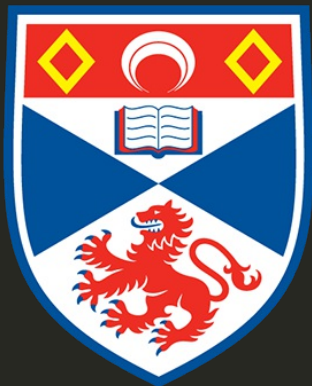


Practical advice



University of
St Andrews

Real survey data is messy

Distance sampling in the Real World

- We've talked a lot about models
- We've also talked about assumptions
- Our example is relatively well-behaved
- What can we do about all the nasty real world stuff?

Some days...



Aims

- Here we want to cover common questions
- Not definitive answers
- Some guidance on where to look for answers

What should my sample size be?

What do we mean by "sample size"?

- Number of animal (groups) recorded
 -
- Number of segments
 -
- Number of segments with observations
 -

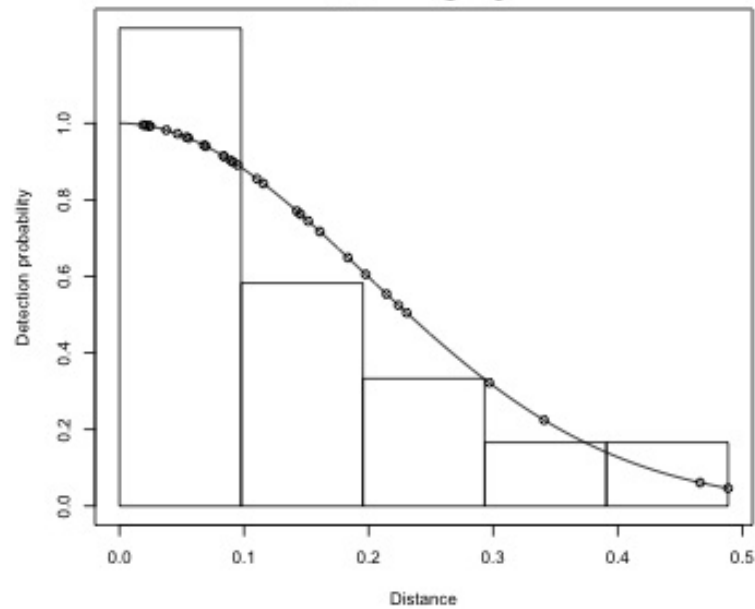
Re-frame

How would we know when we have enough samples?

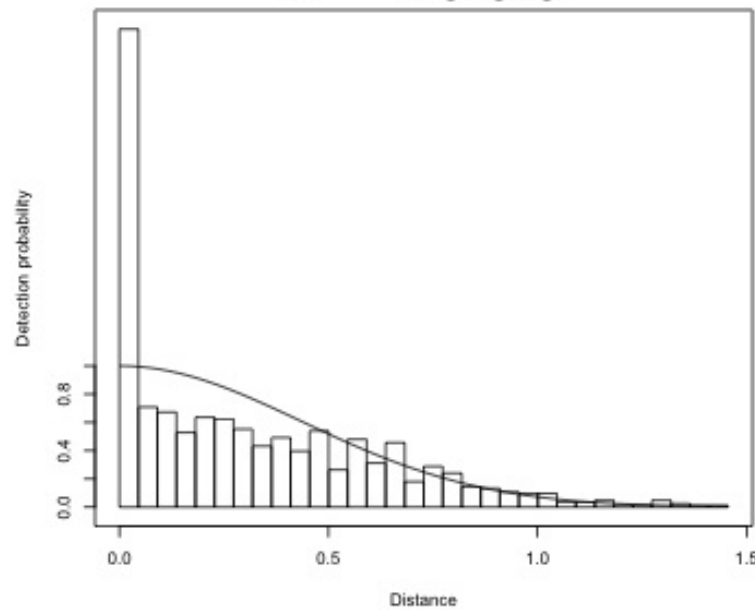
- We don't
- Heavily context-dependent
- Go back to assumptions

"How many data?"

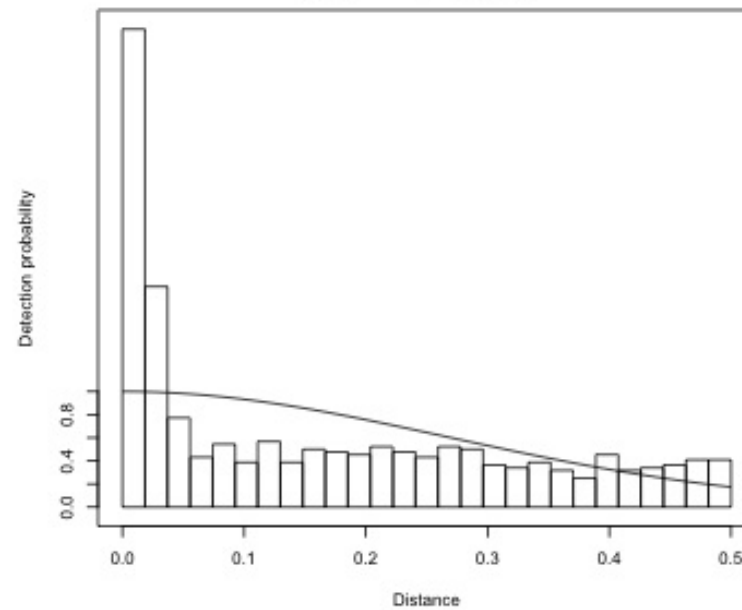
n=30



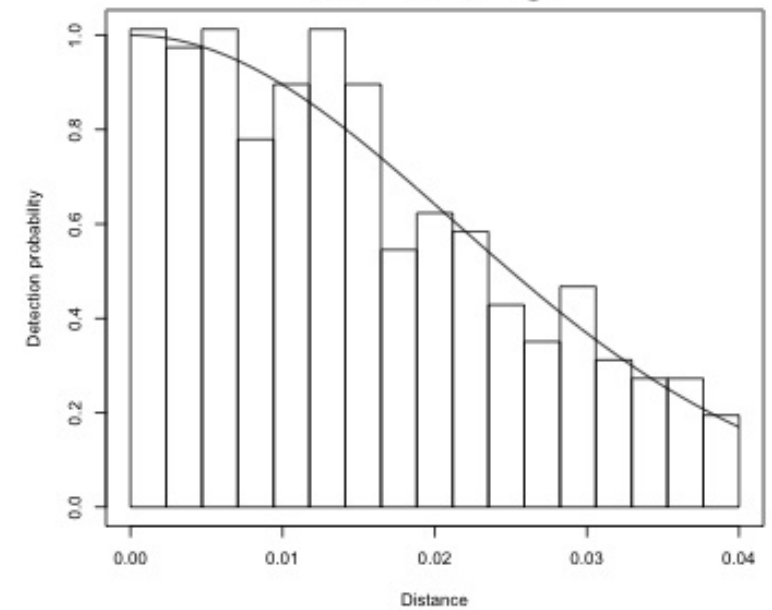
n=1000



n=747



n=273



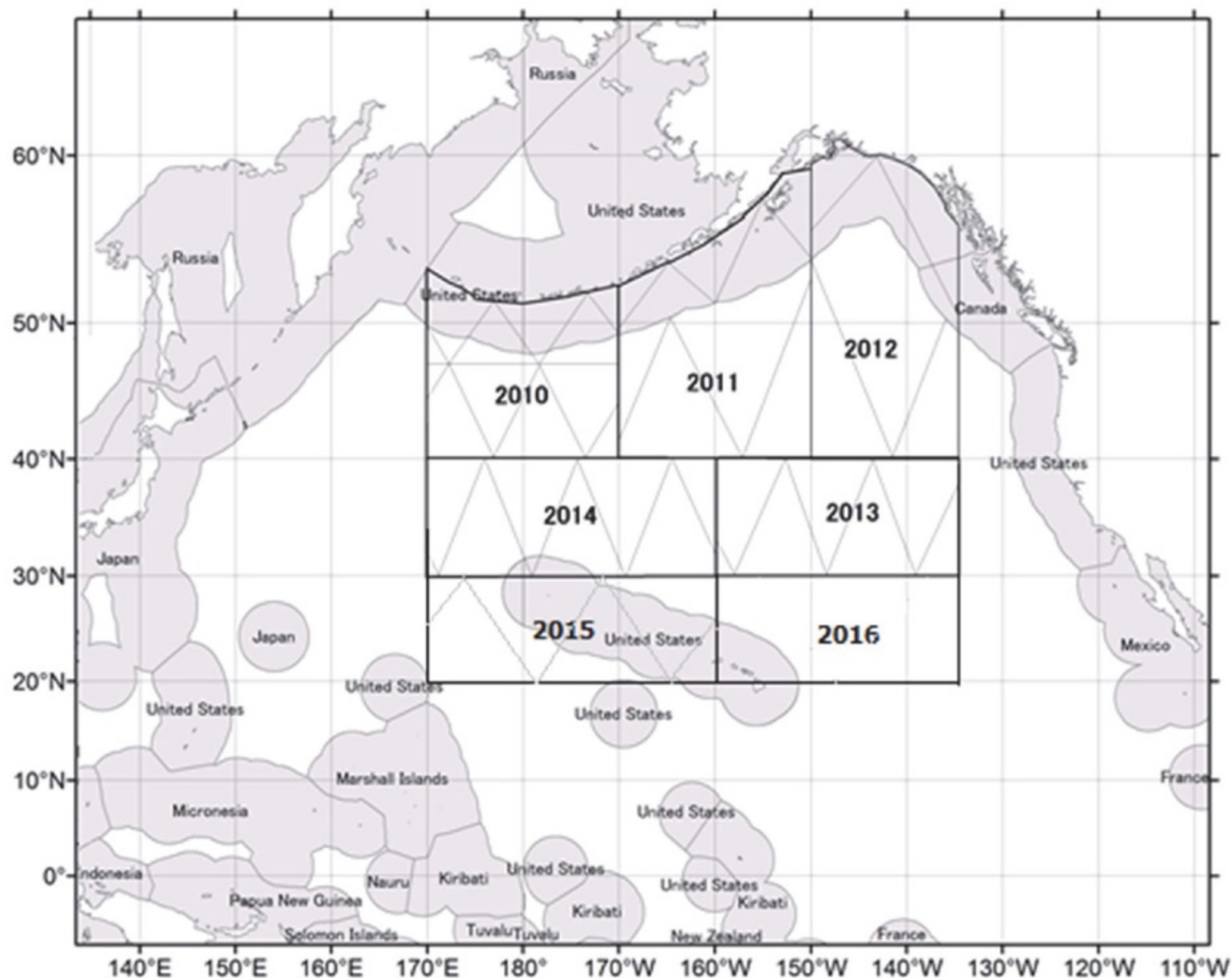
Pilot studies and "you get what you pay for"

- Designing surveys is hard
- Designing surveys is essential
- Better to fail one season than fail for 5, 10 years
- Get information early, get it cheap
 - Inform design from a pilot study

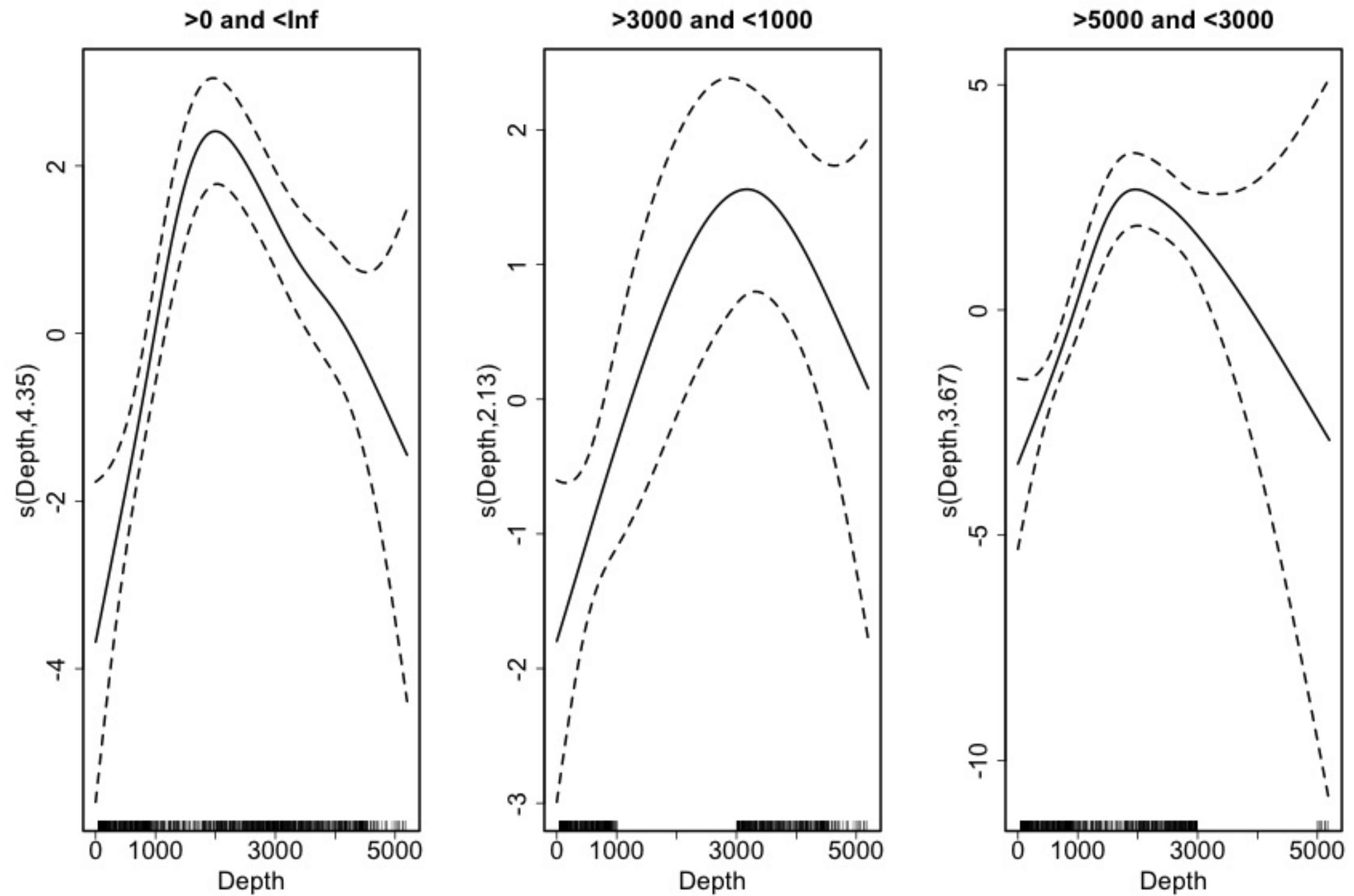
Avoiding rules of thumb

- Think about assumptions
 - Detection function
 - Spatial model
- Think about design
 - Spatial coverage
 - Covariate coverage

Spatial coverage (IWC POWER)

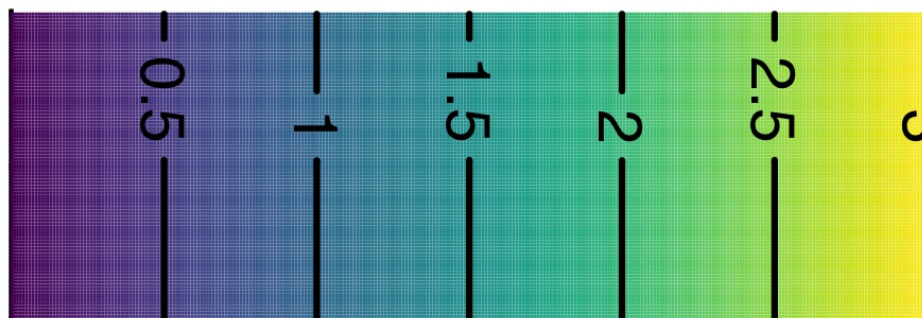
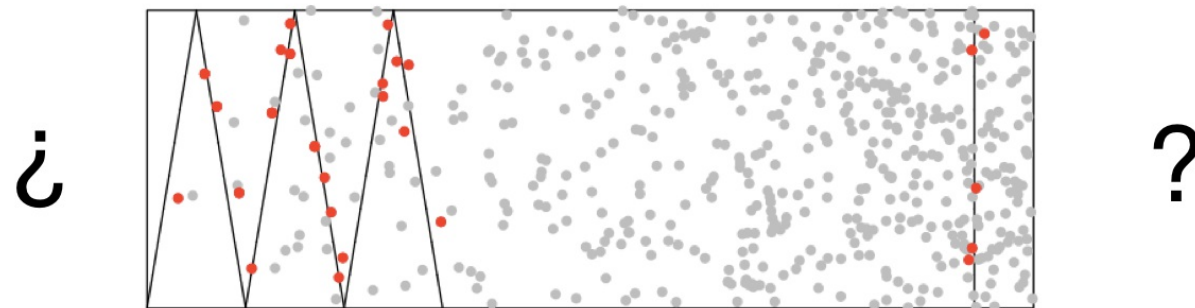


Covariate coverage

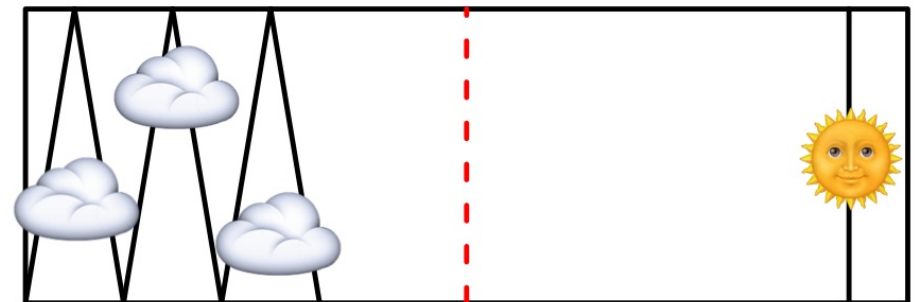


Sometimes things are complicated

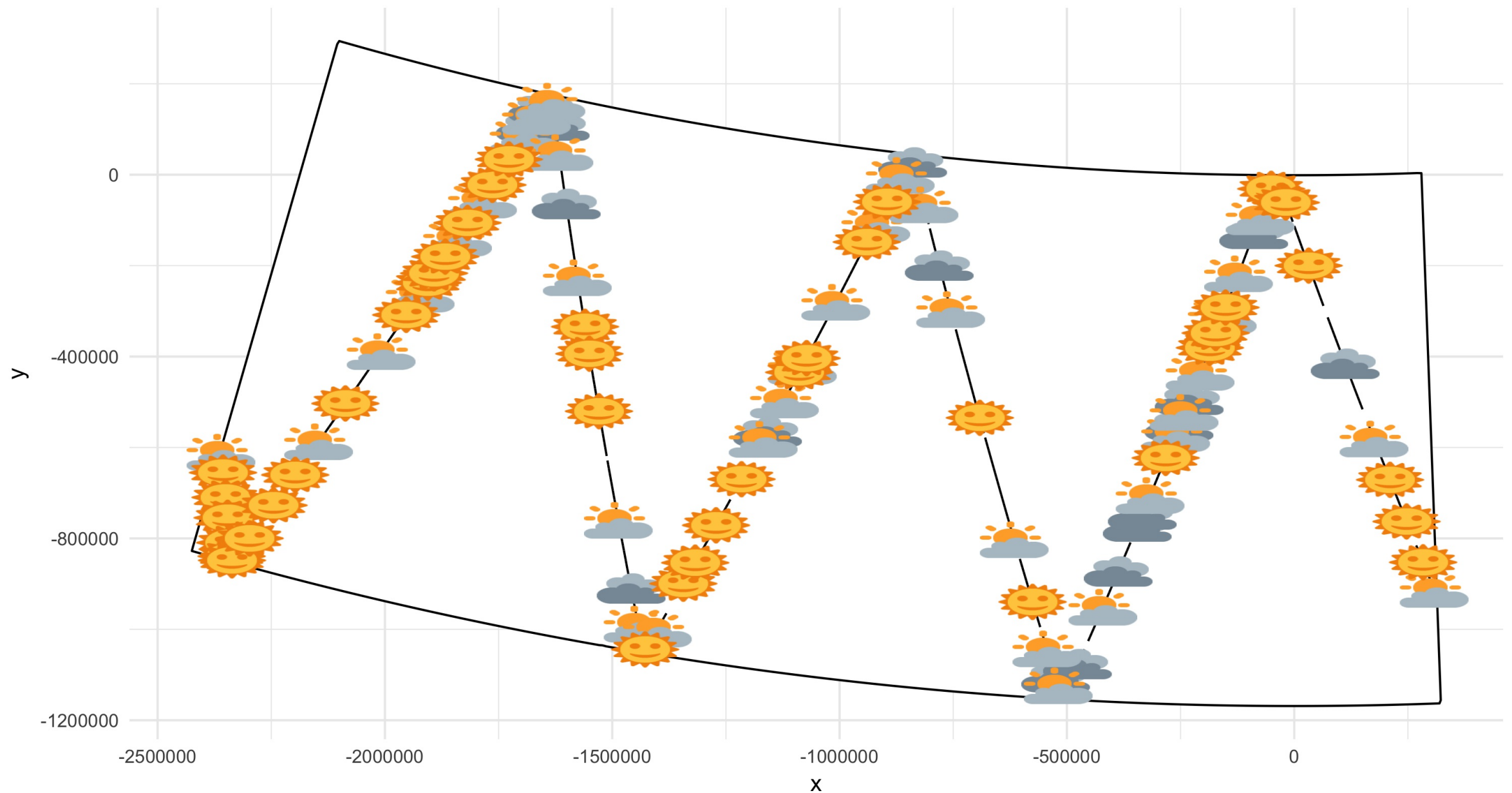
- Weather has a big effect on detectability
- Need to record during survey
- Disambiguate between distribution/detectability
- Potential confounding can be BAD



OR

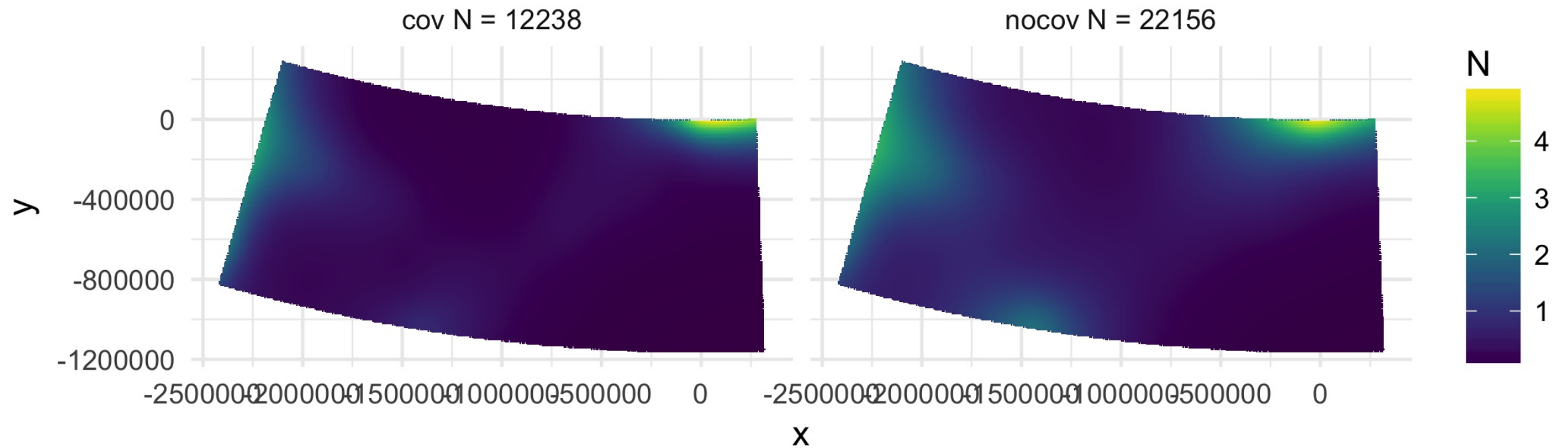


Visibility during POWER 2014



Thanks to Hiroto Murase and co. for this data!

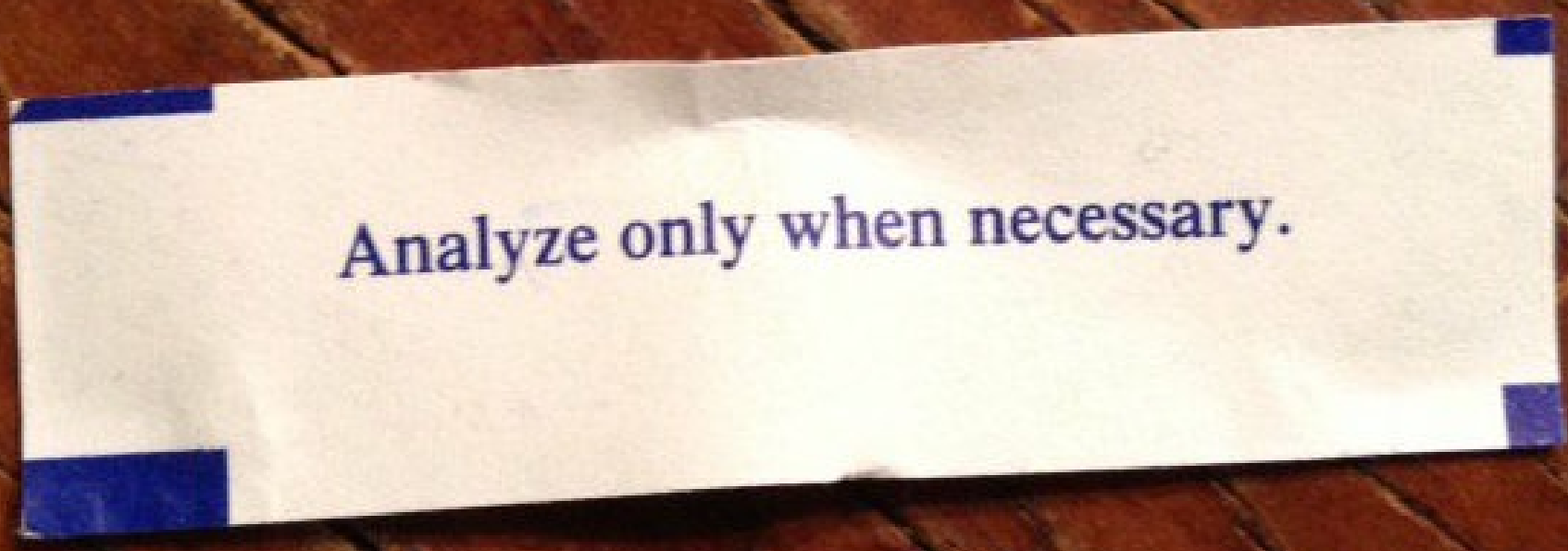
Covariates can make a big difference!



Disappointment

- Sometimes you don't have enough data
- Or, enough coverage
- Or, the right covariates

Sometimes, you can't build a spatial model



Analyze only when necessary.

@kitabeta

"Which of options X, Y, Z is correct?"

Alternatives problem

- When faced with options, try them.
- **Where** does the sensitivity lie?
- What's **really** going on?
- What is your **objective**?

"How big should our segments be?"

Segment size

- If you think it's an issue test it
- Resolution of covariates also important
- Maybe species-/domain-dependent?
- (Solutions on the horizon to avoid this)

"Is our model right?"

Model validation

- Some variety of cross-validation
- Temporal replication
 - Leave out 1 year, fit to others, predict, assess
- Spatial "pseudo-jackknife"
 - Leave out every n^{th} segment, refit, ...
 - (Maybe leave out 2, 3 etc...)

Modelling philosophy

Which covariates should we include?

- Dynamic vs static variables
- Spatial terms? Habitat models?

Getting help

Resources

- Bibliography has pointers to these topics
- Distance sampling Google Group
 - Friendly, helpful, low traffic
 - see distancesampling.org/distancelist.html

Advanced topics

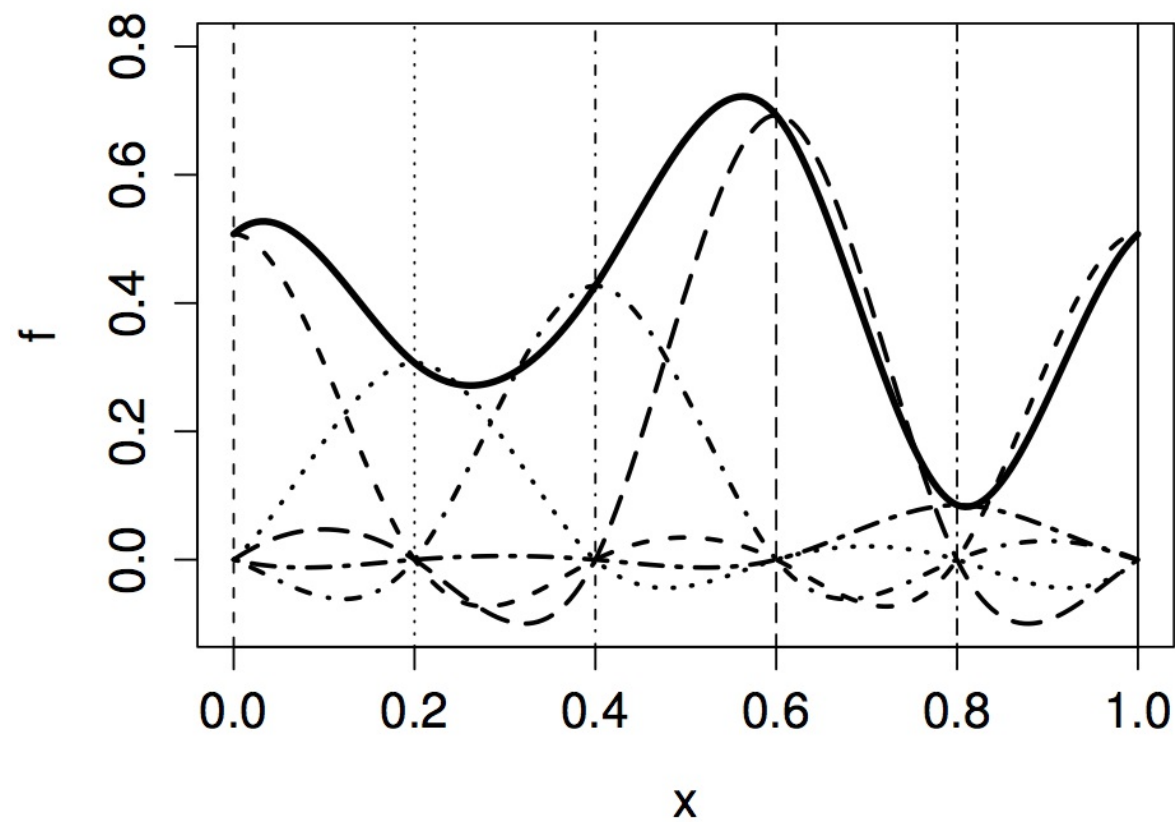
This is a whirlwind tour...

...and some of this is experimental

Smoother zoo

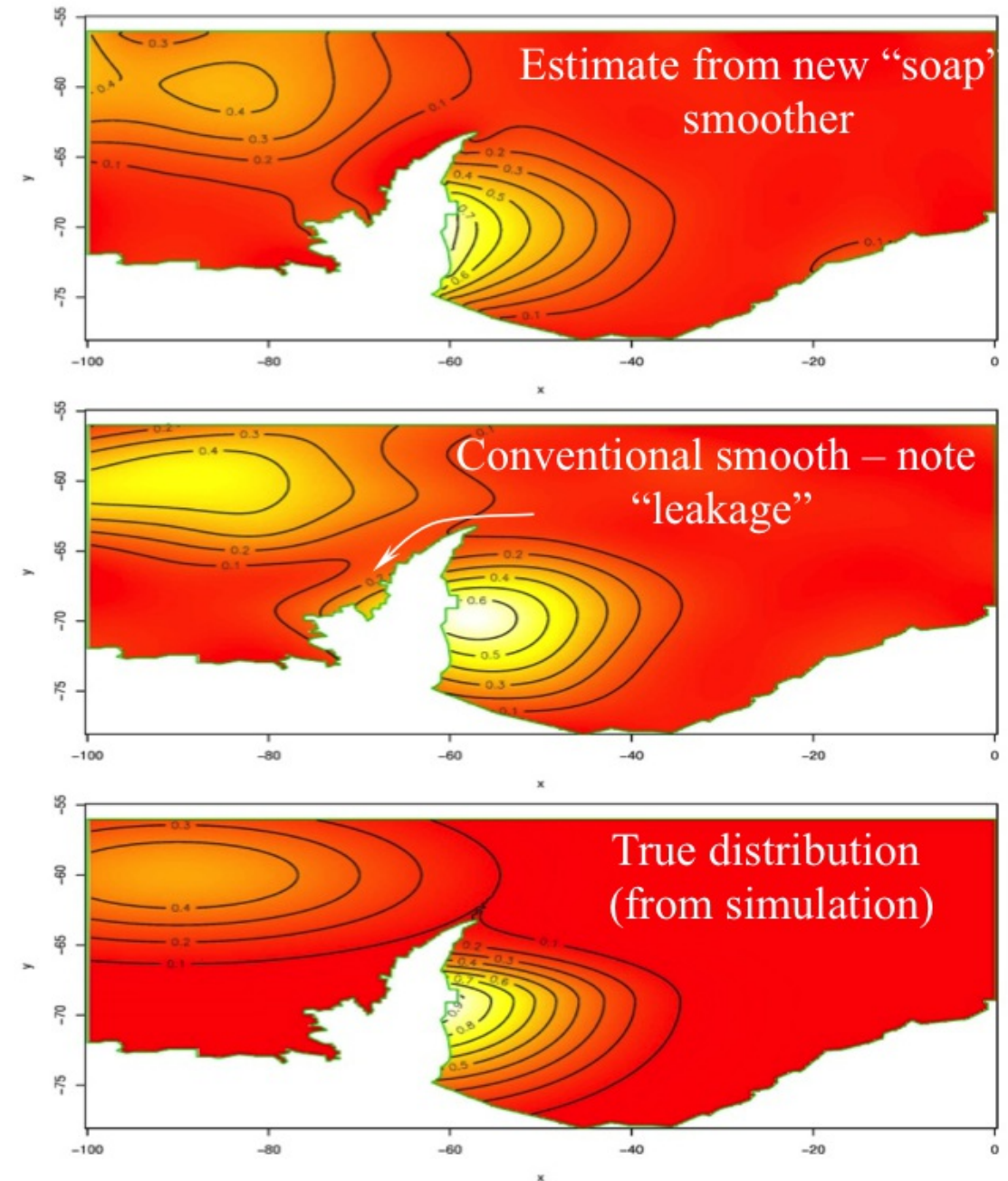
Cyclic smooths

- What if things "wrap around"? (Time, angles, ...)
- Match value and derivative
- Use
- See



Smoothing in complex regions

- Edges are important
- Whales don't live on land
- Bad things happen when we don't account for this
- Include boundary info in smoother
-

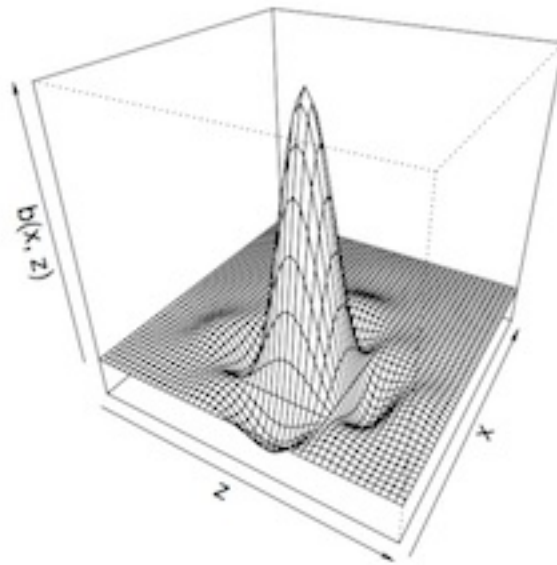
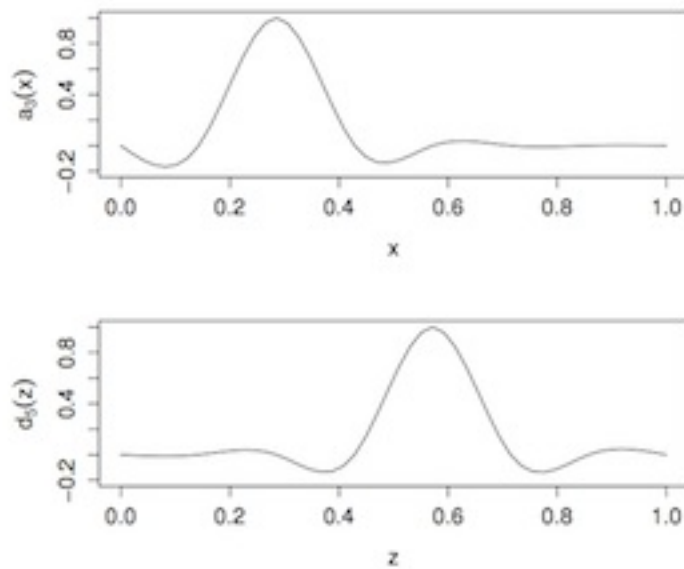


Multivariate smooths

- Thin plate splines are
- 1 unit in any direction is equal
- Fine for space, not for other things

Tensor products

- $s_{x,z}(x, z) = \sum_{k_1} \sum_{k_2} \beta_k s_x(x) s_z(z)$
- As many covariates as you like! (But takes time)
- or (instead of)



Black bears like to sunbathe

Random effects

- normal random effects
- exploits equivalence of random effects and splines
- useful when you just have a “few” random effects
-

Making things faster

Parallel processing

- Some models are very big/slow
- Run on multiple cores
- Use !
- Some constraints in what you can do
- Wood, Goude and Shaw (2015)

Summary

- Lots of complicated problems
- Lots of potential solutions
- (see also "other approaches" mini-lecture)
- Need to get simple things right first
- **Trade assumptions for data**