Photo: Norbert Wu Productions

What spatial statistical model is best for predicting fisheries bycatch risk?

BRIAN STOCK

Thank you!

SIO

- Brice Semmens

SWFSC

- Tomo Eguchi

NWFSC

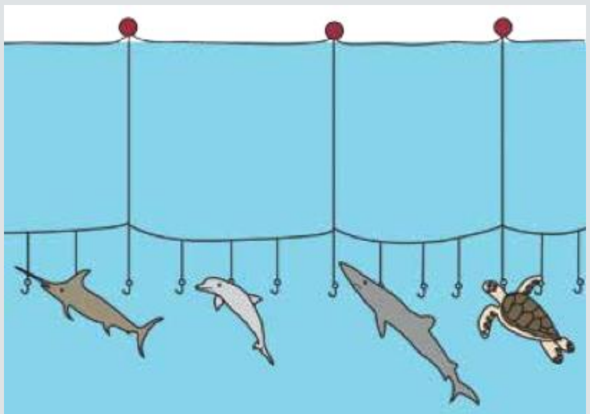
- Eric Ward
- Essential Fish Habitat (Blake Feist)
- West Coast Groundfish Observer Program (Jason Jannot)

PIFSC

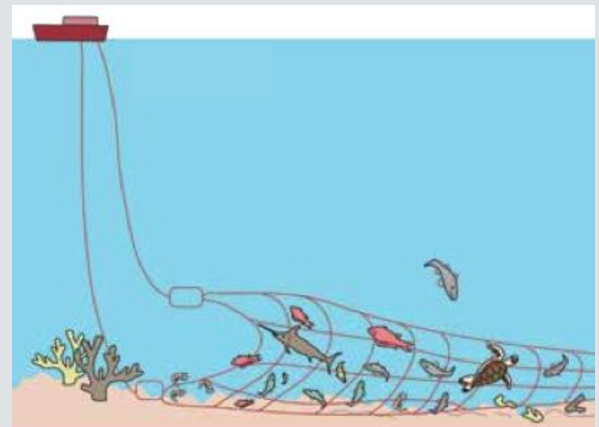
- Hawaii Longline Observer Program (Eric Forney)

“Target” vs. “bycatch”

Longline

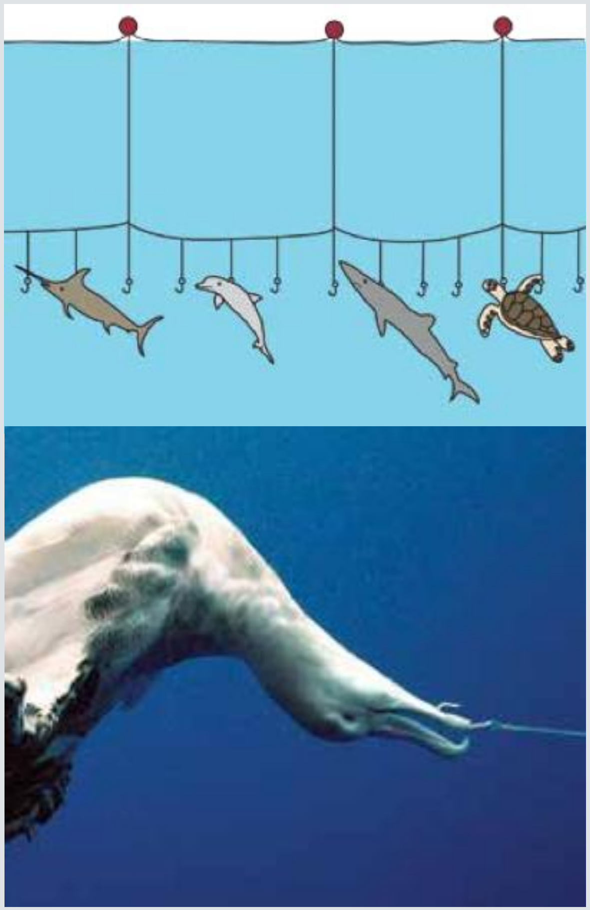


Trawl

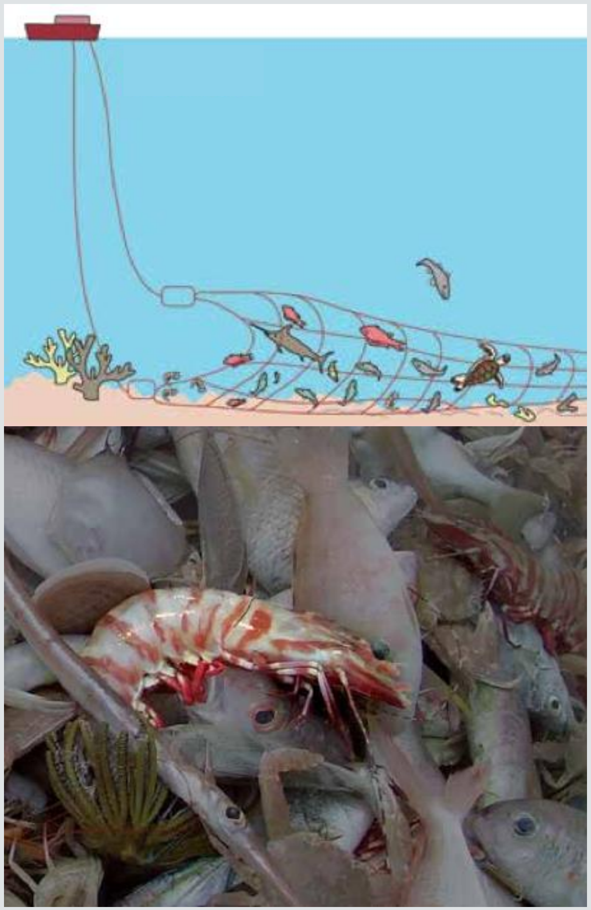


“Target” vs. “bycatch”

Longline



Trawl



Oceana (2014)

Bycatch is a big (spatial) issue

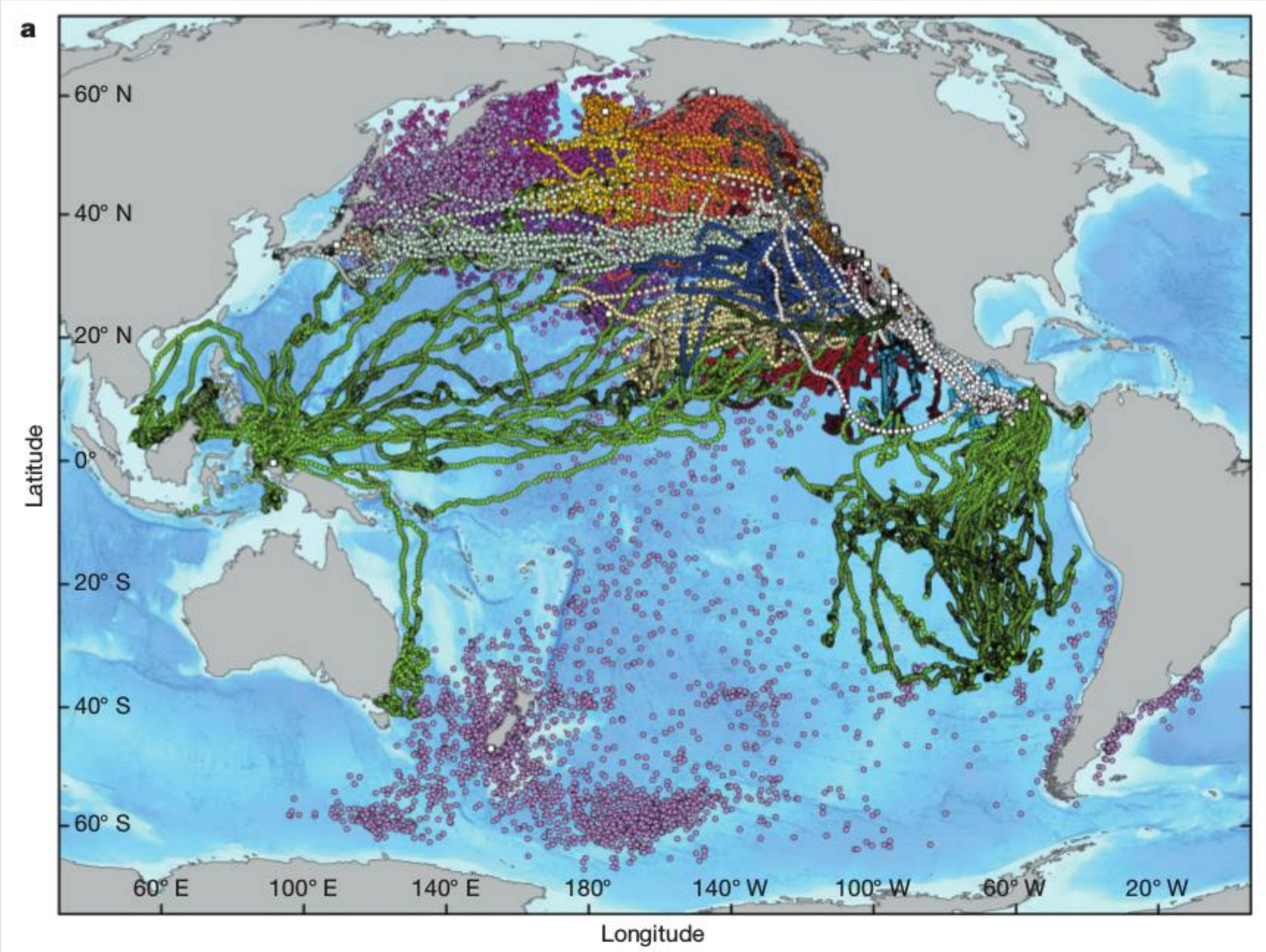
Protected species

Recovering species

Competing fisheries

Unmarketable species

Difficult when they move so much…



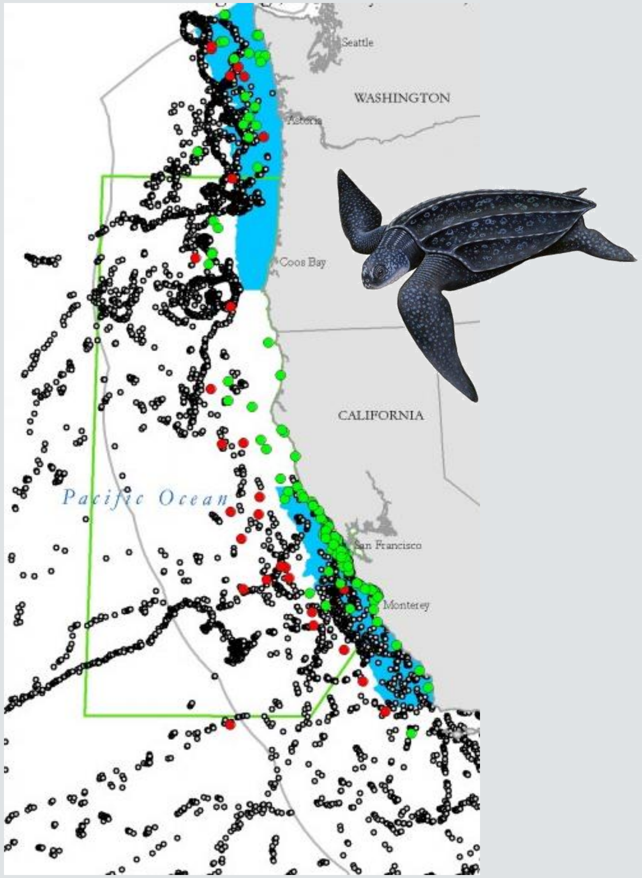
Introduction

Block et al. (2011)

Static vs. dynamic management

Dynamic

Static



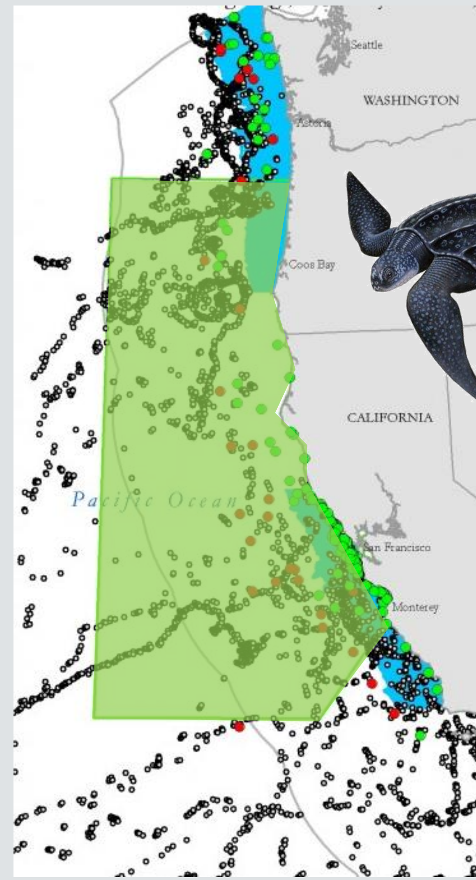
Oceana (2012)

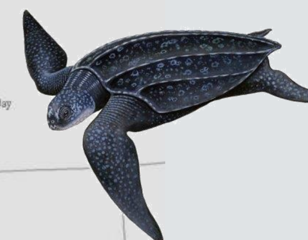
Static vs. dynamic management

Dynamic

1. Effectively protected?
2. Huge loss of fishing area

Static

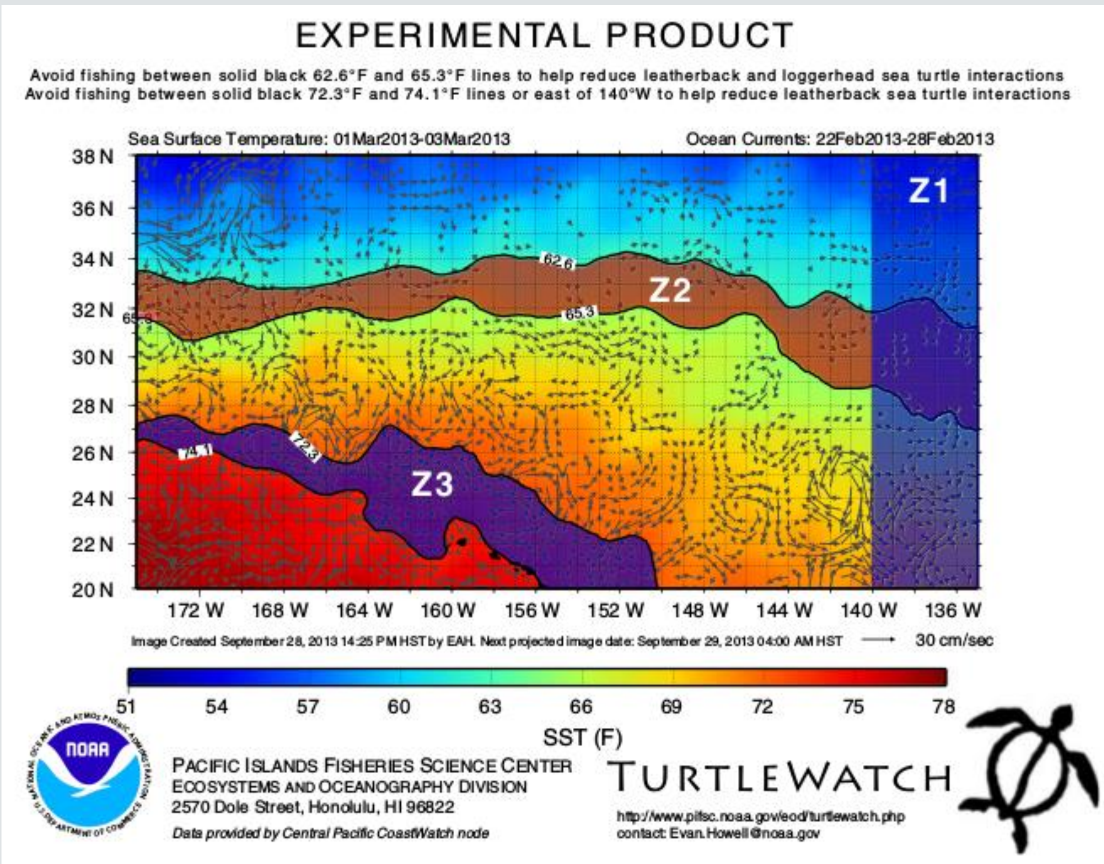




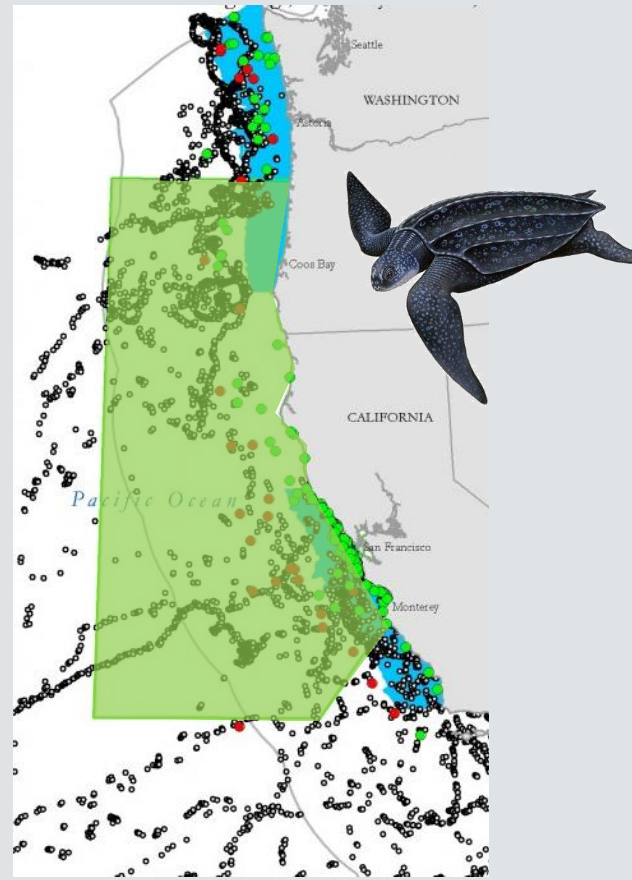
Oceana (2012)

Static vs. dynamic management

Dynamic



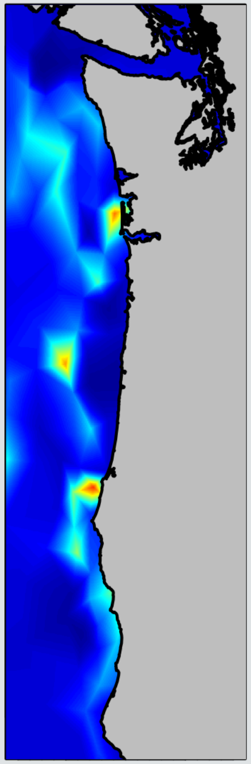
Static



Howell et al. (2008, 2015)

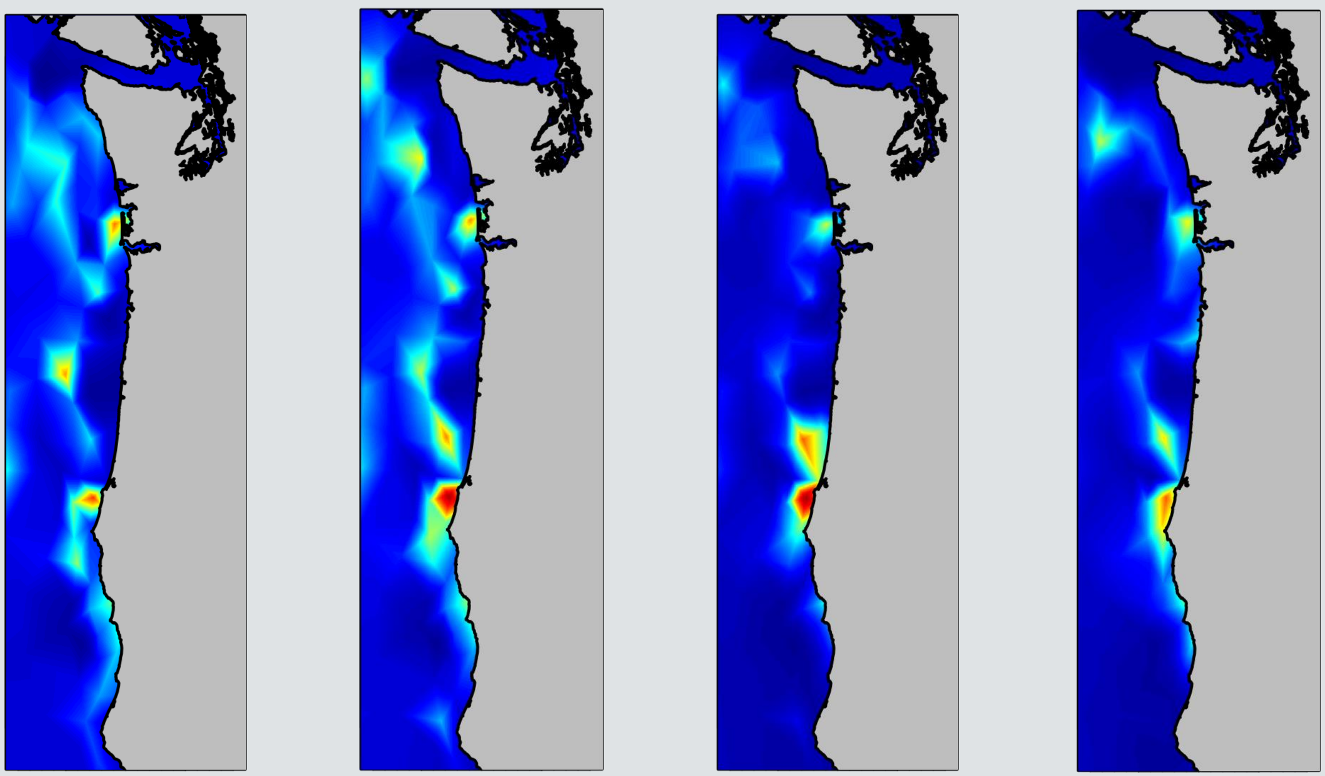
Tools for dynamic management

Need map of bycatch risk



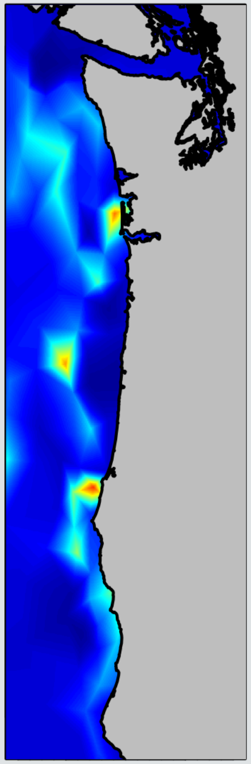
Tools for dynamic management

Need map of bycatch risk



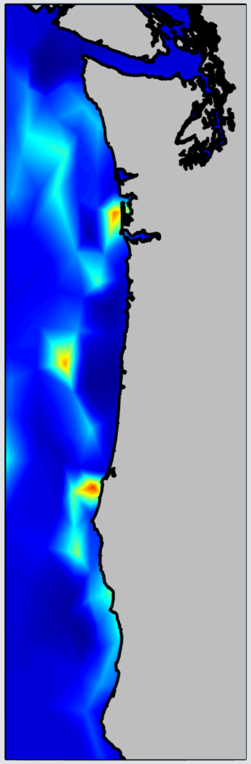
Tools for dynamic management

Need map of bycatch risk

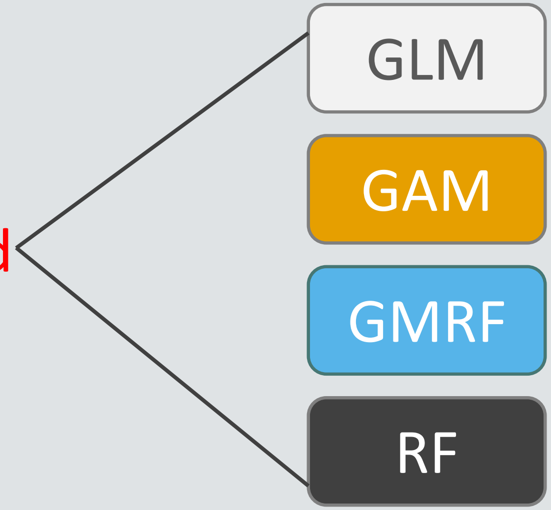


- temperature
- depth
- substrate
- spatial field

Q1: Which spatial model is best?



- temperature
- depth
- substrate
- spatial field



Research question

The data (fisheries observers)

West Coast Groundfish Trawl

- 2002-2013
- 55,835 tows

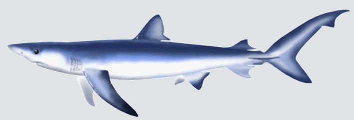
Hawaii Longline

- 1994-2014
- 16,714 sets (swordfish only)

Research question

Q2: Does the answer depend on species?

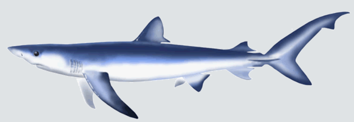
			
Habitat:	Bottom	Rocky bottom	Rocky bottom
Movement:	Med	Low	Low
Bycatch Rate:	29%	18%	0.3%

			
Habitat:	Open ocean	Open ocean	Open ocean
Movement:	High	High	High
Bycatch Rate:	96%	1.4%	0.7%

Q2: Does the answer depend on species?



Habitat:	Bottom	Rocky bottom	Rocky bottom
Movement:	Med	Low	Low
Bycatch Rate:	29%	18%	0.3%



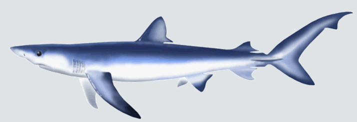
Habitat:	Open ocean	Open ocean	Open ocean
Movement:	High	High	High
Bycatch Rate:	96%	1.4%	0.7%

Research question

Q2: Does the answer depend on species?



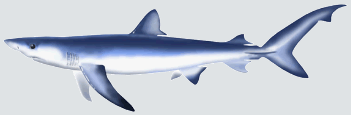
Habitat:	Bottom	Rocky bottom	Rocky bottom
Movement:	Med	Low	Low
Bycatch Rate:	29%	18%	0.3%



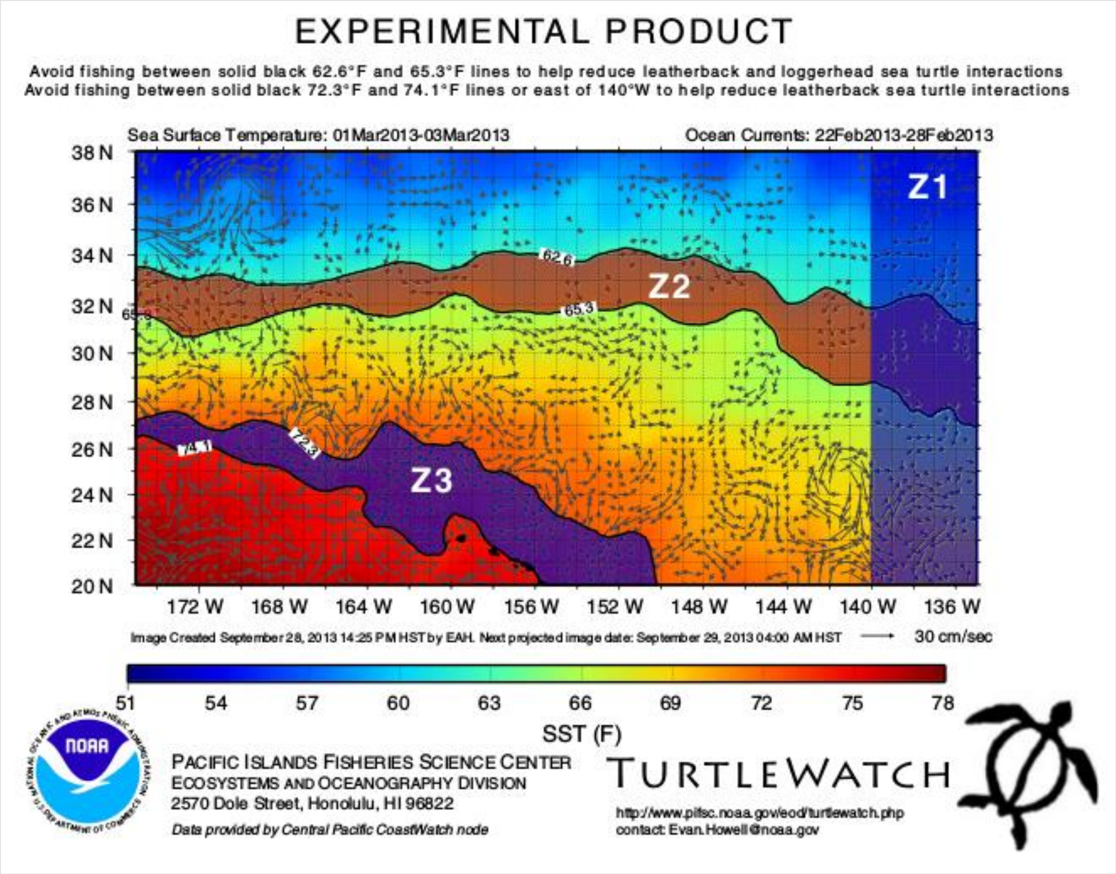
Habitat:	Open ocean	Open ocean	Open ocean
Movement:	High	High	High
Bycatch Rate:	96%	1.4%	0.7%

Q2: Does the answer depend on species?

			
Habitat:	Bottom	Rocky bottom	Rocky bottom
Movement:	Med	Low	Low
Bycatch Rate:	29%	18%	0.3%

			
Habitat:	Open ocean	Open ocean	Open ocean
Movement:	High	High	High
Bycatch Rate:	96%	1.4%	0.7%

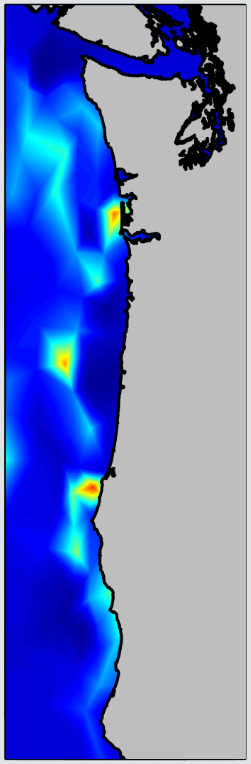
Q3: How much bycatch can they prevent?



Research question

“Species distribution models”

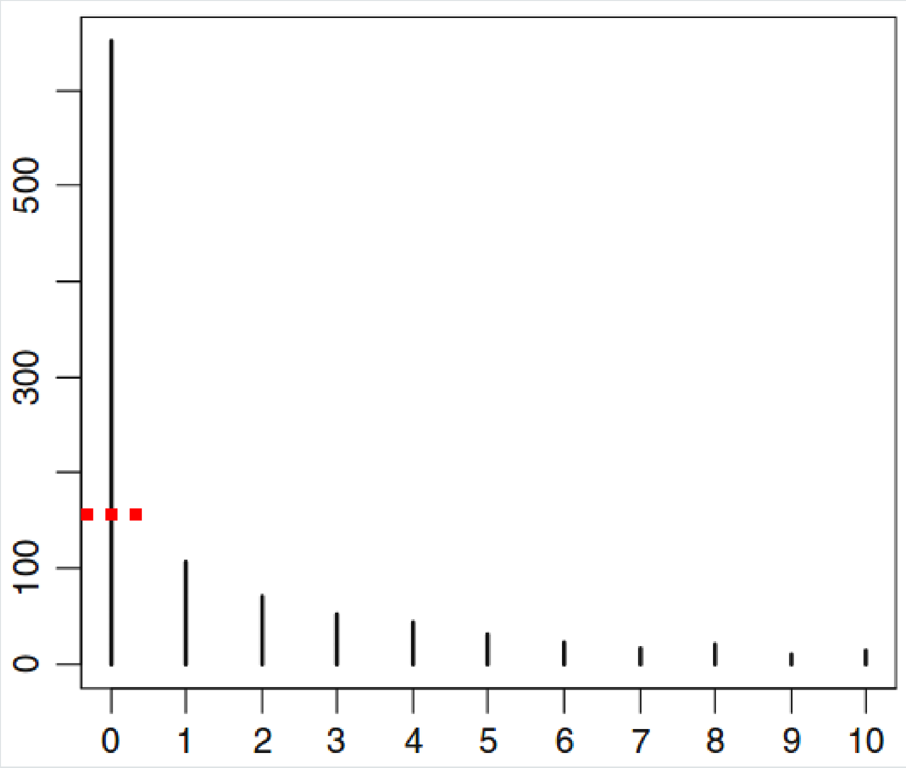
Fundamental ecological question: *where are they*?



- temperature
- depth
- substrate
- spatial field

“Zero-inflated” data

More zeros than expected

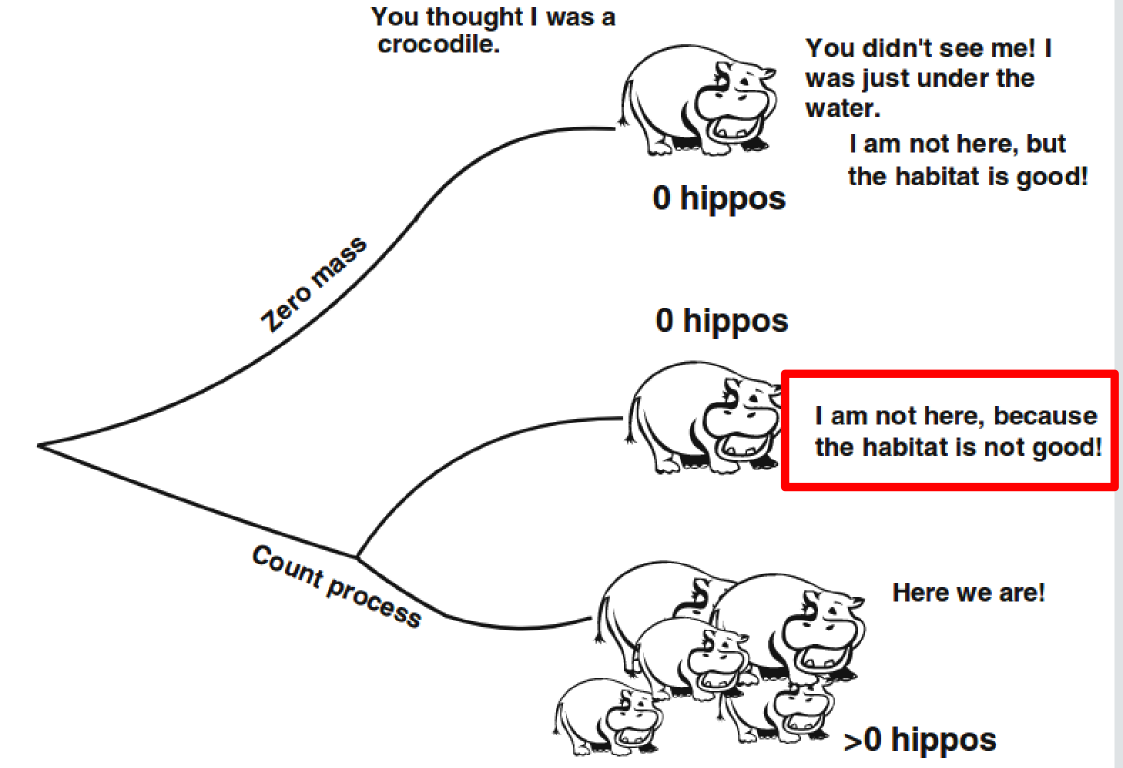


Zuur et al. (2009)

“Zero-inflated” data

Approach 1: Zero-inflated distributions

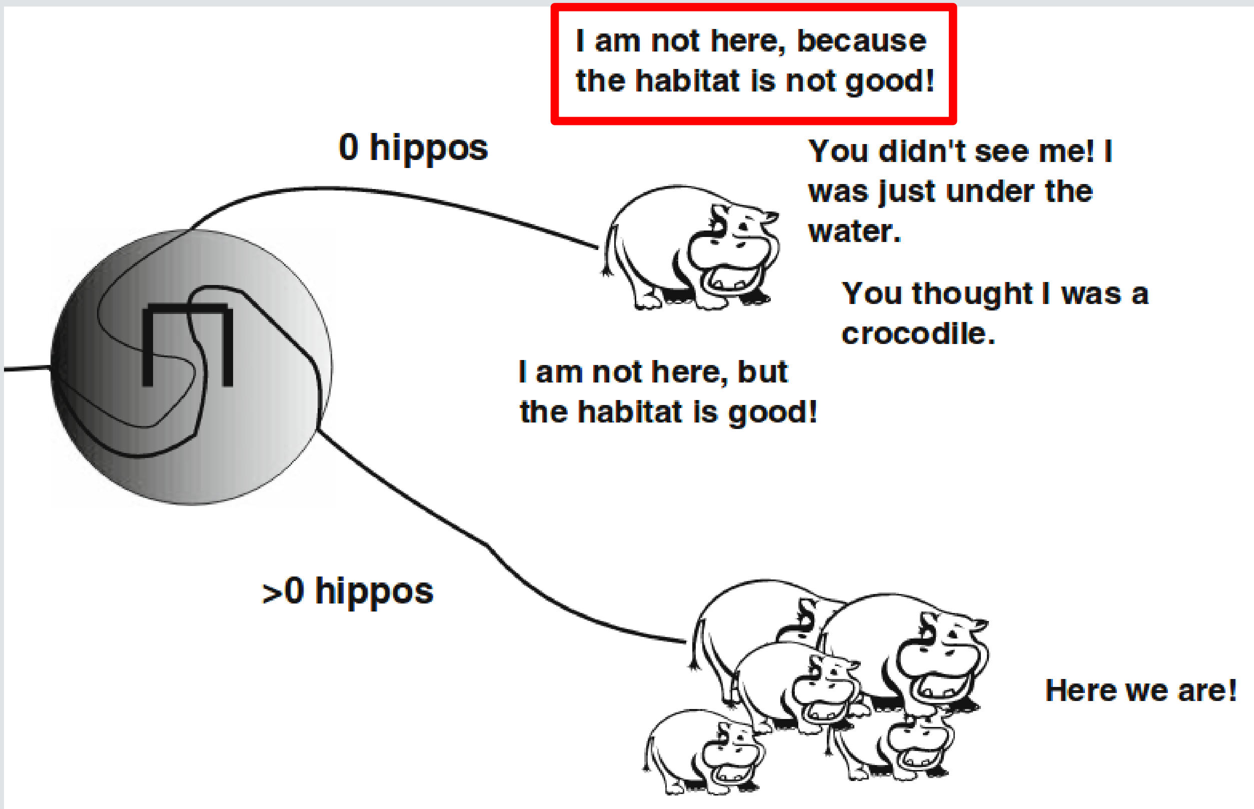
- ZI-Poisson
- ZI-Neg Binomial



Zuur et al. (2009)

“Zero-inflated” data

Approach 2: Delta (hurdle) model



Zuur et al. (2009)

"Zero-inflated" data

Approach 2: Delta (hurdle) model

Binomial

Pr(some bycatch)

$$logit(p_i) = log\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{X_i \beta}$$

$$Y_i \sim Bernoulli(p_i)$$

"Zero-inflated" data

Approach 2: Delta (hurdle) model

Binomial — Pr(some bycatch)

Positive — E(bycatch | some bycatch)

$log(\mu_i) = \boldsymbol{X_i \beta}$

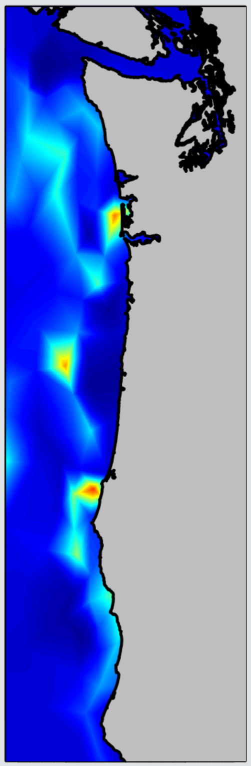
$Y_i \sim Gamma(\mu_i, \nu)$ for $Y_i > 0$

"Zero-inflated" data

Approach 2: Delta (hurdle) model

Binomial	Pr(some bycatch)
Positive	E(bycatch | some bycatch)
Binomial x Positive	E(bycatch)

Q1: Which spatial model is best?



- temperature
- depth
- substrate
- spatial field

GLM

GAM

GMRF

RF

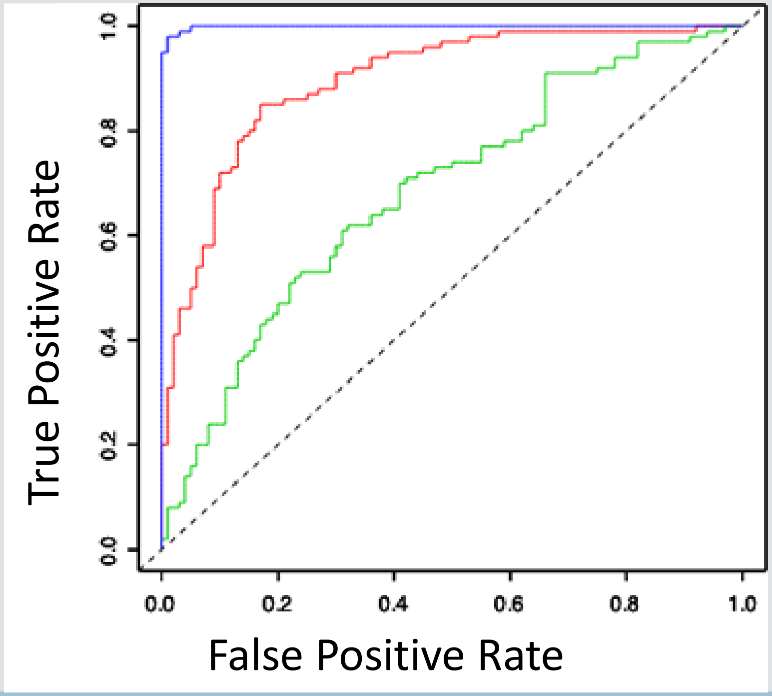
Q1: Which spatial model is best?

Goal: prediction

5-fold cross validation repeated 10x

Binomial

ROC curve (AUC)



ROC curve	AUC
Worthless	0.5
Ok	0.7
Good	0.8
Awesome	0.9+

Methods: evaluation

Q1: Which spatial model is best?

Goal: prediction

5-fold cross validation repeated 10x

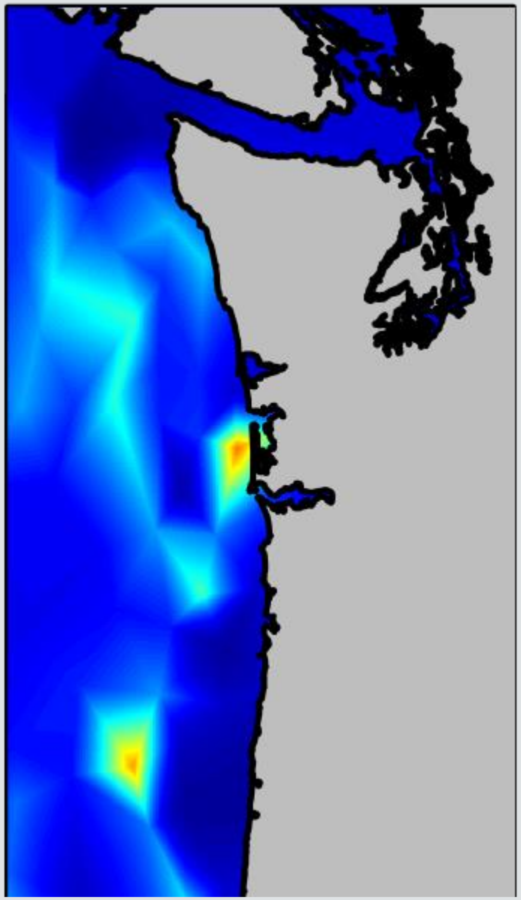
Binomial — AUC

Positive — RMSE, R^2 (pred – obs)

$$\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Methods: evaluation

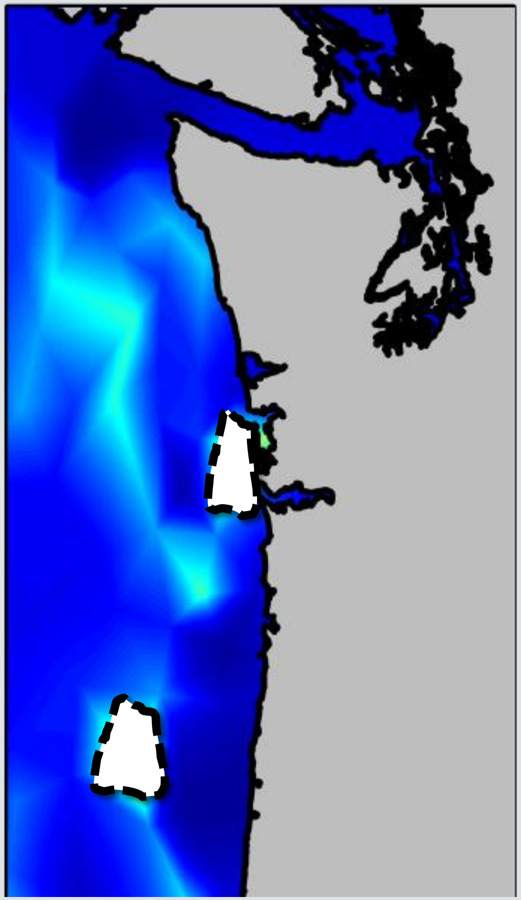
Q3: How much bycatch can they prevent?



Simulate management:

1. Predict bycatch risk at test locations

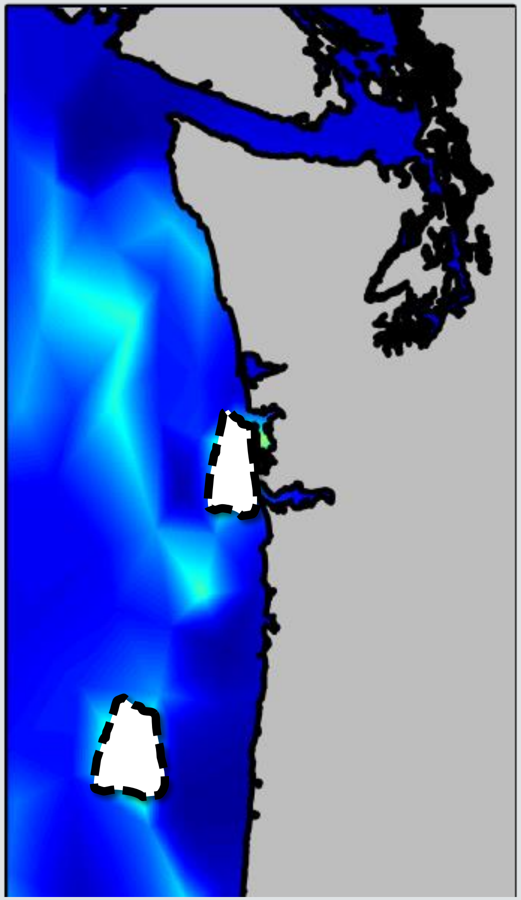
Q3: How much bycatch can they prevent?



Simulate management:

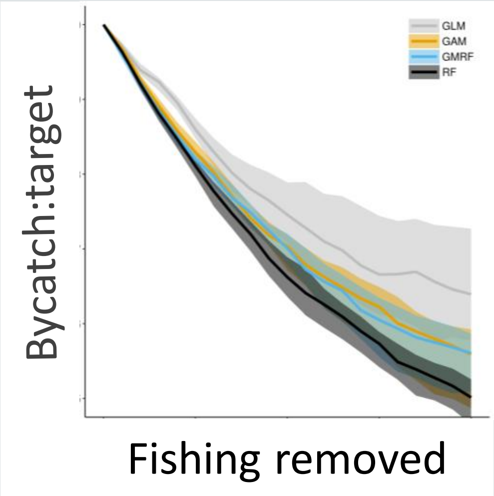
1. Predict bycatch risk at test locations
2. Remove X% of fishing effort with highest bycatch risk

Q3: How much bycatch can they prevent?



Simulate management:

1. Predict bycatch risk at test locations
2. Remove X% of fishing effort with highest bycatch risk
3. Calculate “prevented” bycatch and target catch (bycatch:target ratio)



Methods: evaluation

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

obs ~ environmental predictors (temp, depth, …)

$Y_i \sim Bernoulli(logit^{-1}[\boldsymbol{X_i \beta}])$ Binomial

$Y_i \sim Gamma(e^{\boldsymbol{X_i \beta}}, \nu)$ Positive

Dormann (2007)

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

obs ~ environmental predictors (temp, depth, …)

$$Y_i \sim Bernoulli(logit^{-1}[\boldsymbol{X}_i\boldsymbol{\beta}])$$

Binomial

$$Y_i \sim Gamma(e^{\boldsymbol{X}_i\boldsymbol{\beta}}, \nu)$$

Positive

How much variability can we explain?

- with covariates
- *without spatial locations*

Dormann (2007)

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

obs ~ environmental predictors (temp, depth, …)

$Y_i \sim Bernoulli(logit^{-1}[\boldsymbol{X_i \beta}])$ Binomial

$Y_i \sim Gamma(e^{\boldsymbol{X_i \beta}}, \nu)$ Positive

Problem:

spatial correlation in residuals (Prediction – Observed)

Dormann (2007)

Why does spatial correlation matter?

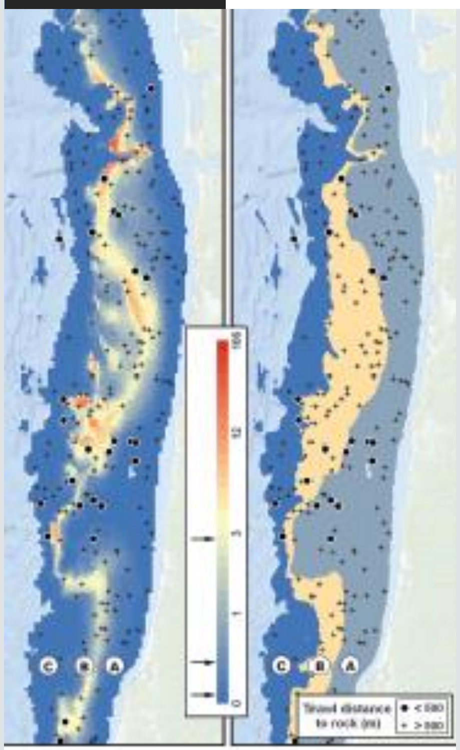
1. Valid statistical inference
   - Observations not independent
   - Lower effective sample size (i.e. CI should be wider)

Dormann (2007)

Why does spatial correlation matter?

2. Get the temporal trend right

Spatial | No space

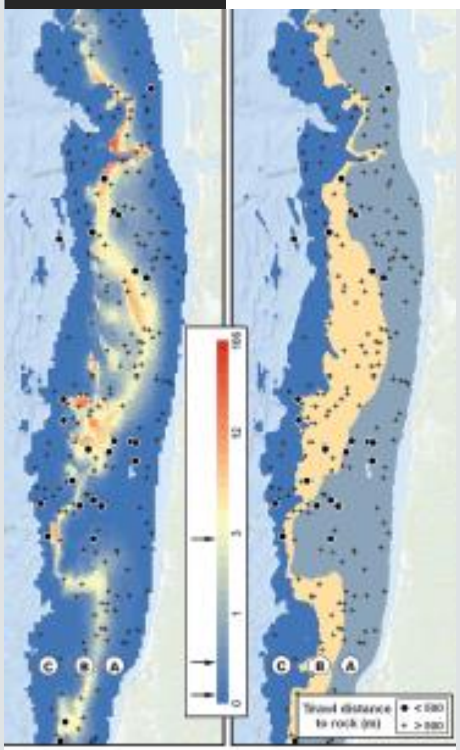


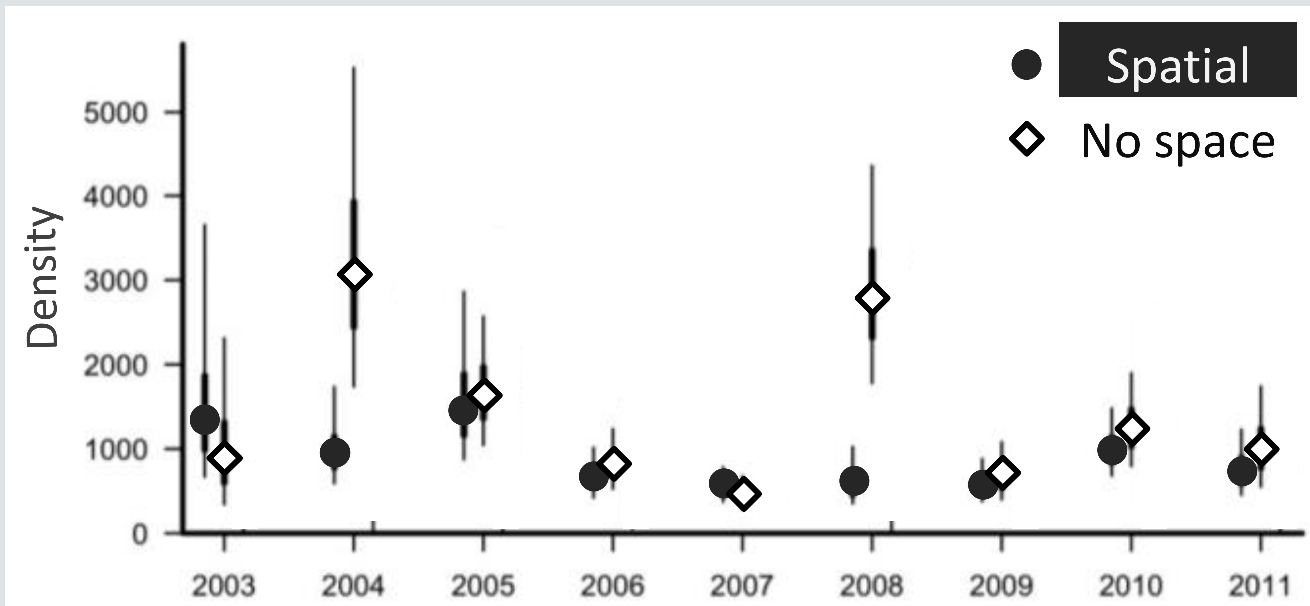
Shelton et al. (2014)

Why does spatial correlation matter?

2. Get the temporal trend right

Spatial No space



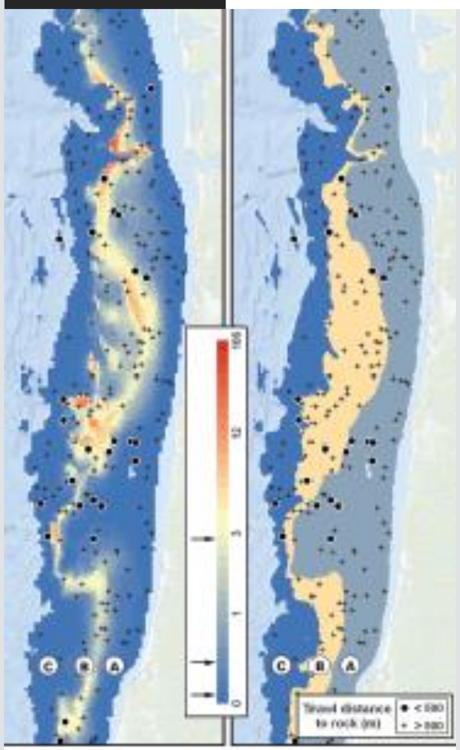


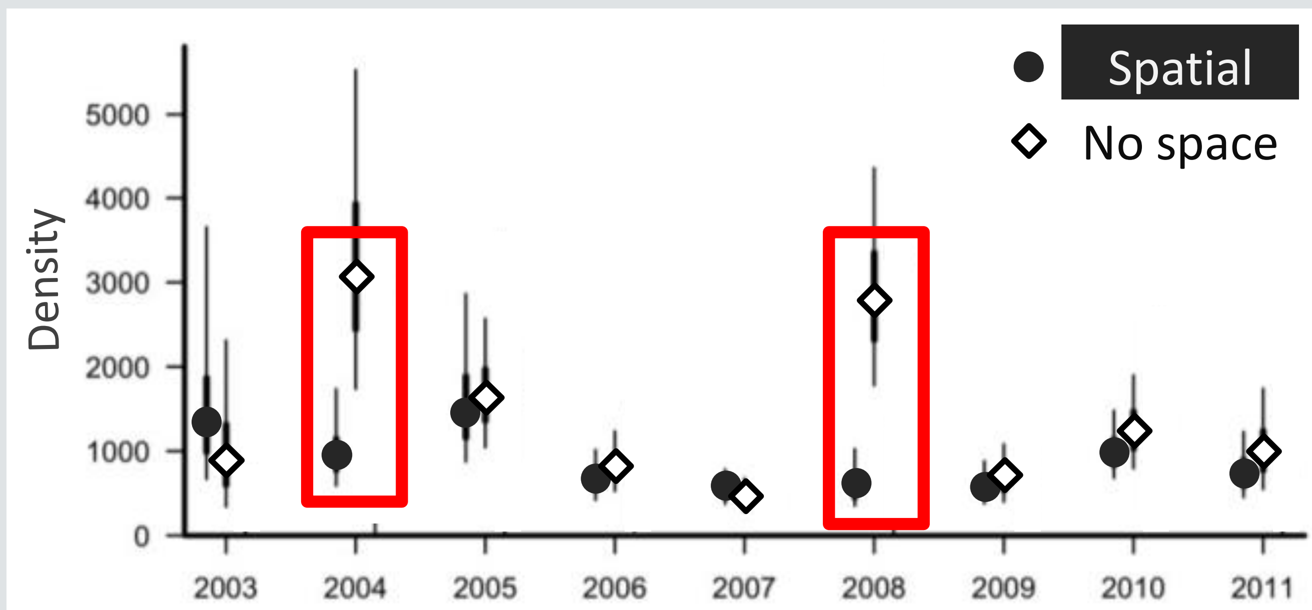
Shelton et al. (2014)

Why does spatial correlation matter?

2. Get the temporal trend right

Spatial | No space

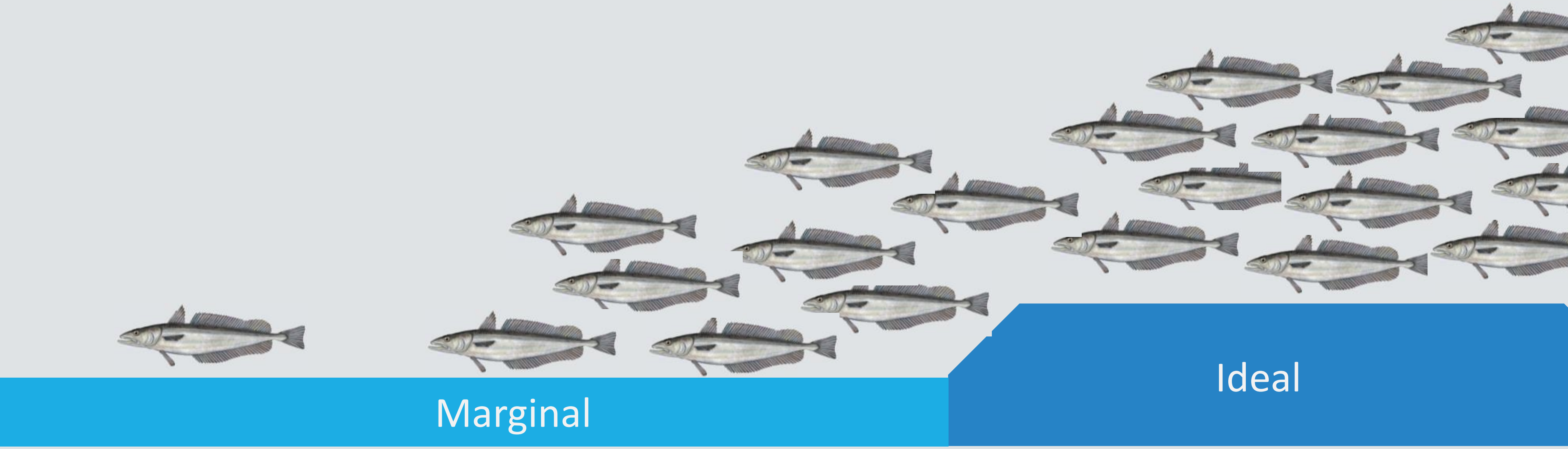




Shelton et al. (2014)

Why does spatial correlation matter?

3. Effect of habitat vs. schooling

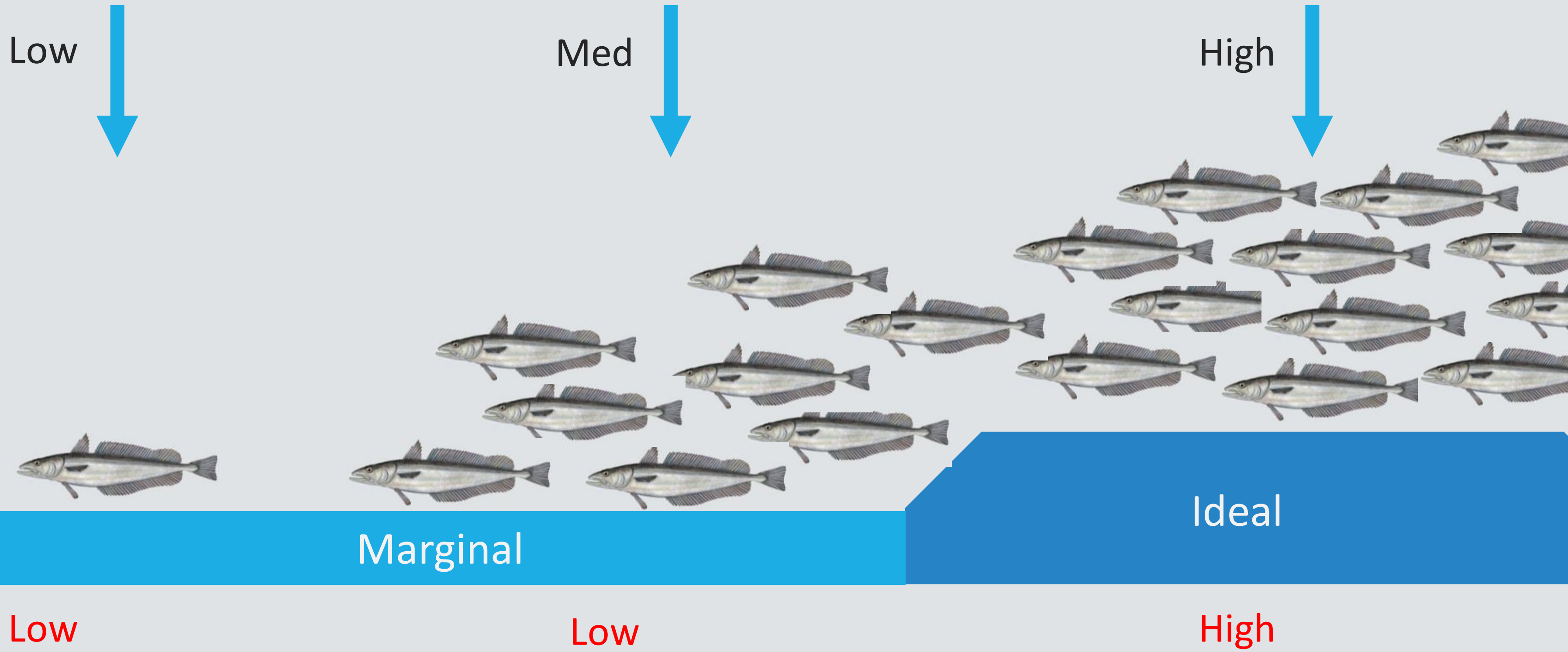


Methods

Agostini et al. (2008)

Why does spatial correlation matter?

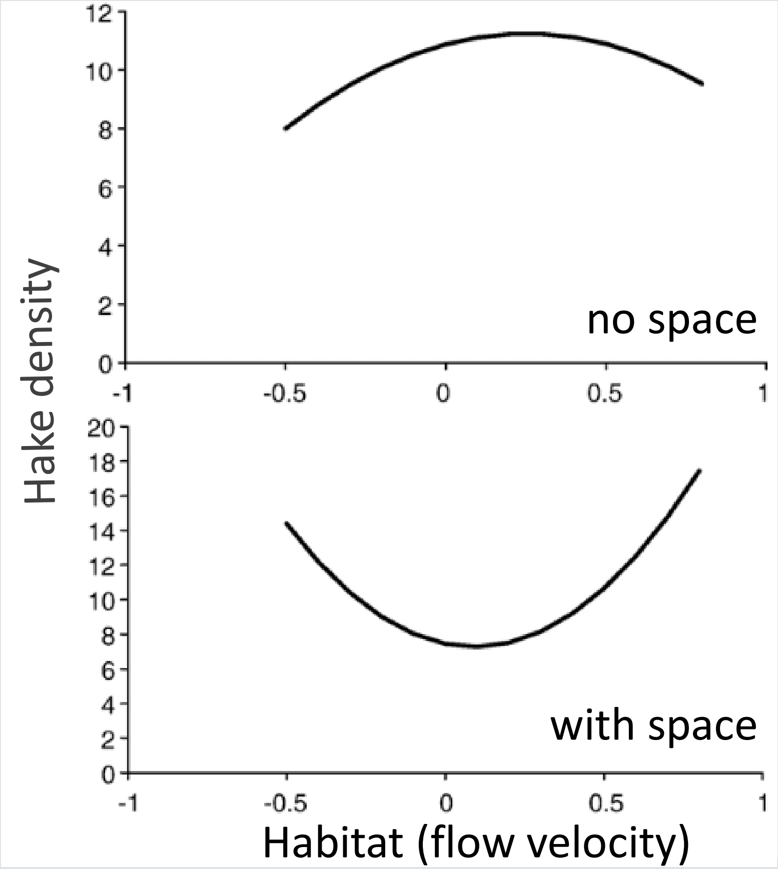
3. Effect of habitat vs. schooling



Agostini et al. (2008)

Why does spatial correlation matter?

3. Effect of habitat vs. schooling



Agostini et al. (2008)

Q1: Which spatial model is best?

GLM

GAM

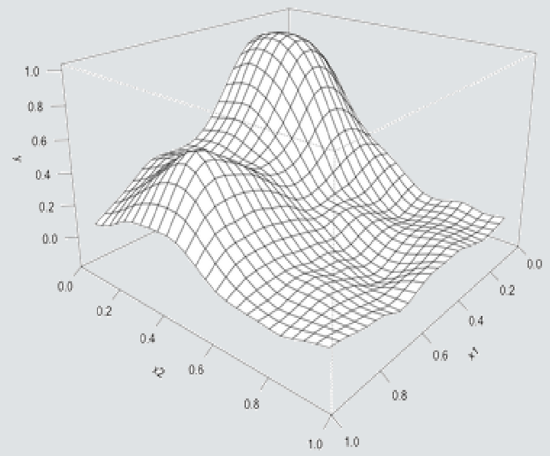
GMRF

RF

Generalized Additive Models

obs ~ environmental predictors + s(lat,lon)

- Common, simple approach
- Parameterized by spline basis functions (*not spatial correlation*)



Wood (2006)

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

Gaussian Markov random field

Lindgren, Rue, & Lindstrom (2011)

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

Gaussian Markov random field

- Models *covariance* as function of spatial locations

  obs $\sim$ environmental predictors $+ \; MVN(0, \boldsymbol{\Sigma})$

Lindgren, Rue, & Lindstrom (2011)

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

Gaussian Markov random field

- Models *covariance* as function of spatial locations

  obs ~ environmental predictors $+ MVN(0, \mathbf{\Sigma})$

Problem…

- $\mathbf{\Sigma}$ has $O(N^2)$ elements
- Computations scale as $O(N^3)$ from $|\mathbf{\Sigma}|$ and $\mathbf{\Sigma}^{-1}$

Lindgren, Rue, & Lindstrom (2011)

Q1: Which spatial model is best?

GLM

GAM

GMRF

RF

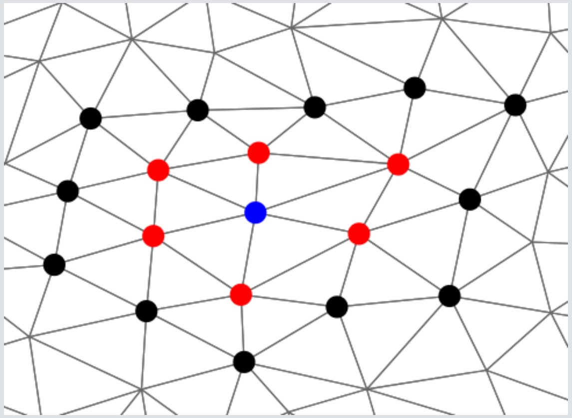
Gaussian Markov random field

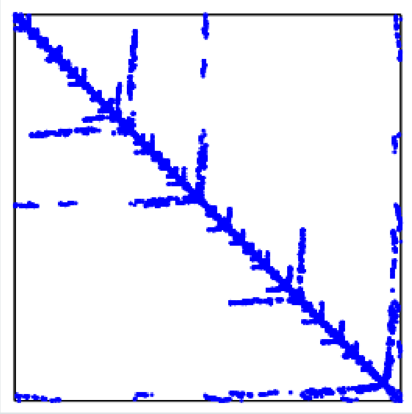
- Models *covariance* as function of spatial locations

  obs $\sim$ environmental predictors + $MVN(0, \boldsymbol{\Sigma})$

Solution:

- correlation = 0 for "far away" points → sparse matrix





Methods

Lindgren, Rue, & Lindstrom (2011)

Q1: Which spatial model is best?

GLM

GAM

GMRF

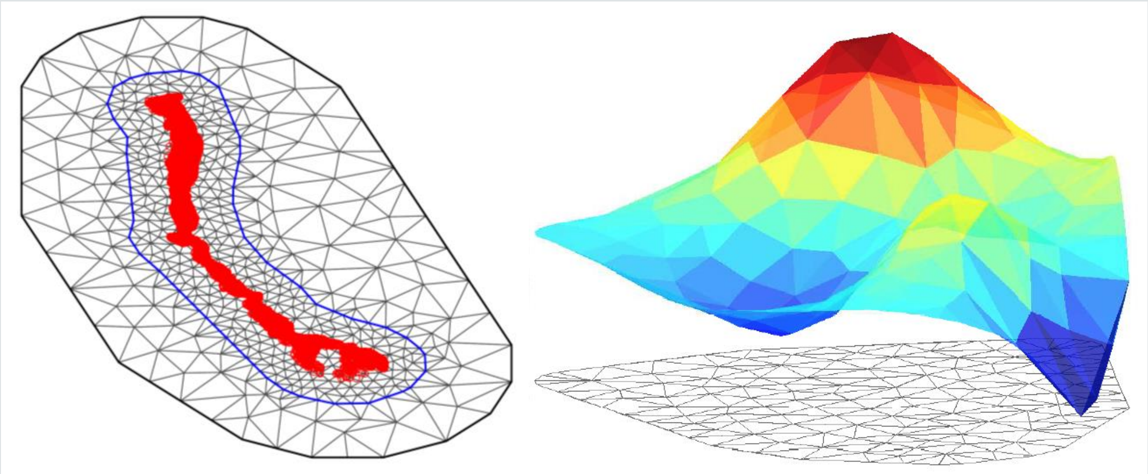
RF

Gaussian Markov random field

- Models *covariance* as function of spatial locations

  obs $\sim$ environmental predictors $+ MVN(0, \boldsymbol{\Sigma})$

- Discrete approximation of continuous space



Lindgren, Rue, & Lindstrom (2011)

Q1: Which spatial model is best?

GLM

GAM

GMRF

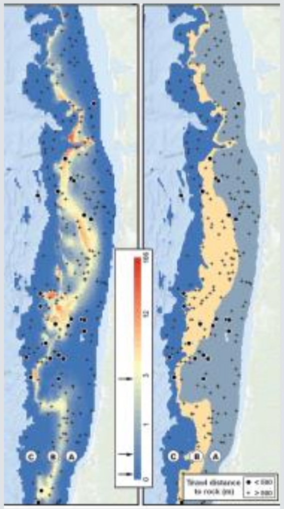
RF

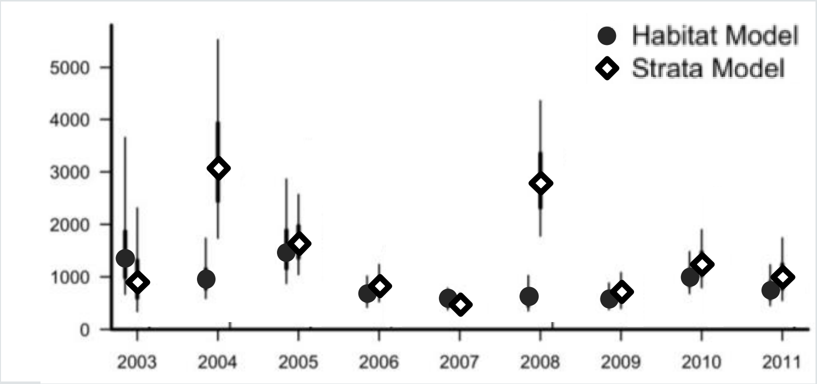
Gaussian Markov random field

- Models *covariance* as function of spatial locations

  obs $\sim$ environmental predictors $+$ $MVN(0, \boldsymbol{\Sigma})$

- Increasing adoption in fisheries





Shelton et al. (2014)

Q1: Which spatial model is best?

GLM

GAM

GMRF

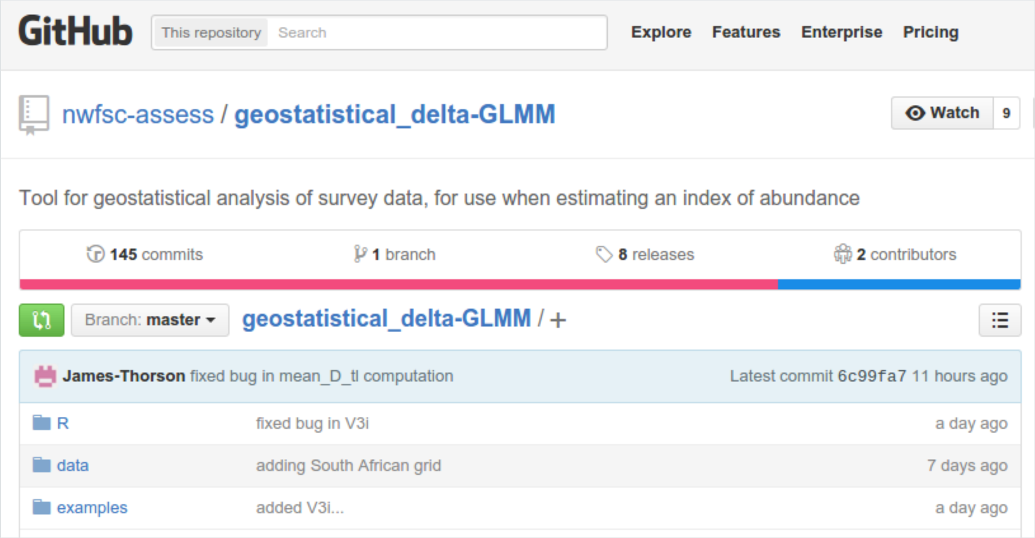
RF

Gaussian Markov random field

- Models *covariance* as function of spatial locations

  obs $\sim$ environmental predictors $+\ MVN(0, \boldsymbol{\Sigma})$

- Increasing adoption in fisheries



Methods

Thorson et al. (2015)

Q1: Which spatial model is best?

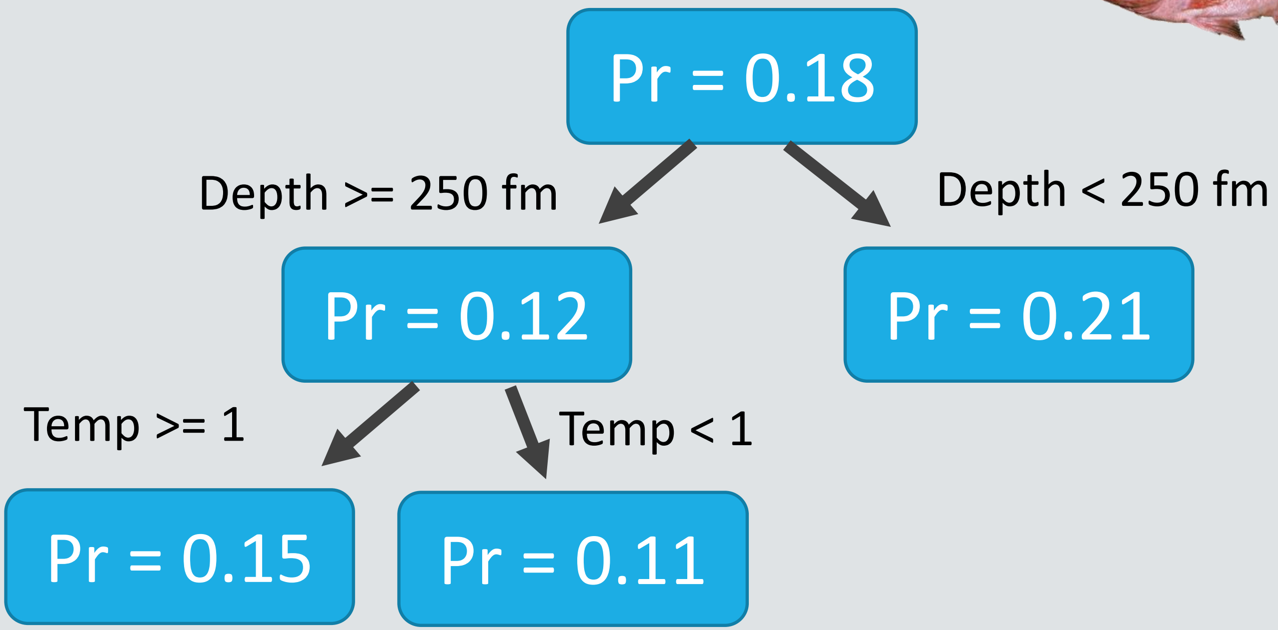
GLM

GAM

GMRF

RF

Random Forest



Breiman (2001)

Q1: Which spatial model is best?

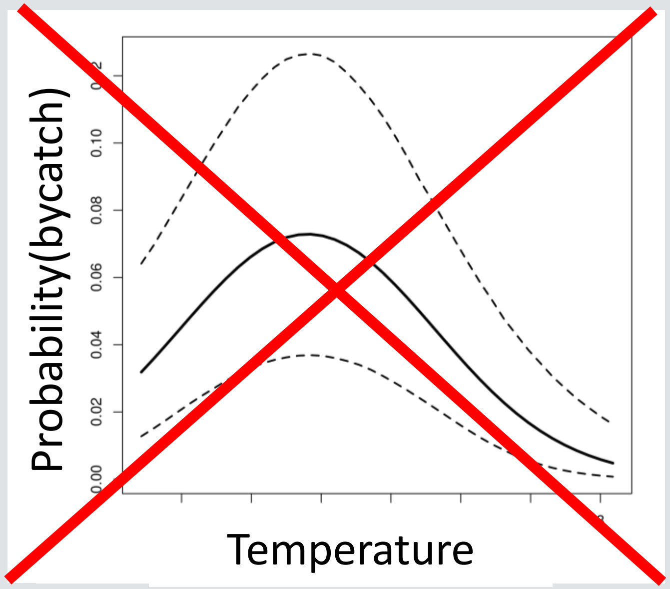
GLM

GAM

GMRF

RF

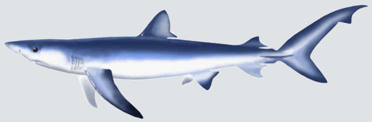
- Machine learning, designed for prediction
- "Black box"
  - Predictor-bycatch relationships not modeled
  - No spatial field (add LAT, LON)



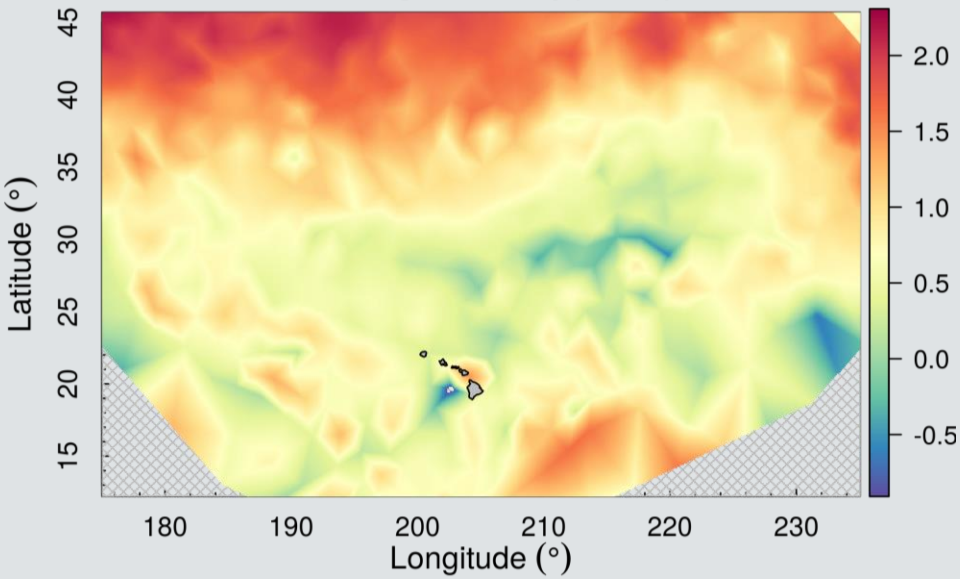
Methods

Breiman (2001)

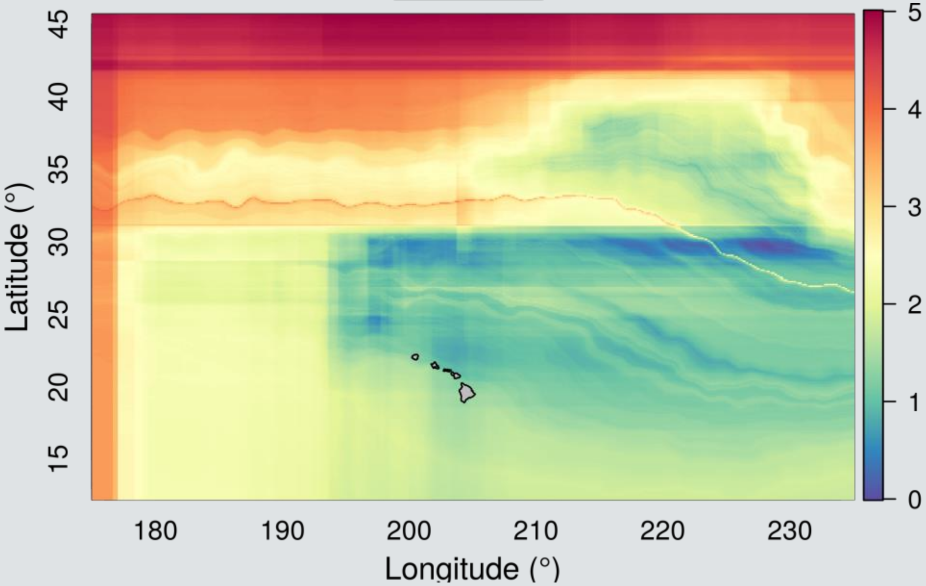
Bycatch risk maps



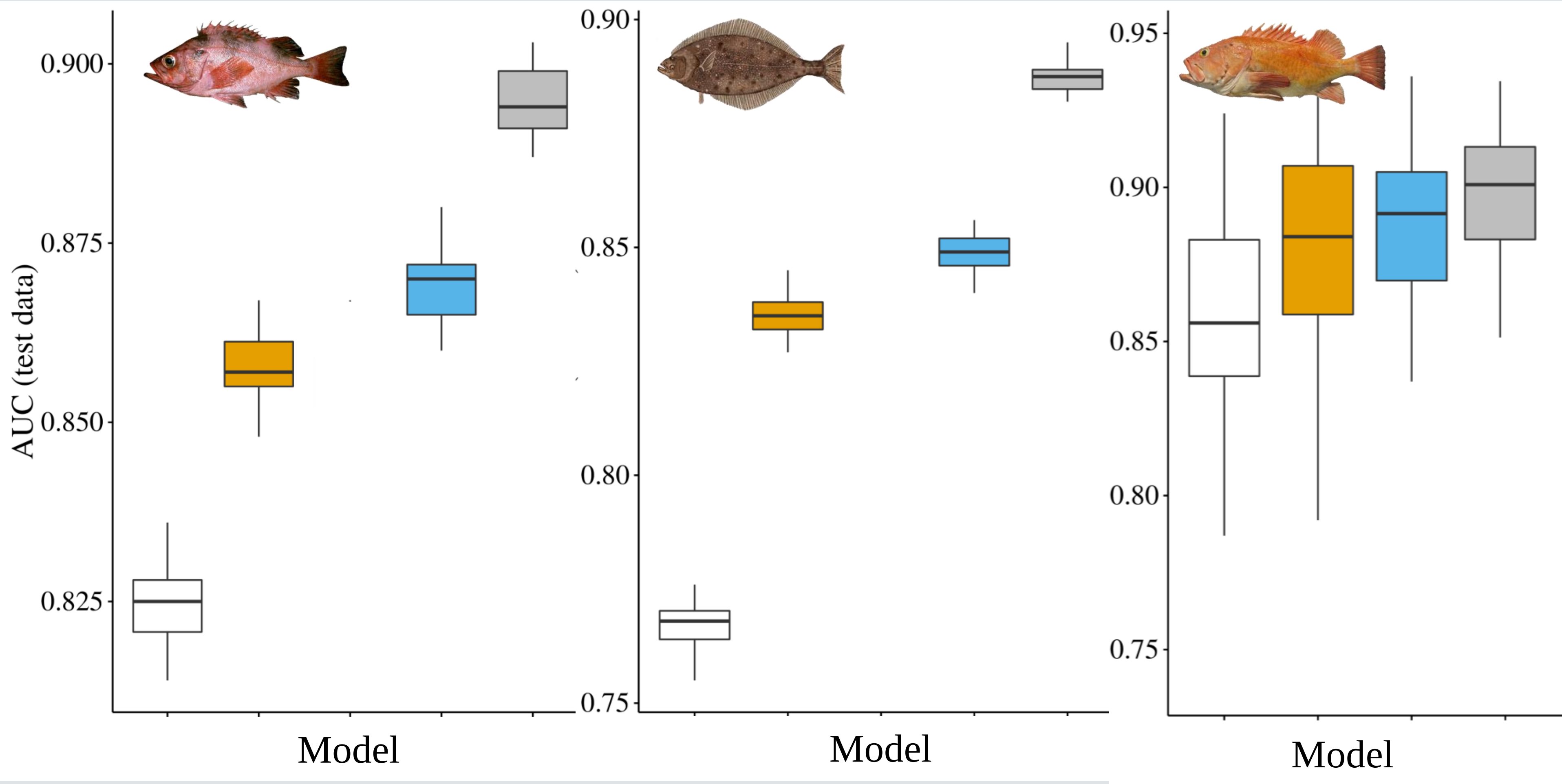
GMRF



RF



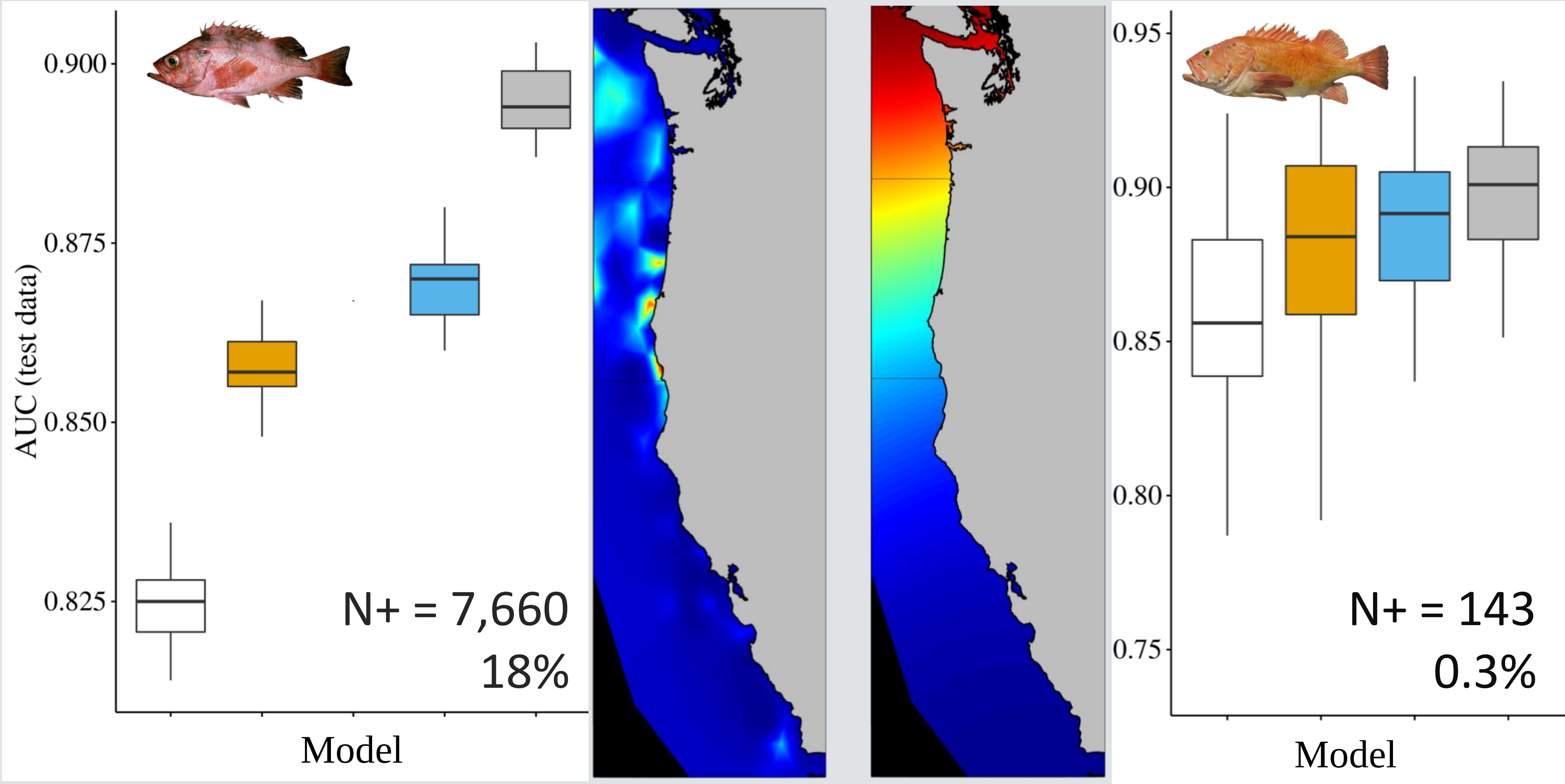
Generally: GLM < GAM < GMRF < RF



Results Binomial

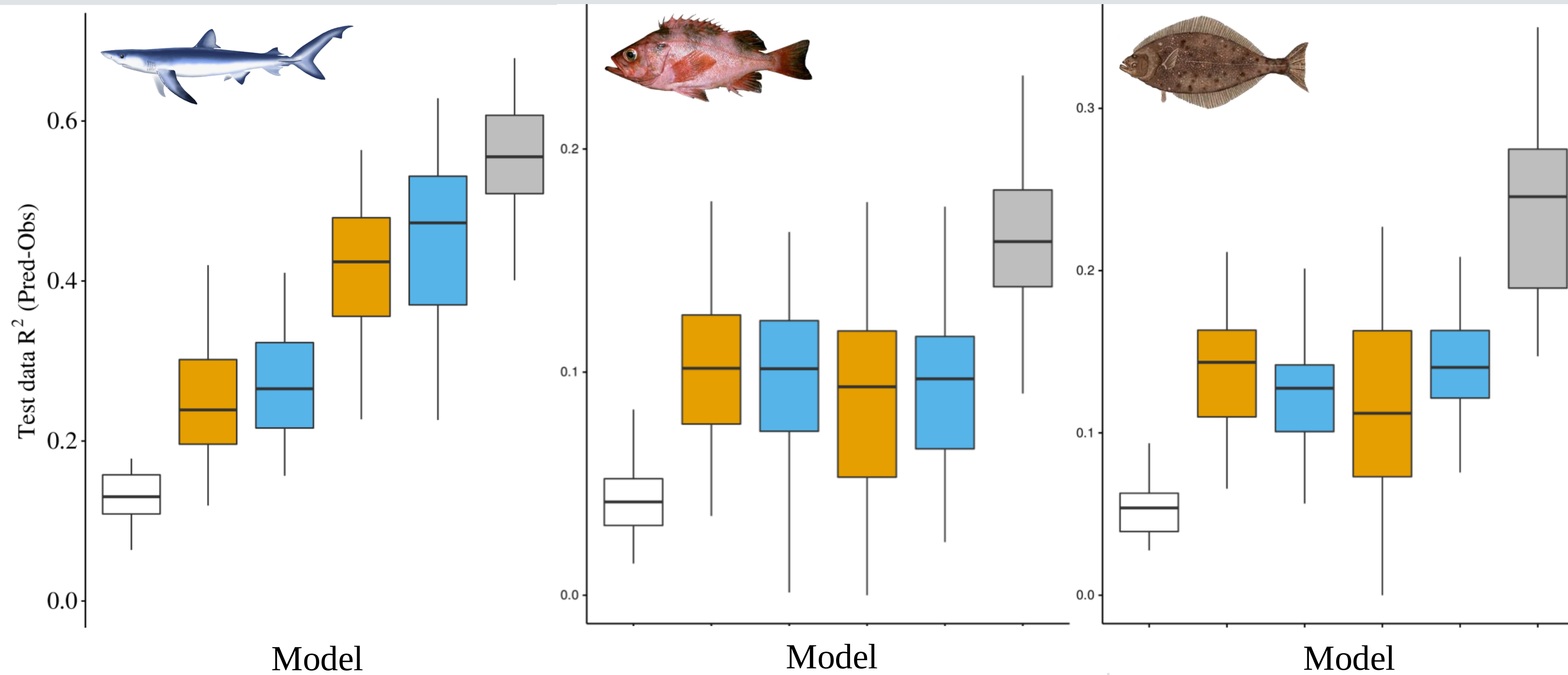
Generally: GLM < GAM < GMRF < RF

Less clear for rarer species



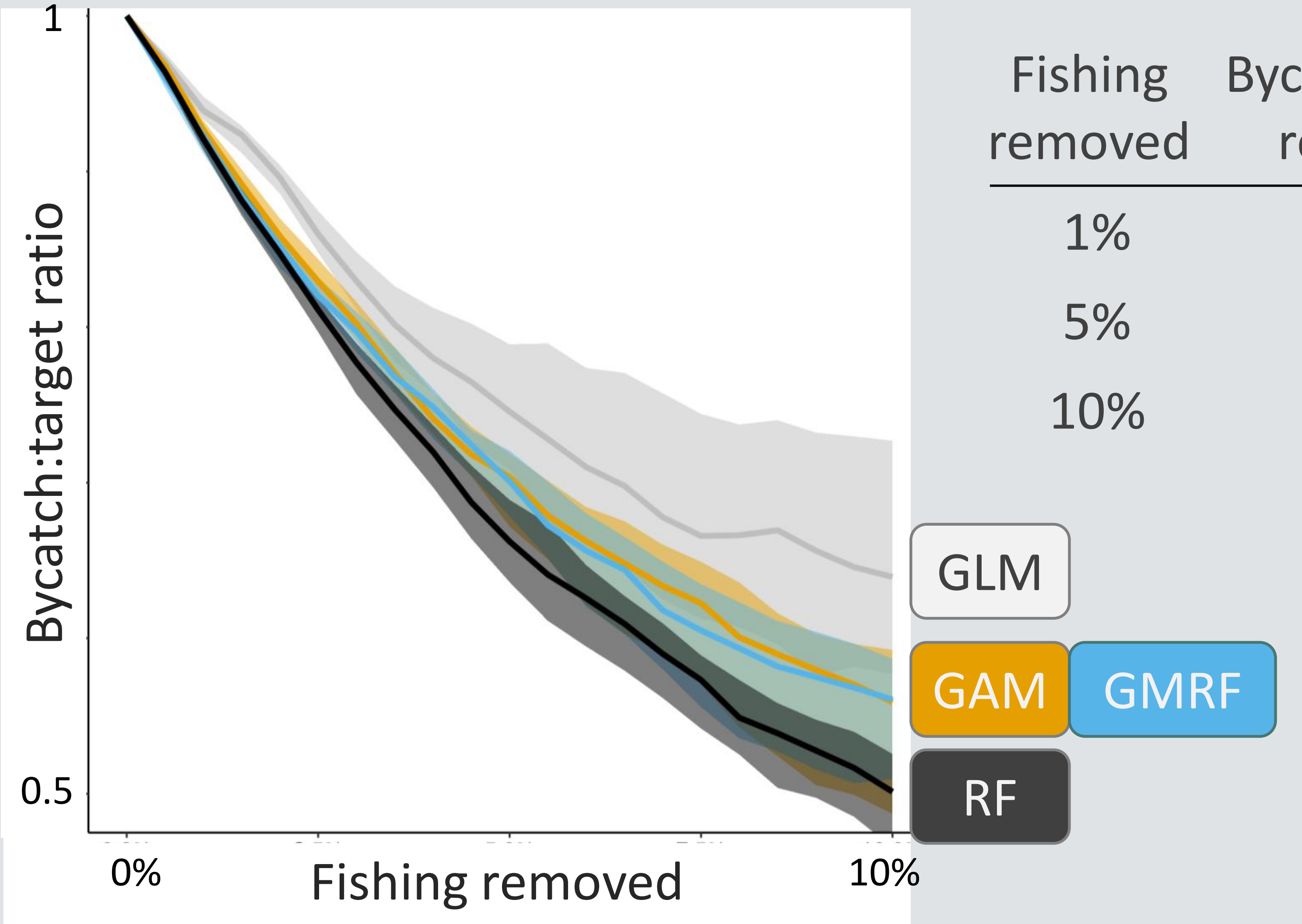
Results Binomial

Generally: GLM < GAM < GMRF < RF



Results Positive

Q3: How much bycatch can they prevent?



Fishing removed	Bycatch:target reduction
1%	8%
5%	34%
10%	50%

Conclusions

Q1: Which spatial model best predicts bycatch?

GLM < GAM < GMRF < RF

Q2: Does the answer depend on species?

No, RF had consistent advantage

(larger for species with higher bycatch rates)

Q3: How much bycatch can they prevent?

Enough to consider using them in management

Discussion

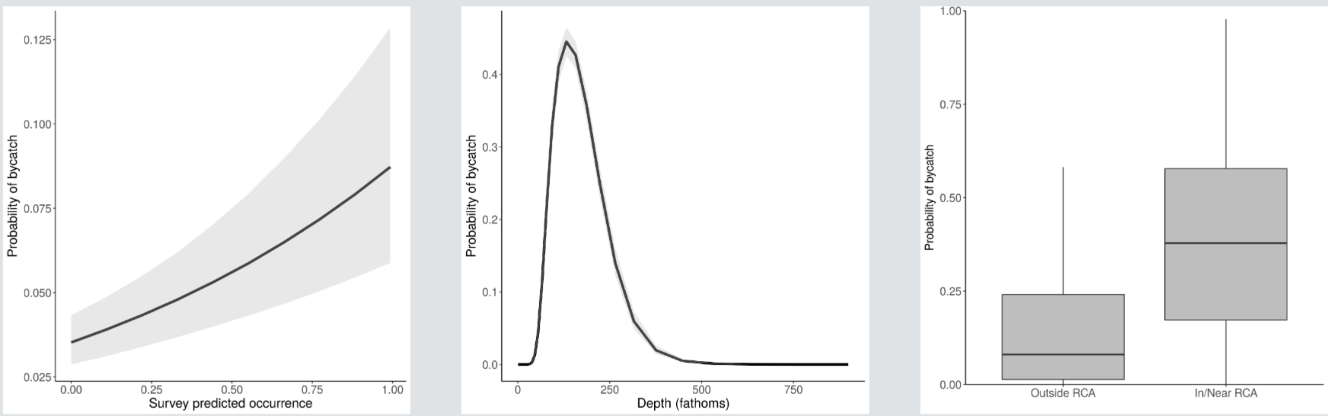
If the goal is purely *prediction*:

GLM < GAM < GMRF < RF

…but if we care about *inference on processes* affecting bycatch:

Covariate effects

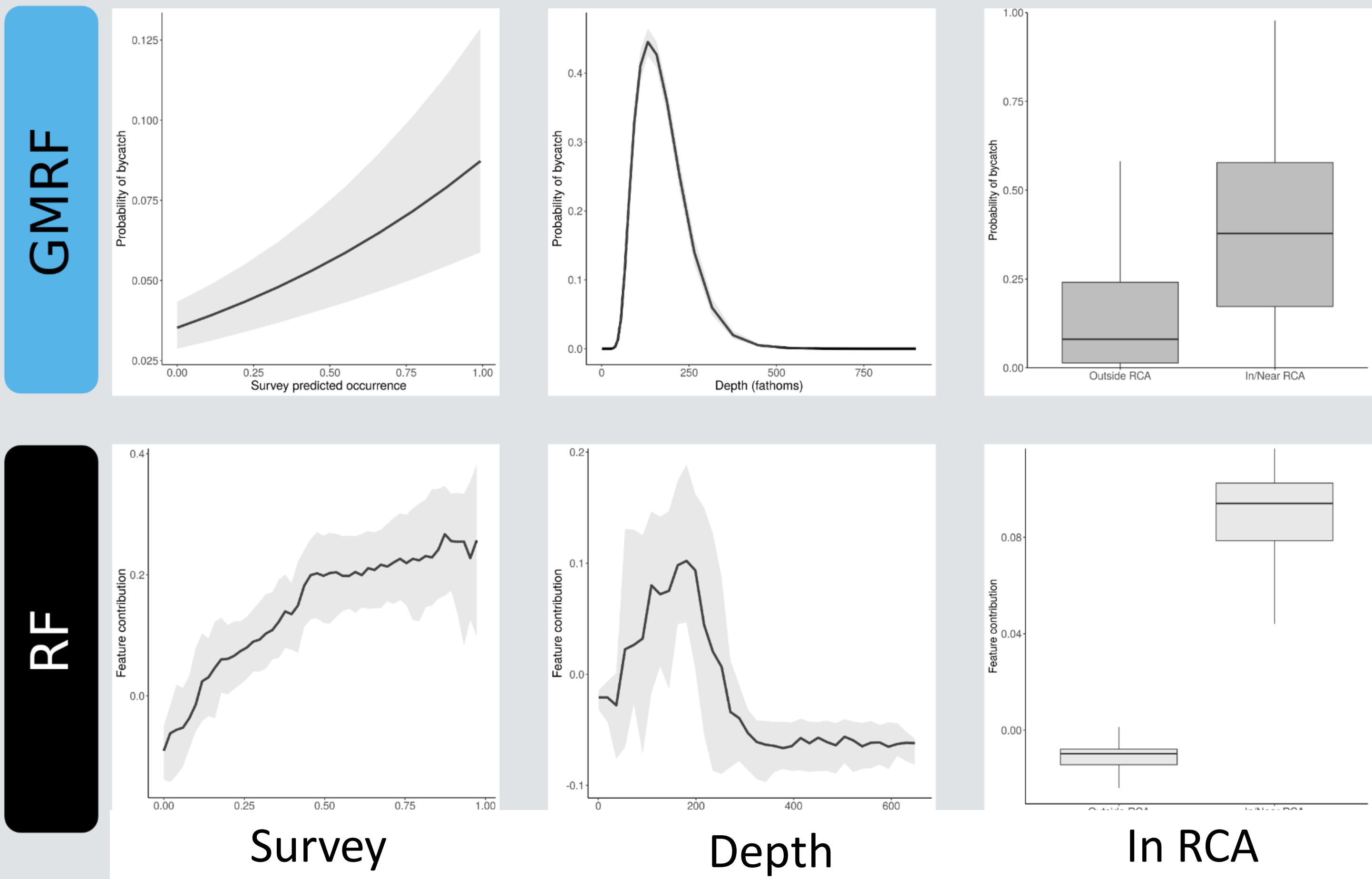
GMRF



Palczewksa (2013), Welling (2016)

Covariate effects

Are random forests really “black boxes”?

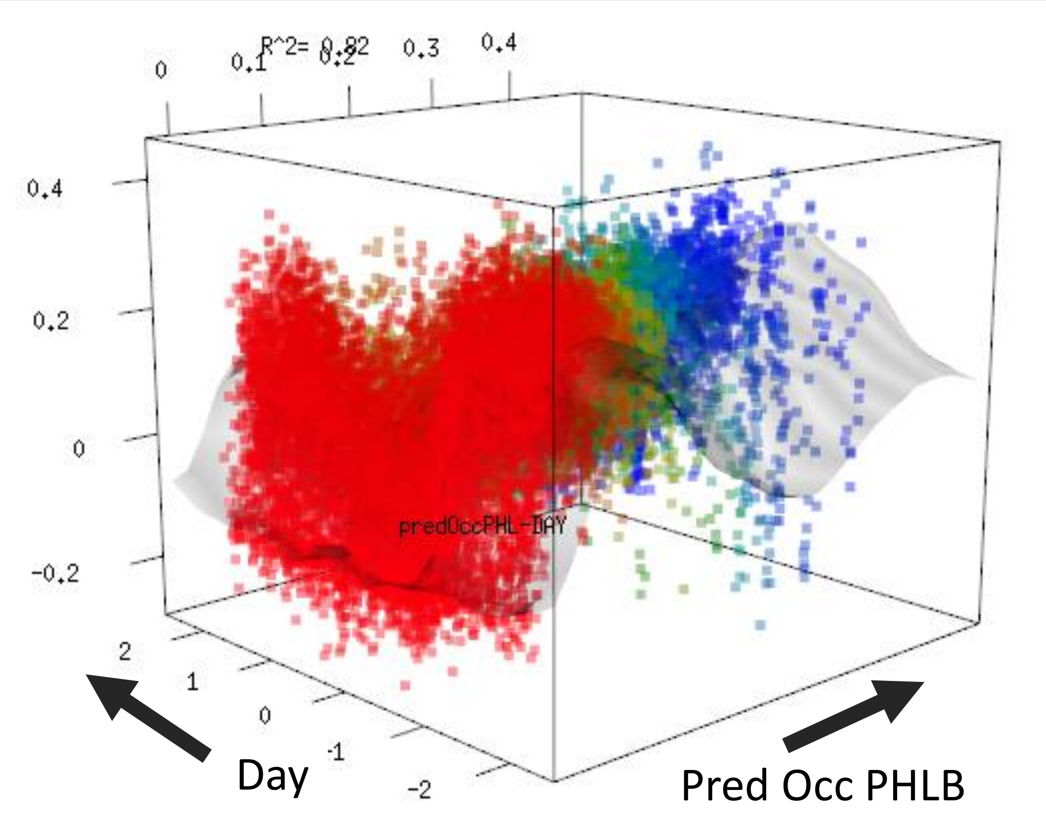


Palczewksa (2013), Welling (2016)

Can random forests do *better*?

RF

Identifying covariate interactions



Discussion

Thank you!

SIO
- Brice Semmens

SWFSC
- Tomo Eguchi

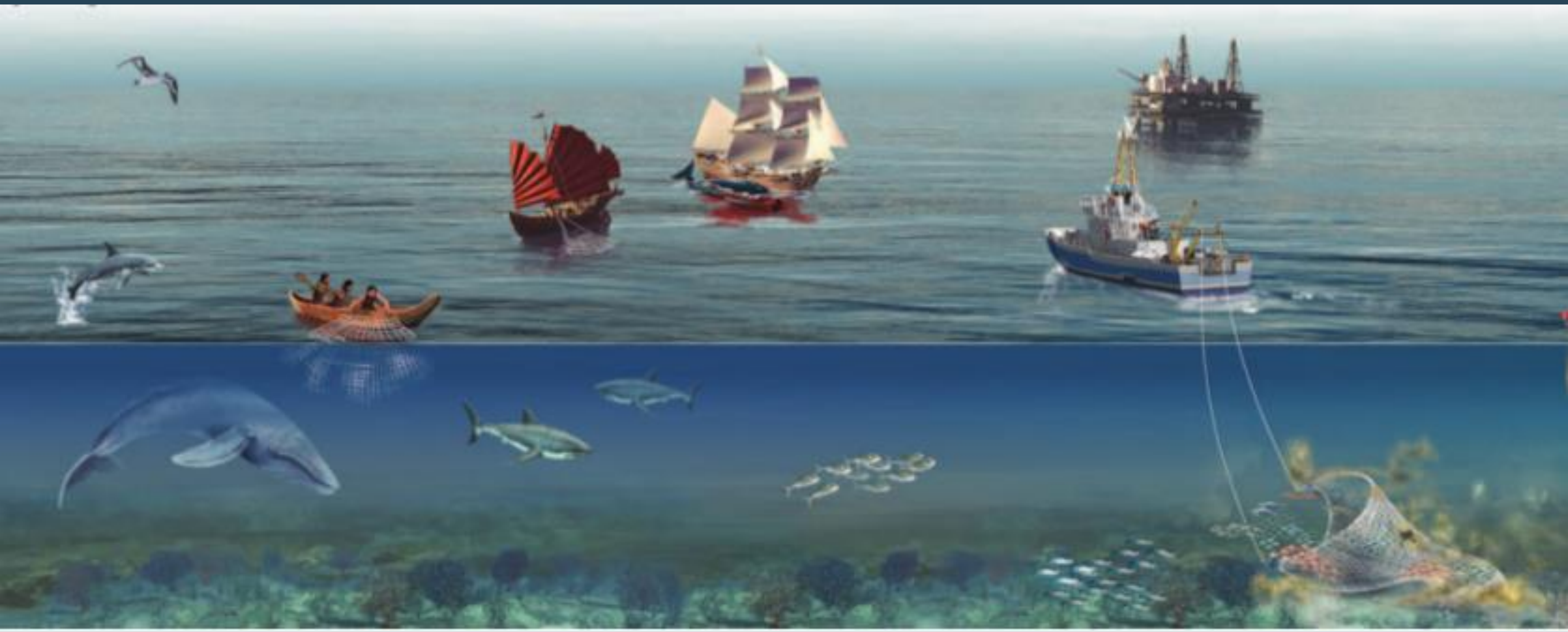
NWFSC
- Eric Ward
- Essential Fish Habitat (Blake Feist)
- West Coast Groundfish Observer Program (Jason Jannot)

PIFSC
- Hawaii Longline Observer Program (Eric Forney)

XSEDE
Extreme Science and Engineering Discovery Environment

Research opportunities in applied math/statistics and fisheries science

We can easily harvest too many fish



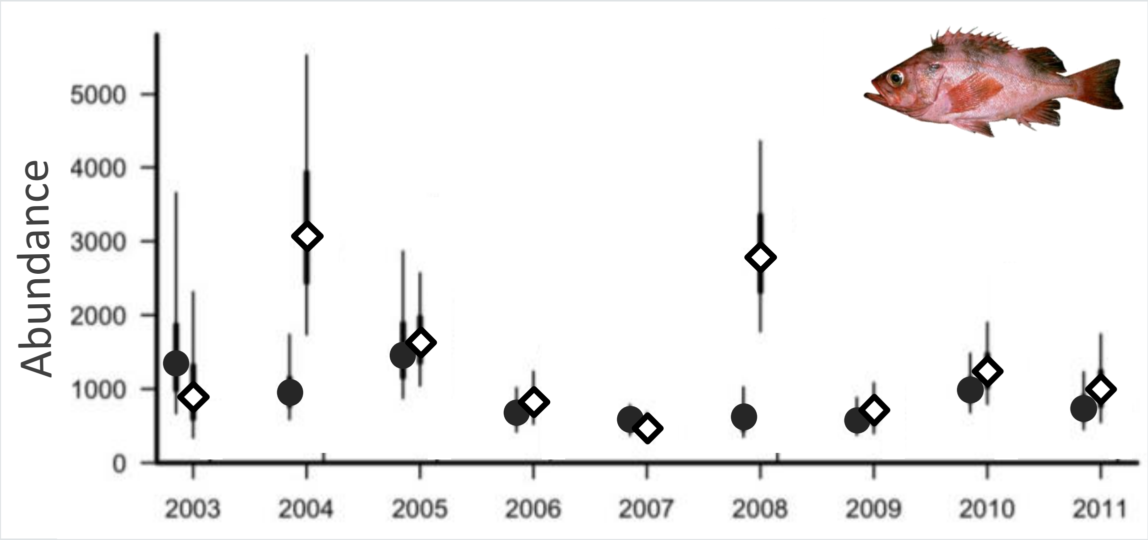
McCauley et al. 2015

We use models in management

1. Sustainable harvest → need to assess populations

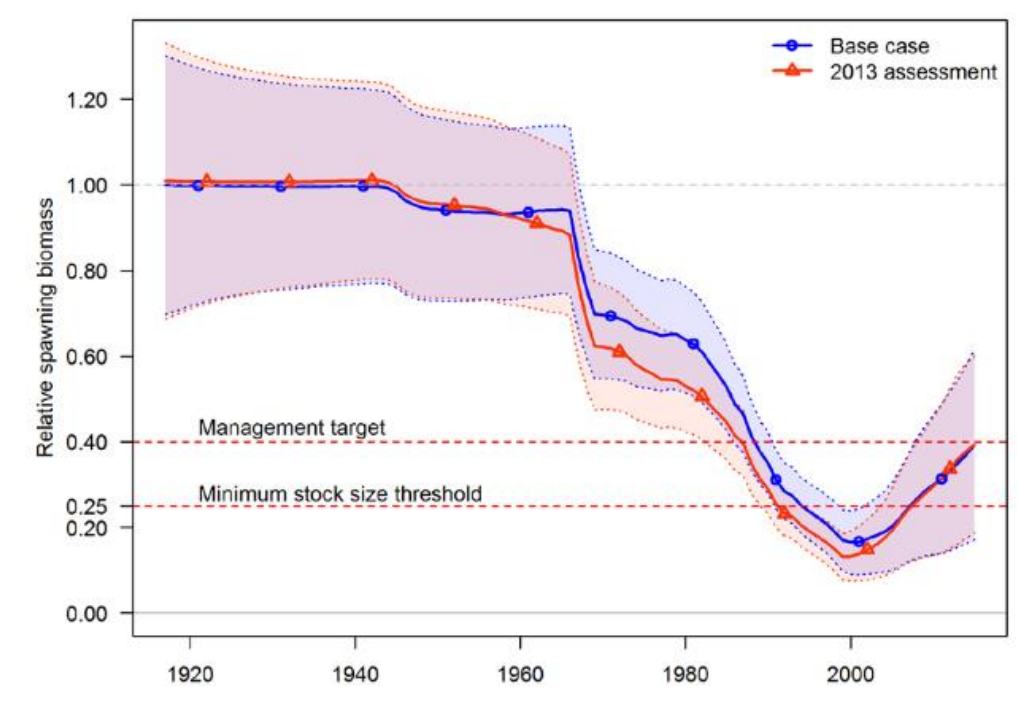
We use models in management

1. Sustainable harvest → need to assess populations
2. Primarily, *how many* and *where*



Build & test population models

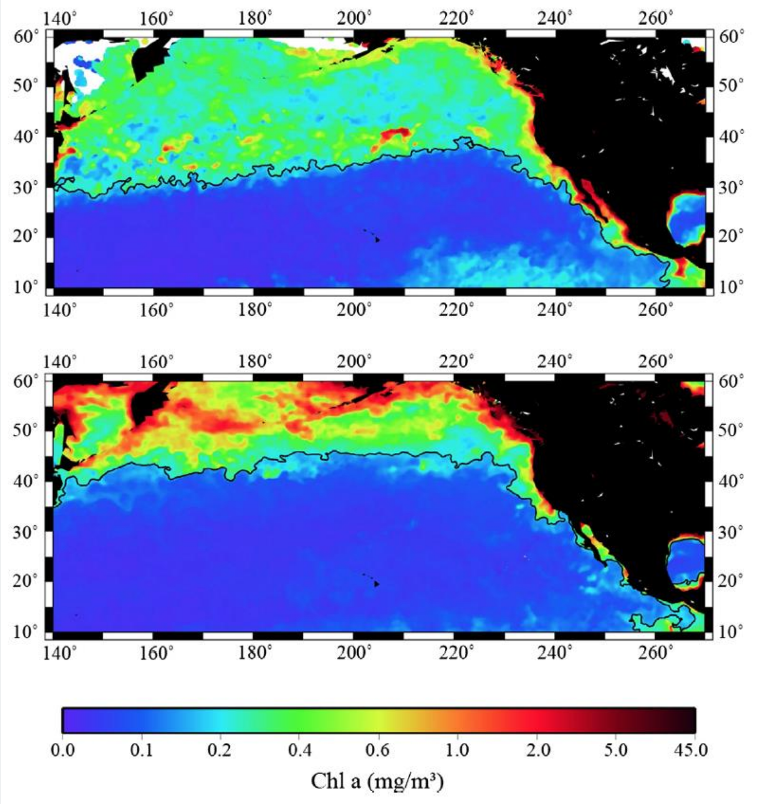
1. Stock assessment
2. Simulate alternative harvest strategies



Work with physics/climate modelers

What are the effects on fish of:

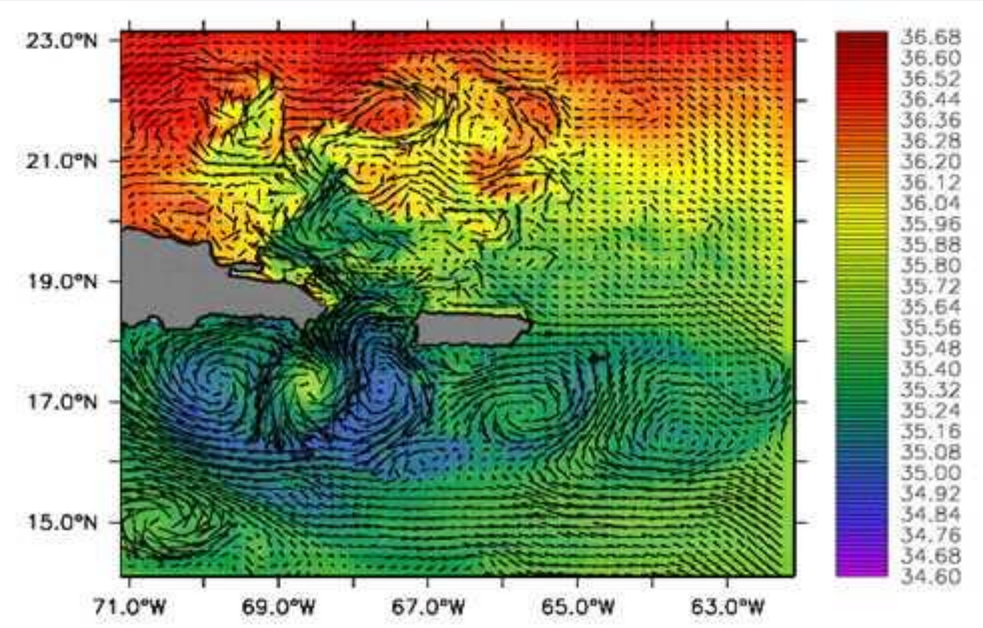
1. Ocean productivity?



Work with physics/climate modelers

What are the effects on fish of:

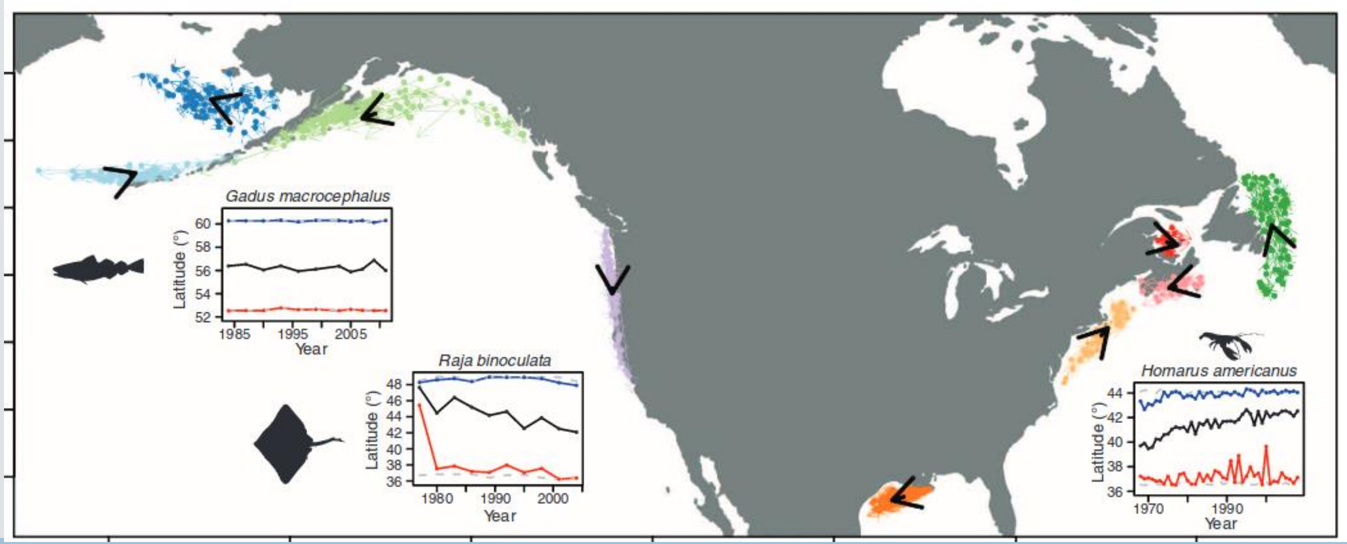
1. Ocean productivity?
2. Dispersal of eggs and larvae?



Work with physics/climate modelers

What are the effects on fish of:

1. Ocean productivity?
2. Dispersal of eggs and larvae?
3. Range shifts?



How to gauge model performance?

Goal: prediction

5-fold cross validation repeated 10x

Binomial	ROC curve (AUC)
Positive	RMSE

West Coast Groundfish covariates

Binomial

Positive

$\sim$ sst + sst^2 +

depth + depth^2 +

distance to rocky substrate +

size of rocky patch +

in Rockfish Conservation Area +

predicted occurrence (survey) +

day of year +

spatial field

Shelton et al. (2014)

Hawaii Longline covariates

Binomial

Positive

~ sst + sst^2 +

day of year +

spatial field

Shelton et al. (2014)

RF

+ Better at prediction

+ More complex covariate relationships (incl. interactions)

+ Much quicker to set up and run (~2 min vs. 5-15 hours)

+ Not just a “black box”?

GMRF

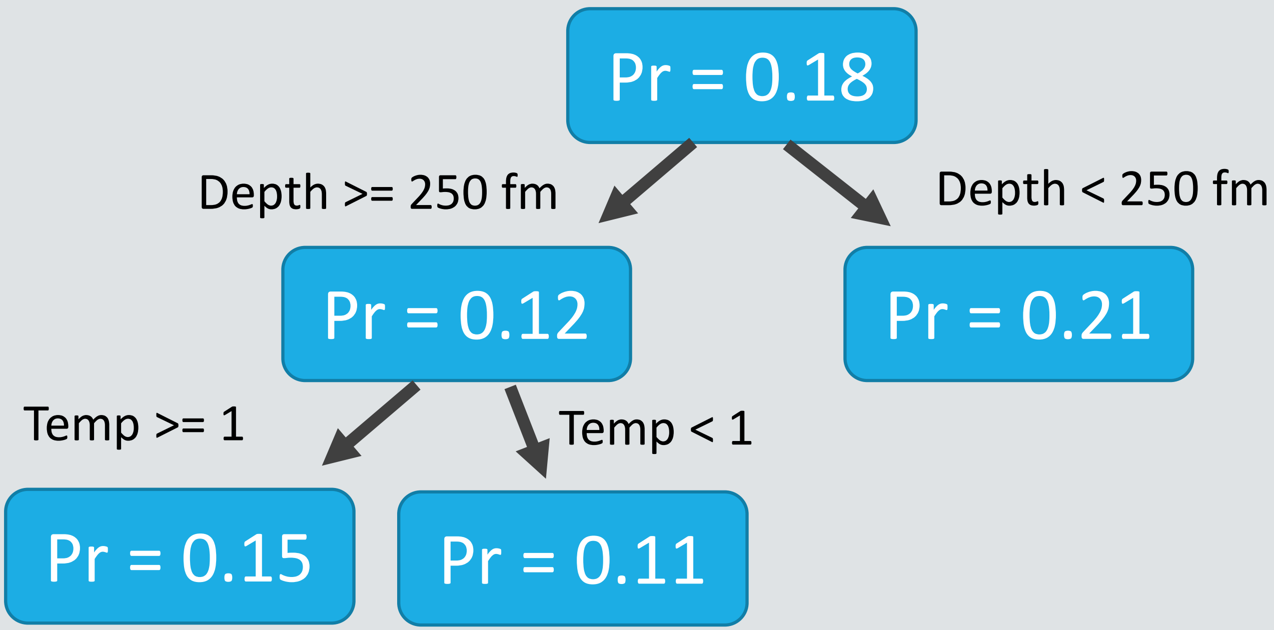
+ Statistical inference, marginal posteriors for covariate effects

+ Ability to include observation error

Covariate effects

RF

What is a “feature contribution”??

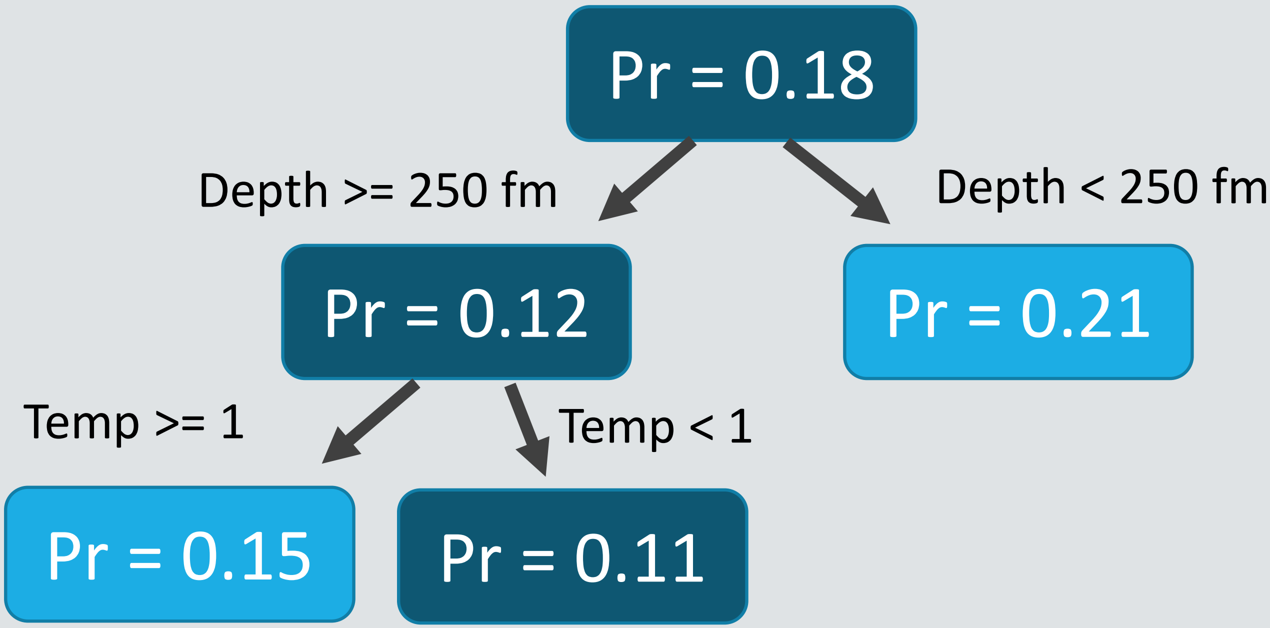


Palczewksa (2013), Welling (2016)

Covariate effects

RF

What is a “feature contribution”??



$$\text{Prediction}_i = 0.11 = 0.18 - 0.06\ (\text{Depth}) - 0.01\ (\text{Temp})$$

Palczewksa (2013), Welling (2016)