Predictions and variance

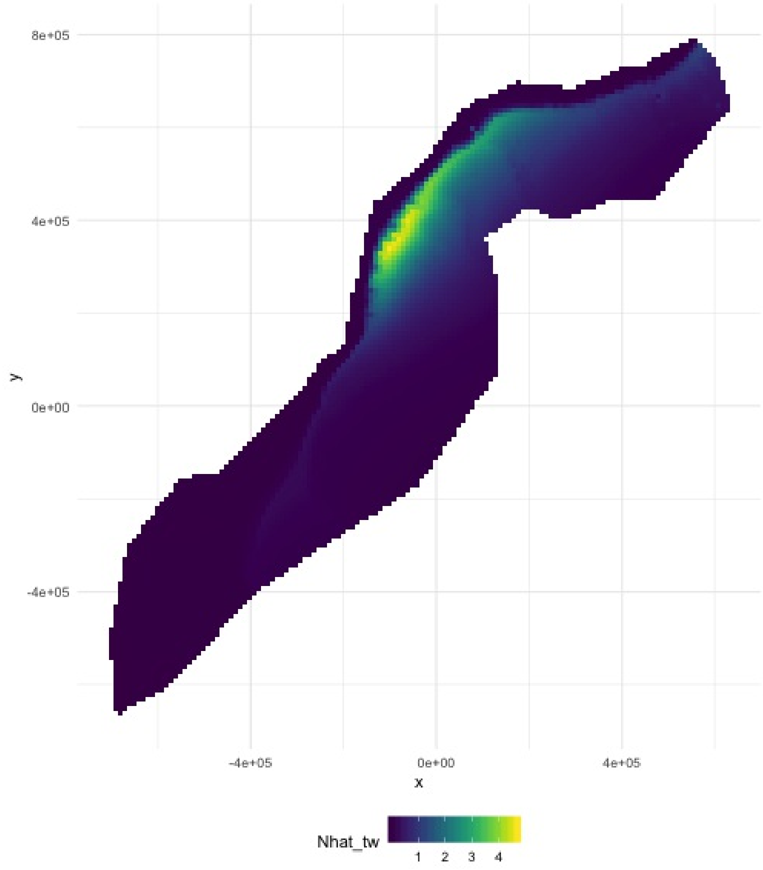
University of St Andrews

So far...

- Build, check & select models for detectability
- Build, check & select models for abundance
- Make some ecological inference about smooths
- **What about predictions?**

Let's talk about maps

What does a map mean?



- Grids!
- Cells are abundance estimate
- "snapshot"
- Sum cells to get abundance
- Sum a subset?

Going back to the formula

Count model (j observations):

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

Predictions (index r):

$$\hat{n}_r = A_r \exp\left[\hat{\beta}_0 + \hat{s}(y_r) + \hat{s}(\text{Depth}_r)\right]$$

Need to "fill-in" values for A_r , y_r and Depth_r .

Predicting

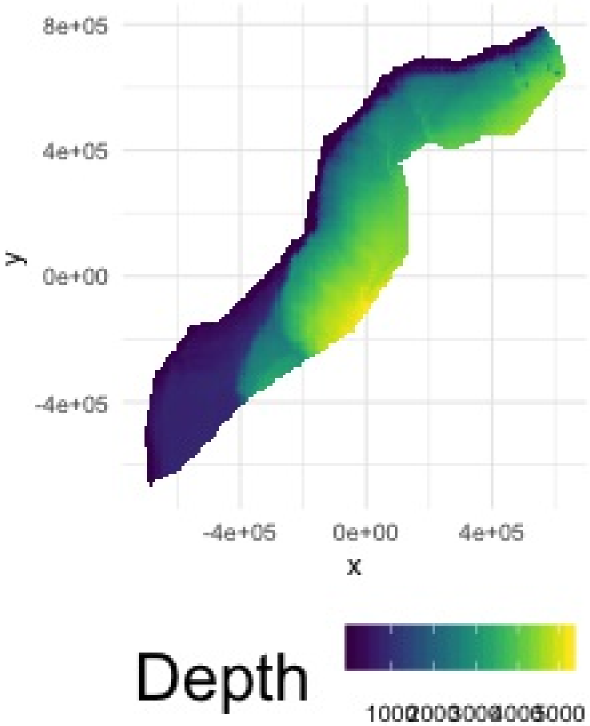
- With these values can use in R
-

Prediction data

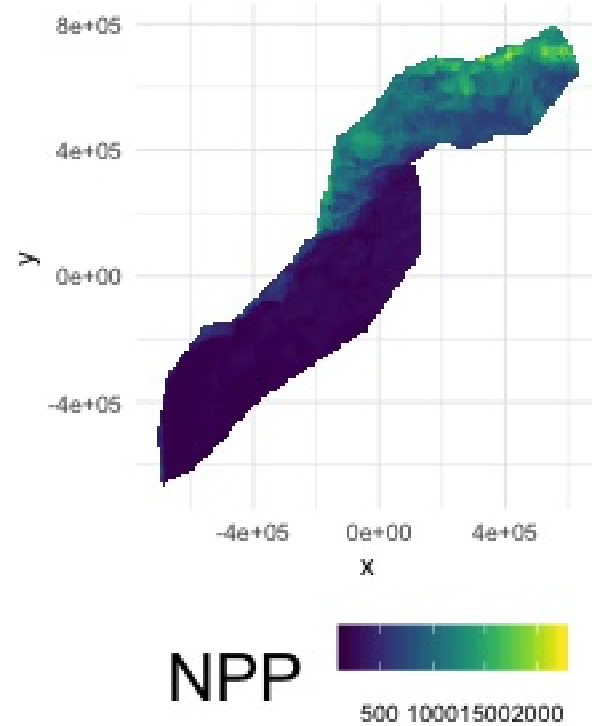
A quick word about rasters

- We have talked about rasters a bit
- In R, the is king
- Fortunately exists
- Make our "stack" and then convert to

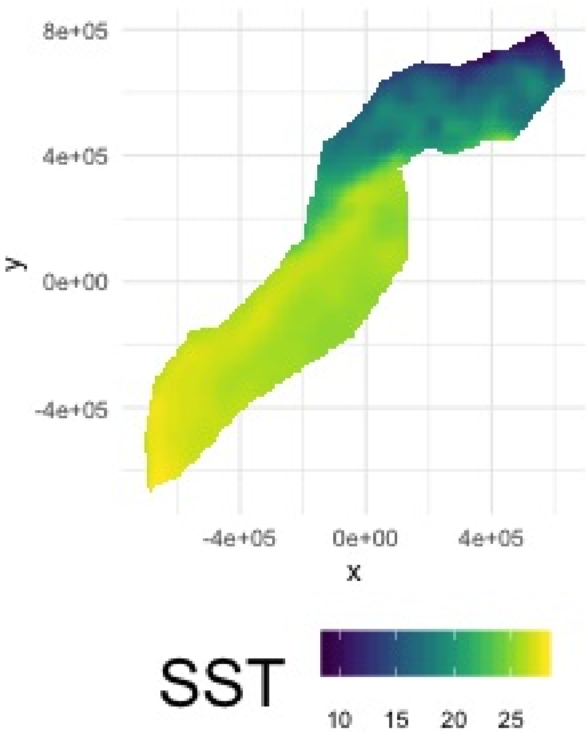
Predictors



8e+05
4e+05
0e+00
-4e+05
y
-4e+05
0e+00
4e+05
x
Depth



8e+05
4e+05
0e+00
-4e+05
y
-4e+05
0e+00
4e+05
x
NPP
500 1000 1500 2000

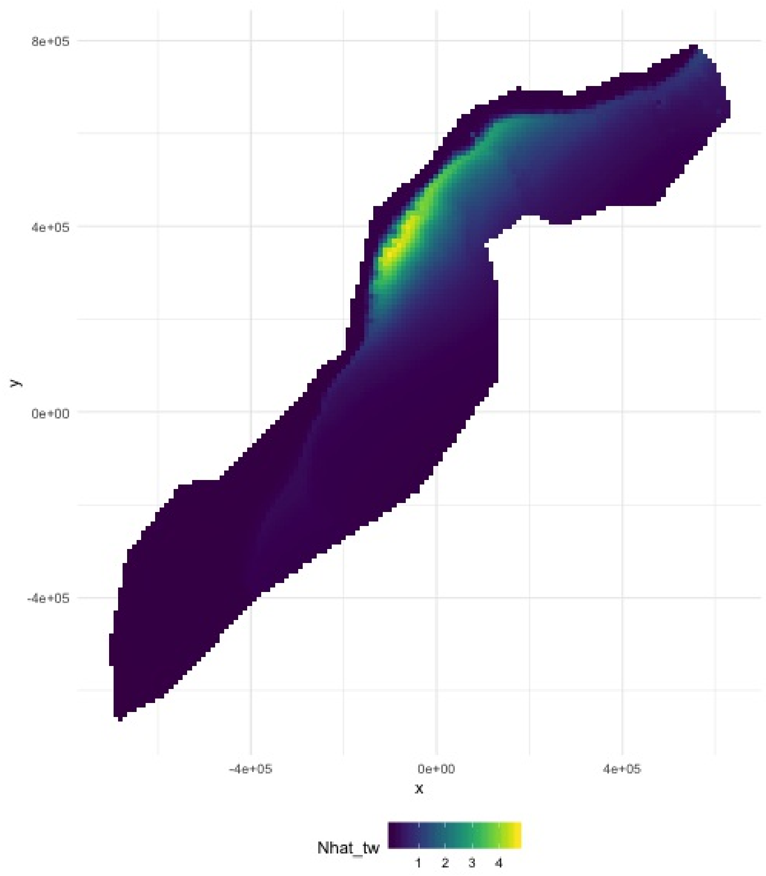


8e+05
4e+05
0e+00
-4e+05
y
-4e+05
0e+00
4e+05
x
SST
10
15
20
25

Making a prediction

- Add another column to the prediction data
- Plotting then easier (in R)

Maps of predictions



Total abundance

Each cell has an abundance, sum to get total

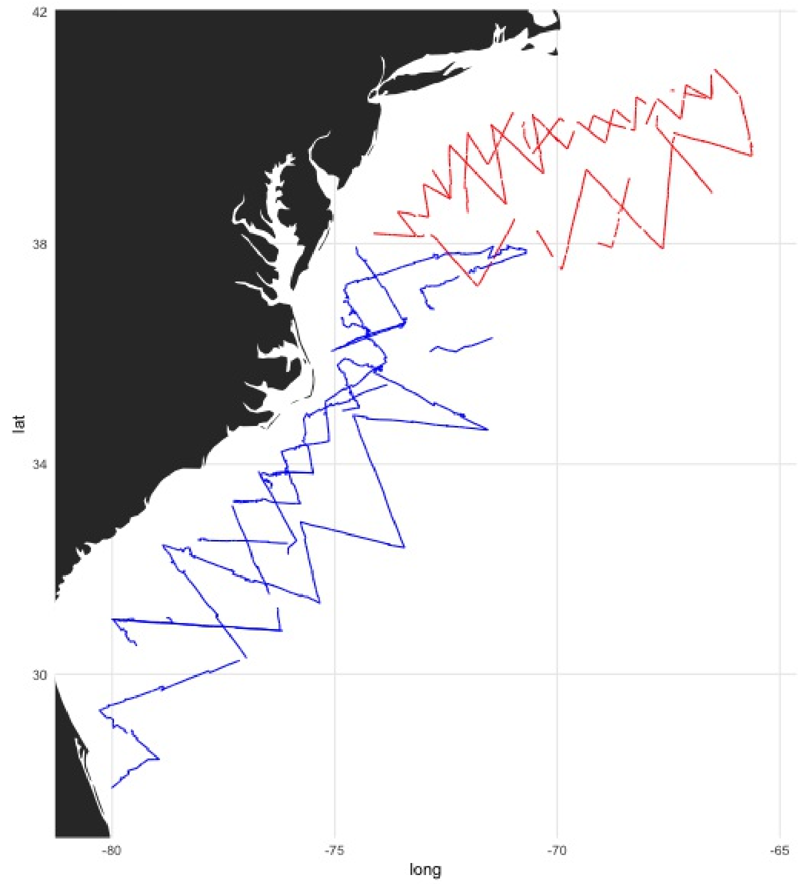
Subsetting

R subsetting lets you calculate "interesting" estimates:

Extrapolation

What do we mean by extrapolation?

- Predicting at values outside those observed
- What does "outside" mean?
  - between transects?
  - outside "survey area"?

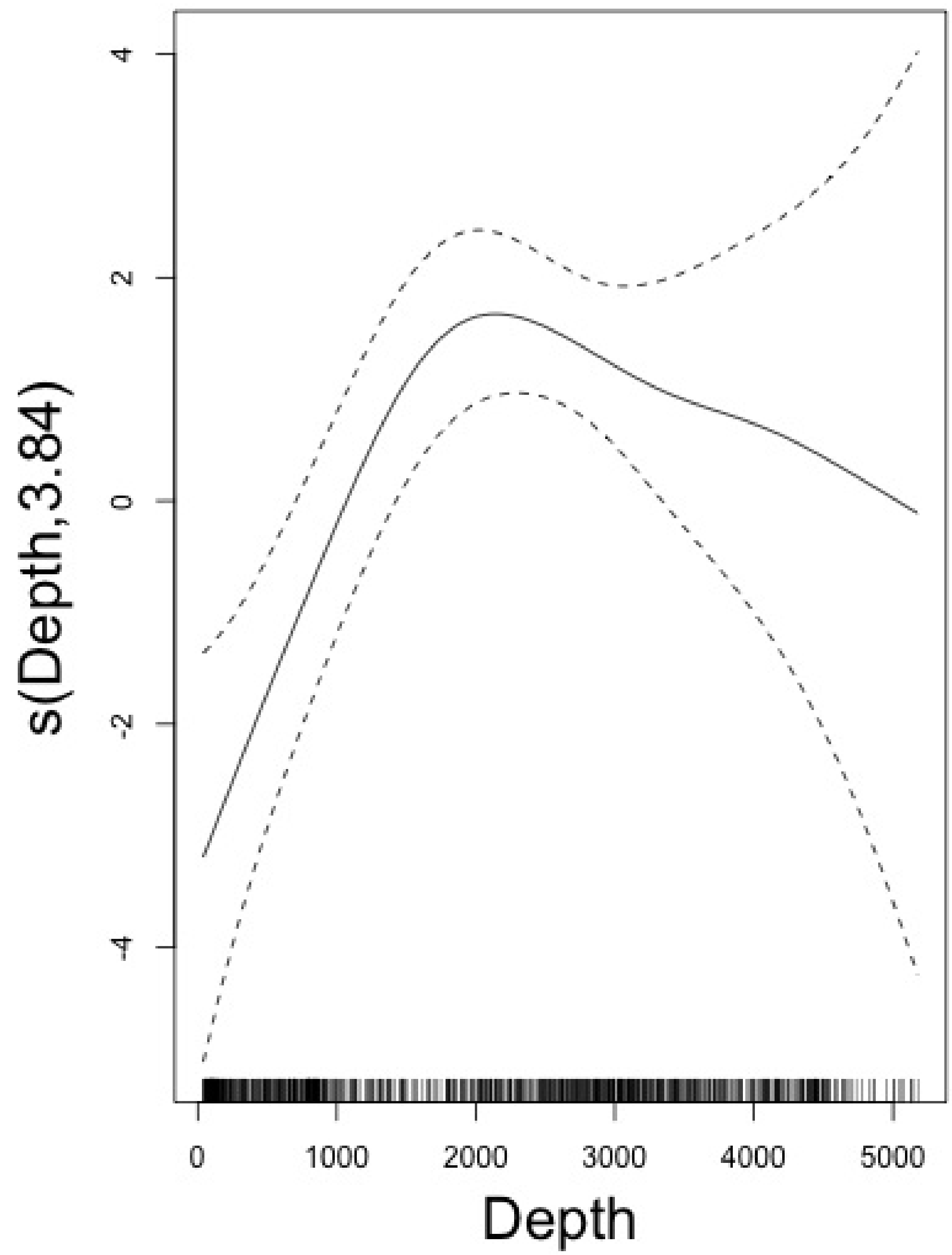


Temporal extrapolation

- Models are temporally implicit (mostly)
- Dynamic variables change seasonally
- Migration can be an issue
- Need to understand what the predictions **are**

Extrapolation

- Extrapolation is fraught with issues
- Want to be predicting "inside the rug"
- In general, try not to do it!
- (Think about variance too!)



Recap

- Using
- Getting "overall" abundance
- Subsetting
- Plotting in R
- Extrapolation (and its dangers)

Estimating variance

Now we can make predictions

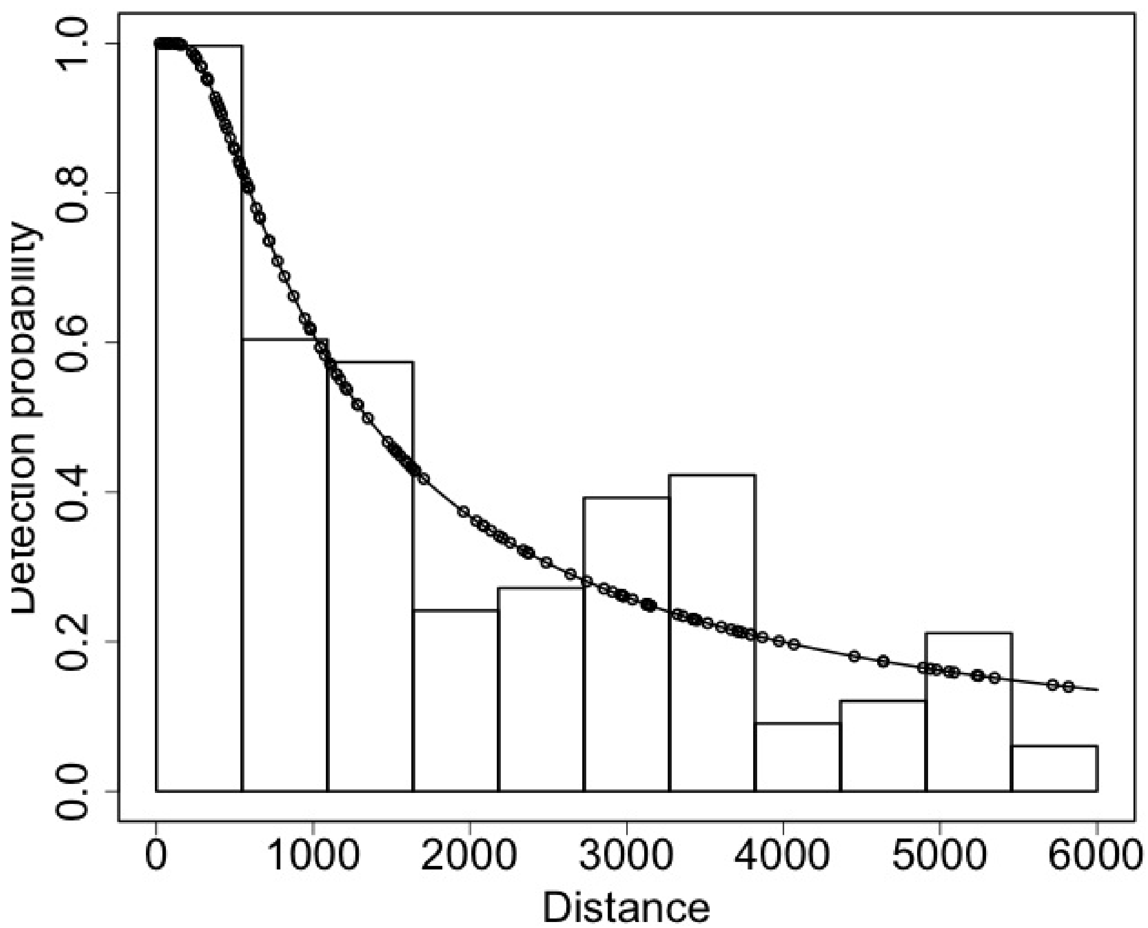
Now we are dangerous.

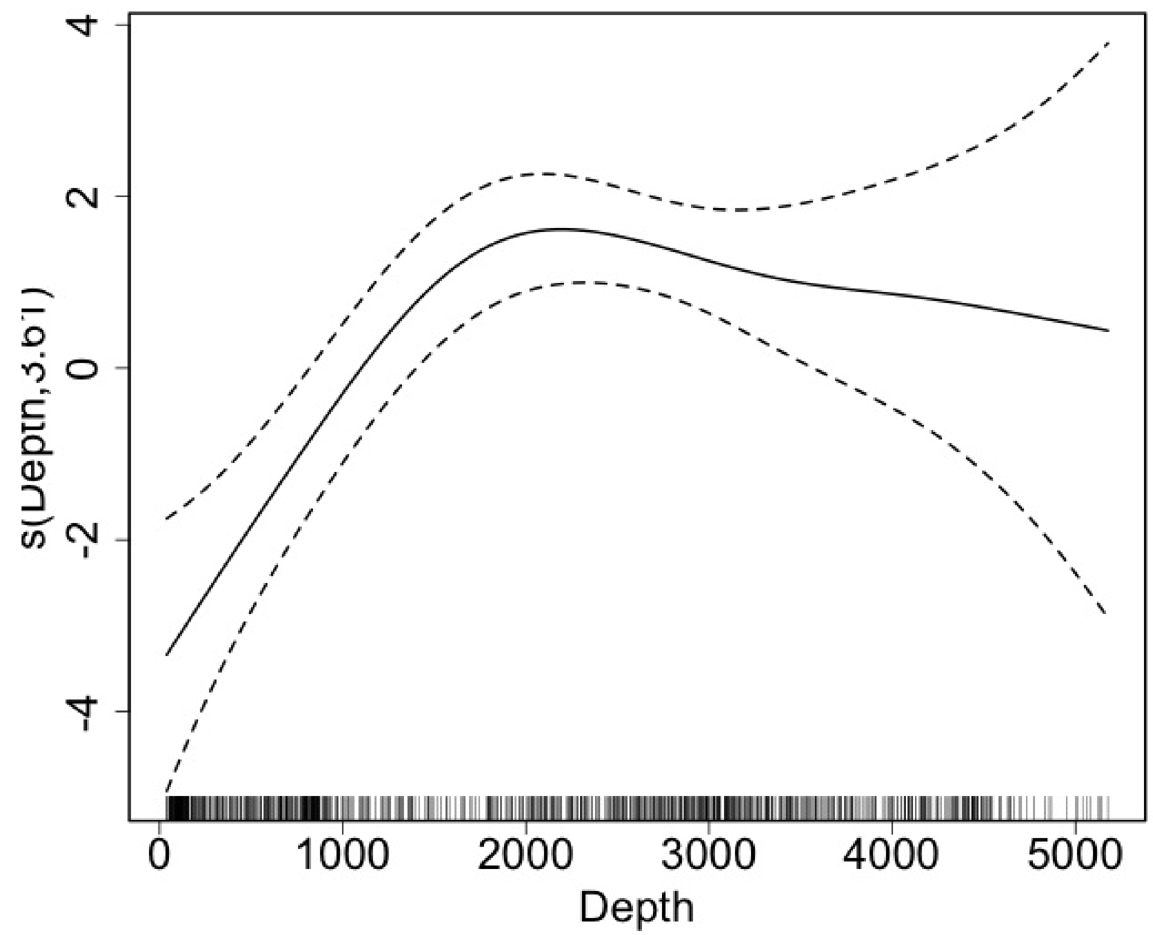
Predictions are useless without uncertainty

Where does uncertainty come from?

Sources of uncertainty

- Detection function parameters
- GAM parameters

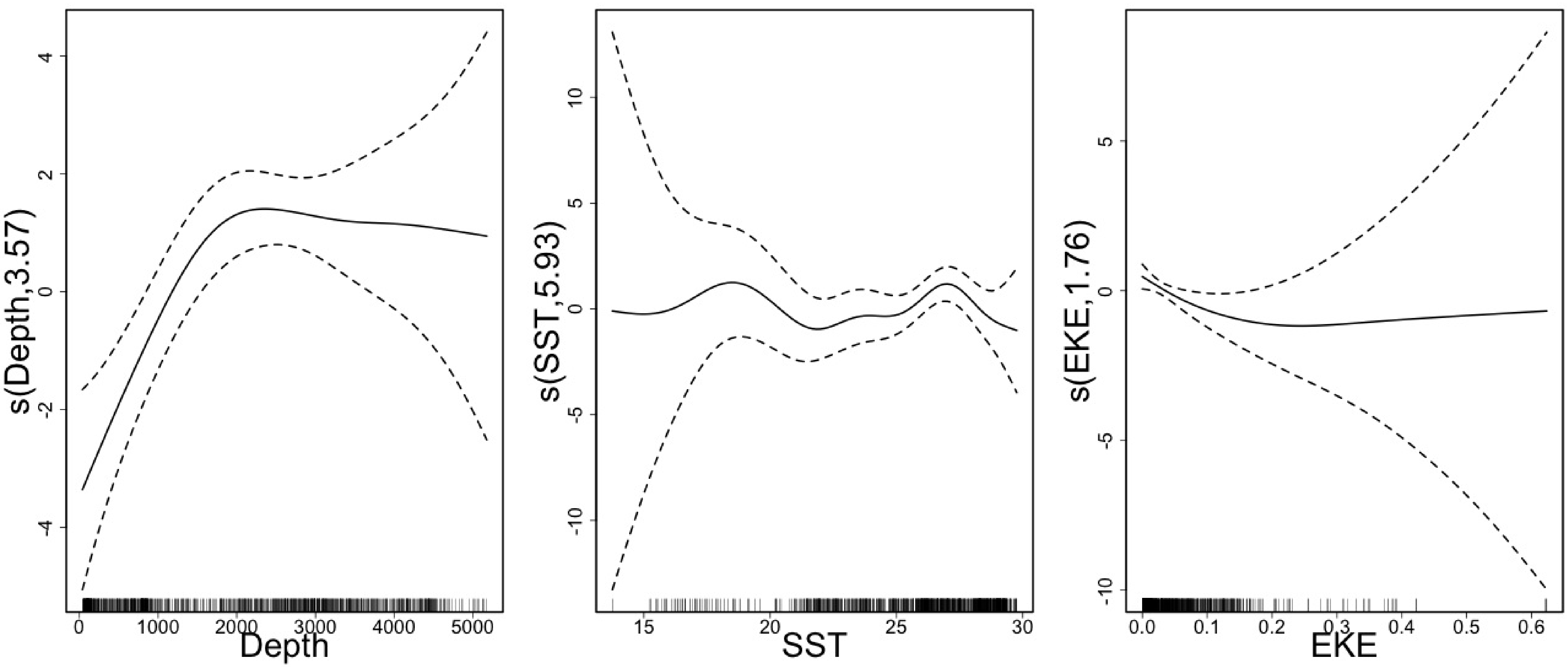




Let's think about smooths first

Uncertainty in smooths

- Dashed lines are +/- 2 standard errors
- How do we translate to $\hat{N}$?



Back to bases

- Before we expressed smooths as:
  - $s(x) = \sum_{k=1}^{K} \beta_k b_k(x)$
- Theory tells us that:
  - $\boldsymbol{\beta} \sim N(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}})$
  - where $\mathbf{V}_{\boldsymbol{\beta}}$ is a bit complicated
  - (derived from the smoother matrix)

Predictions to prediction variance (roughly)

- "map" data onto fitted values $\mathbf{X}\hat{\boldsymbol{\beta}}$
- "map" prediction matrix to predictions $\mathbf{X}_p\hat{\boldsymbol{\beta}}$
- Here $\mathbf{X}_p$ need to take smooths into account
- pre-/post-multiply by $\mathbf{X}_p$ to "transform variance"
  - $\Rightarrow \mathbf{X}_p^T \mathbf{V}_{\boldsymbol{\beta}} \mathbf{X}_p$
  - link scale, need to do another transform for response

Adding in detection functions

GAM + detection function uncertainty

(Getting a little fast-and-loose with the mathematics)

$$\text{CV}^2\left(\hat{\text{N}}\right) \approx \text{CV}^2\left(\text{GAM}\right) + \text{CV}^2\left(\text{detection function}\right)$$

Not that simple...

- Assumes detection function and GAM are **independent**
- **Maybe** this is okay?
- (Probably not true?)

Variance propagation

- Include the detectability as term in GAM
- Random effect, mean zero, variance of detection function
- Uncertainty "propagated" through the model
- Details in bibliography (too much to detail here)
- Under development
- (Can cover in special topic)

That seemed complicated...

R to the rescue

In R...

- Functions in to do this
-
  - assumes spatial model and detection function are independent
-
  - propagates uncertainty from detection function to spatial model
  - only works for models (more or less)

Variance of abundance

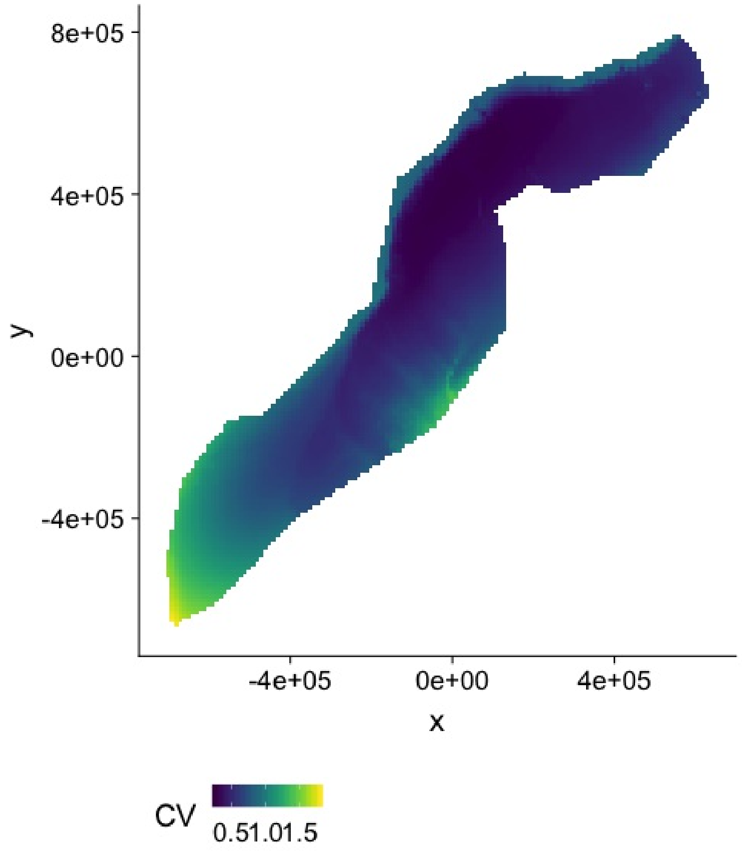
Using

Plotting - data processing

- Calculate uncertainty per-cell
- thinks is one "region"
- Need to split data into cells (using)
- (Could be arbitrary sets of cells, see exercises)
- Need and of cells for plotting

Plotting (code)

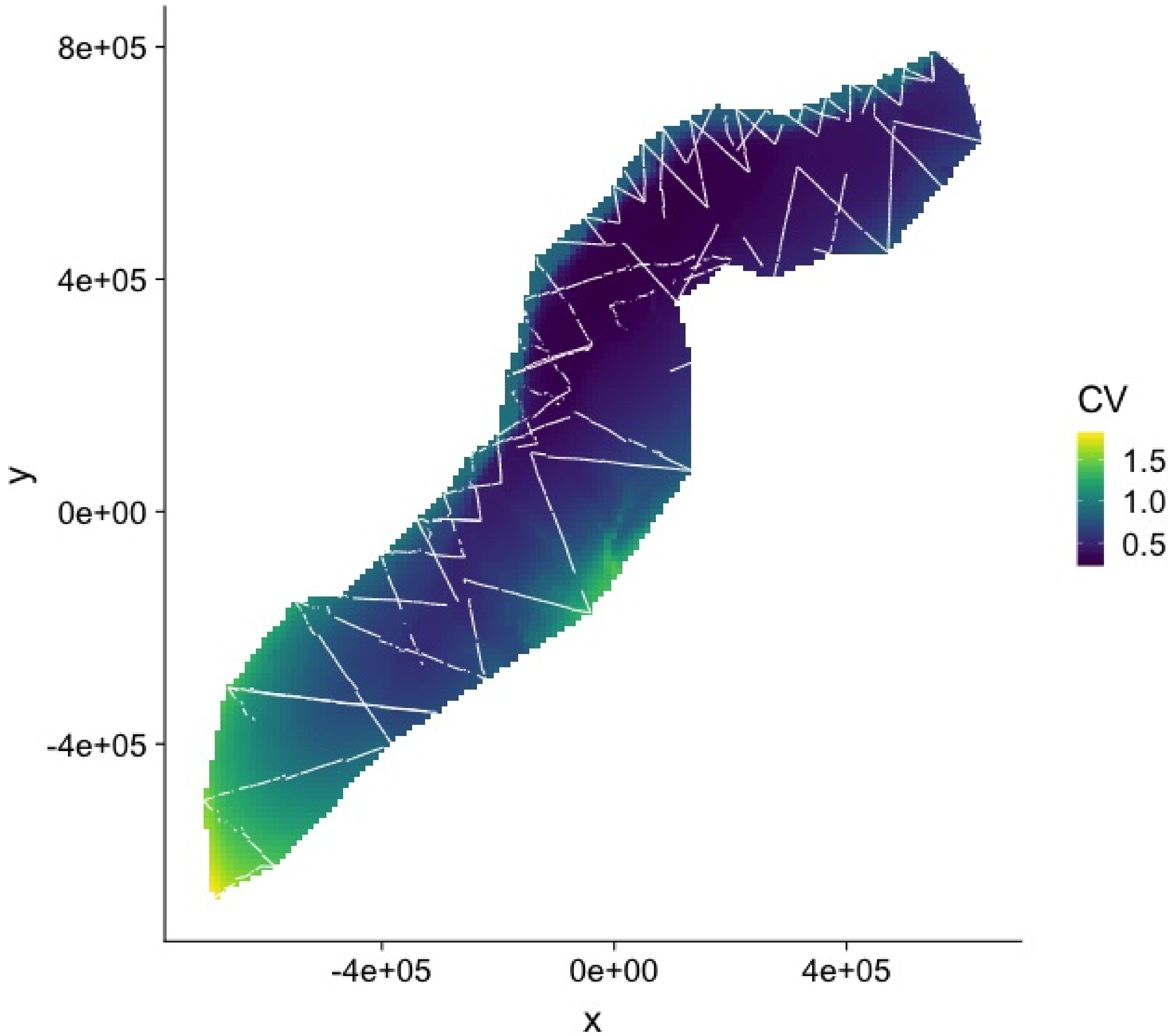
CV plot



Interpreting CV plots

- Plotting coefficient of variation
- Standardise standard deviation by mean
- $CV = se(\hat{N})/\hat{N}$ (per cell)
- Can be useful to overplot survey effort

Effort overplotted



Big CVs

- Here CVs are "well behaved"
- Not always the case (huge CVs possible)
- These can be a pain to plot
- Use in R to make categorical variable
  - e.g. or somesuch

Recap

- How does uncertainty arise in a DSM?
- Estimate variance of abundance estimate
- Map coefficient of variation

Let's try that!