Using machine learning techniques to estimate pelagic species distributions under novel environmental conditions in the California Current system

Barbara Muhling, Carlos Gaitan, Elliott Hazen, Stephanie Brodie, Michael Jacox, James Smith, Desiree Tommasi, Toby Auth, Rick Brodeur

UNIVERSITY OF CALIFORNIA SANTA CRUZ

Climate change and transboundary fish stocks

Climate change → *Species redistributions* → *Management challenges*

ClimateChangePost

Mackerel migrating to the north: the first climate change related conflict in European politics?

October 9th, 2017

Hakai magazine — Coastal science and societies

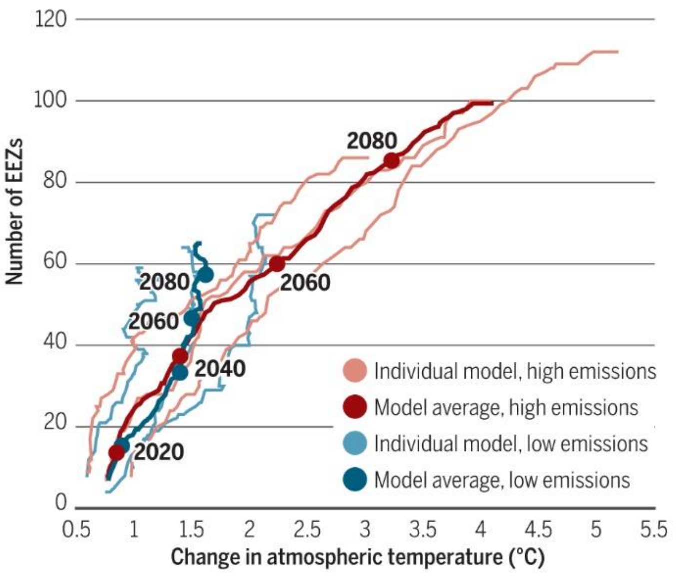
International Fish Fights on the Rise

A new report shows that there is increasing competition between countries for access to seafood.

Pinsky et al. 2018

The number of EEZs with new transboundary stocks increases with global temperature

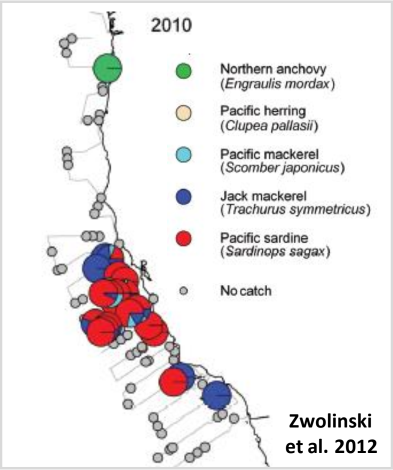
The extent of warming and number of EEZs were greater under a high–greenhouse gas emissions scenario (RCP 8.5, red) and lower under a low-emissions scenario (RCP 2.6, blue). See supplementary materials.



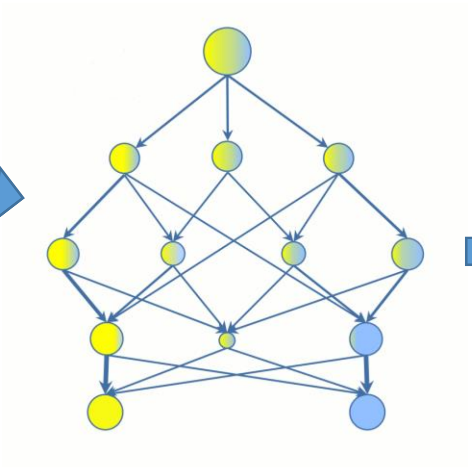
Species distribution models (SDMs)

- Quantify relationships between species distribution and oceanographic environment
- Many methods available, including some machine learning techniques

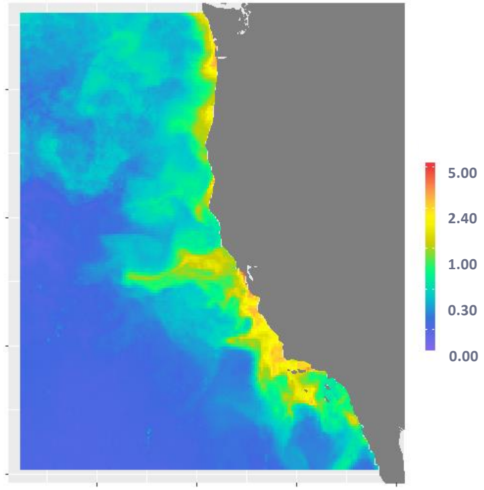
Biological observations



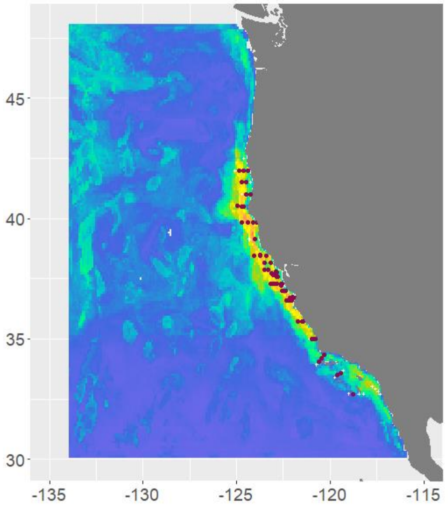
Species distribution model



Environmental predictors



Spatial predictions

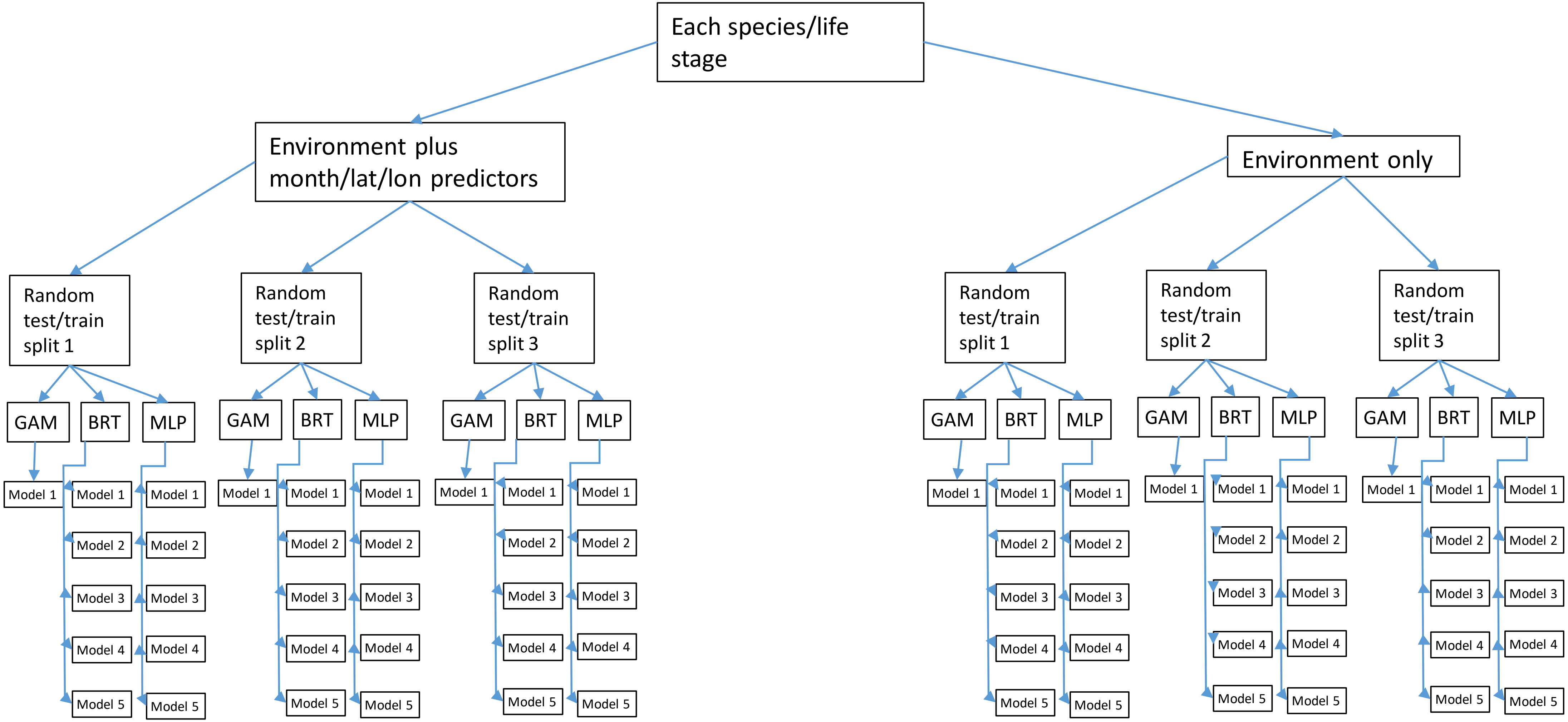


Approach

- Test performance of three different types of SDM during novel environmental conditions
- Sardine (*Sardinops sagax*) and anchovy (*Engraulis mordax*) in the California Current
  - Adults from trawl surveys
  - Larvae from CalCOFI plus other surveys (Auth, Brodeur et al.)
- Biological data split into three sections:
  - Model **training**: to build the SDM (2002 – 2013)
  - Model **testing**: to determine the best configuration for the SDM (2002 – 2013)
  - Model **validation**: marine heatwave years (2014 – 2016)
- Historical test/train split:
  1. 50% of data used for model training, 50% for model testing, split determined randomly, repeated 3 times

Approach - 2

- Two binomial SDMs built for each species/life stage and method:
  - One including only environmental and stock size predictors
  - One also including latitude, longitude, and month
- Three SDM methods:
  - Generalized Additive Models (GAMs) in the *mgcv* package
    - Number of knots (k) allowed to vary from 3 - 7, to keep partial relationships biologically reasonable
  - Boosted Regression Trees (BRTs) in the *gbm* package
    - Tree complexity allowed to vary from 3 – 9, number of trees from 1000 – 3000
  - Multilayer Percepton (MLP) neural networks in the *neuralnet* package
    - One hidden layer. Number of neurons in hidden layer allowed to vary from 1 – 10
    - Resilient backpropagation with weight backtracking algorithm
- The best SDM configuration was chosen based on highest AUC against model testing data



i.e. 66 SDMs per species/life stage

Environmental predictors

ROMs Variable	Biological Relevance
Moon phase	Foraging behavior and depth distribution
Sea surface temperature	Metabolic processes, thermal niche
SD of sea surface temperature	Dynamic temperature variability
Sea surface height	Mesoscale current and eddy features
SD of sea surface height	Dynamic mesoscale feature variability
Eastward surface current flow	Inshore/offshore transport
Eastward surface wind stress	Nearshore dynamics, retention
Northward surface current flow	Alongshore transport
Northward surface wind stress	Upwelling proxy
Wind stress curl	Tendency for convergence/divergence at surface
Eddy kinetic energy	Eddy dynamics
Isothermal layer depth	Depth of surface mixing
Bulk buoyancy frequency	Stratification and stability in upper water column

DYNAMIC HABITAT USE OF ALBACORE AND THEIR PRIMARY PREY SPECIES IN THE CALIFORNIA CURRENT SYSTEM

BARBARA MUHLING, STEPHANIE BRODIE,
OWYN SNODGRASS, DESIREE TOMMASI
University of California, Santa Cruz
Institute for Marine Science
Santa Cruz, CA
ph: (858) 546-7197
Barbara.Muhling@noaa.gov

BARBARA MUHLING, OWYN SNODGRASS,
HEIDI DEWAR, DESIREE TOMMASI, JOHN CHILDERS
NOAA Southwest Fisheries Science Center
San Diego, CA

STEPHANIE BRODIE, MICHAEL JACOX
NOAA Southwest Fisheries Science Center
Monterey, CA

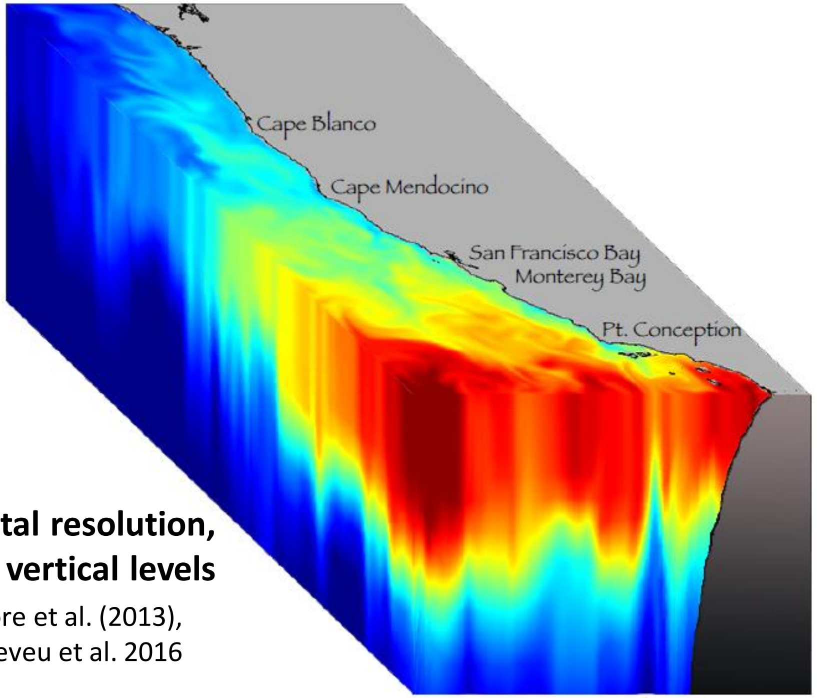
MICHAEL JACOX
NOAA Earth System Research Laboratory
Boulder, CO

CHRISTOPHER A. EDWARDS
Ocean Sciences Department
University of California, Santa Cruz, CA

YI XU
Department of Fisheries and Oceans
Delta, British Columbia, Canada

STEPHANIE SNYDER
Thomas More University,
Crestview Hills, KY

- Sourced from data-assimilative ROMS
- Surface chlorophyll also included from satellite observations



0.1 horizontal resolution,
42 vertical levels

See Moore et al. (2013),
Neveu et al. 2016

frontiers in Marine Science

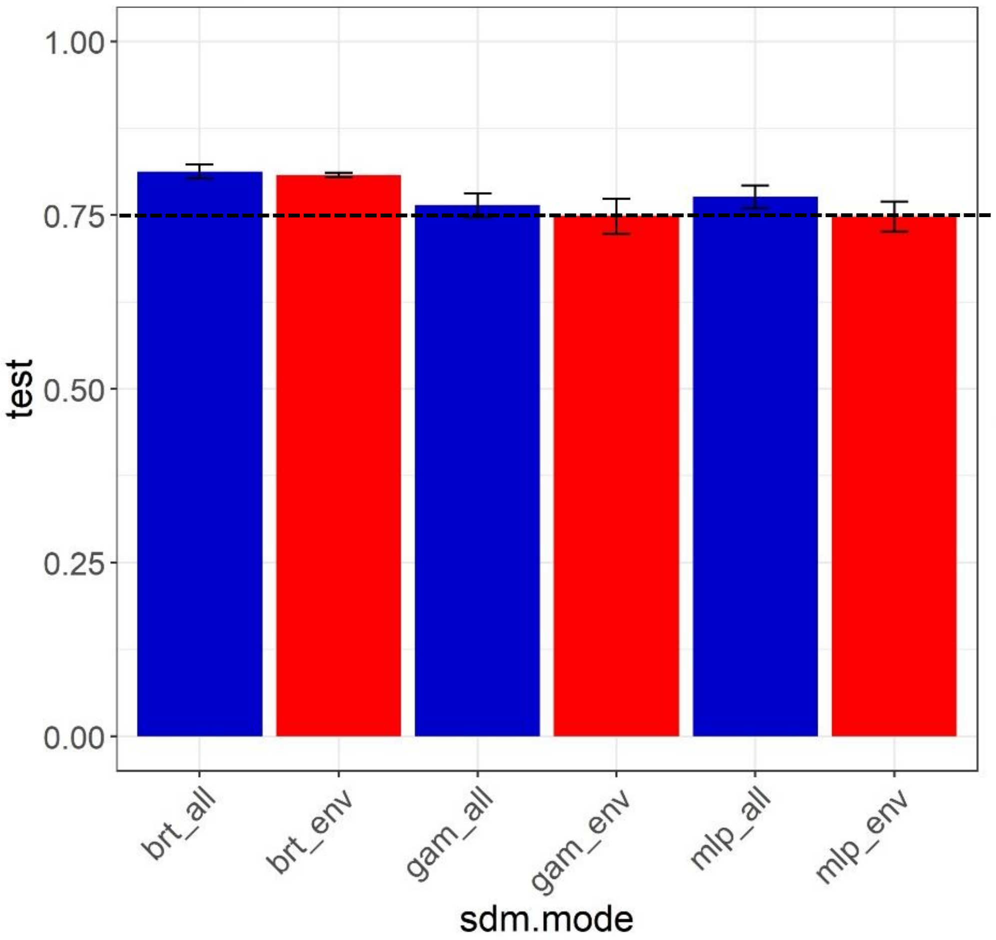
Integrating Dynamic Subsurface Habitat Metrics Into Species Distribution Models

Stephanie Brodie[1,2], Michael G. Jacox[1,2,3], Steven J. Bograd[1,2], Heather Welch[1,2], Heidi Dewar[4], Kylie L. Scales[5], Sara M. Maxwell[6], Dana M. Briscoe[1], Christopher A. Edwards[1], Larry B. Crowder[7], Rebecca L. Lewison[8] and Elliott L. Hazen[1,2]*

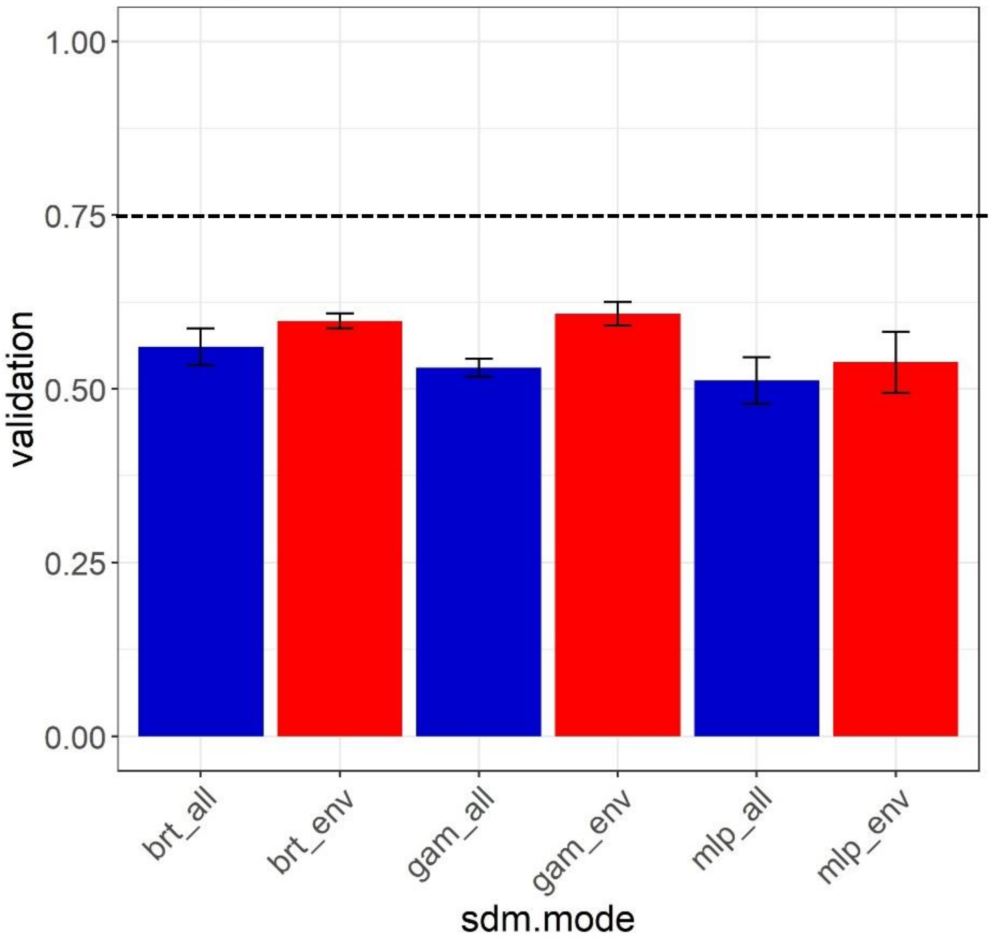
Fit to unseen test and validation data

- Example showing AUCs for adult sardine from trawl surveys
- All SDMs did well when tested against unseen test data from 2002 - 2013
- In contrast, all SDMs did quite poorly against the marine heatwave years (2014 – 2016)

unseen test data (2002 – 2013)

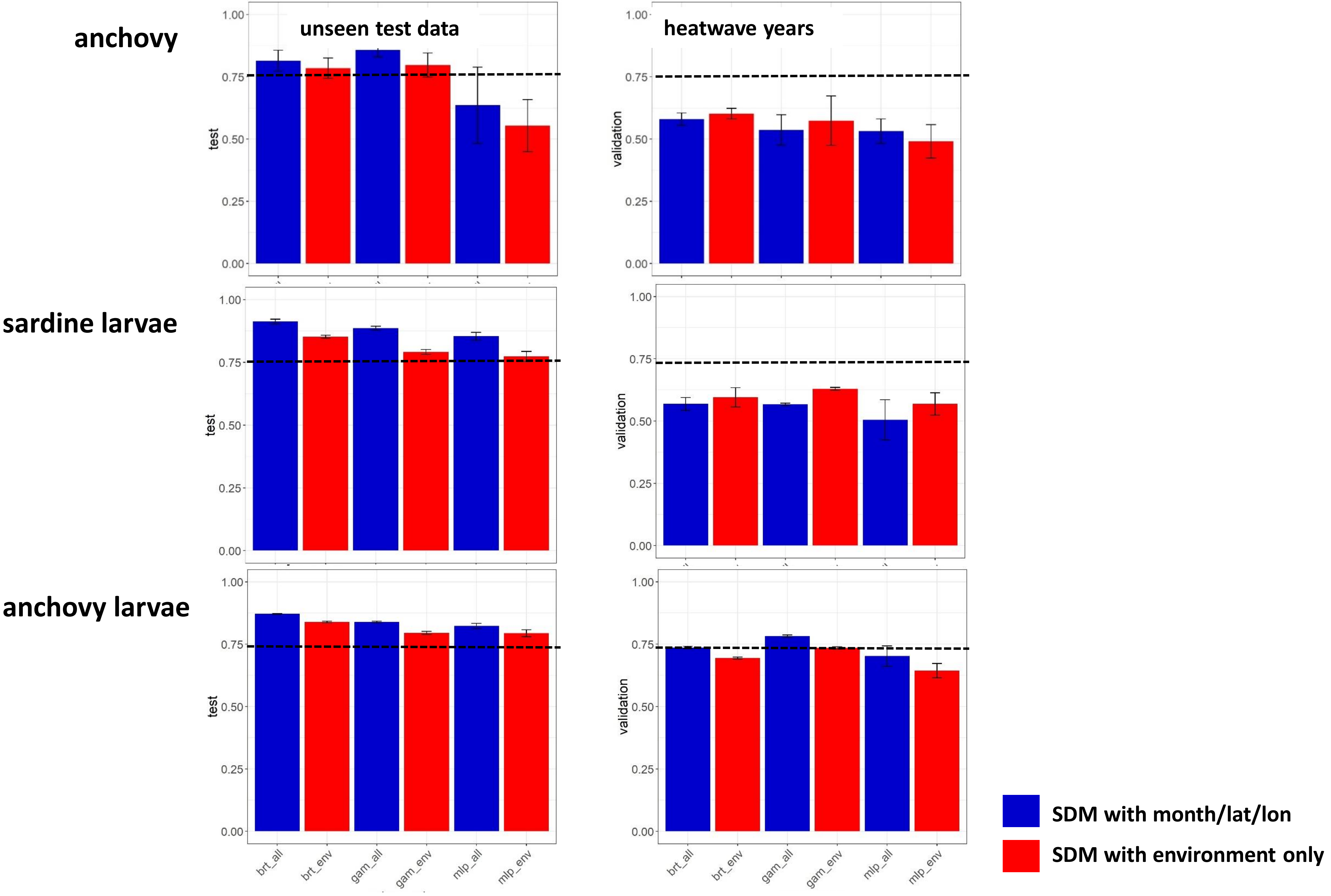


heatwave years (2014 – 2016)



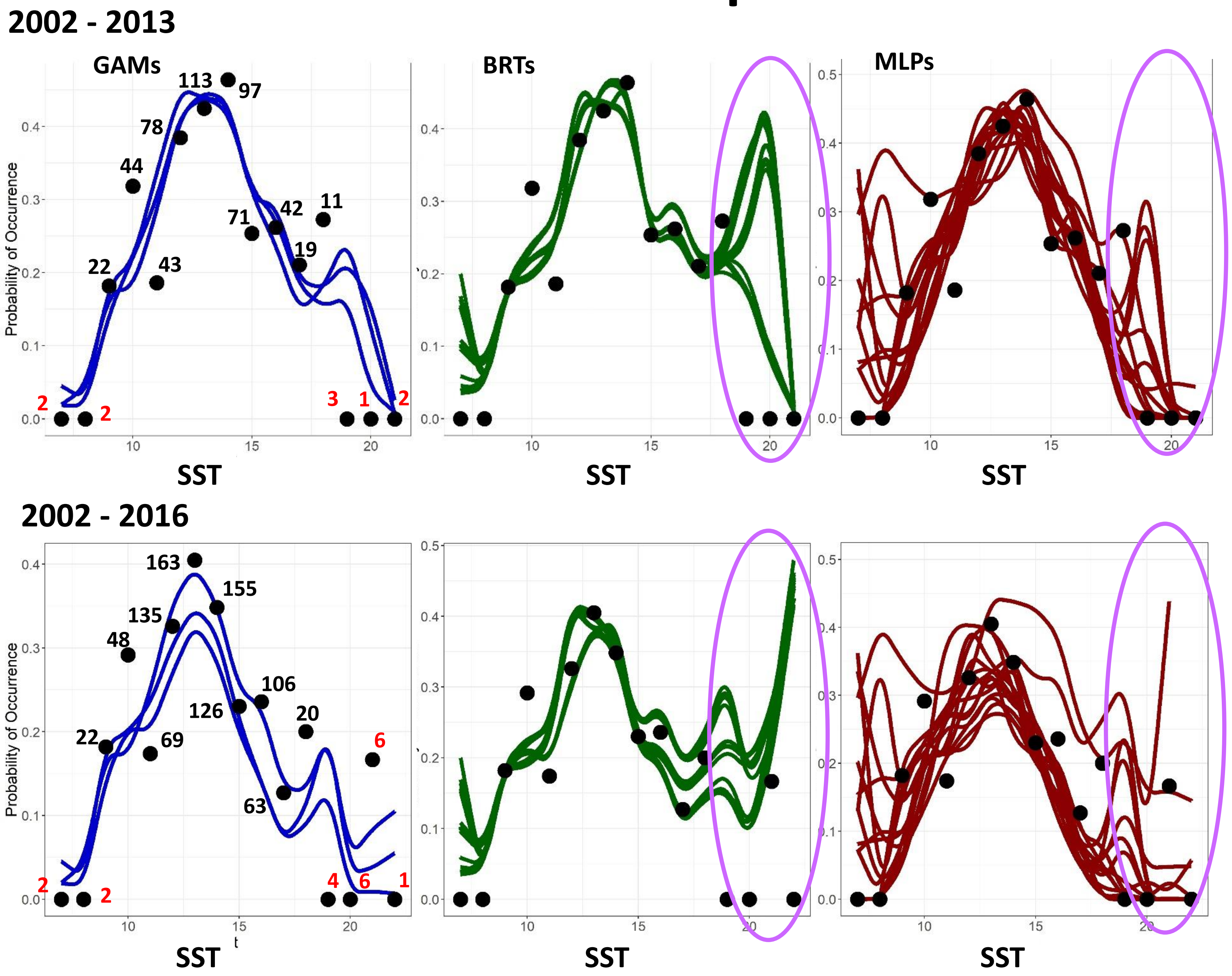
SDM with month/lat/lon

SDM with environment only



anchovy
sardine larvae
anchovy larvae
unseen test data
heatwave years
test
validation
1.00
0.75
0.50
0.25
0.00
brt_all
brt_env
gam_all
gam_env
mlp_all
mlp_env
SDM with month/lat/lon
SDM with environment only

Model extrapolation: adult sardine



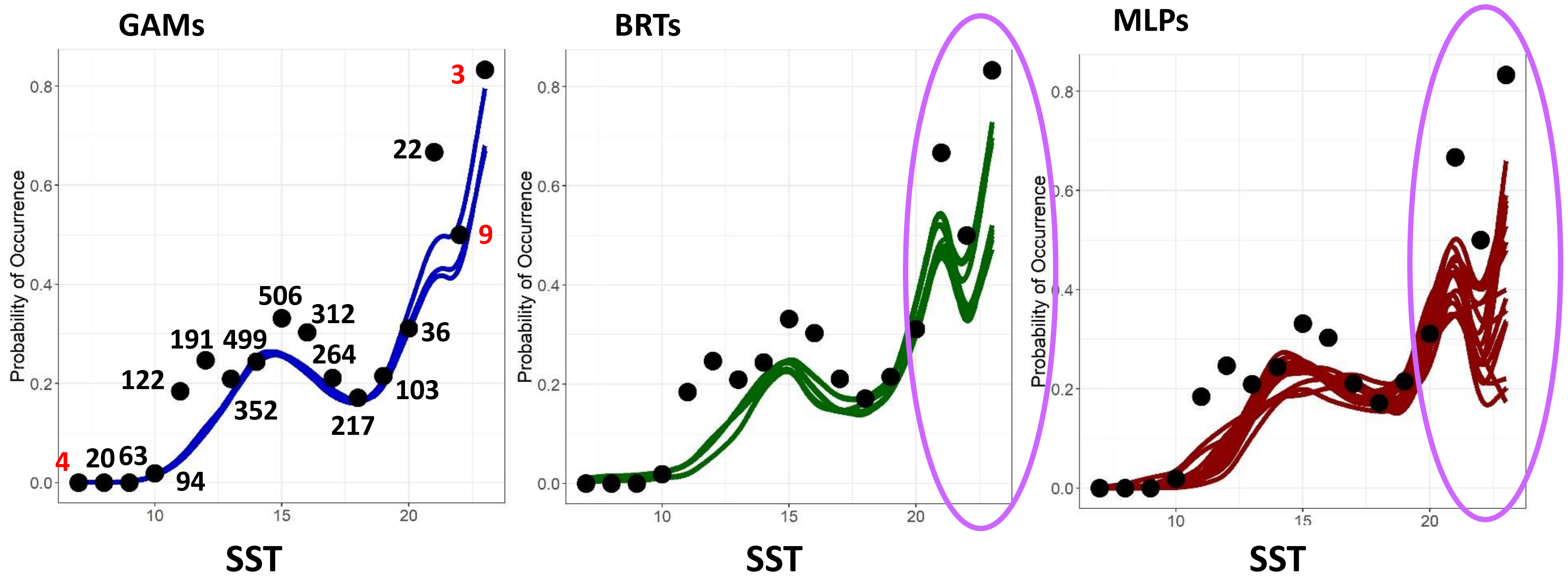
- Uncertainty in model responses at high temperatures for years 2002 – 2014 is magnified during marine heatwave years, especially in BRTs and MLPs
- Partially due to low sample sizes near limits

Observed

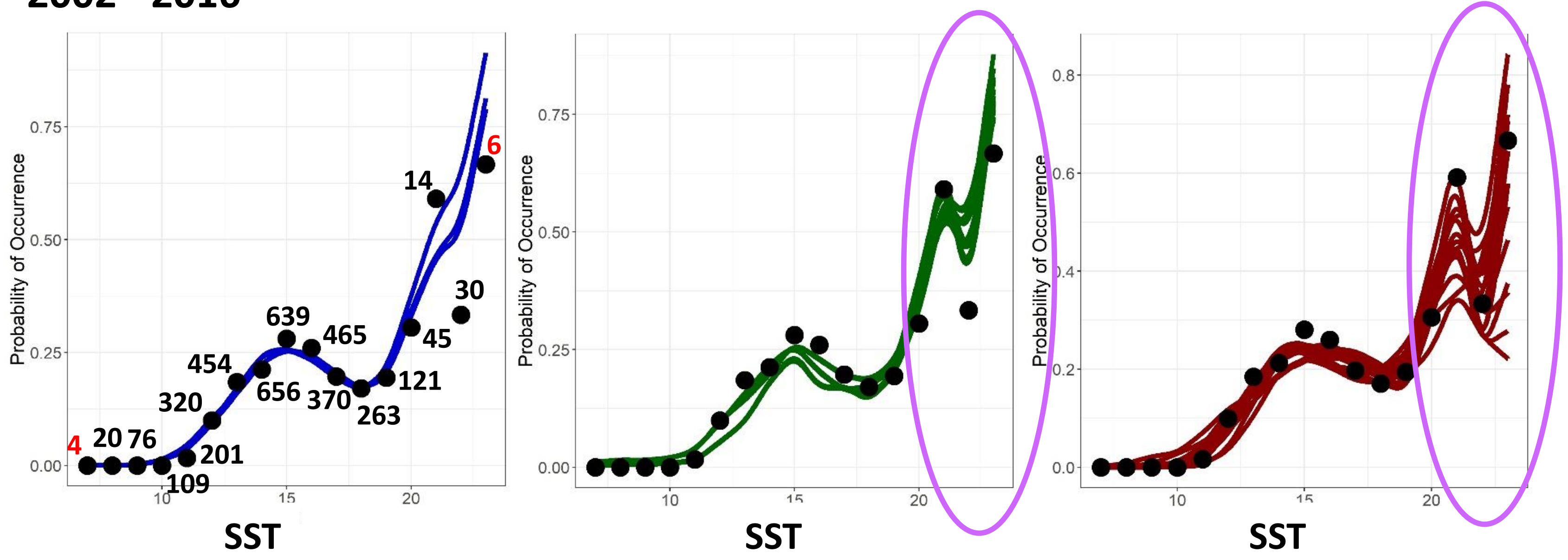
Predicted

Model extrapolation: larval anchovy

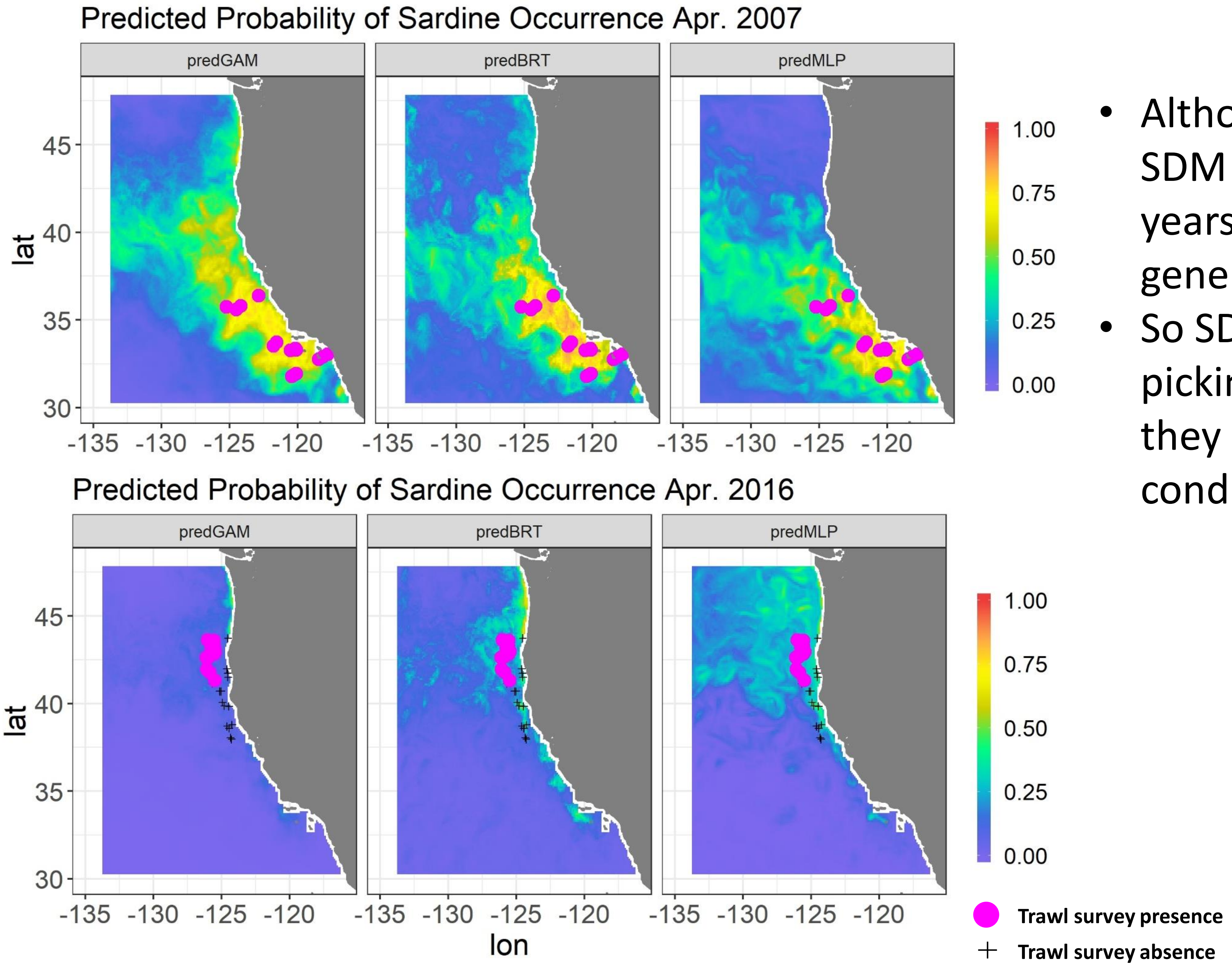
2002 - 2013



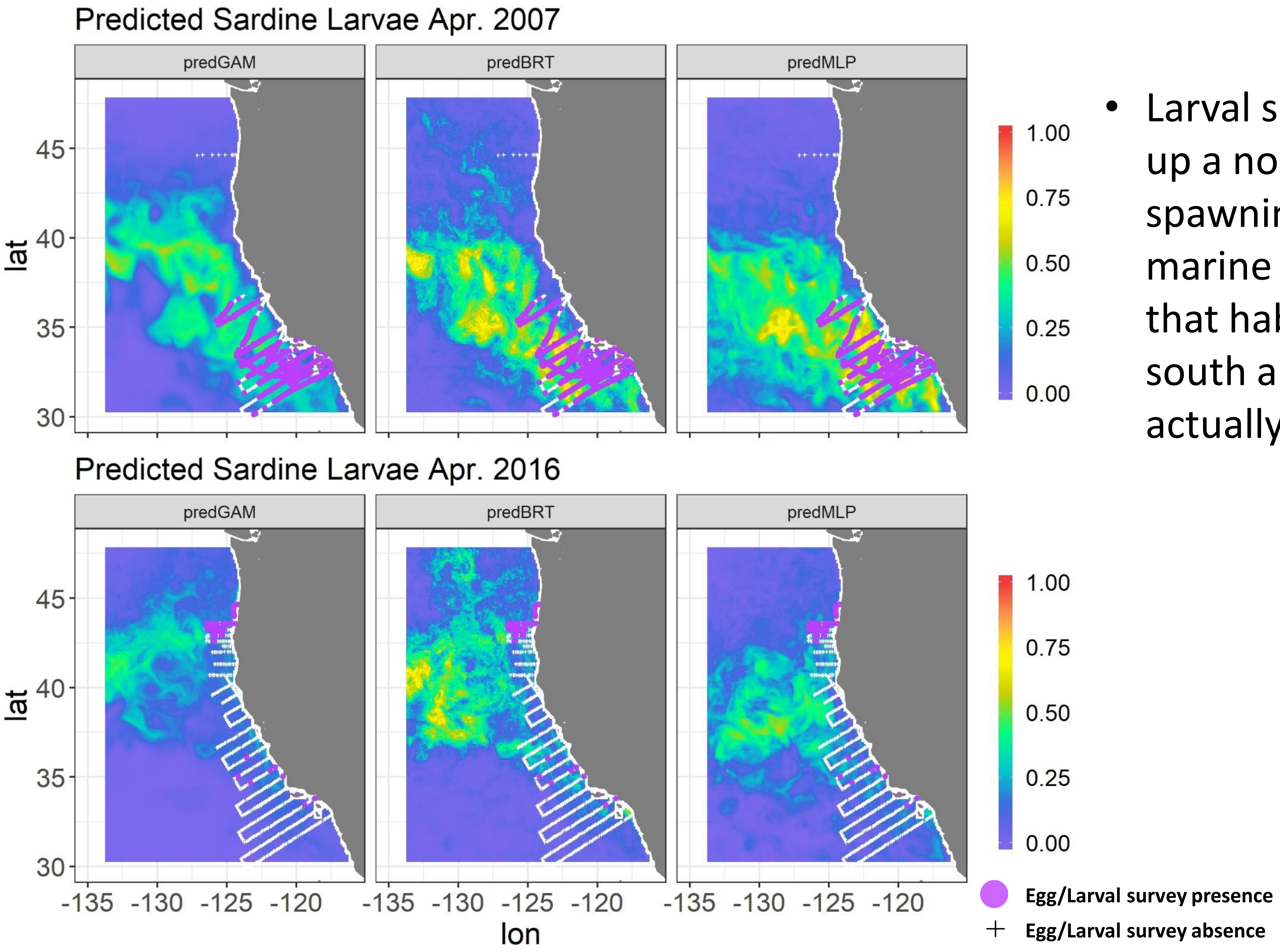
2002 - 2016



- But models do much better when historical relationships are more linear, and more data available near limits



- Although AUCs low for sardine SDM during marine heatwave years, they did capture the general movement north
- So SDMs may still be useful for picking up general trends, even if they lose skill under novel conditions



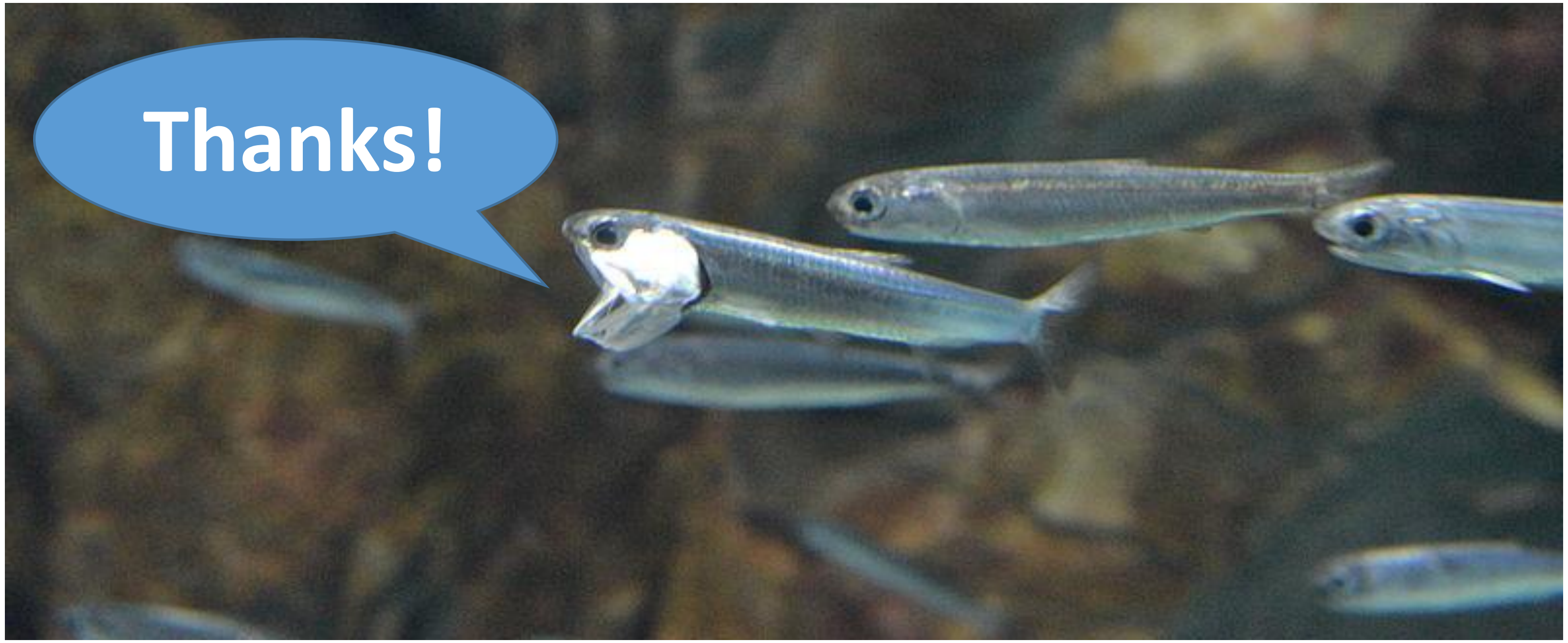
- Larval sardine SDMs also picked up a northward movement of spawning activity during the marine heatwave, but predicted that habitat would be further south and offshore than it actually was

Conclusions

- All SDMs lost a lot of skill when extrapolated to new environmental conditions
  - But some still picked up useful trends
- GAMs and BRTs generally did better than MLPs
- Larval models (especially anchovy) did better than adult models
  - Perhaps due to stronger and more linear associations with temperature
  - And higher number of observations near distribution limits

Next steps

- Random Forests?
- Downsampling to reduce zero-inflation
- Compare to simple SST niche model? Or to hybrid correlative-mechanistic models?
- Other species?
- Future projections!
- Suggestions and comments are welcome



www.future-seas.com