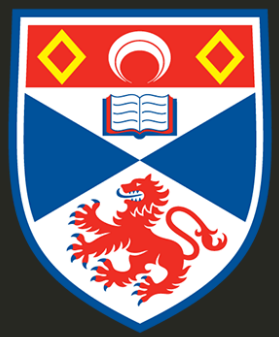
Extras
and
advanced topics



University of
St Andrews

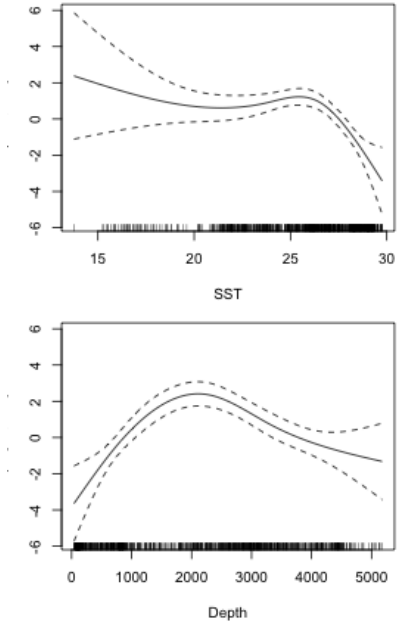
More complicated effects

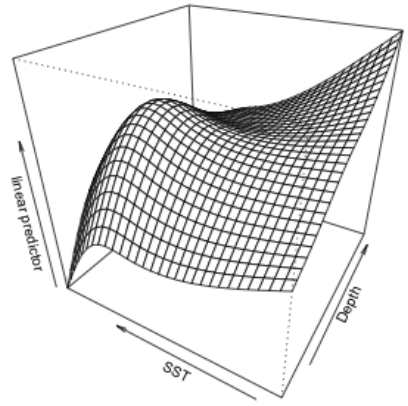
`s(x,y)` doesn't always work

- Only works for `bs="tp"` or `bs="ts"`
- Covariates are isotropic
- What if we wanted to use lat/long?
- Or, more generally: interactions between covariates?

Enter `te()`

- We can built interactions using `te()`
- Construct 2D basis from 2 1D bases
- 💭 "marginal 1Ds, join them up"





Using `te()`

Just like `s()`:

```
dsm_te <- dsm(count ~ te(Depth, SST),
              ddf.obj=df_hr,
              observation.data=obs, segment.data=segs,
              family=tw())
```

summary

```
##
## Family: Tweedie(p=1.282)
## Link function: log
##
## Formula:
## count ~ te(Depth, SST) + offset(off.set)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.3862     0.2831  -72.02   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                 edf Ref.df     F p-value
## te(Depth,SST) 11.79  14.03 7.104  <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.117   Deviance explained = 36.6%
## -REML = 387.64  Scale est. = 4.5541    n = 949
```

Things to fiddle with

- Setting `k=` 2 ways:
  - `k=5`: 5 for all covariates (total $5 * 5 = 25$)
  - `k=c(3,5)`: per basis, in order (total $3 * 5 = 15$)
- Setting `bs=` 2 ways:
  - `bs="tp"`: tprs for all bases
  - `bs=c("tp", "tp")`: tprs per basis

Pulling `te()` apart: `ti()`

- Can we look at the components of the `te()`
- `te(x, y) = ti(x, y) + ti(x) + ti(y)`

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

summary

```
##
## Family: Tweedie(p=1.281)
## Link function: log
##
## Formula:
## count ~ ti(Depth, SST) + ti(Depth) + ti(SST) + offset(off.set)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.4337     0.2868  -71.25   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                 edf Ref.df      F  p-value
## ti(Depth,SST) 2.295  2.794  2.068    0.124
## ti(Depth)     3.477  3.817 16.905  < 2e-16 ***
## ti(SST)       3.175  3.505  8.492 4.08e-06 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.114   Deviance explained =   36%
```

Space x time

- We had a 2d spatial model, add time?
  - `te(x, y, year)` ?
- `d=` groups covariates
  - `te(x, y, year, d=c(2, 1))` gives `x, y` smooth and `year` smooth tensor
- (Assuming default `k=` and `bs=` for bases above)

Fiddling

- Often fewer temporal replicates
  - Fewer years than unique locations
  - `k=` smaller for temporal covariate?
- Use cubic spline basis for time?
  - simpler basis, even knot placement
- When using `ti()` arguments (`k`, `bs`) need to match up between terms
  - if `k=3` for `Depth` in one term it needs to be that in all terms

Other effects

Random effects

- "Simple" random slope/random intercept models
- `s(..., bs="re")`
- **think** about what these models mean

Factor-smooth interactions

- What if we only have a few "years"?
- What if we don't think the "years" are smooth?
  - (Before/after?)
- Terms like `s(Depth, by=year)` change the smooth by year
- also `s(Depth, year, bs="fs")` (lots of ways to specify)
- see Pedersen et al. (2019) for more on these models

Availability

Availability

- Is an animal *available* to be detected
- e.g., diving marine mammals
- Primitive way to do this in `dsm`
- `availability=` for each segment (only for `count` models)
- Active research area!

g(0), MRDS etc

Mark-recapture distance sampling

- Will be able to include these models in next `dsm` release
- Only independent observer (`"io"`) and trial (`"trial"`) modes supported
- Example here

Combining multiple surveys

Combining multiple surveys

- What about combining aerial/shipboard data?
- Different detection functions
- Again, next `dsm` release allows this
- Fitting complicated models example

Finally...

Recent developments

- New `dsm` out in the next few weeks!
- Fitting DSMs in JAGS/Nimble
- DenMod project has produced lots of methodology
- Society for Marine Mammalogy meeting December

Extra bits

Deviance explained, explained

Deviance explained, explained

- Avoid R^2 (see these notes for more info)
- But what about deviance explained?
- First, what is it?

$$D = -2(l_s - l)$$

where $\mathcal{L}_s$ is the *saturated* log likelihood and $\mathcal{L}$ is the likelihood of our model.

- Saturated means the "best" model we can get, one parameter per data point.
- So meaning is it's relative to the best we can do *for this model*

Deviance explained, explained

- `mgcv` reports "Deviance explained" as a percentage

$$D_{\%} = 100(l_s - l)/l_s$$

- Problem: for different models (with different numbers of parameters) l_s is different
- So are we making fair comparisons?
- AIC is simpler and easier to think about!

More info on deviance for GAMs

More difficulties with explanatory power

- Low (<60%) deviance is common. But why?
- Sampling a temporally variable system
- Revisiting the same place multiple times, we might get zero counts twice and then one large count.
- What should the model make of this?
- Without explicit temporal model, it tries to average
- So prediction will be a "medium" count, bad prediction for the zeros and the large counts
- No one is happy!
- See observed vs. expected diagnostics etc

That's all folks!