Making predictions

David L Miller

So far...

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

What predictions do we want to make?

- Abundance estimates
- Maps of abundance
- These are related

Let’s talk about maps

What does a map mean?

- Each cell is an abundance estimate
- Whole map is a “snapshot”
- Sum all the cells to get the overall abundance
- Sum a subset to get a stratified estimate

Going back to the formula

Model:

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

Predictions (index r):

$$n_r = A_r \exp\left[\beta_0 + s(y_r) + s(\mathrm{Depth}_r)\right]$$

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

Predicting

- With these values can use `predict` in R
- `predict(model, newdata=data)`

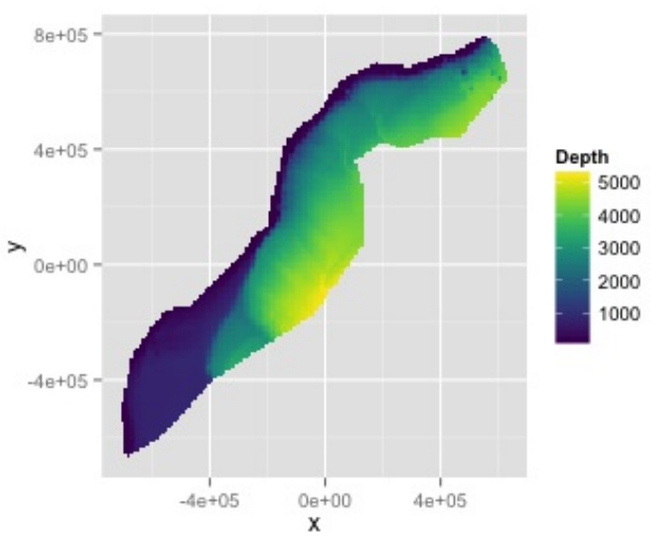
Rasters

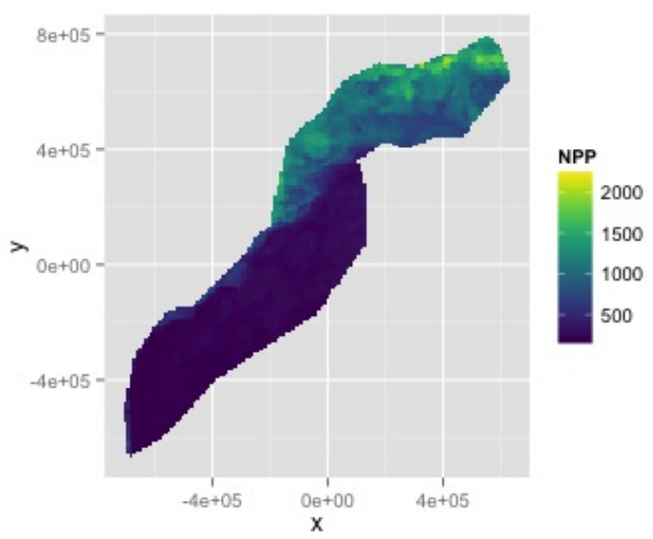
- Jason has talked about rasters a bit
- In R, the `data.frame` is king
- Fortunately `as.data.frame` exists
- Make our "stack" and then convert to `data.frame`

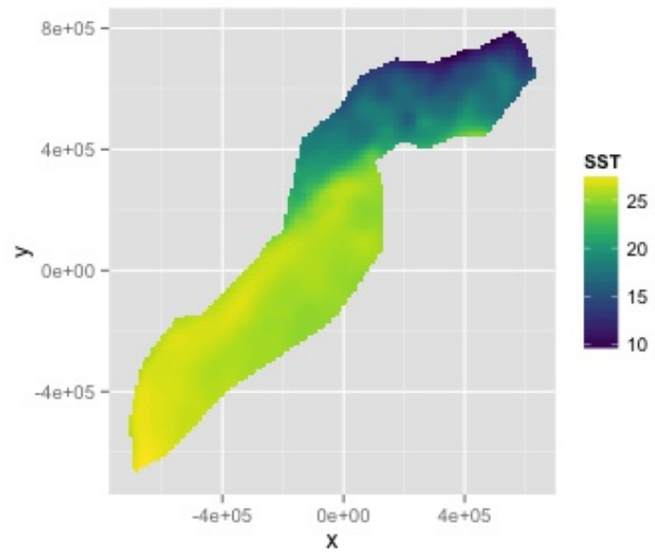
Prediction data

	x	y	Depth	SST	NPP	off.set
126	547984.6	788254	153.59825	9.049170	1462.521	1e+08
127	557984.6	788254	552.31067	9.413981	1465.410	1e+08
258	527984.6	778254	96.81992	9.699239	1429.432	1e+08
259	537984.6	778254	138.23763	9.727216	1424.862	1e+08
260	547984.6	778254	505.14386	9.880866	1379.351	1e+08
261	557984.6	778254	1317.59521	10.091471	1348.544	1e+08

Prediction data plotted





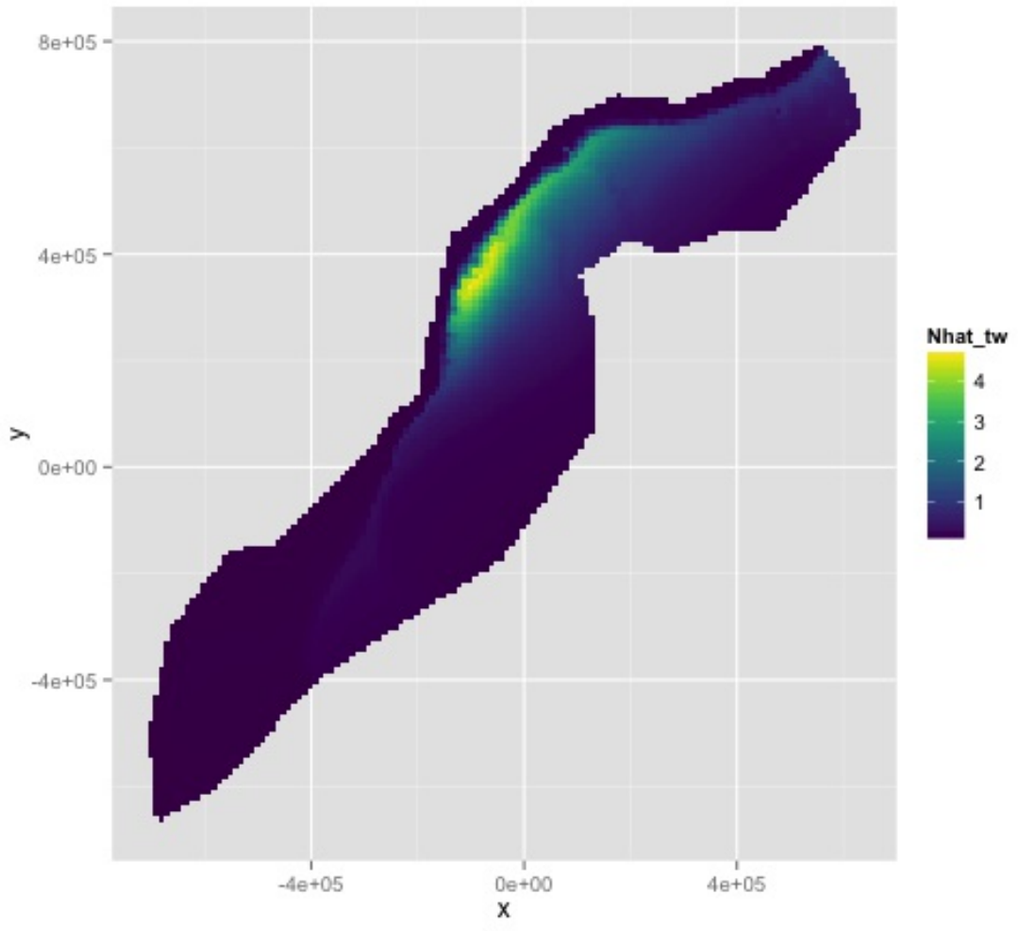


Making a prediction

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

```
predgrid$Nhat_tw <- predict(dsm_all_tw_rm, predgrid)
```

Maps of predictions




```
p <- ggplot(predgrid) +

geom_tile(aes(x=x,y=y,fill=Nhat_tw))
 +
      scale_fill_viridis() +
      coord_equal()
print(p)
```

Total abundance

Each cell has an abundance, sum to get total

```
sum(predict(dsm_all_tw_rm, predgrid))
```

```
[1] 2491.864
```

Subsetting

R subsetting lets you calculate “interesting” estimates:

```
# how many sperm whales at depths less than 2500m?
sum(predgrid$Nhat_tw[predgrid$Depth <= 2500])
```

```
[1] 1006.271
```

```
# how many sperm whales North of 0?
sum(predgrid$Nhat_tw[predgrid$x>0])
```

```
[1] 1383.744
```

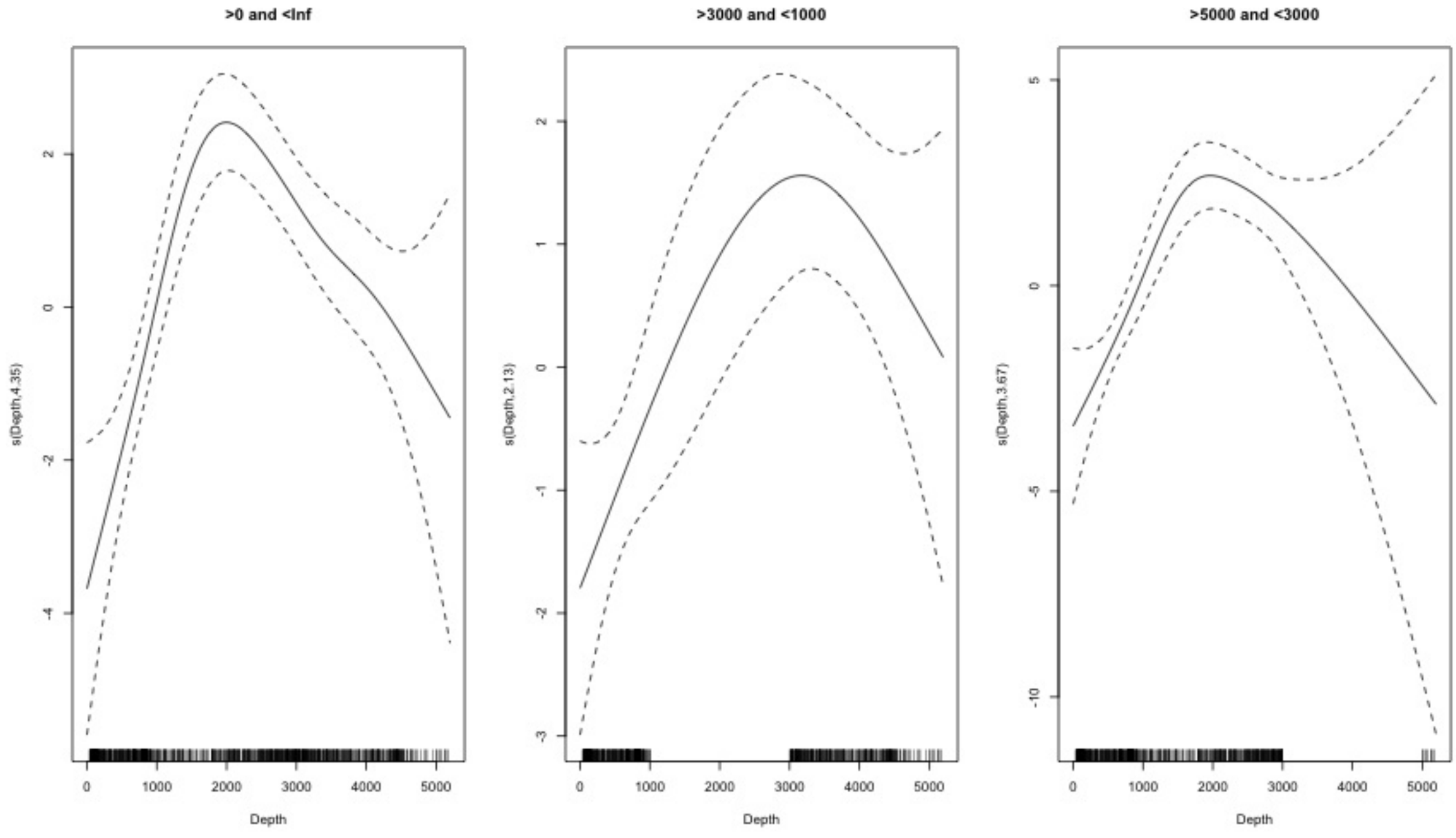
Extrapolation

DANGER WILL ROBINSON, DANGER

What do we mean by extrapolation?

- Predicting at values outside those observed
- What does “outside” mean?
- Multidimensional problem

”Outside”


>0 and <Inf
>3000 and <1000
>5000 and <3000
s(Depth,4.35)
s(Depth,2.13)
s(Depth,3.67)
Depth
Depth
Depth

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
- In general, try not to do it!
- Want to be predicting “inside the rug”
- More on this in the “advanced” lecture

Recap

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