Multivariate smoothing & model selection

University of St Andrews

Recap

- How GAMs work
- How to include detection info
- Simple spatial-only models
- How to check those models

Univariate models are fun, but...

Ecology is not univariate

- Many variables affect distribution
- Want to model the **right** ones
- Select between possible models
  - Smooth term selection
  - Response distribution
- Large literature on model selection

Models with multiple smooths

Adding smooths

- Already know that is our friend
- Can build a big model…

```
```

Now we have a huge model, what do we do?

Term selection

- Two popular approaches (use p-values)

Stepwise selection - path dependence

All possible subsets - computationally expensive (fishing?)

p-values

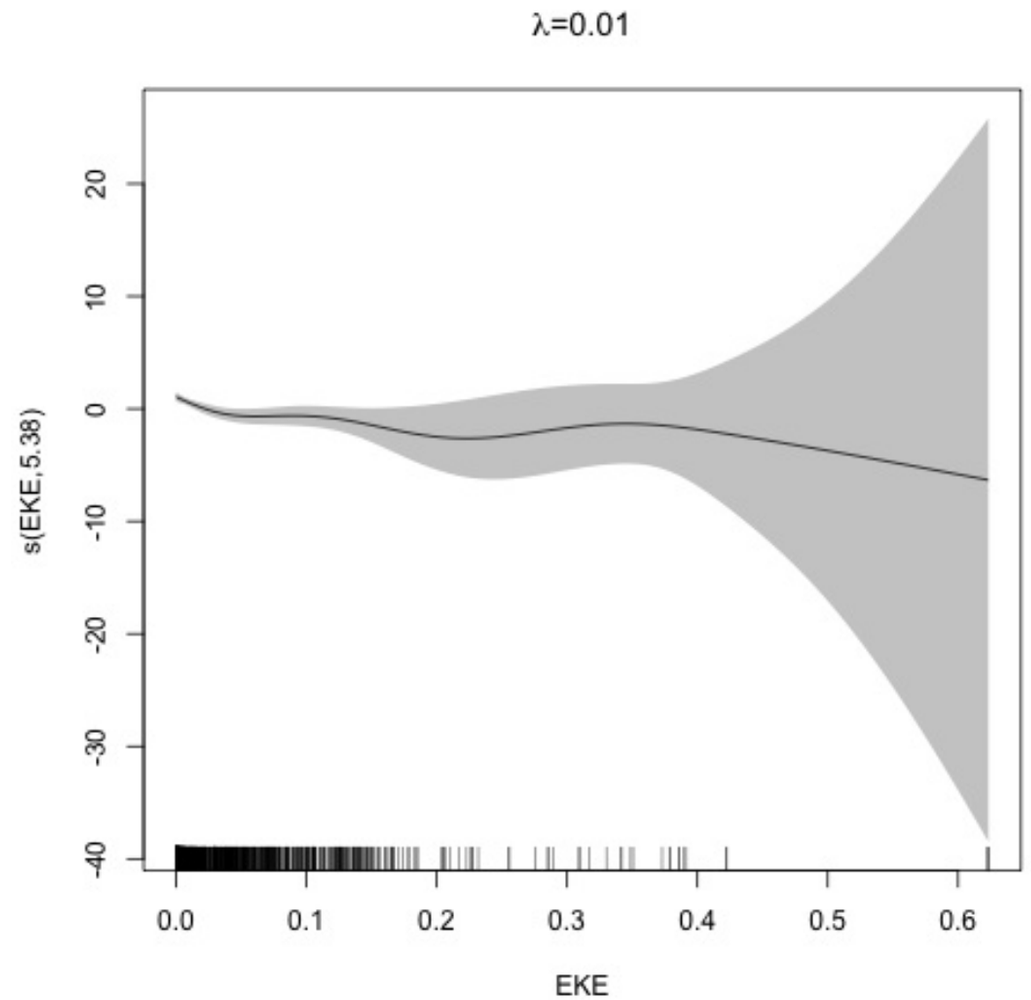
- Test for of a smooth
- They are **approximate** for GAMs (but useful)
- Reported in

p-values example

Path dependence is an issue here

Term selection during fitting

- Penalty already removes complexity from model
- What about using it to remove the whole term?

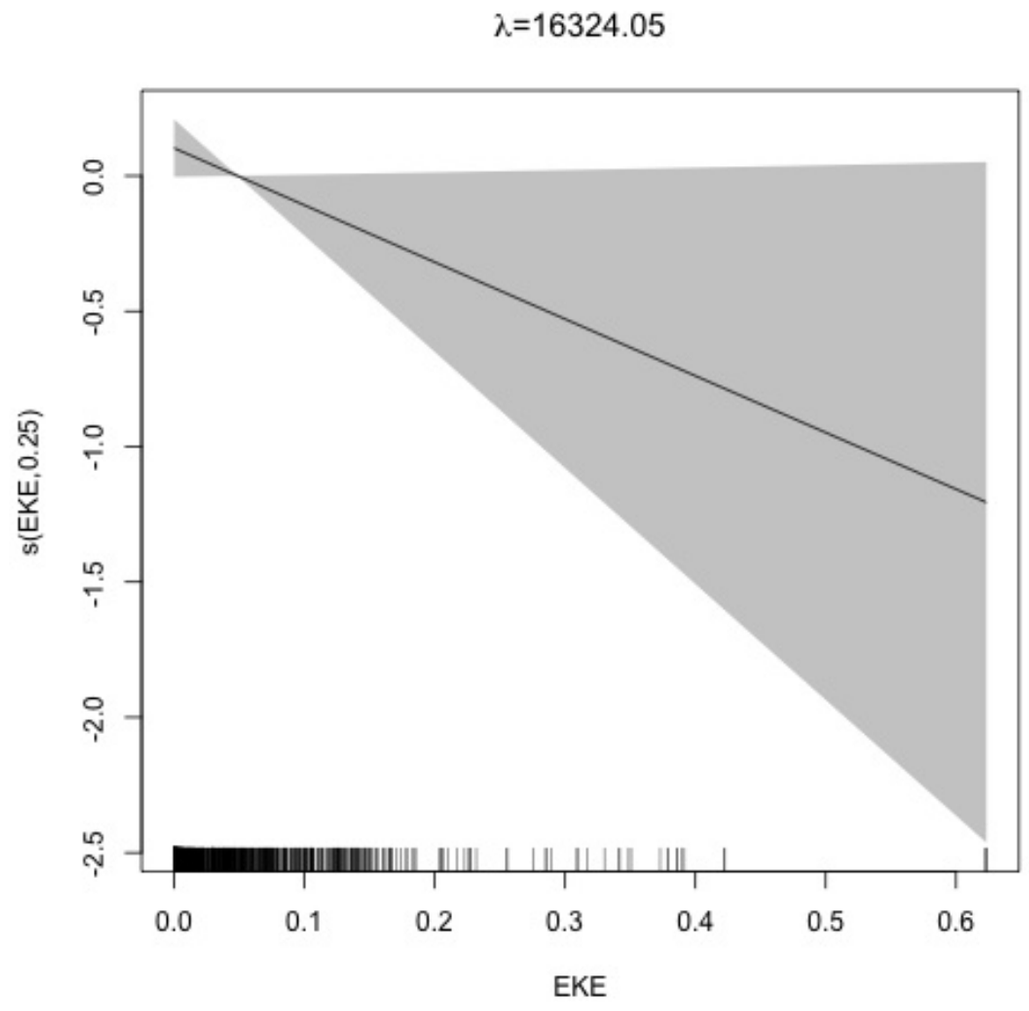


Shrinkage approach

- Basis -
  thin plate splines

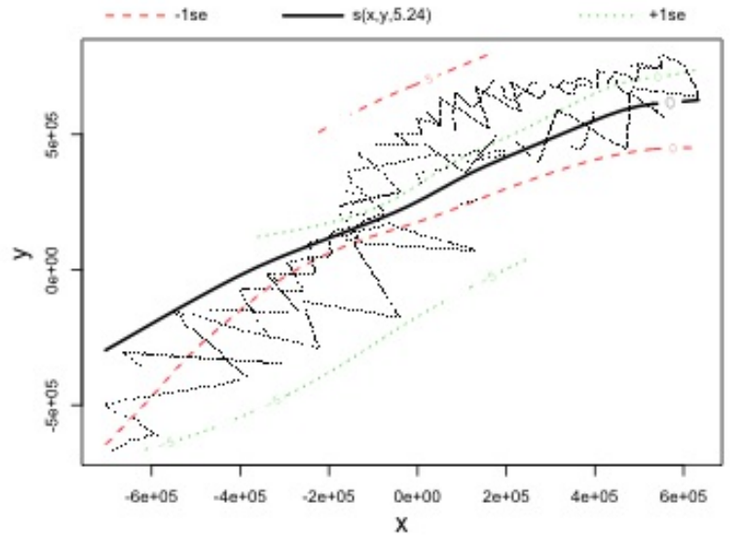
- remove the wiggles **then**
  remove the "linear" bits

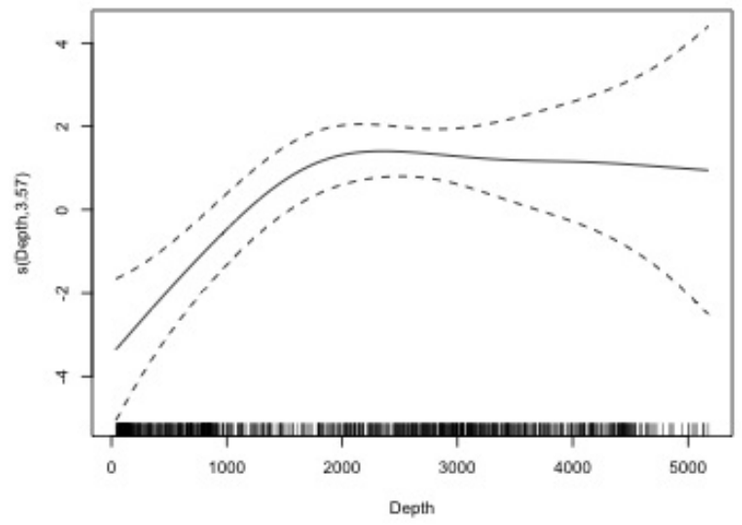
- nullspace should be shrunk
  less than the wiggly part

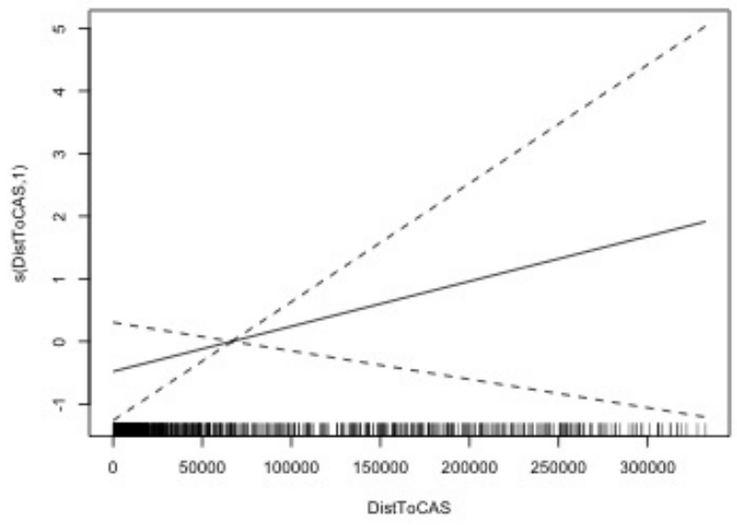


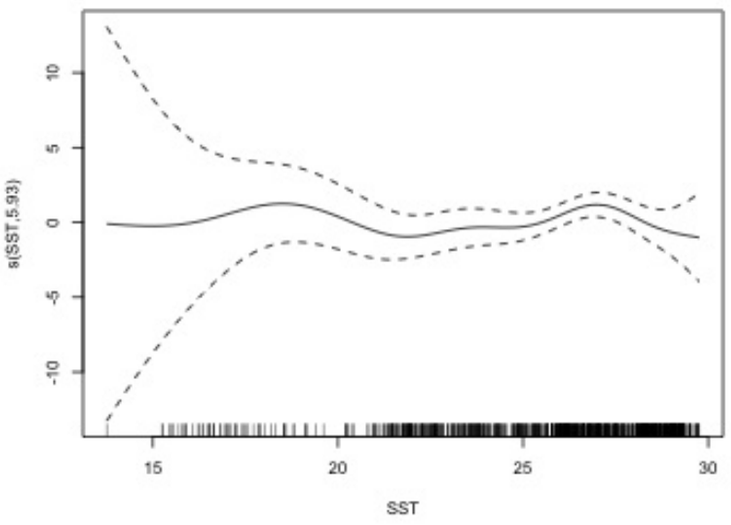
Shrinkage example

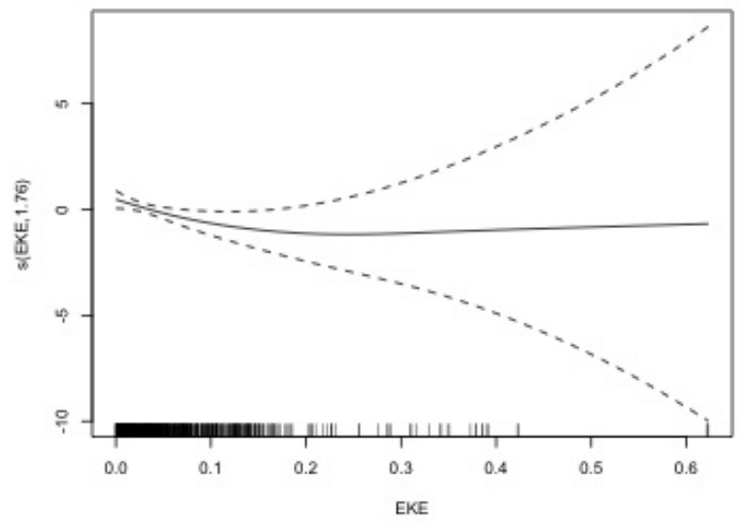
Model with no shrinkage

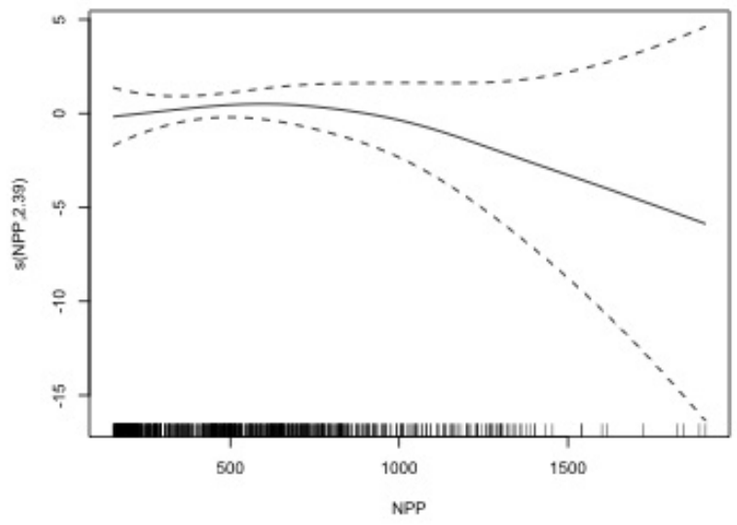




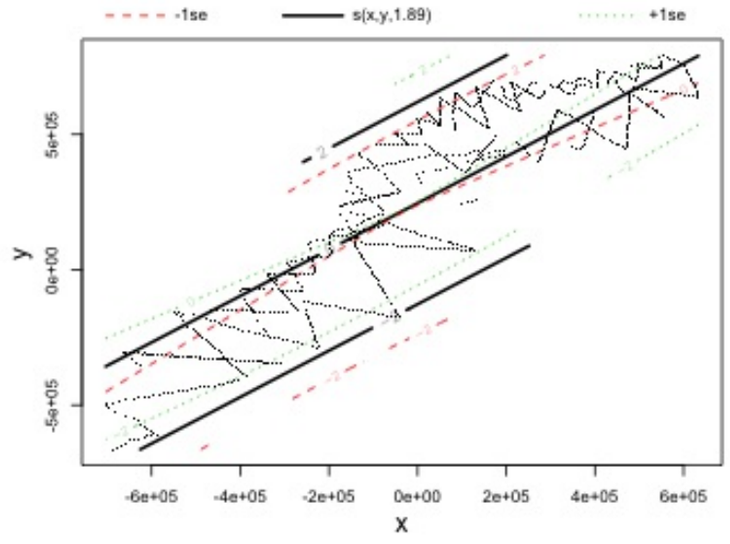


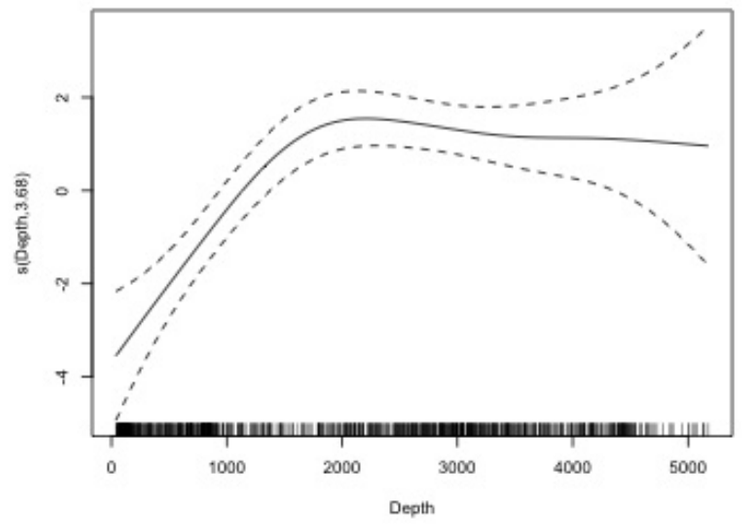


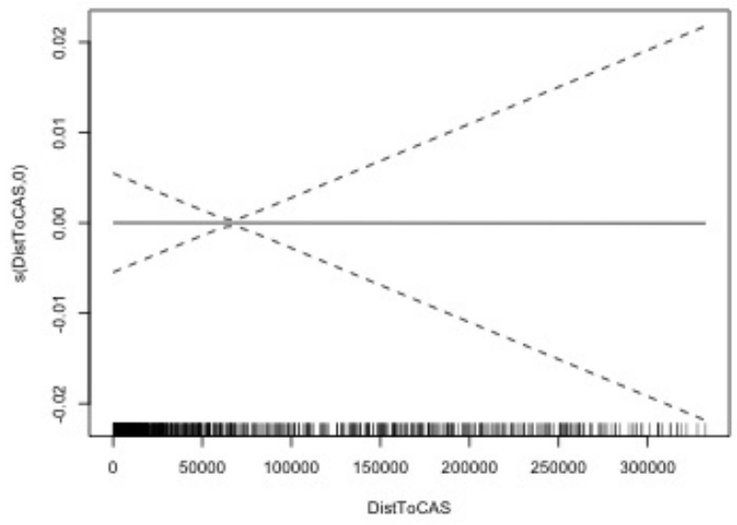


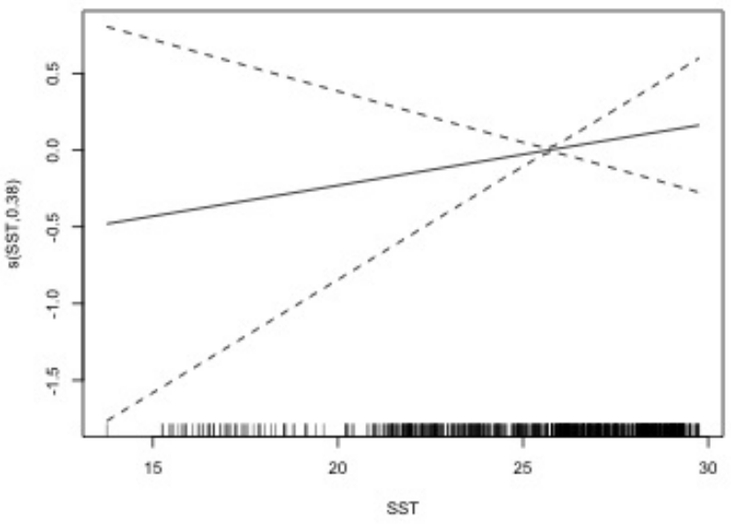


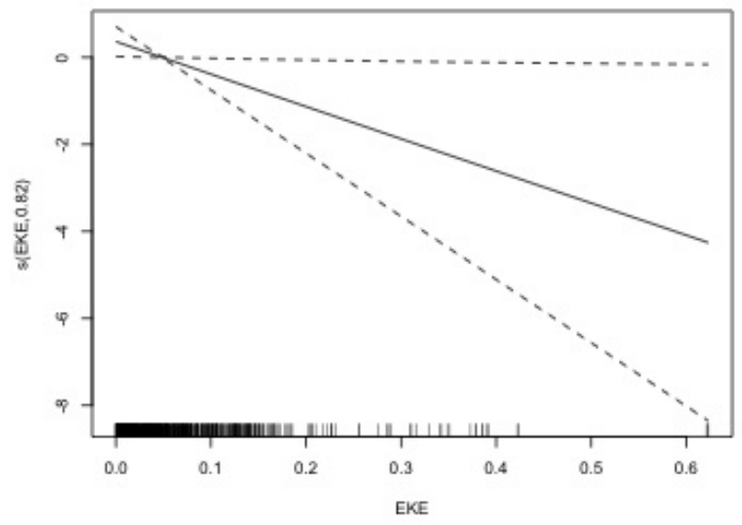
… with shrinkage

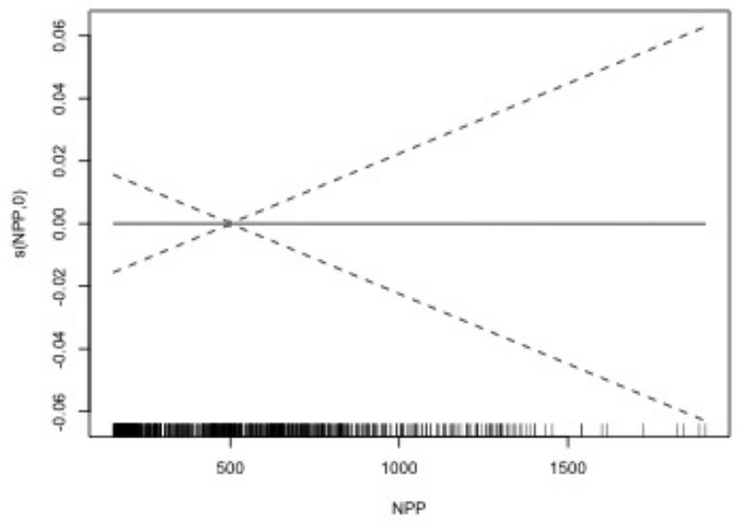








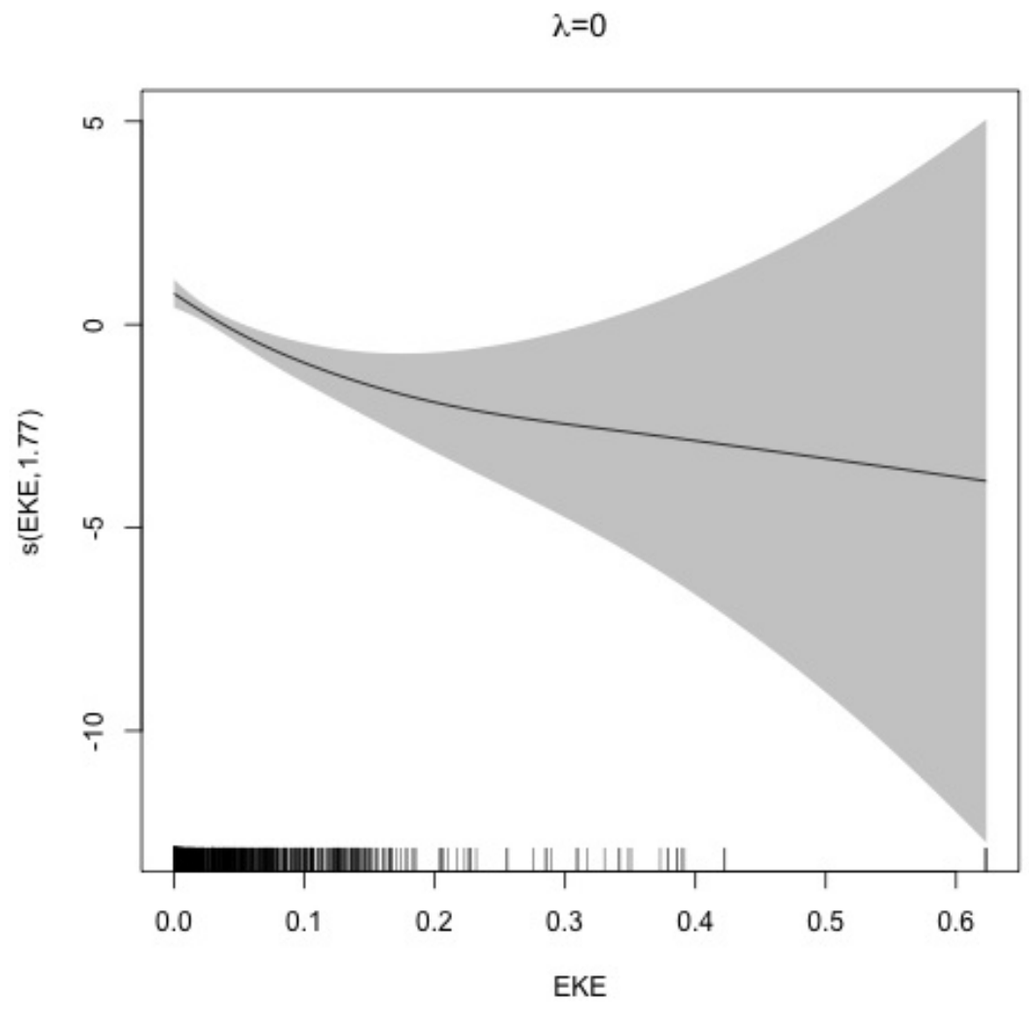




Shrinkage in action

Alternative: extra penalty

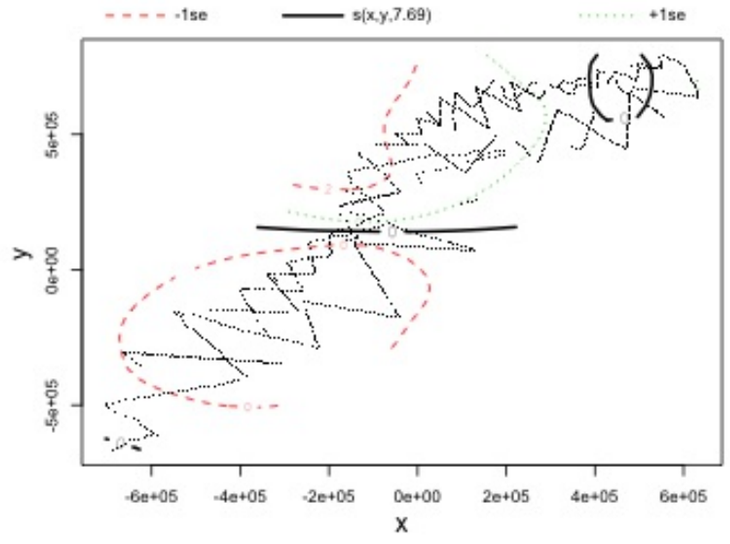
- -
  extra penalty
- no assumption of how much to shrink the nullspace
- "no assumption" might not be such a good idea?

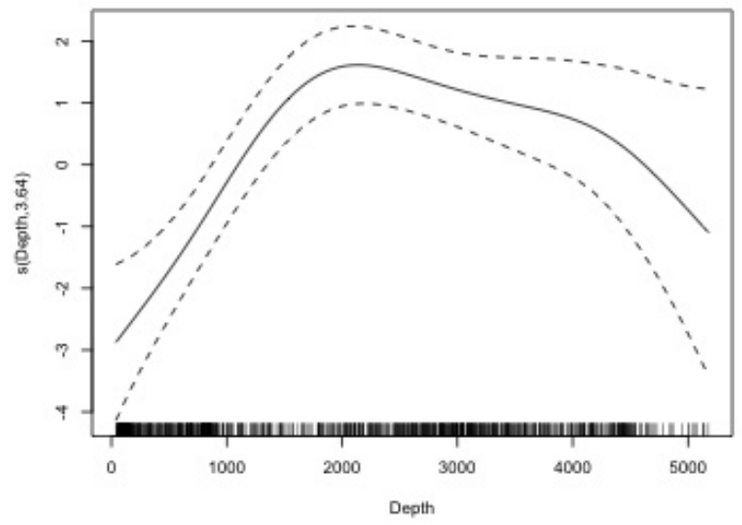


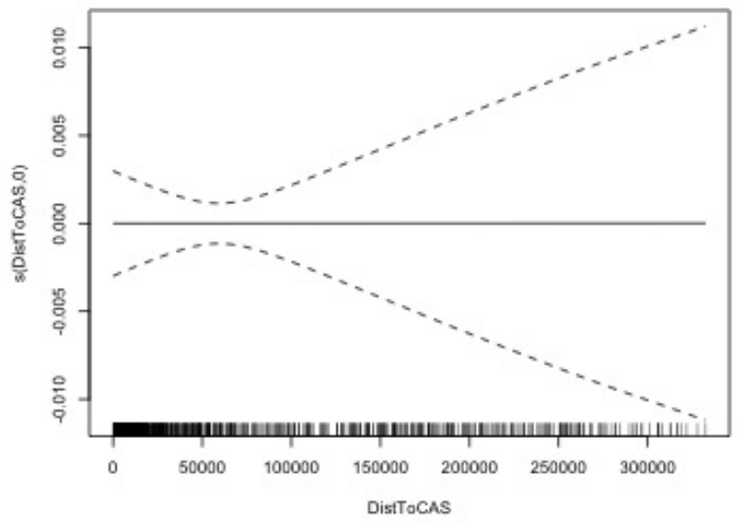
Extra penalty example

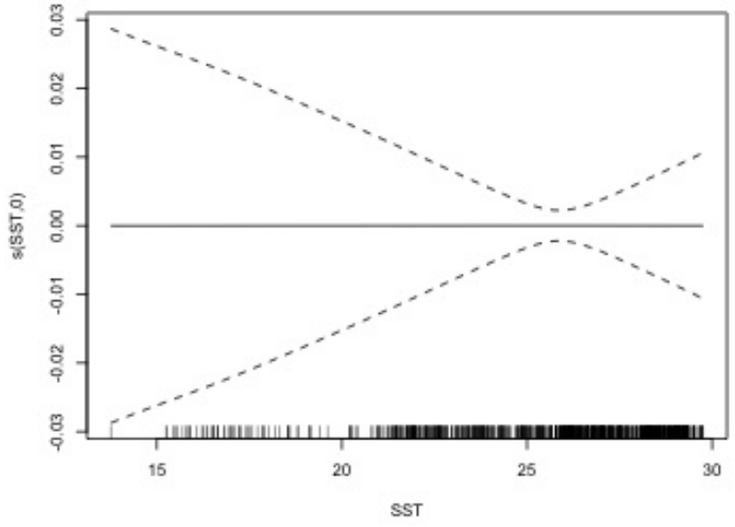
Extra penalty example

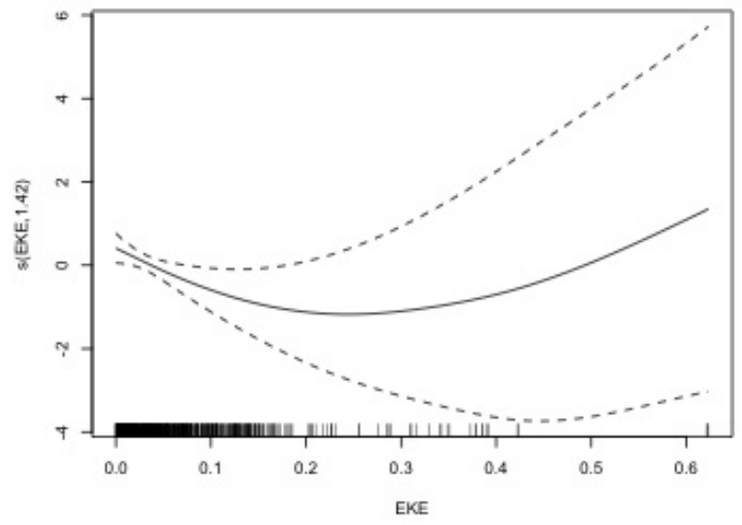
Extra penalty example

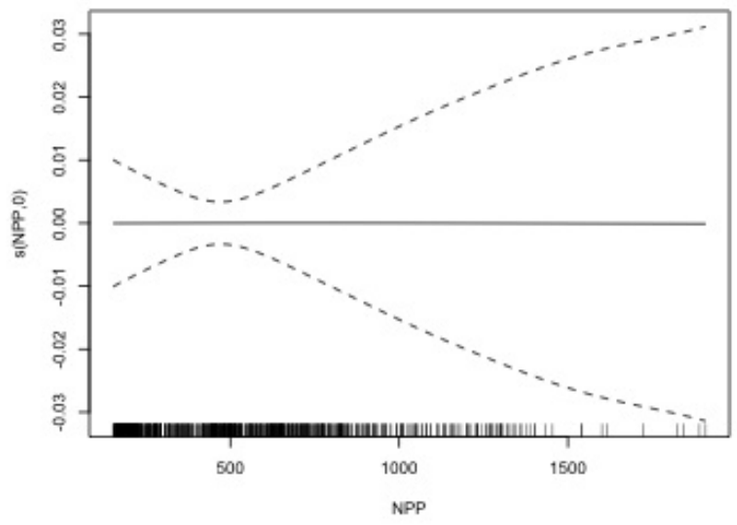












EDF comparison

	allterms	select	ts
s(x,y)	5.2361	7.6936	1.8875
s(Depth)	3.5677	3.6449	3.6794
s(DistToCAS)	1.0001	0.0000	0.0001
s(SST)	5.9270	0.0002	0.3827
s(EKE)	1.7628	1.4174	0.8196
s(NPP)	2.3929	0.0002	0.0004

Double penalty can be slow

- Lots of smoothing parameters to estimate

How do we select smooth terms?

1. Look at EDF
    - Terms with EDF<1 may not be useful (can we remove?)
2. Remove non-significant terms by p-value
    - Decide on a significance level and use that as a rule

(In some sense leaving "shrunk" terms in is more "consistent" in terms of variance estimation, but can be computationally annoying)

Comparing models

Comparing models

- Usually have >1 option
- How can we pick?
- Even if we have 1 model, is it any good?

Nested vs. non-nested models

- Compare with
  - nested models
- What about vs.
  - don't want to have all these in the model
  - not nested models

Measures of "fit"

- Two listed in
  - Deviance explained
  - Adjusted R^2 (not useful)
- Deviance is a generalisation of R^2
- Highest likelihood value (model) minus estimated model value
- (These are usually not very high for DSMs)

AIC

- Can get AIC from our model
- Comparison of AIC fine (but not the end of the story)

```
```

```
```

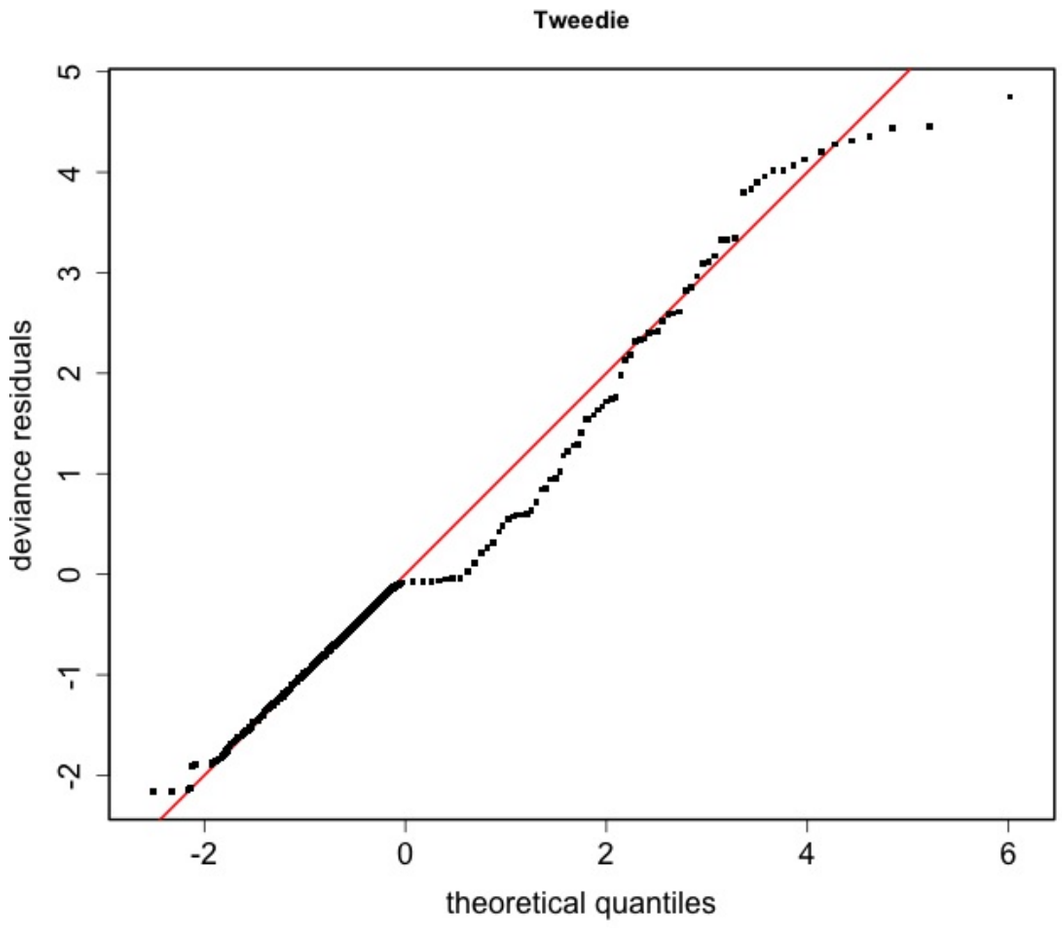
A quick note about REML scores

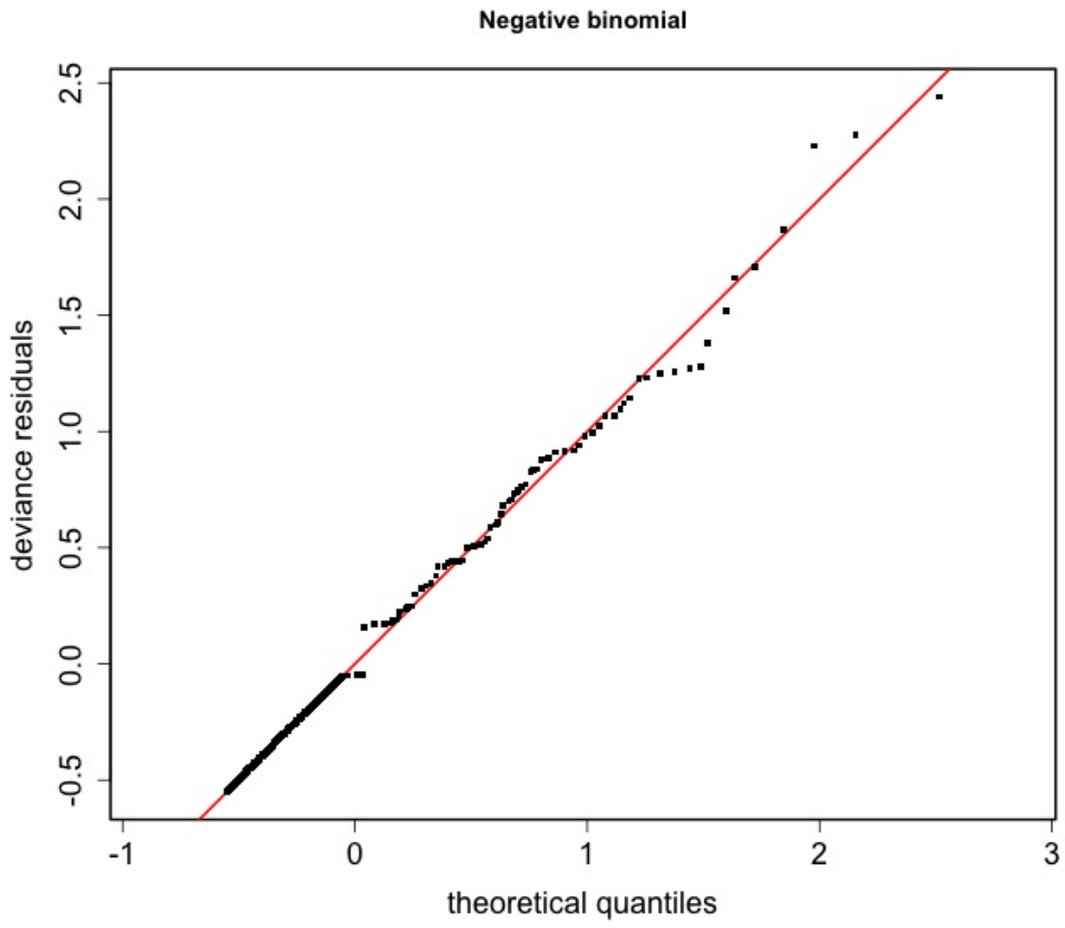
- Use REML to select the smoothness
- Can also use the score to do model selection
- **BUT** only compare models with the same fixed effects
  - (i.e., same "linear terms" in the model)
- $\Rightarrow$ **All terms** must be penalised
  - or

Selecting between response distributions

Goodness of fit tests

- Q-Q plots
- Closer to the line == better



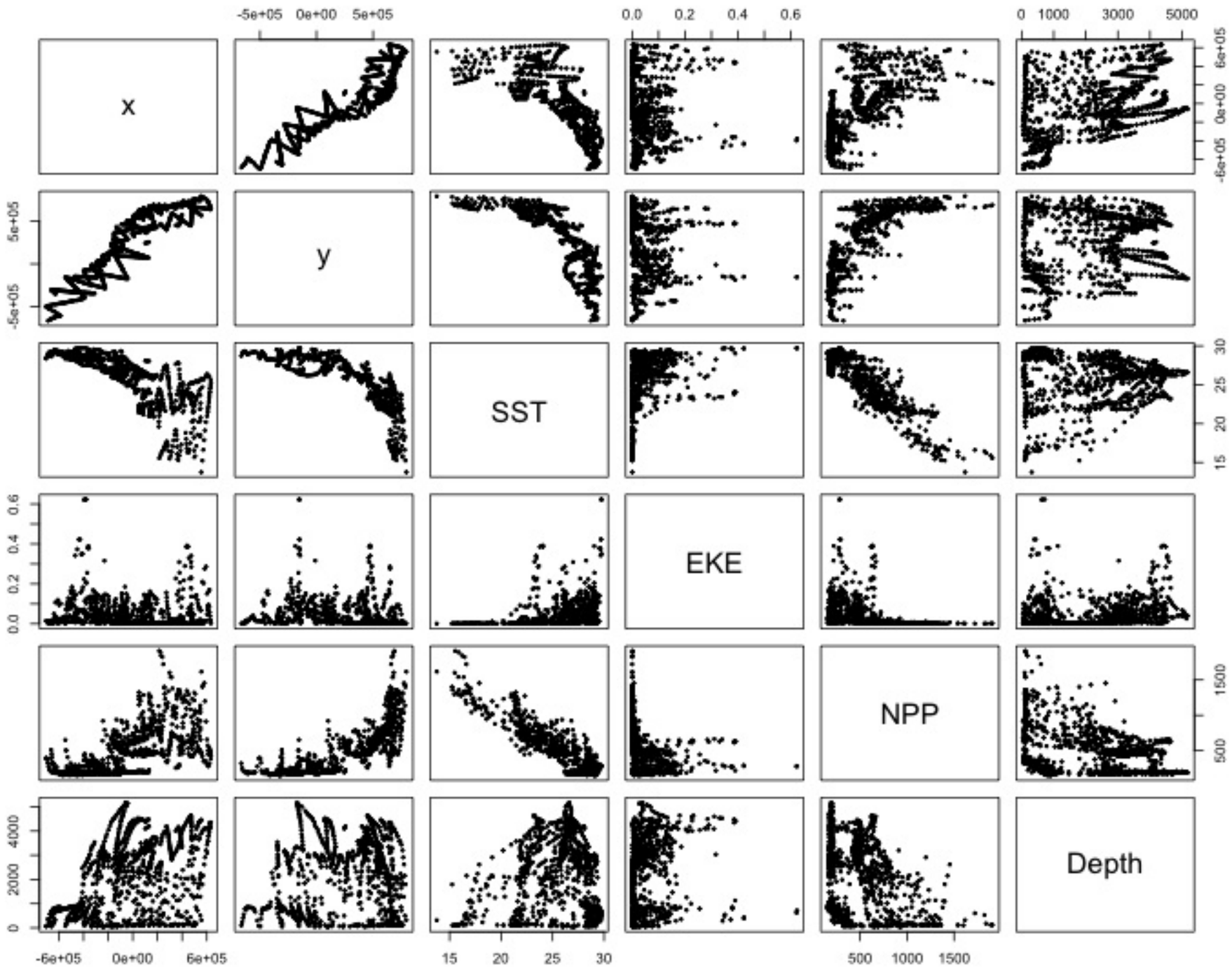


Tobler's first law of geography

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

Tobler (1970)

Implications of Tobler's law



Covariates are not only correlated (linearly)...

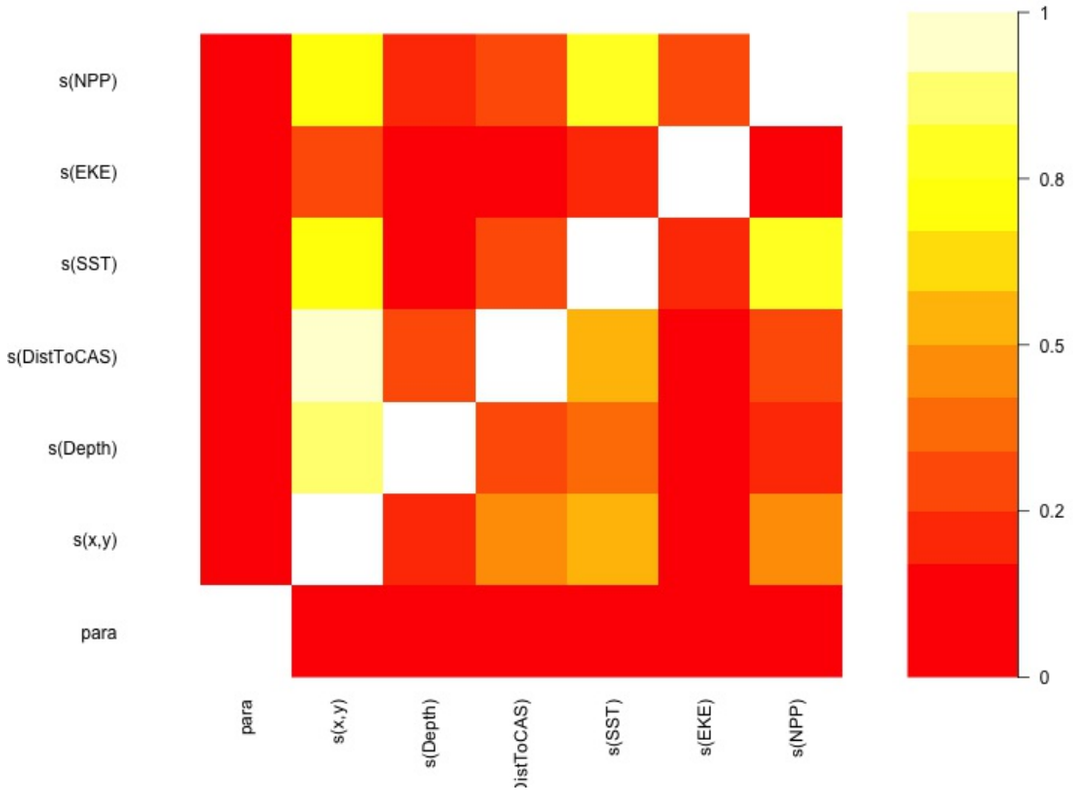
...they are also "concurve"

"How much can one smooth be approximated by one or more other smooths?"

Concurvity (model/smooth)

Concurvity between smooths

Visualising concurvity between terms



- Previous matrix output visualised
- High values (yellow) = BAD

Path dependence

Sensitivity

- General path dependency?
- What if there are highly concurve smooths?
- Is the model is sensitive to them?

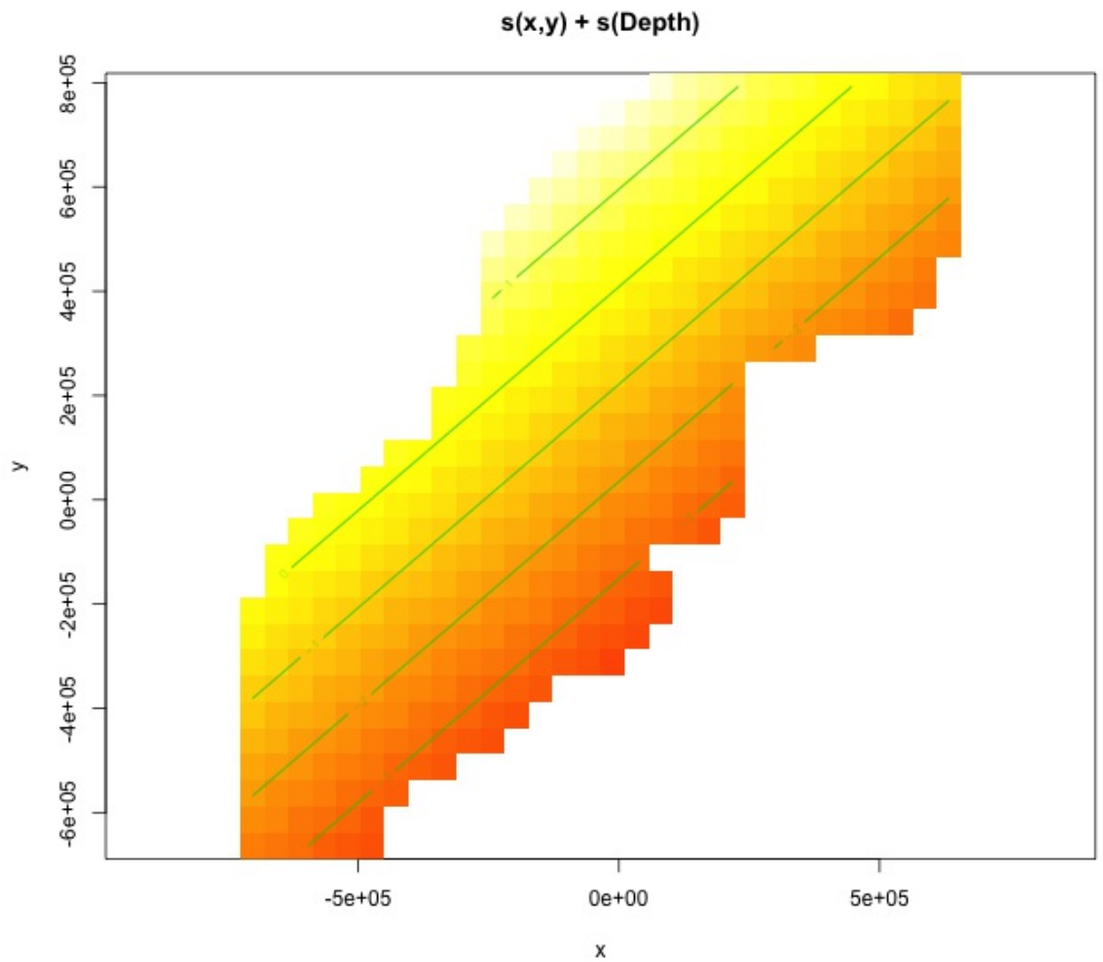
What can we do?

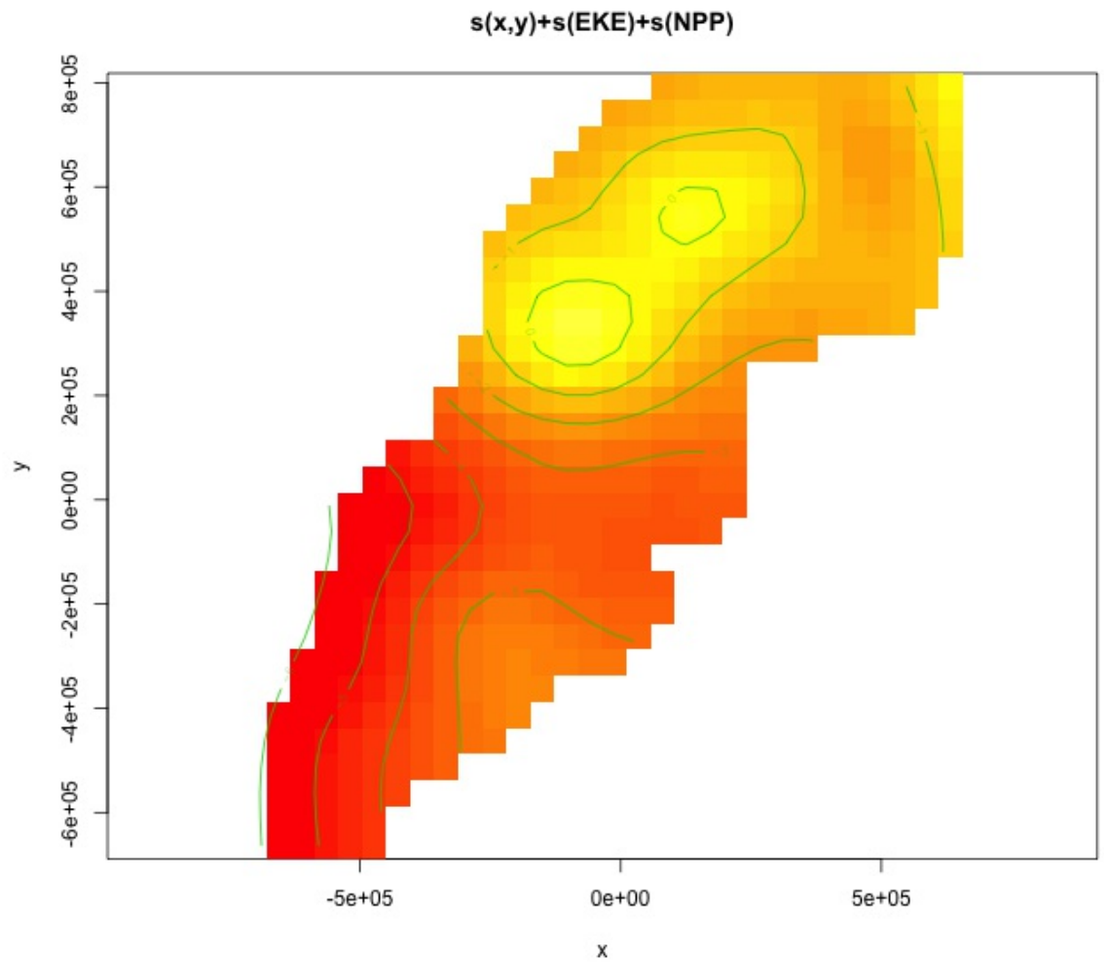
- Fit variations excluding smooths
  - Concurve terms that are excluded early on
- Appendix of Winiarski et al (2014) has an example

Sensitivity example

- and are highly concurve (0.9067)
- Refit removing first

Comparison of spatial effects

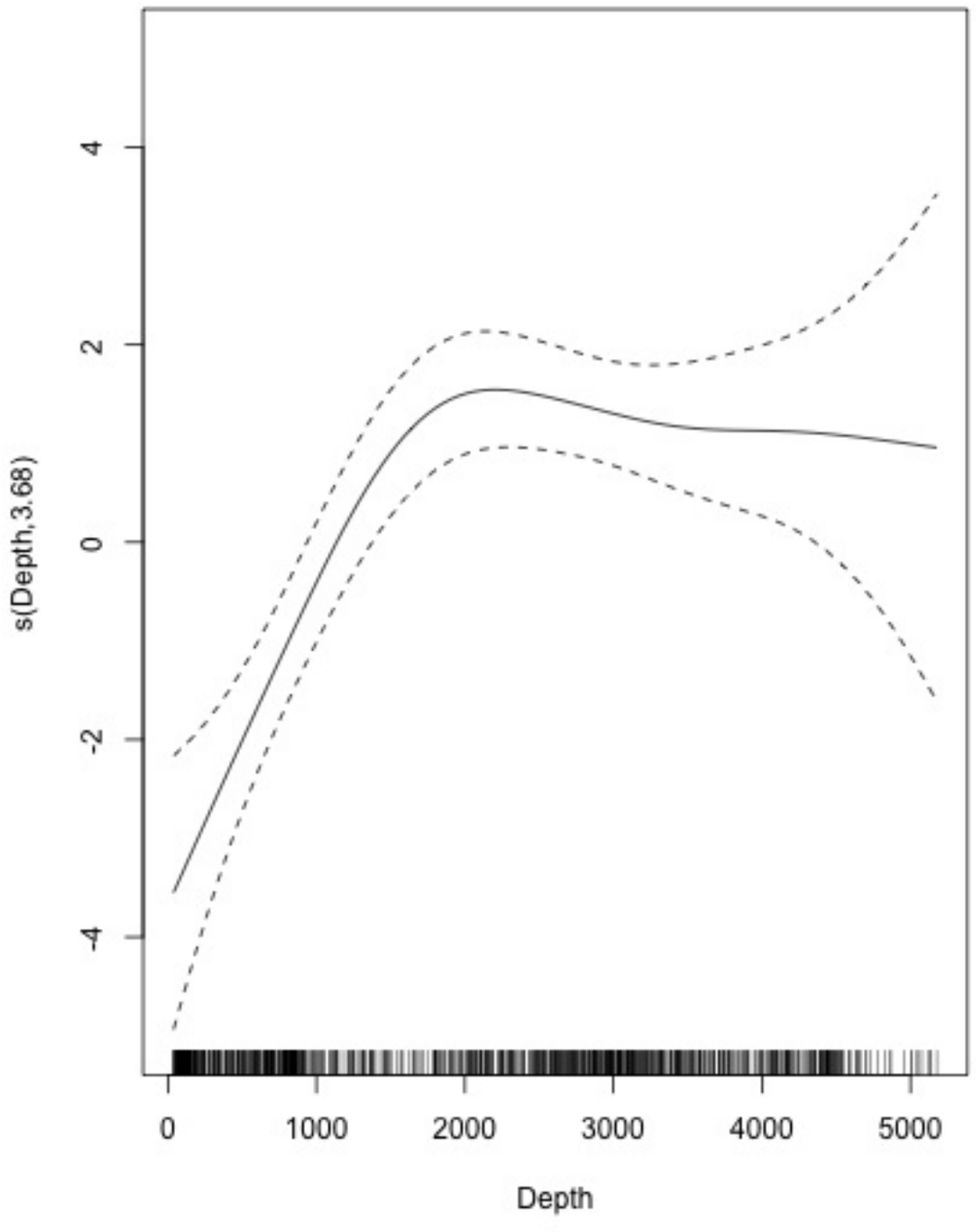


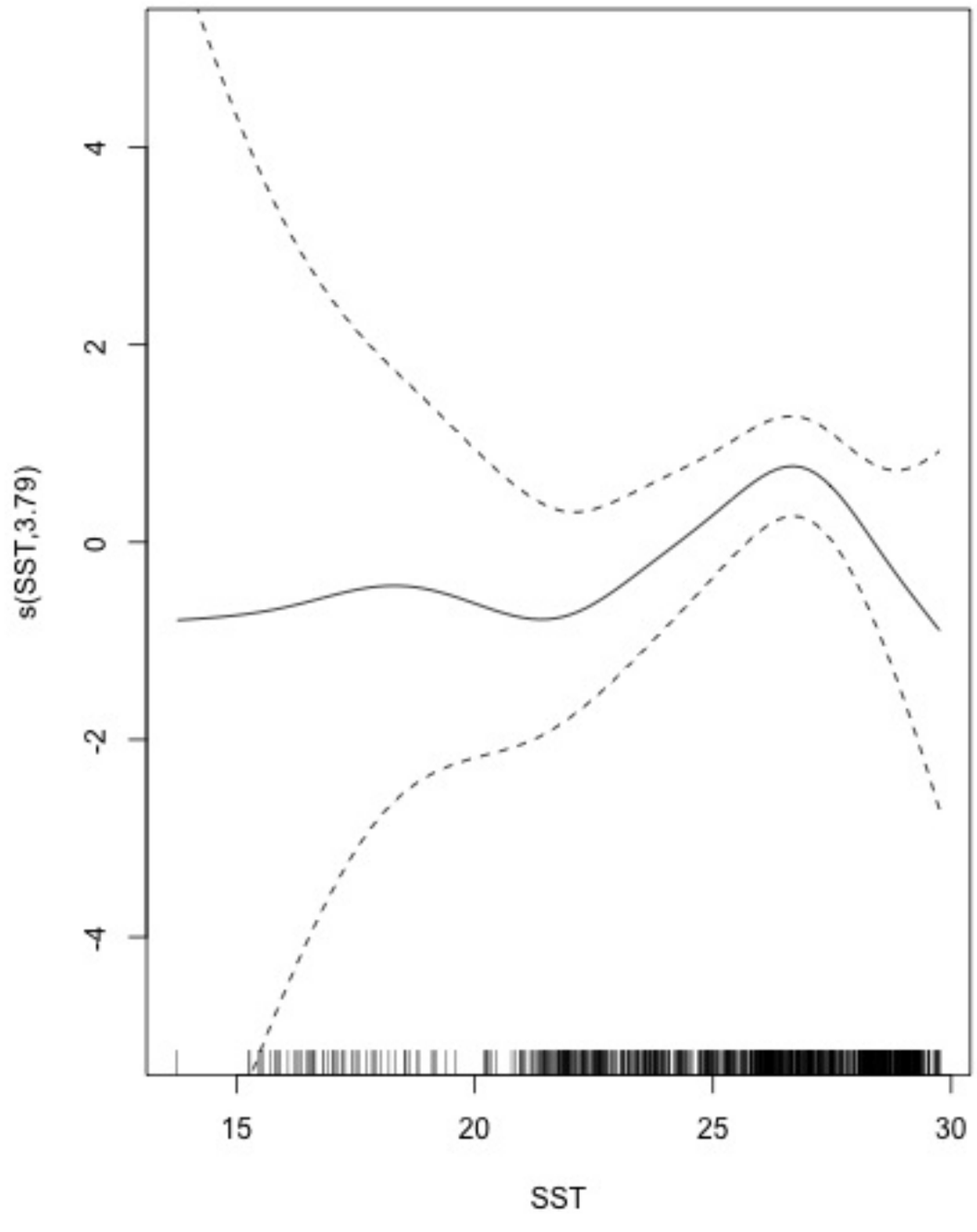


Sensitivity example

- Refit removing and …

Comparison of depth smooths





Comparing those three models...

Model	AIC	Deviance
all	1225.822	35.10
s(x, y) + s(EKE) + s(NPP)	1248.172	34.54
s(SST) + s(Depth)	1228.106	39.28

- "Full" model still explains most deviance
- No depth model requires spatial smooth to "mop up" extra variation
- We'll come back to this when we do prediction

Recap

Recap

- Adding smooths
- Removing smooths
  - p-values
  - shrinkage/extra penalties
- Comparing models
- Comparing response distributions
- Sensitivity