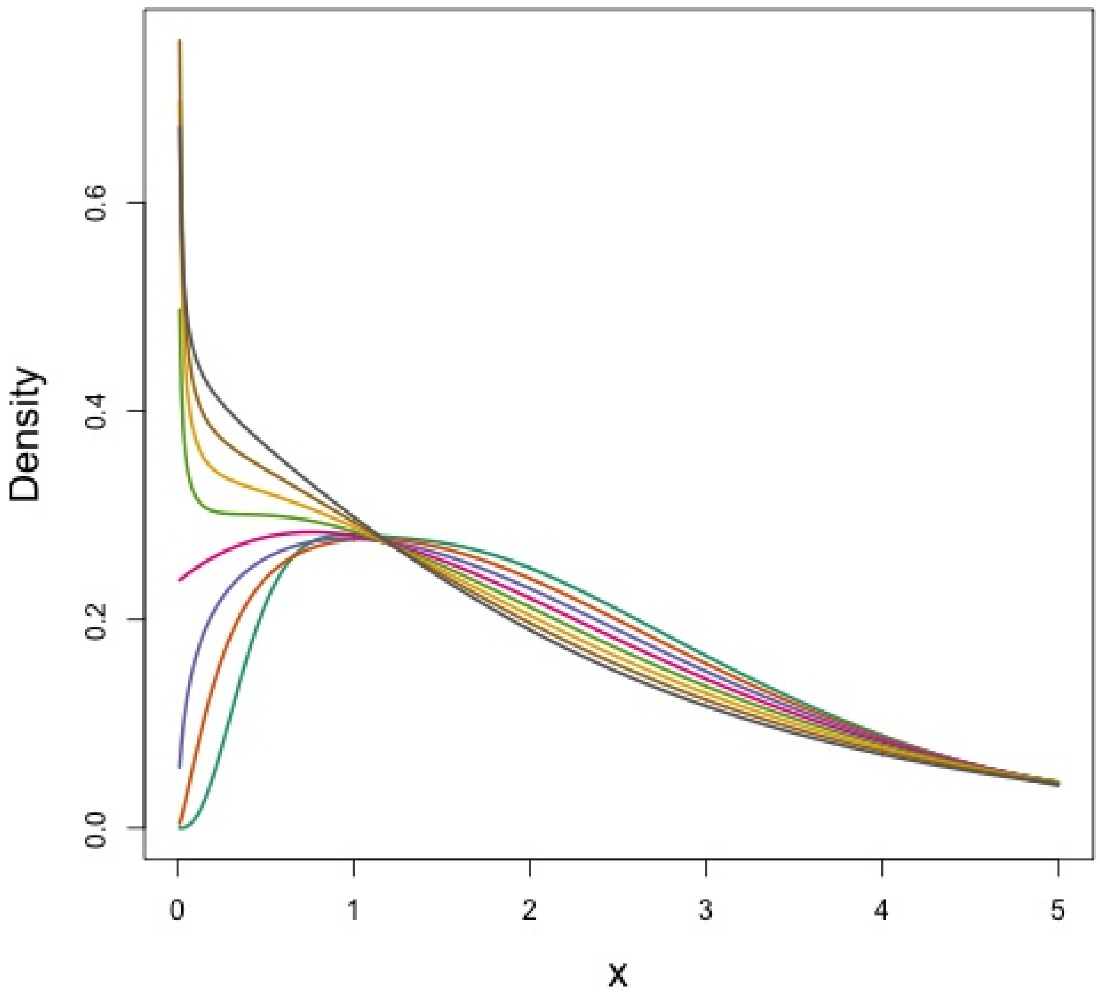
- Response is a count (not not always integer)
- Often, it's mostly zero (that's complicated)
- Want response distribution that deals with that
- Flexible mean-variance relationship

Negative binomial



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

Tweedie distribution



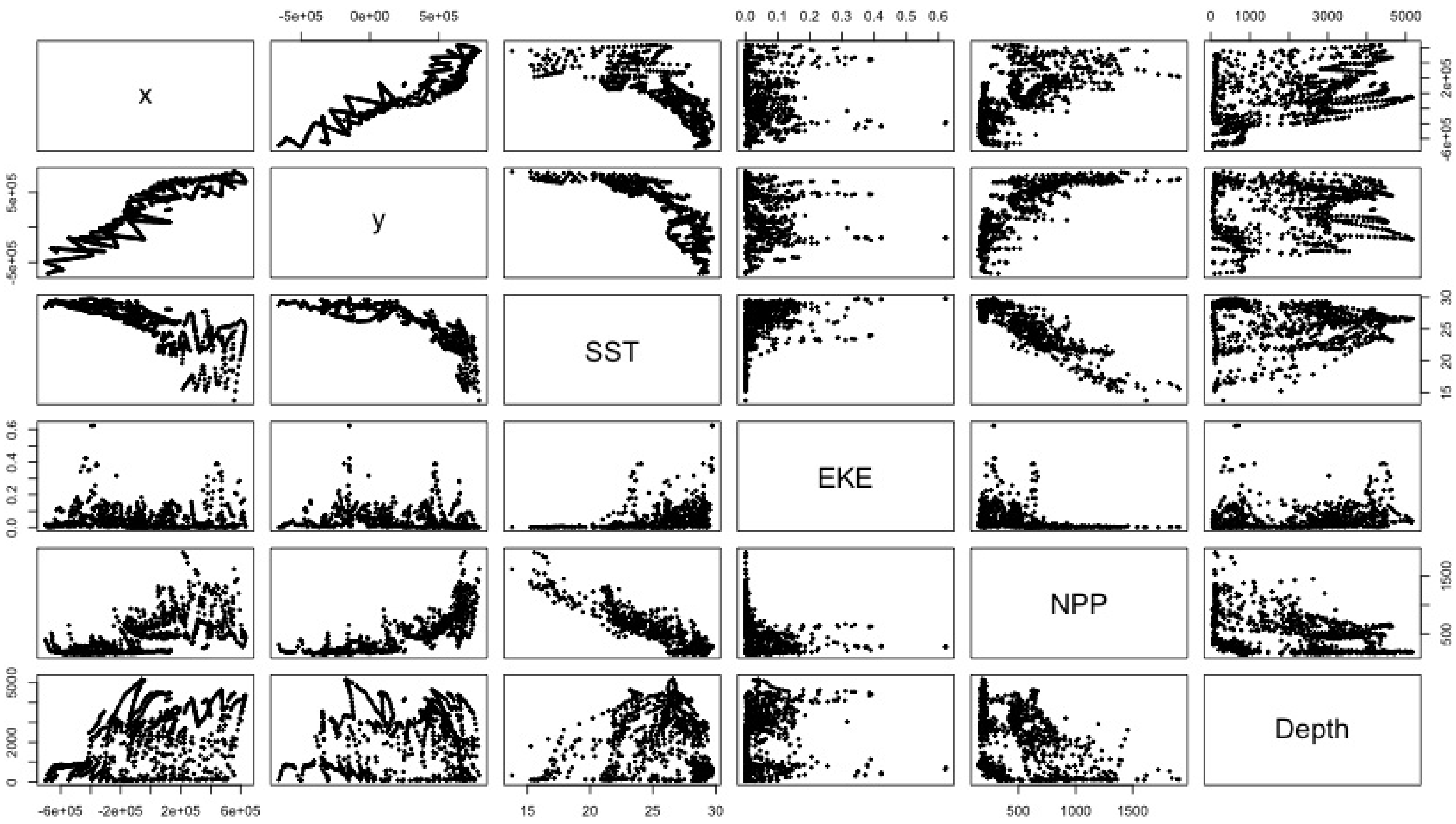
- $\mathrm{Var}(\mathrm{count}) = \varphi(\mathrm{count})^q$
- Common distributions are sub-cases:
  - $q = 1 \Rightarrow$ Poisson
  - $q = 2 \Rightarrow$ Gamma
  - $q = 3 \Rightarrow$ Normal
- We are interested in $1 < q < 2$
- (here $q = 1.2, 1.3, \ldots, 1.9$)

Tobler's first law of geography

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

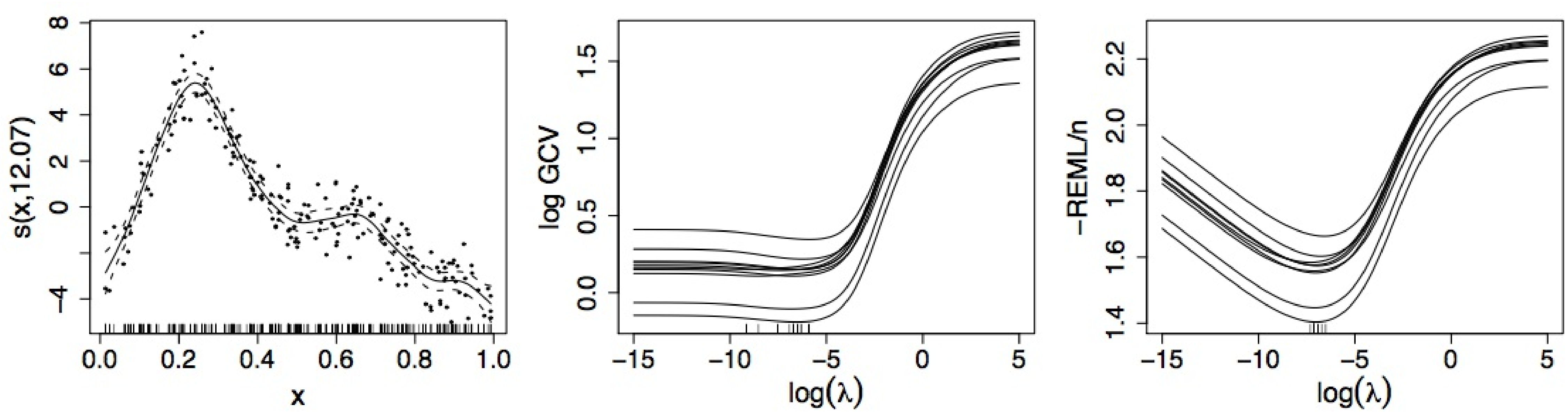
Tobler (1970)

Implications of Tobler's law



What can we do about this?

- Careful inclusion of terms
- Test for sensitivity (lots of models)
- Fit models using robust criteria (REML/ML)
- Test for concurvity (`mgcv::concurvity`, `dsm::vis.concurvity`)



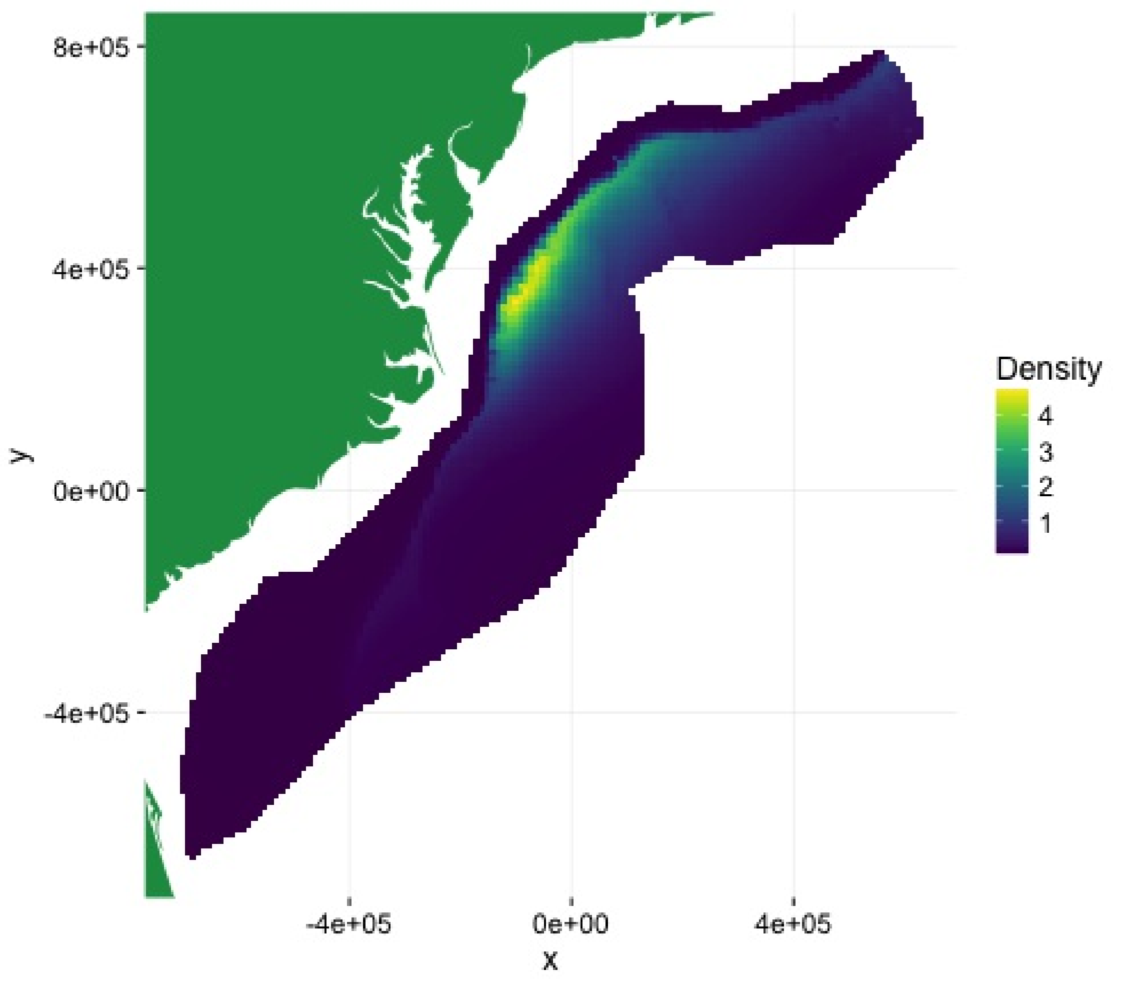
Term selection

- (approximate) p values (Marra & Wood, 2012)
  - path dependence issues
- shrinkage methods (Marra & Wood, 2011)
- ecological-level term selection
  - *which* biomass measure?
  - include spatial smooth or not?

Sideline: GAMs are Bayesian models

- Generally:
  - penalties are improper prior precision matrices
  - (nullspace gives improper priors)
- Using shrinkage smoothers:
  - *proper* priors
  - empirical Bayes interpretation

Predictions over arbitrary areas

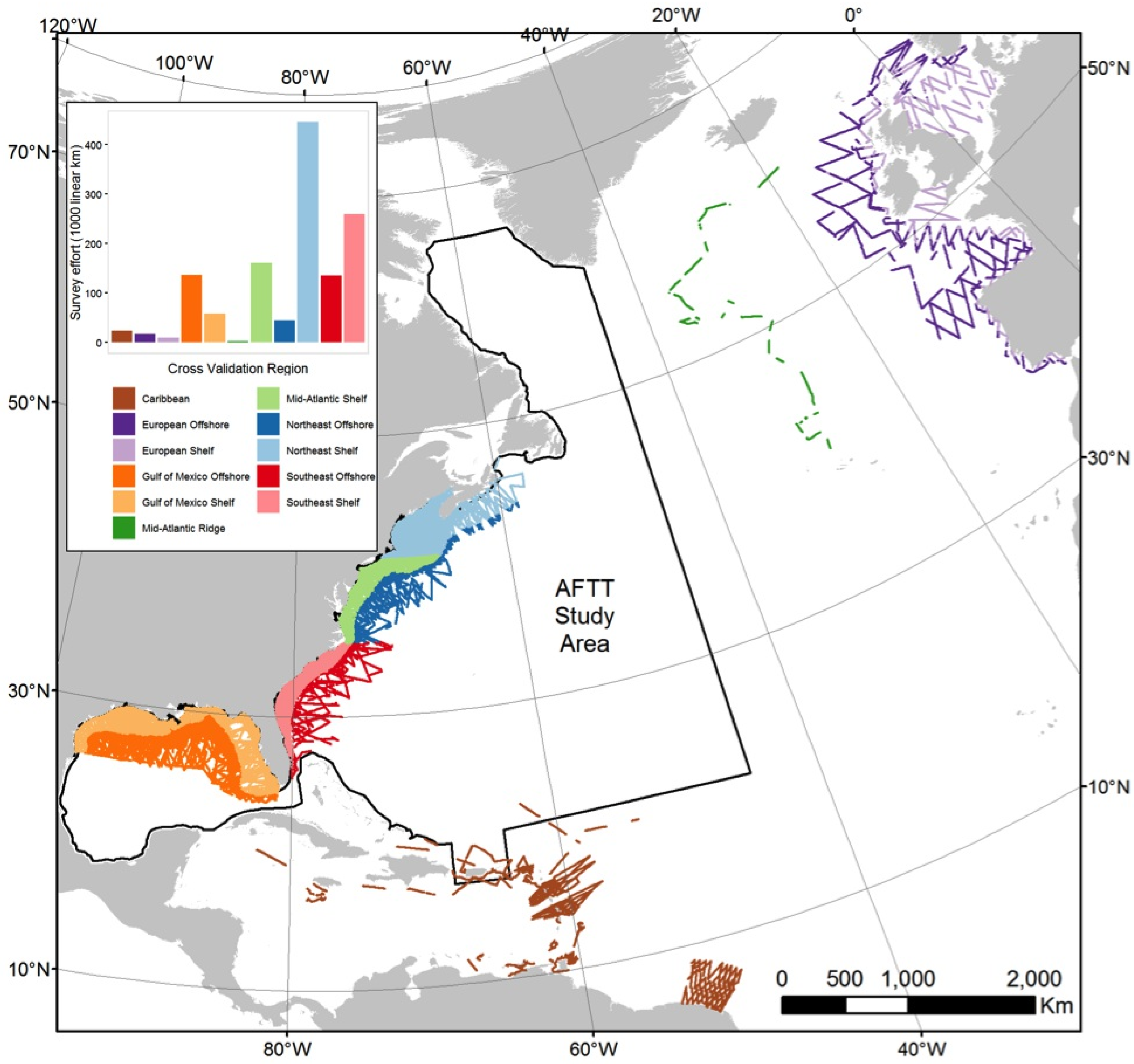


- Don't want to be restricted in where to predict
  - Predict within survey area
  - Extrapolate outside (with caution)
- Working on a grid of cells

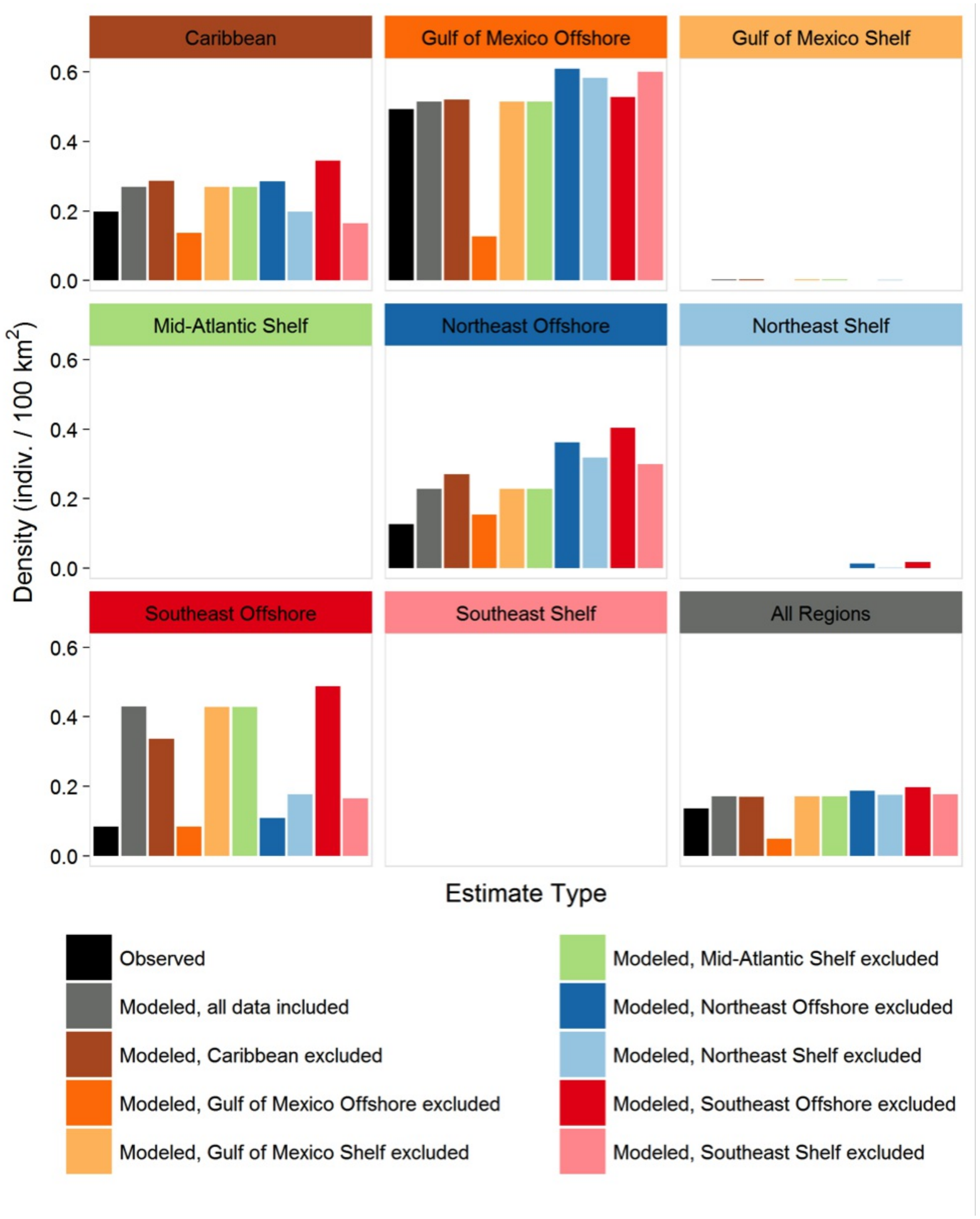
Cross-validation

- How well does the model reproduce what we saw?
- Leave out one area, re-fit model, predict to new data
- Wenger & Olden (2012) have good spatial examples

Cross-validation example



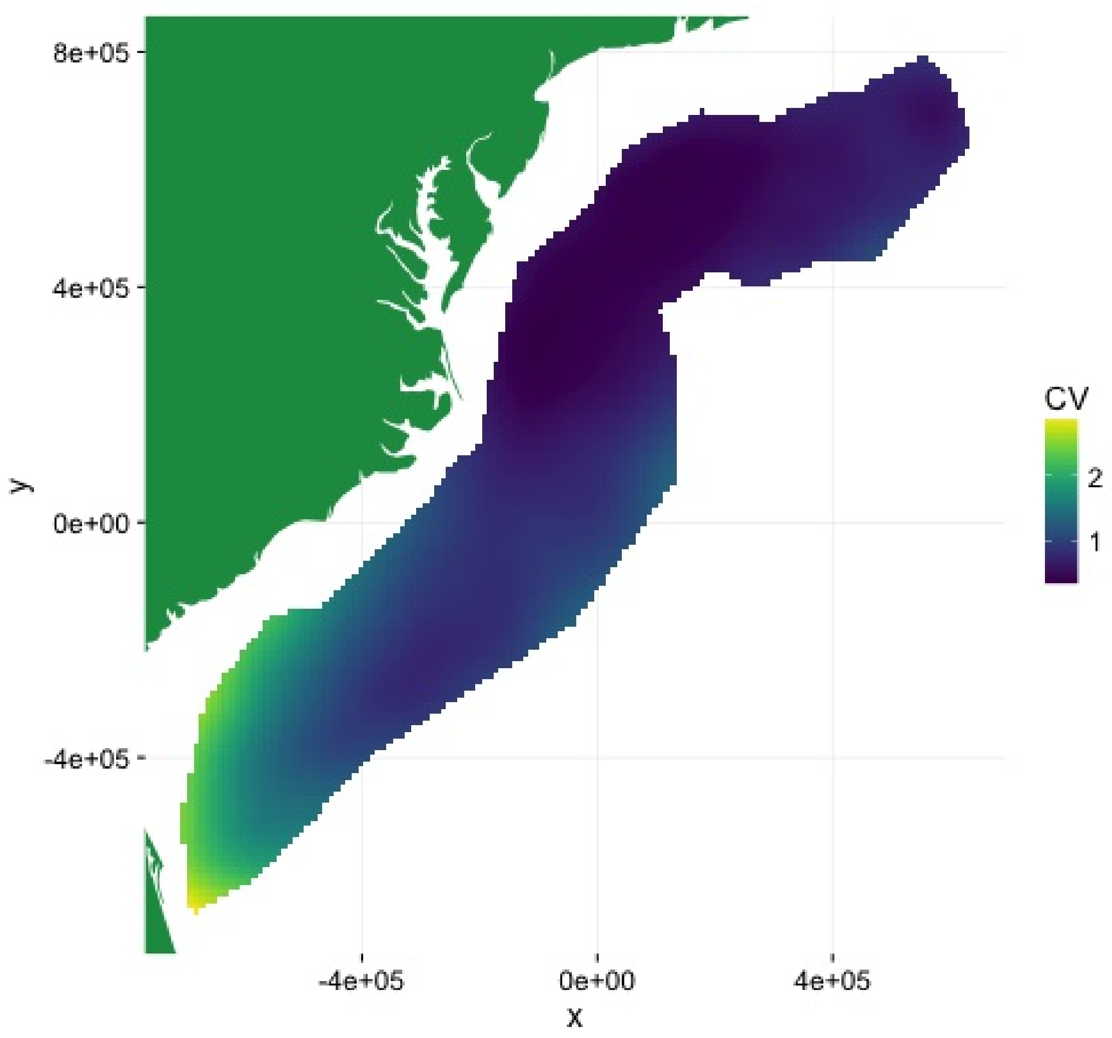
Cross-validation example



Estimating variance

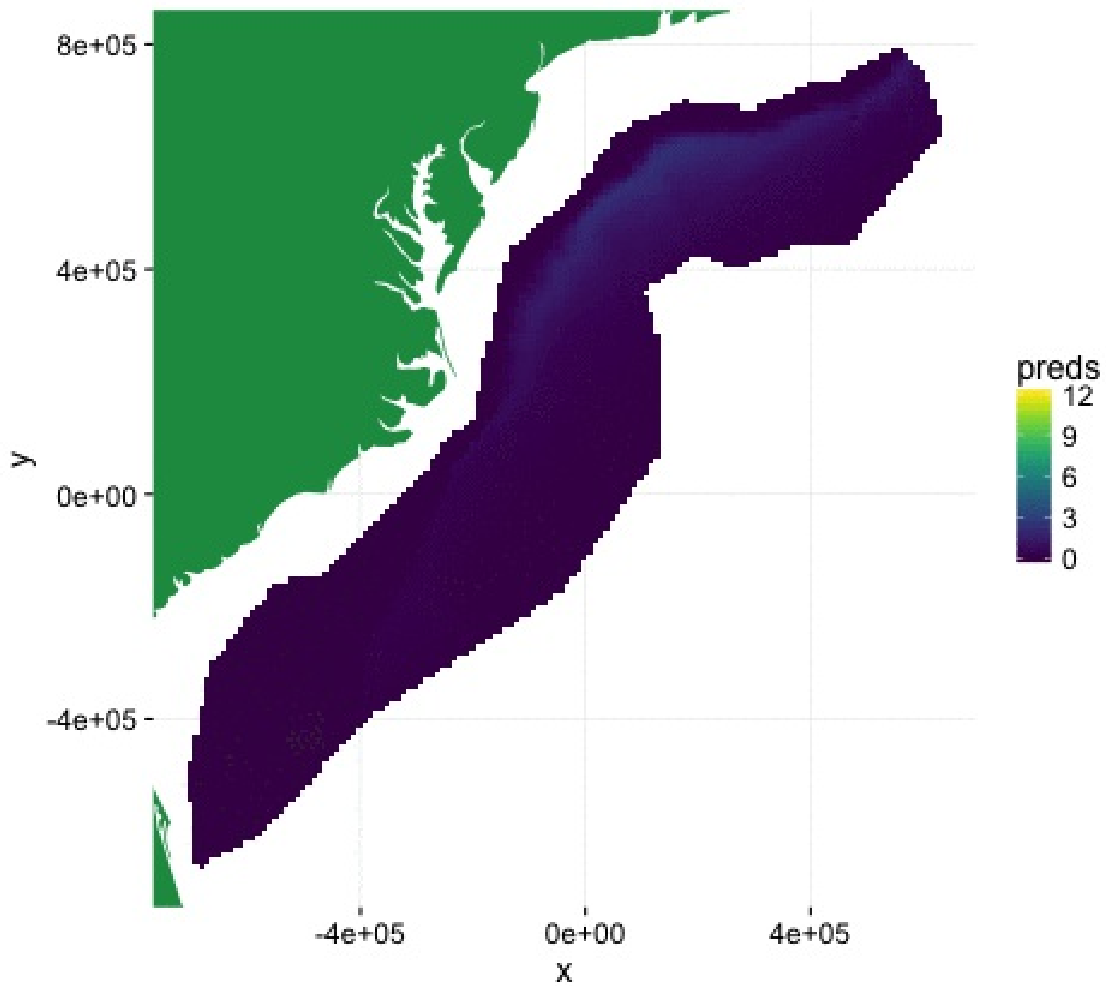
- Uncertainty from:
  - detection function parameters
  - spatial model
- Need to propagate uncertainty!
  - Methods in `dsm`
  - Bravington, Hedley & Miller (in prep)

Plotting uncertainty



- Maps of coefficient of variation
- CV for given stratum (better)
- Visualisation is **hard**

Communicating uncertainty



- Are animations a good way to do this?
- Simulate from posterior parameter distribution
- $\boldsymbol{\beta} \sim \mathrm{N}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$
- Some features (e.g. shelf, N-S gradient) stick out

I am going to stop talking very soon

2 (or more)-stage models

- Not “cool” (statistically), but…
- Multi-stage models are handy!
- Understand and **check** each part
- Split your modelling efforts amongst people

Conclusions

- This methodology is general
  - Bears, birds, beer cans, Loch Ness monsters…
- Models are flexible!
  - Linear things, smooth things, random effect things (and *more*)
- If you know GLMs, you can get started with DSMs
  - Mature theoretical basis, still lots to do
- Active user community, active software development

Resources

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2013 doi: 10.1111/2041-210X.12105

Spatial models for distance sampling data: recent developments and future directions

David L. Miller[1]*, M. Louise Burt[2], Eric A. Rexstad[2] and Len Thomas[2]

[1]*Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881, USA; and* [2]*Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, St Andrews KY16 9LZ, UK*

distancesampling.org/R/

distancesampling.org/workshops/duke-spatial-2015/

Thanks!

Slides w/ references available at converged.yt

References

Burt, M. L., Borchers, D. L., Jenkins, K. J., & Marques, T. A. (2014). Using mark-recapture distance sampling methods on line transect surveys. Methods in Ecology and Evolution, 5(11).

Dunn, P. K., & Smyth, G. K. (1996). Randomized Quantile Residuals. Journal of Computational and Graphical Statistics, 5(3).

Hedley, S. L., & Buckland, S. T. (2004). Spatial models for line transect sampling. Journal of Agricultural, Biological, and Environmental Statistics, 9(2).

Marques, T. A., Thomas, L., Fancy, S. G., & Buckland, S. T. (2007). Improving estimates of bird density using multiple-covariate distance sampling. The Auk, 124(4).

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. Computational Statistics and Data Analysis, 55(7).

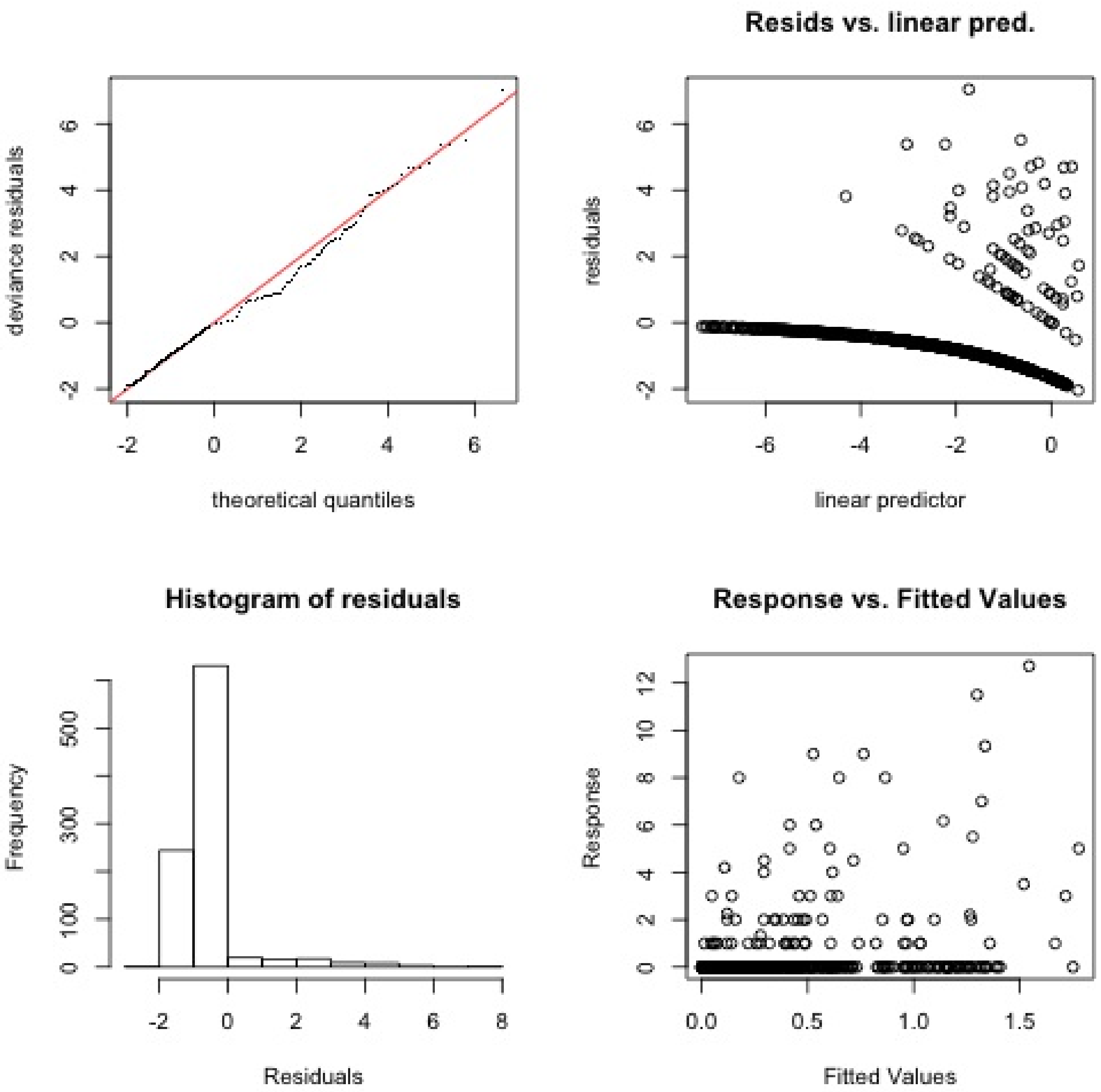
Marra, G., & Wood, S. N. (2012). Coverage Properties of Confidence Intervals for Generalized Additive Model Components. Scandinavian Journal of Statistics, 39(1).

Wenger, S.J. and Olden, J.D. (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3, 260–267.

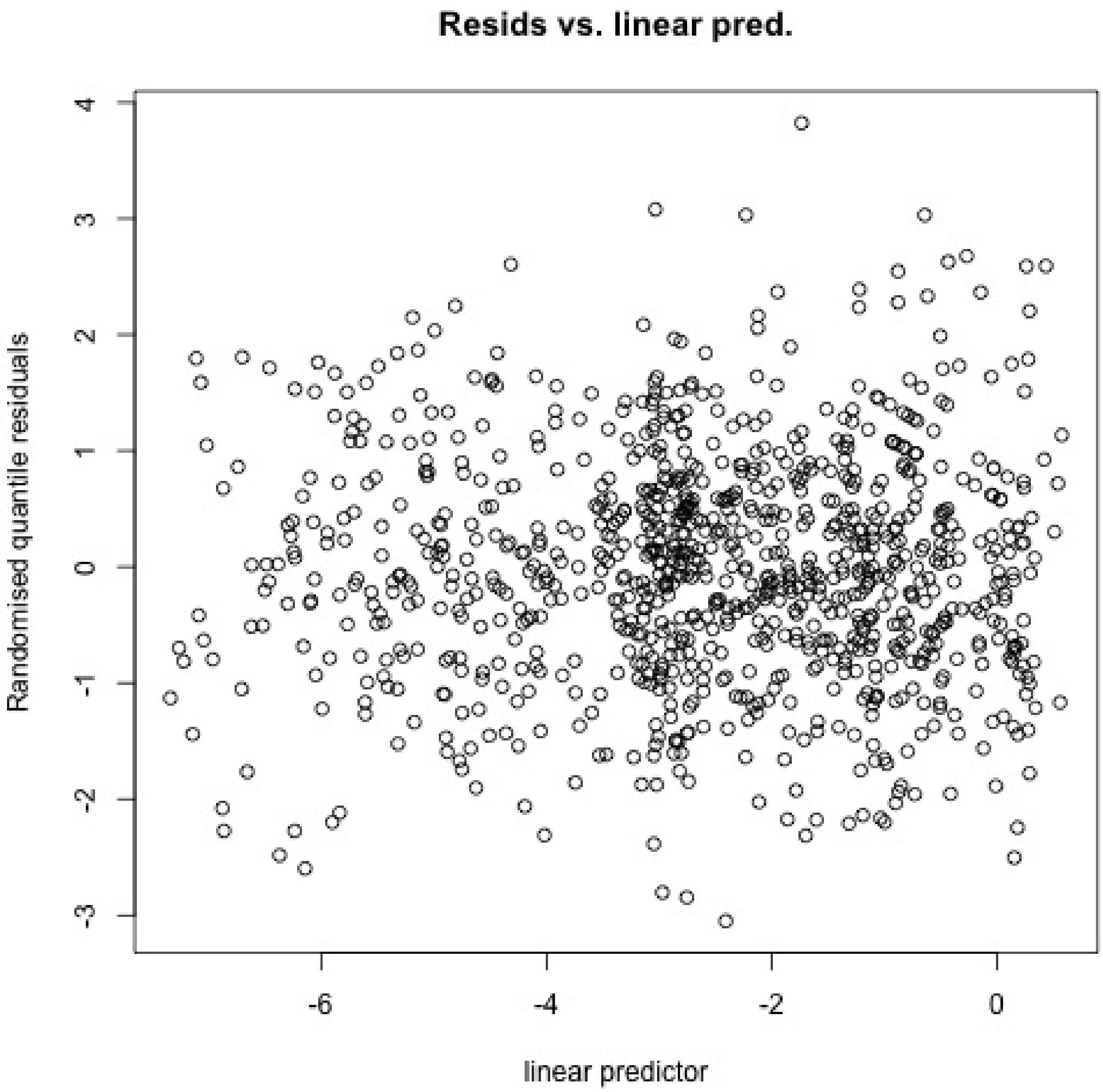
Handy awkward question answers

Don’t throw away your residuals!

gam.check


Resids vs. linear pred.
Histogram of residuals
Response vs. Fitted Values
deviance residuals
theoretical quantiles
residuals
linear predictor
Frequency
Residuals
Response
Fitted Values

rqgam.check (Dunn and Smyth, 1996)



Penalty matrix

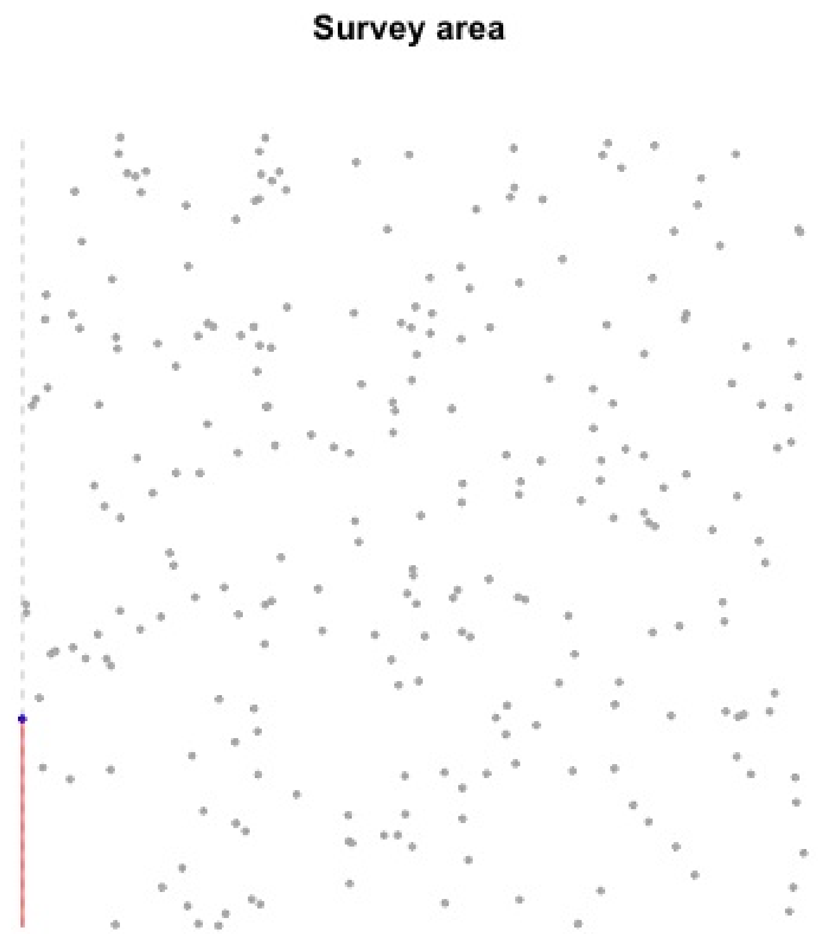
- For each b_k calculate the penalty
- Penalty is a function of β
  - $\lambda\beta^T S\beta$
- S calculated once
- smoothing parameter (λ) dictates influence

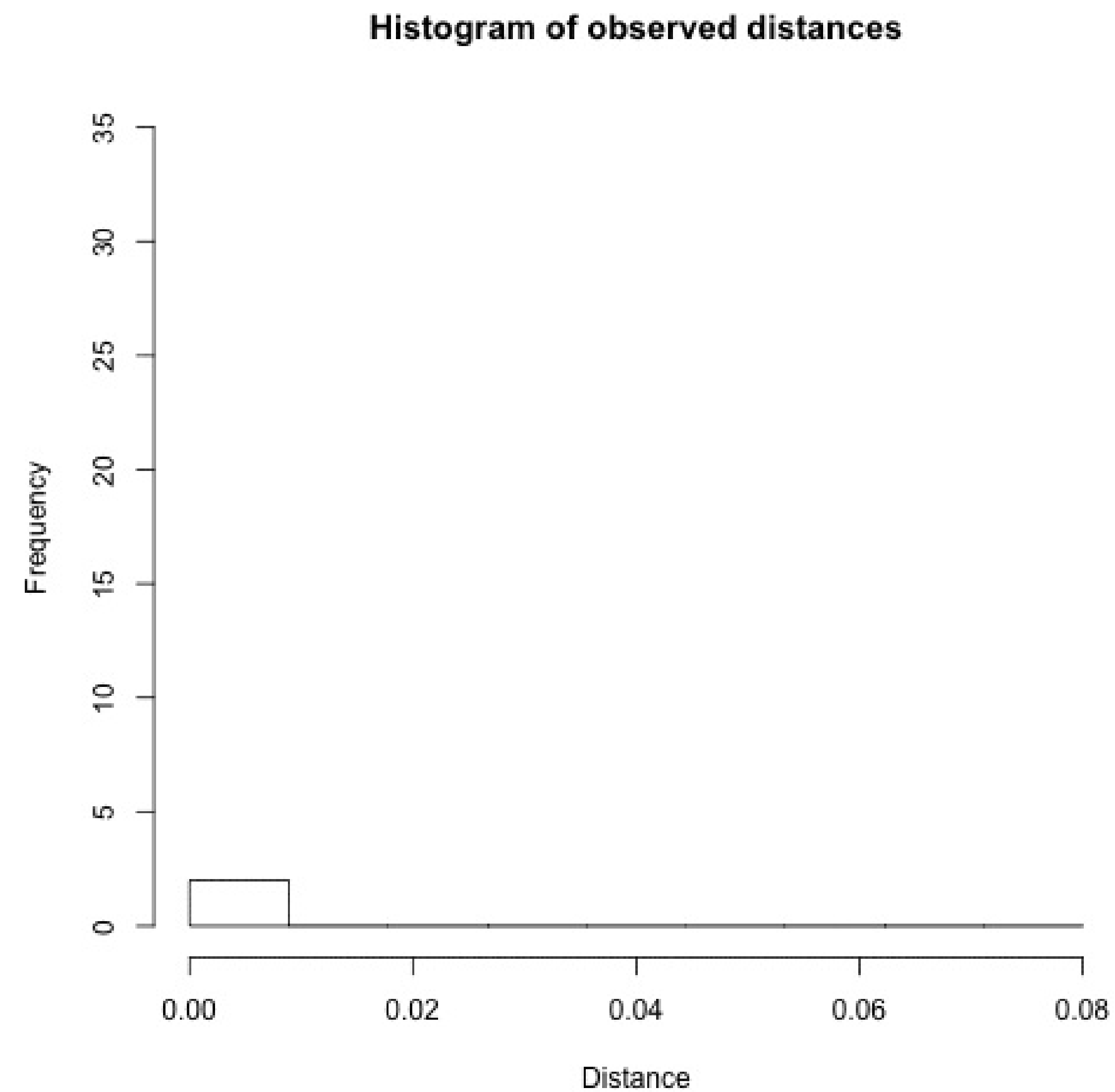
How wiggly are things?

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

Let’s talk about detectability

Detectability





Distance sampling

- "Fit to the histogram"
- Model:

$$\mathbb{P}\,[\text{animal detected} \mid \text{animal at distance y}] = g(y;\boldsymbol{\theta})$$

- Calculate the average probability of detection:

$$\hat{p} = \frac{1}{w}\int_0^w g(y;\hat{\boldsymbol{\theta}})dy$$

Distance sampling (extensions)

- Covariates that affect detectability (Marques et al, 2007)
- Perception bias ($g(0) < 1$) (Burt et al, 2014)
- Availability bias (Borchers et al, 2013)
- Detection function formulations (Miller and Thomas, 2015)
- Measurement error (Marques, 2004)

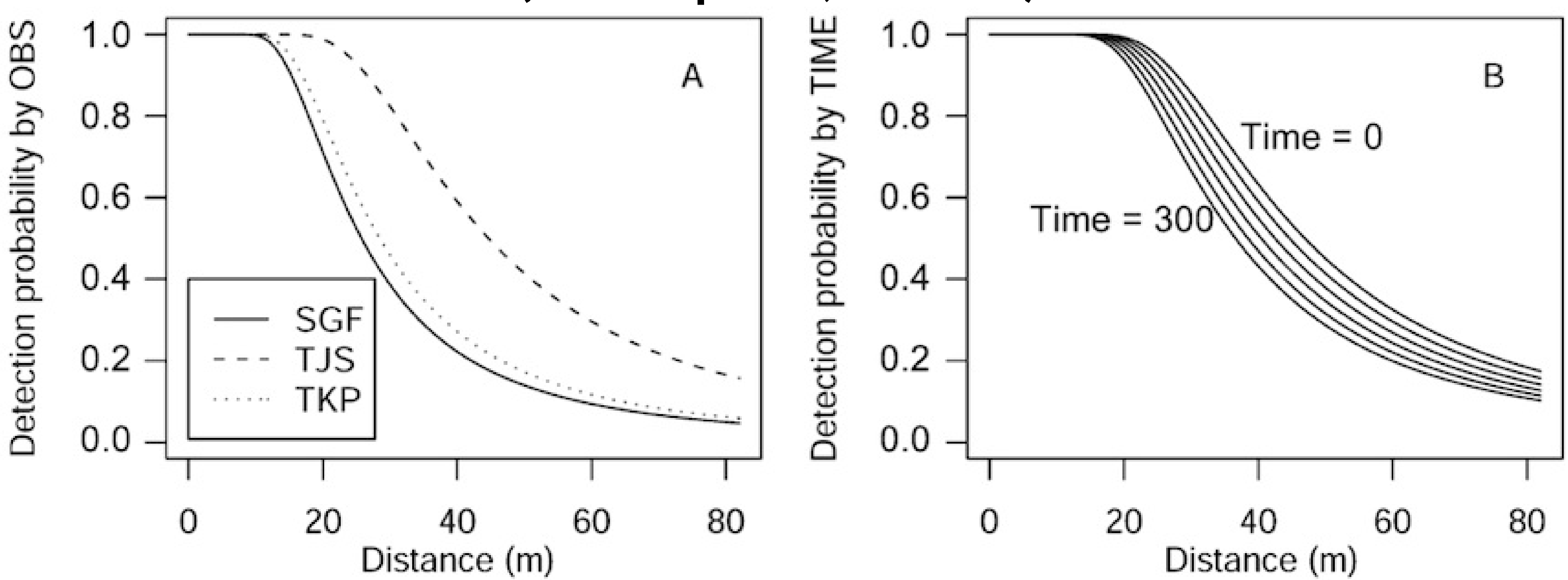


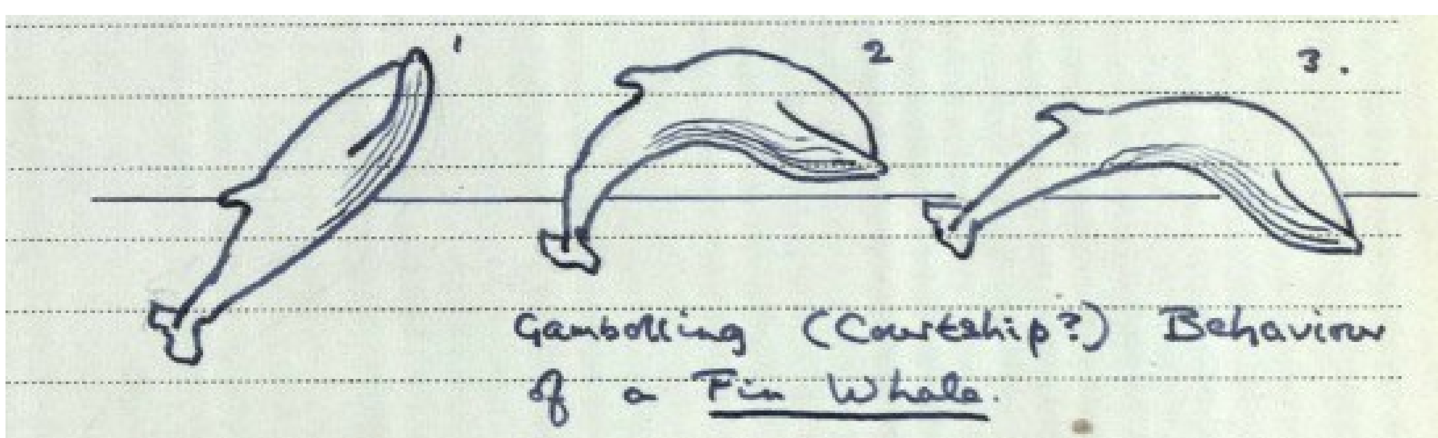
Figure from Marques et al (2007)

That's not really how the ocean works...

Availability

We can only see whales at the surface

- What proportion of the time are they there?
  - Acoustics
  - Tags (DTAGs etc)
  - Behavioural studies
- Fixed correction to $\hat{p}$?
- Model via fancy Markov models (Borchers et al, 2013)



Picture from University of St Andrews Library Special Collections