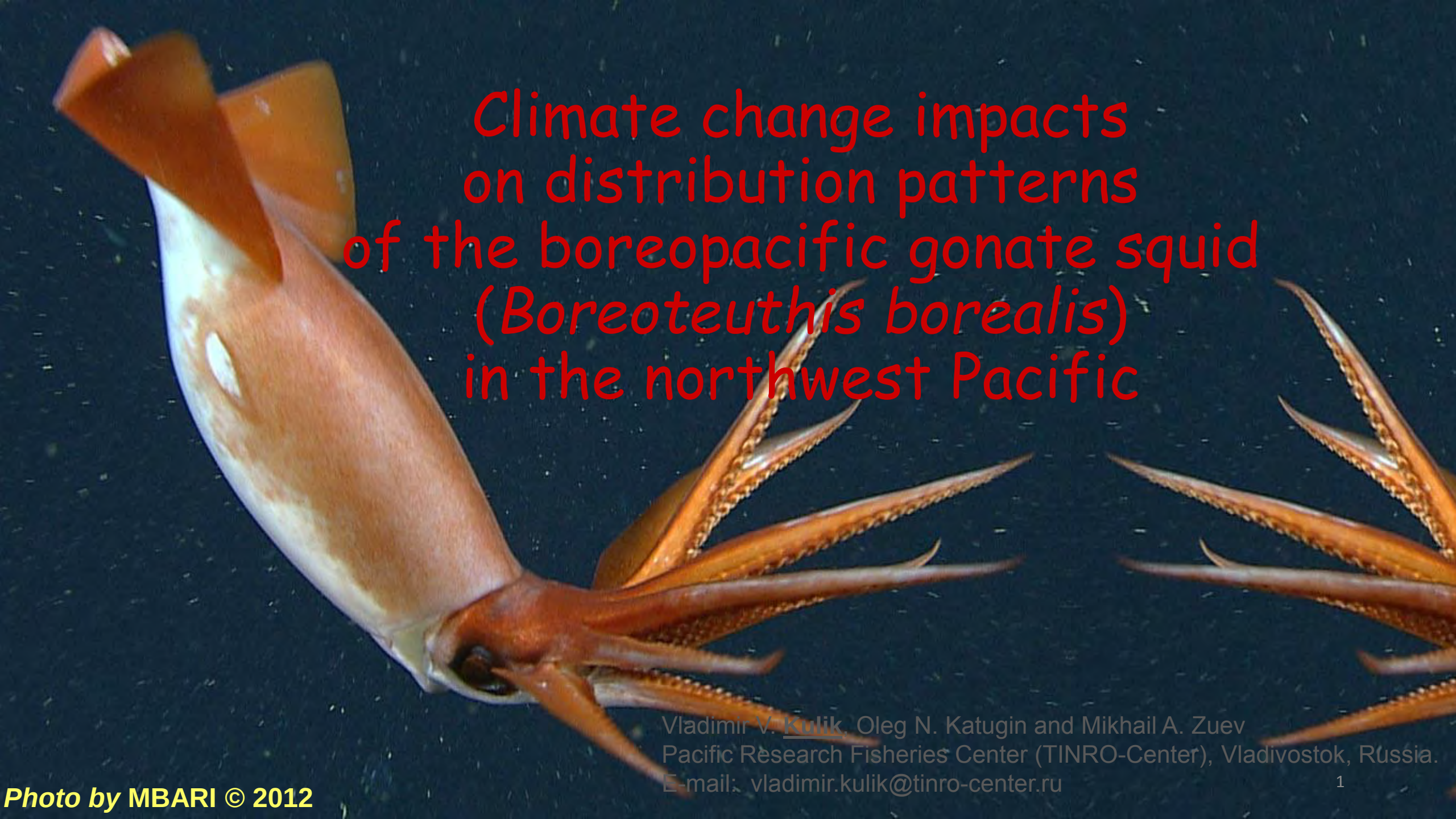


Climate change impacts on distribution patterns of the boreopacific gonate squid (*Boreoteuthis borealis*) in the northwest Pacific

Vladimir V. Kulik, Oleg N. Katugin and Mikhail A. Zuev
Pacific Research Fisheries Center (TINRO-Center), Vladivostok, Russia.
E-mail: vladimir.kulik@tinro-center.ru

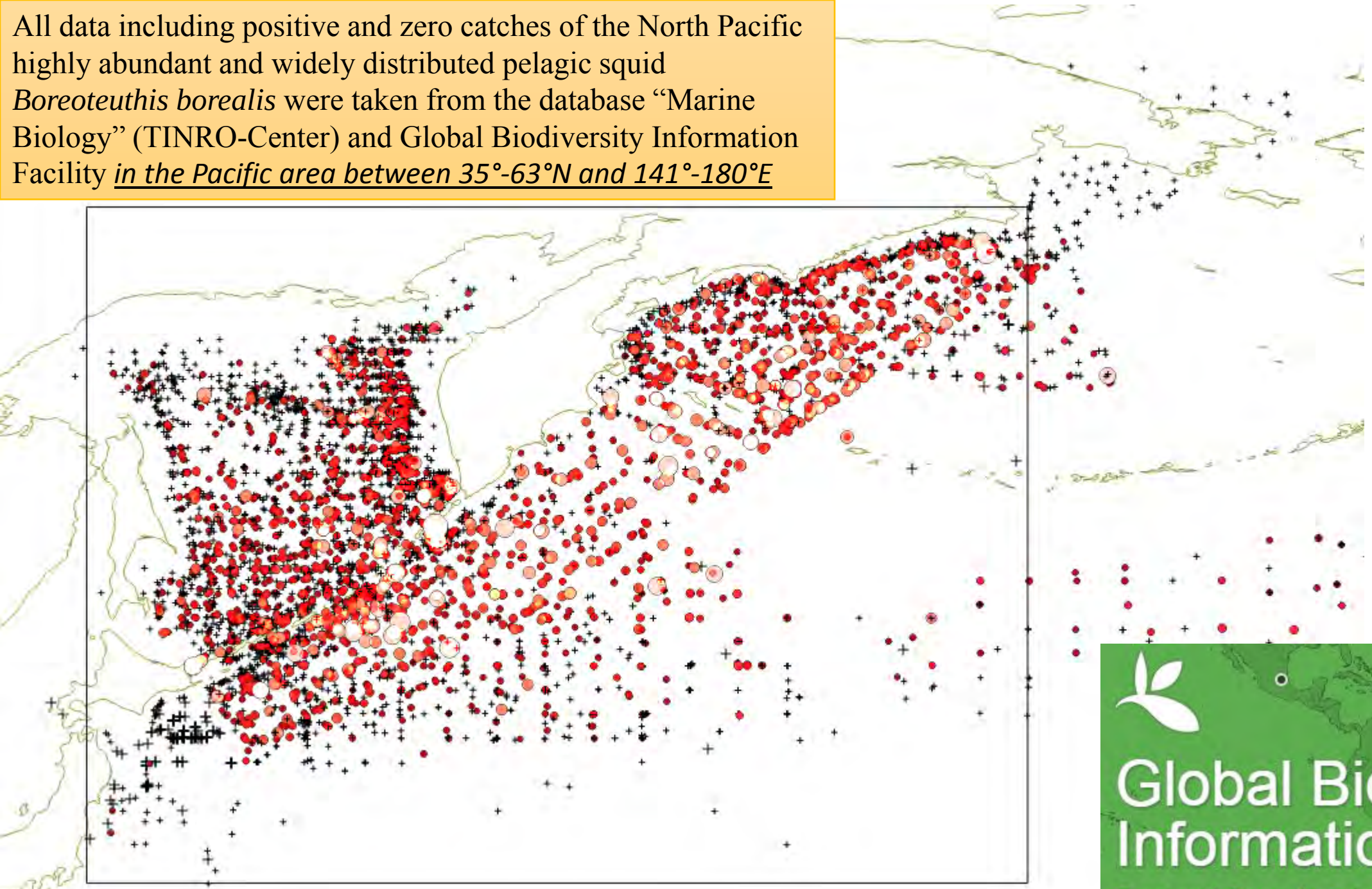
Boreoteuthis borealis (=Gonatopsis borealis)

FAO Names

En - Boreopacific gonate squid, Fr - Encornet boréopacifique, Sp - Gonalura pacificoboreal.
3Alpha Code: GTE Taxonomic Code: 3211401001

1) ***highly abundant*** and ***widely distributed*** pelagic squid in the North Pacific with the exception of the Japan/East Sea (JES), where it was found only once;
2) ***high rate of occurrence*** throughout the geographic range and ***high biomass*** make this squid (known also for its intensive vertical migrations from the upper layers down to the meso- and bathypelagic zone) very important in ***vertical energy transfer***, similar to the mesopelagic fish, which also migrate to the upper pelagic layers at nighttime;
3) expected ***susceptibility to changes*** in oceanic pelagic environment associated with climate variability and change…

All data including positive and zero catches of the North Pacific highly abundant and widely distributed pelagic squid *Boreoteuthis borealis* were taken from the database “Marine Biology” (TINRO-Center) and Global Biodiversity Information Facility *in the Pacific area between 35°-63°N and 141°-180°E*

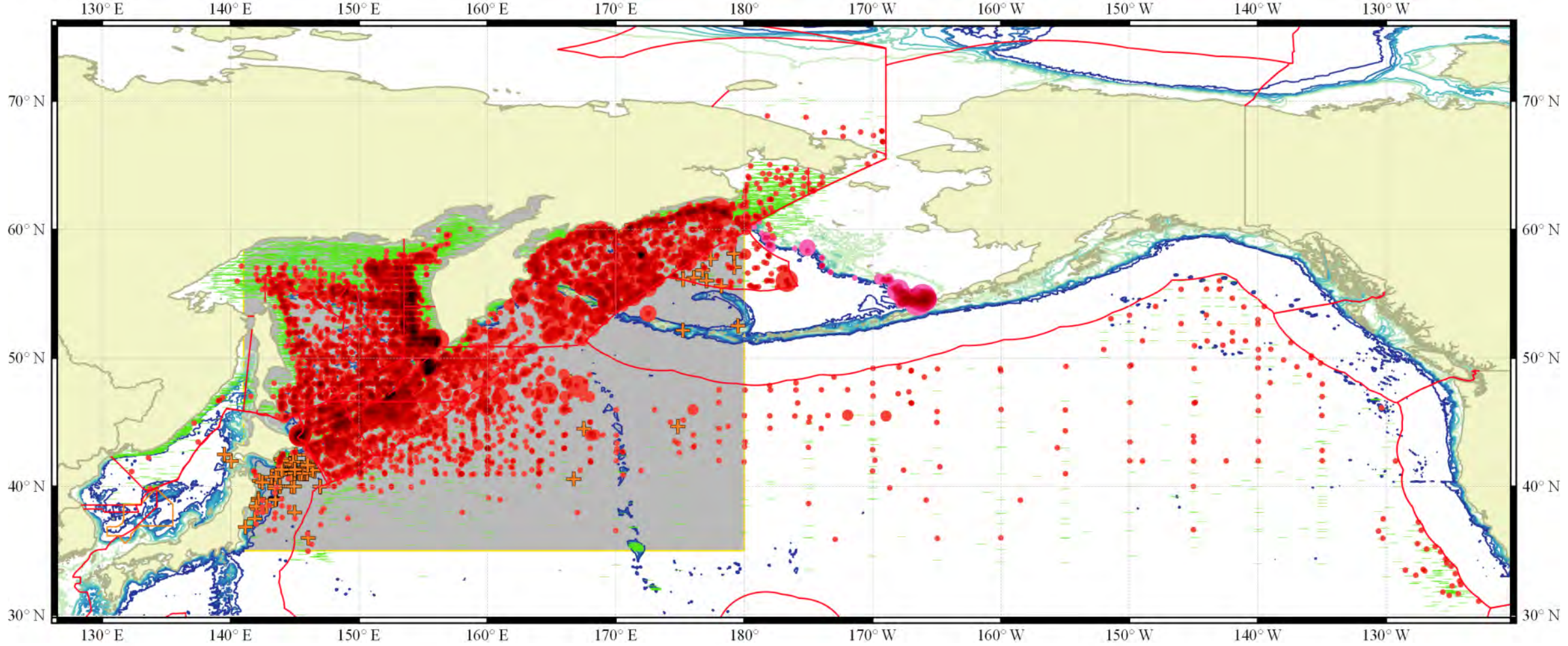


Zero catches are not shown here!

Positive ONLY

Global Biodiversity Information Facility

Free and Open Access to Biodiversity Data



All data we have and
Selected region in grey color.
Zeros (“-”) only from those trips where squid really occurred!

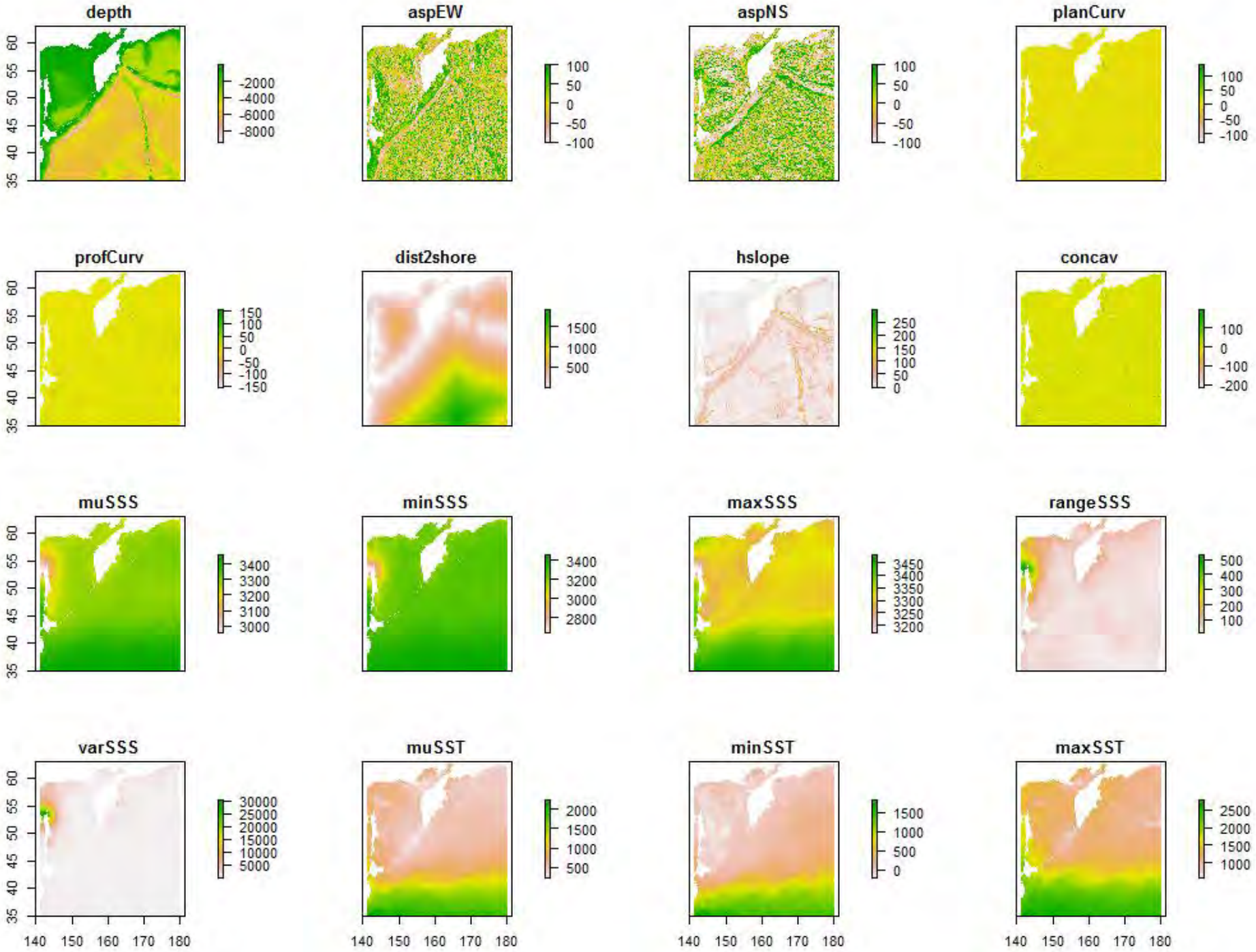
Major objectives:

Analyze long-term wide-scale database for ***Boreoteuthis borealis***, and estimate possible changes in the squid distribution patterns with respect to climate change projections using species distribution model approach…

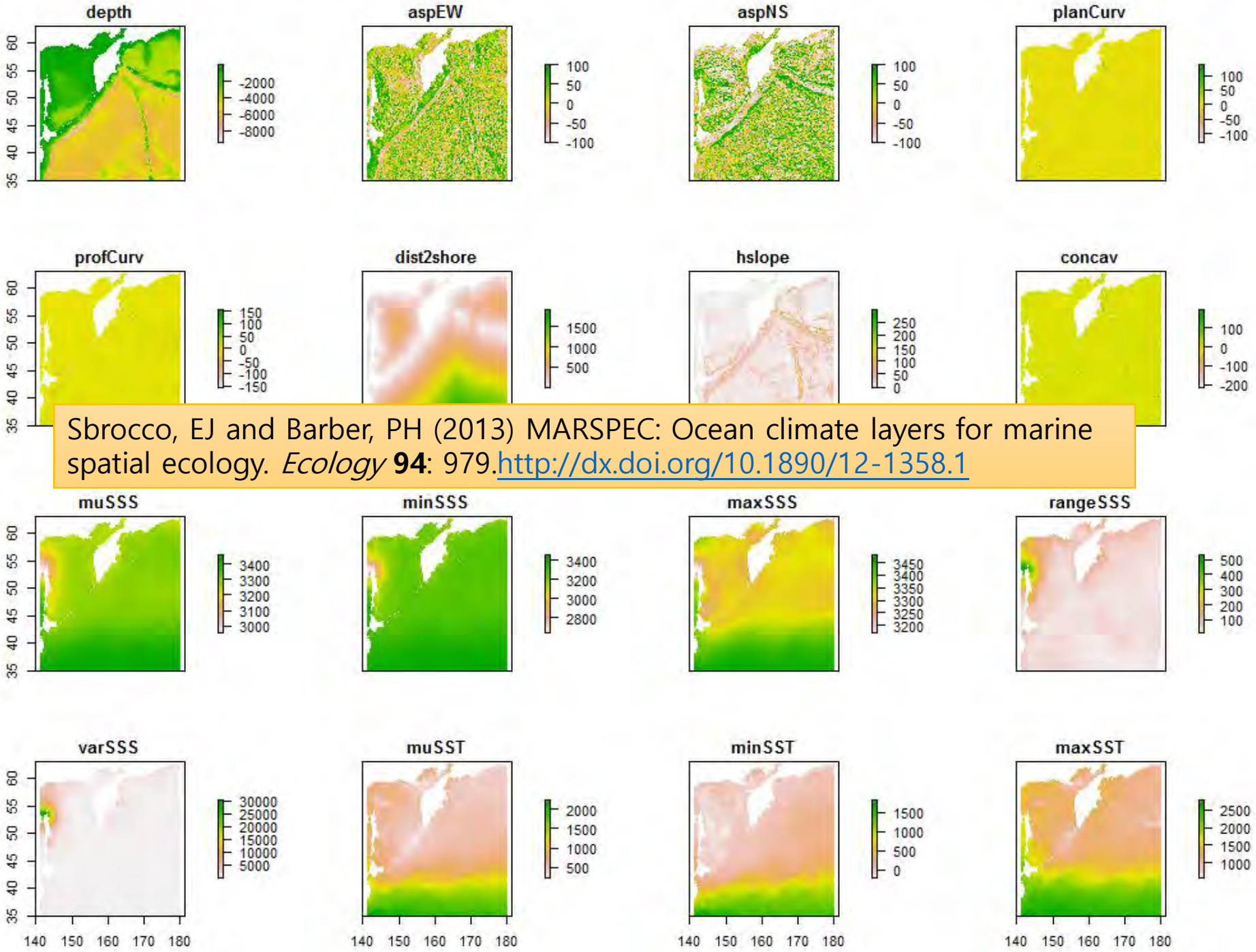
Estimate model performance using 5 different methods:

- Maximum Entropy (**MaxEnt**),
- Generalized Additive Models (**GAM**),
- **Random Forest**,
- Support Vector Machines (**SVM**), and
- Bayesian SDMs using Gaussian processes (**GRaF**).

Make predictions on the basis of future values of sea surface temperature (SST) and salinity (SSS) obtained from **A1B, A2, B1** Intergovernmental Panel on Climate Change scenarios for the late 21st century available in “Bio-ORACLE” database.



The first group of SDMs utilized 10 bioclimatic variables (annual mean, range, variance, and extreme values for temperature and salinity) and geophysical variables (bathymetry and its complexity: components of aspect, slope, concavity of the seafloor, and plan and profile curvature, as well as distance from shore) obtained from "**MARSPEC**" database.

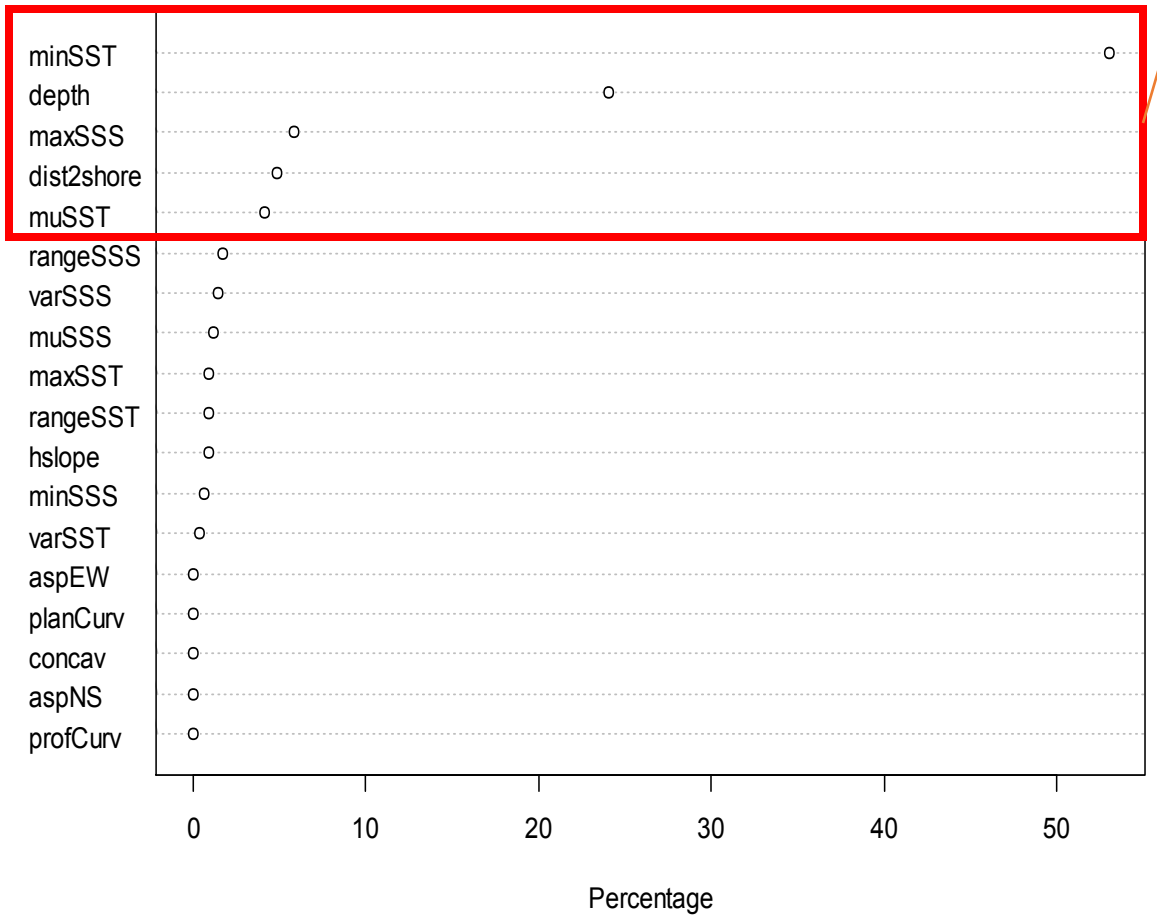


Sbrocco, EJ and Barber, PH (2013) MARSPEC: Ocean climate layers for marine spatial ecology. *Ecology* **94**: 979.http://dx.doi.org/10.1890/12-1358.1

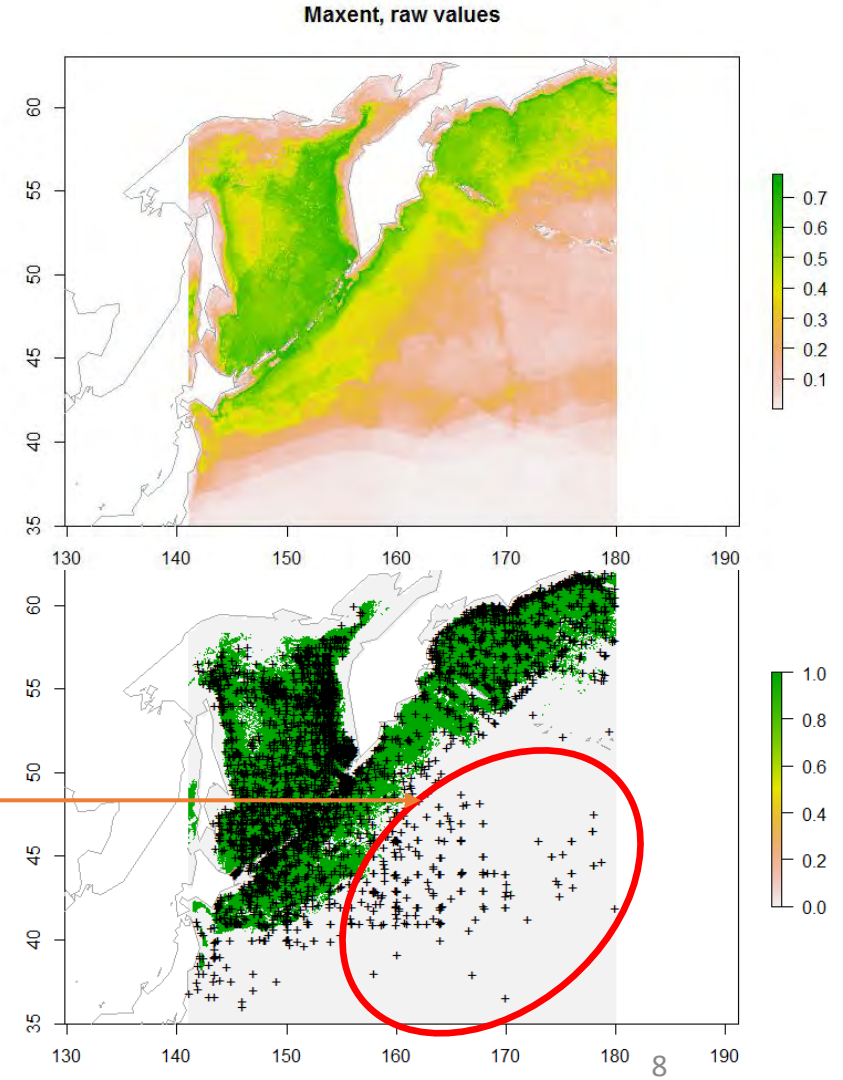
The first group of SDMs utilized 10 bioclimatic variables (annual mean, range, variance, and extreme values for temperature and salinity) and geophysical variables (bathymetry and its complexity: components of aspect, slope, concavity of the seafloor, and plan and profile curvature, as well as distance from shore) obtained from "**MARSPEC**" database.

MaxEnt

Maximum Entropy (Phillips et al., 2006) is the most widely used SDM algorithm. Elith et al. (2010) provided an explanation of the algorithm (and software) geared towards ecologists.



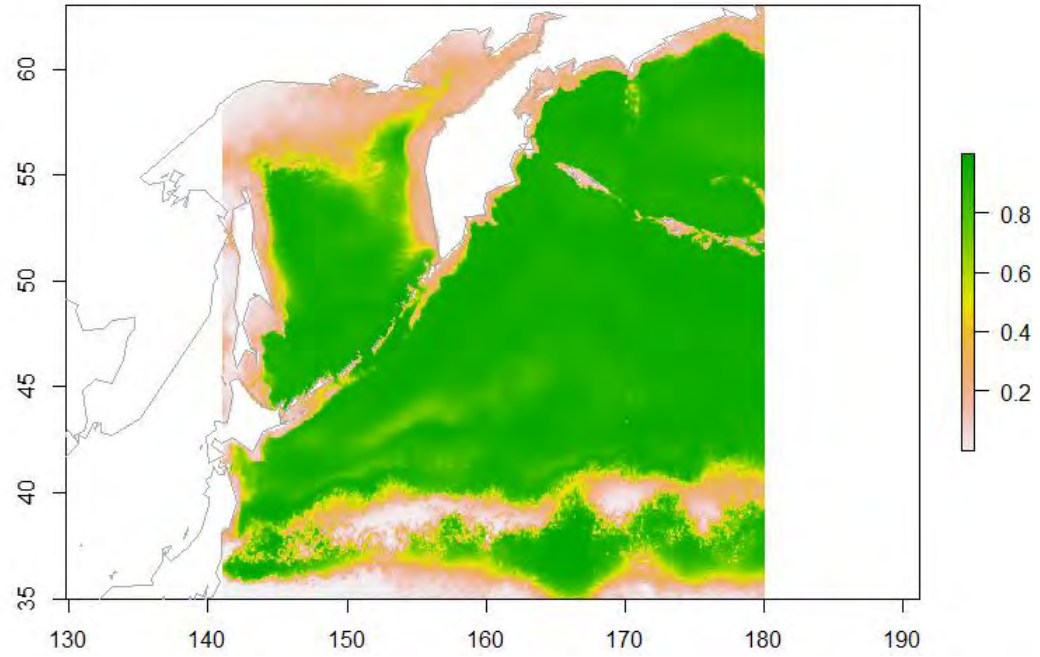
These variables were used in GAM and SVM

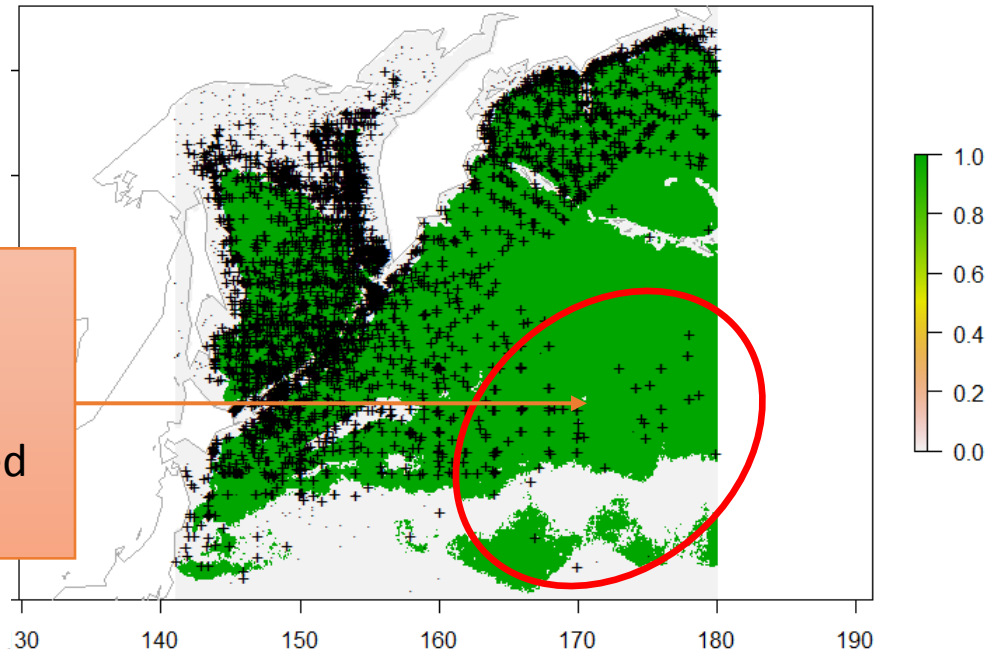


These out of sample (test) occurrences were modeled very poorly

GAM

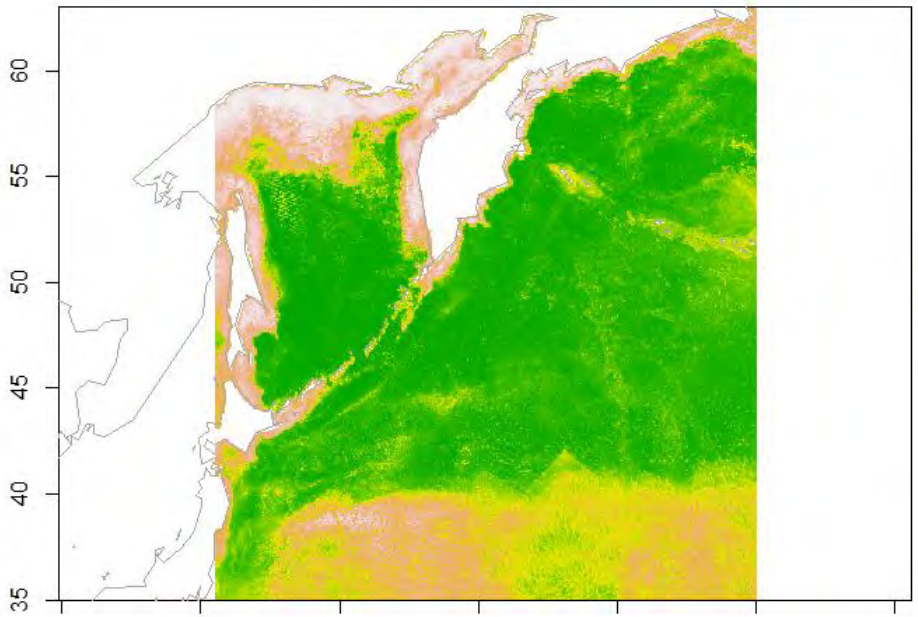
```
Family: binomial
Link function: logit

Formula:
pa ~ s(minSST) + s(depth) + s(maxSSS) + s(dist2shore) + s(muSST)

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.23152    0.04267   28.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                edf Ref.df Chi.sq  p-value
s(minSST)     8.998  9.000  30.45 0.000364 ***
s(depth)      7.082  8.128 918.96  < 2e-16 ***
s(maxSSS)     7.702  8.557  53.43 2.21e-08 ***
s(dist2shore) 8.088  8.731  37.77 2.33e-05 ***
s(muSST)      7.204  8.147  40.83 2.39e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘

R-sq.(adj) =  0.351   Deviance explained = 30.6%
UBRE = -0.13242  Scale est. = 1         n = 5958

n presences    : 1379
n absences     : 606
```



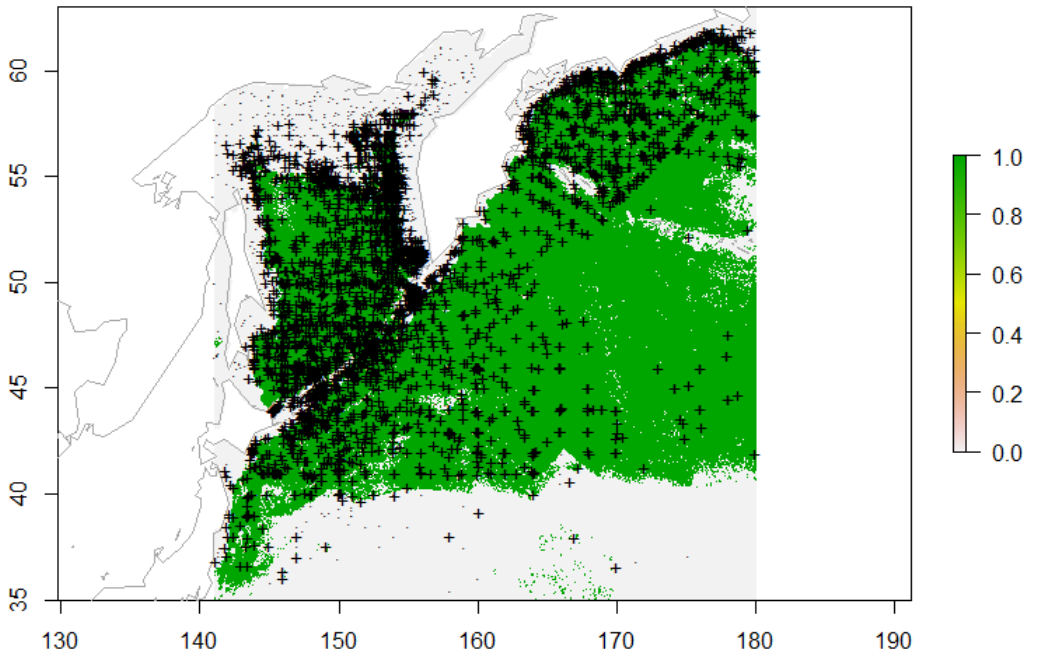


These out of sample (test) occurrences were modeled much better

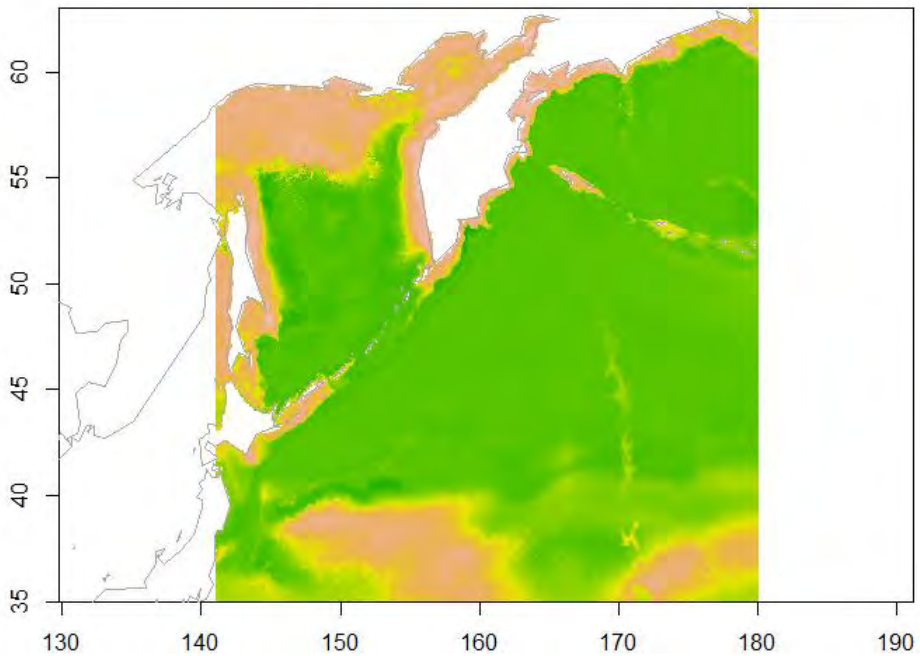
Random Forest, regression

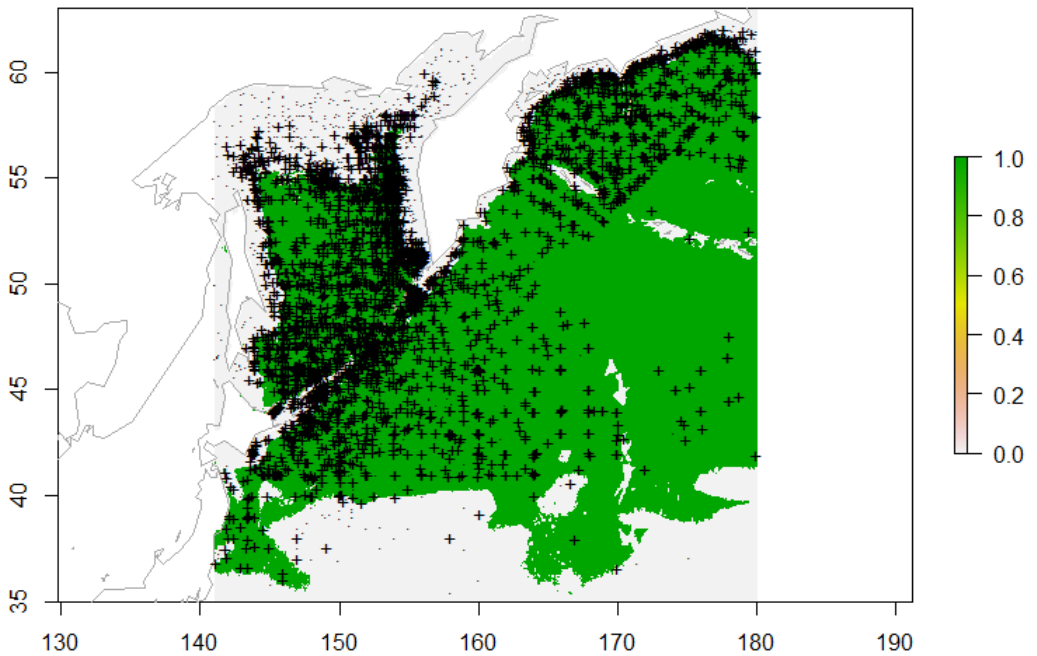


presence/absence

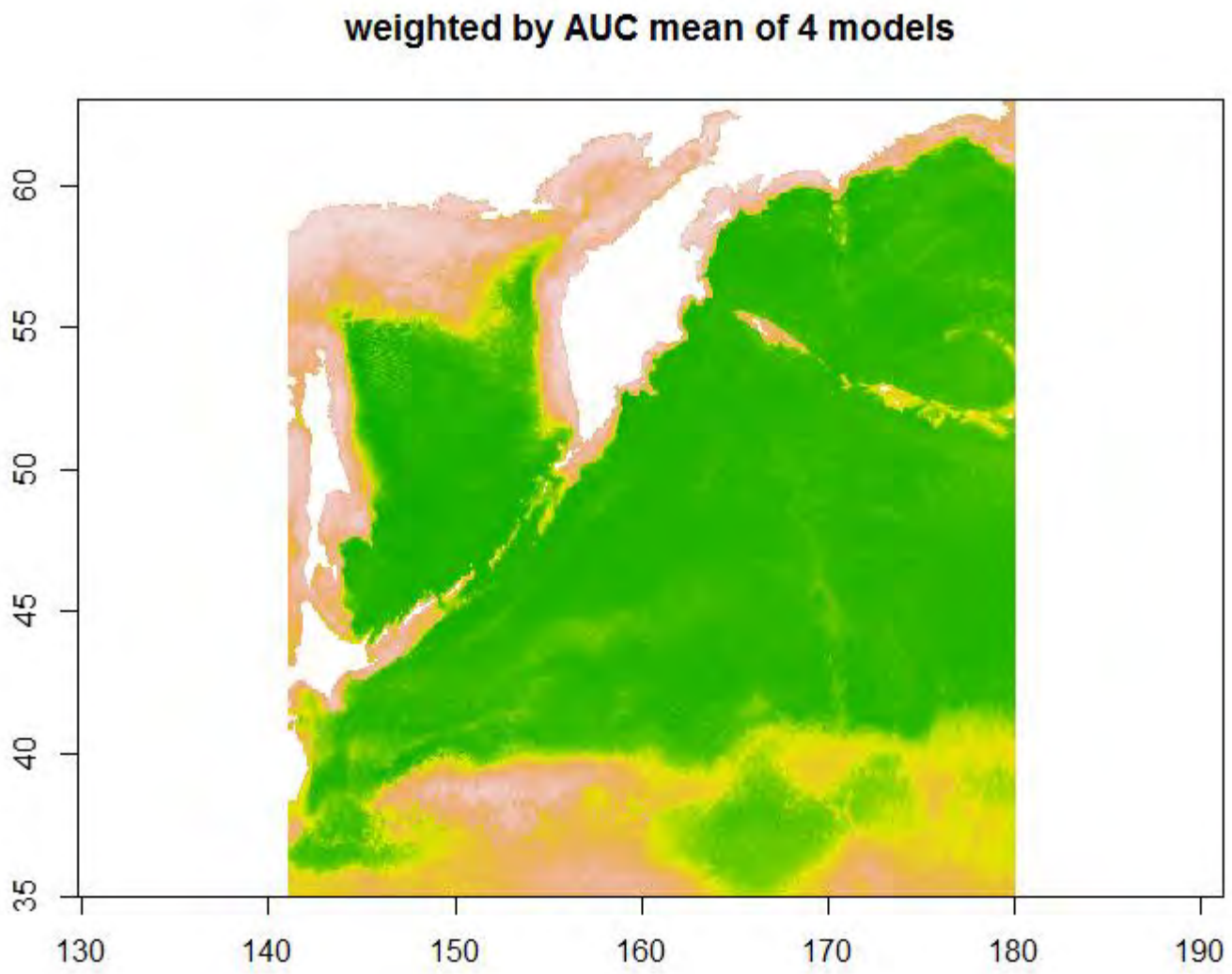


Support Vector Machine

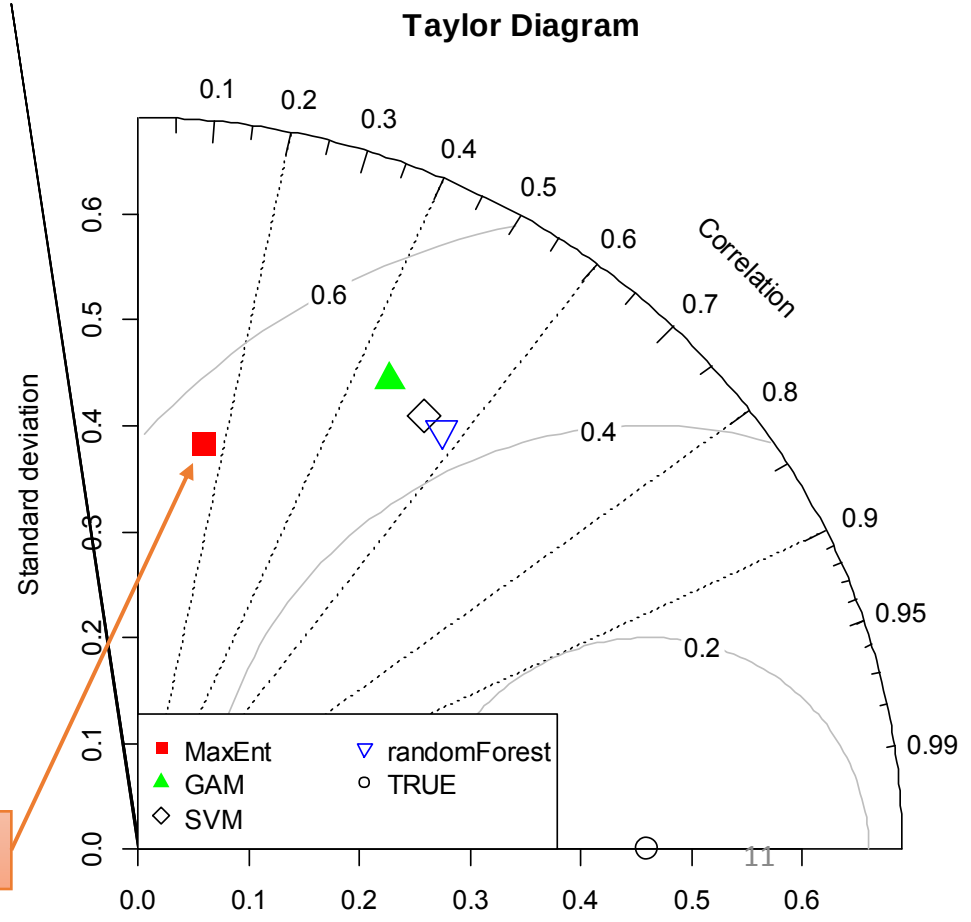




Mean distribution from these four SDMs, weighted by AUC, appears close to known distribution patterns of the squid (***Boreoteuthis borealis***)

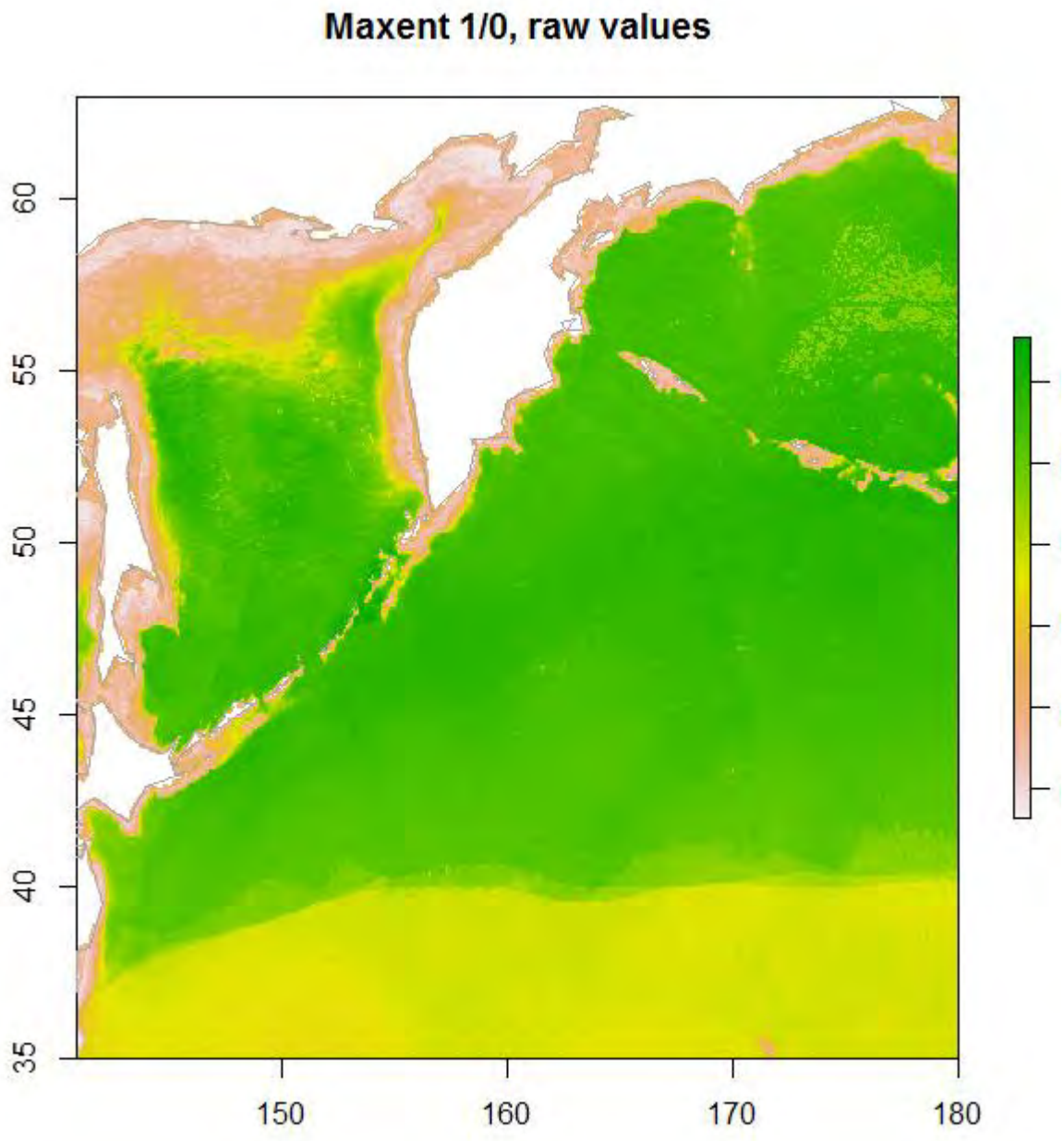


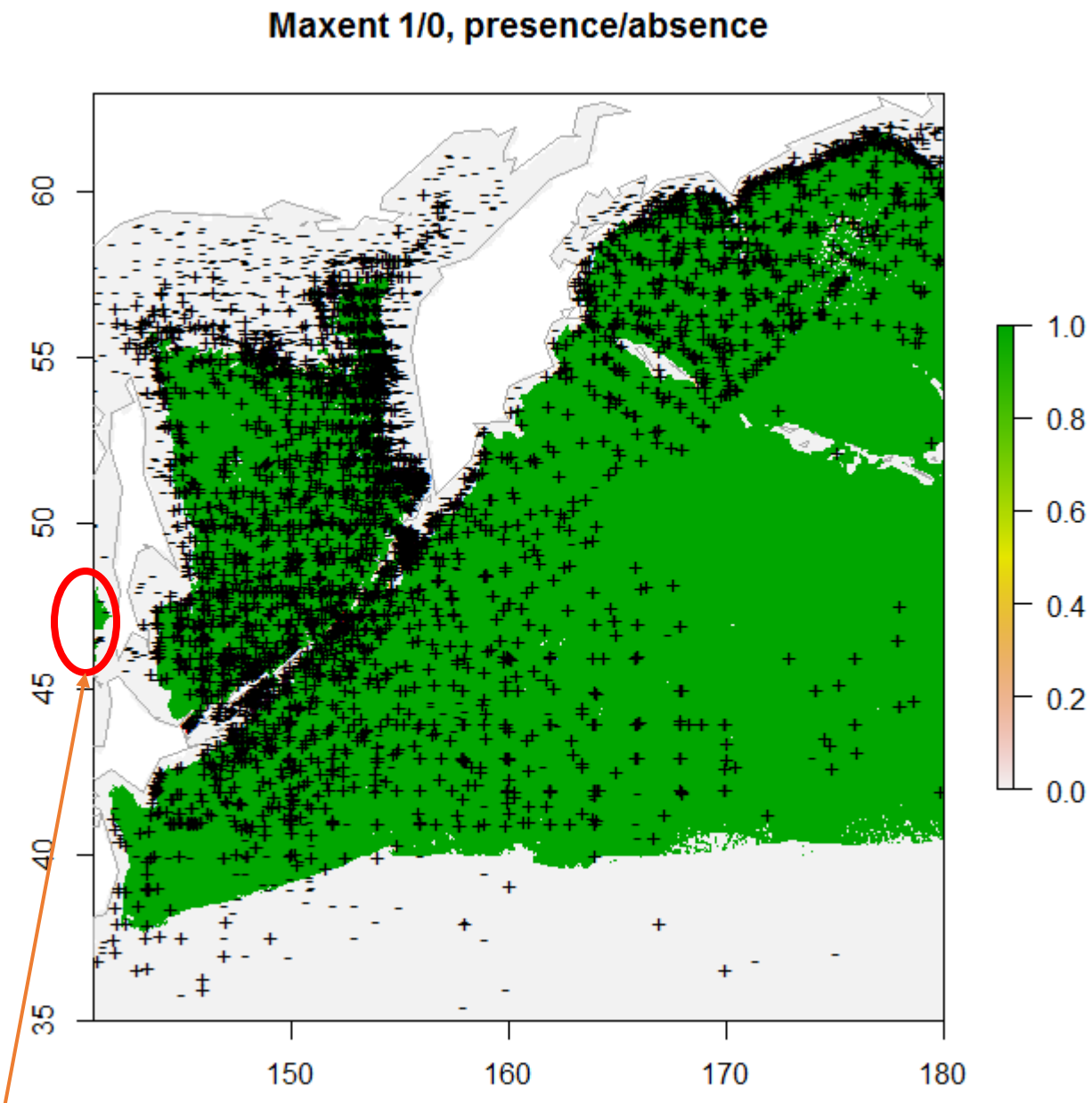
model	*AUC*	*cor*	*max TPR+TNR*
GAM	0.839	0.554	0.833
SVMs	0.831	0.574	0.794
randomForest	0.862	0.619	0.687



Could we do better than that?

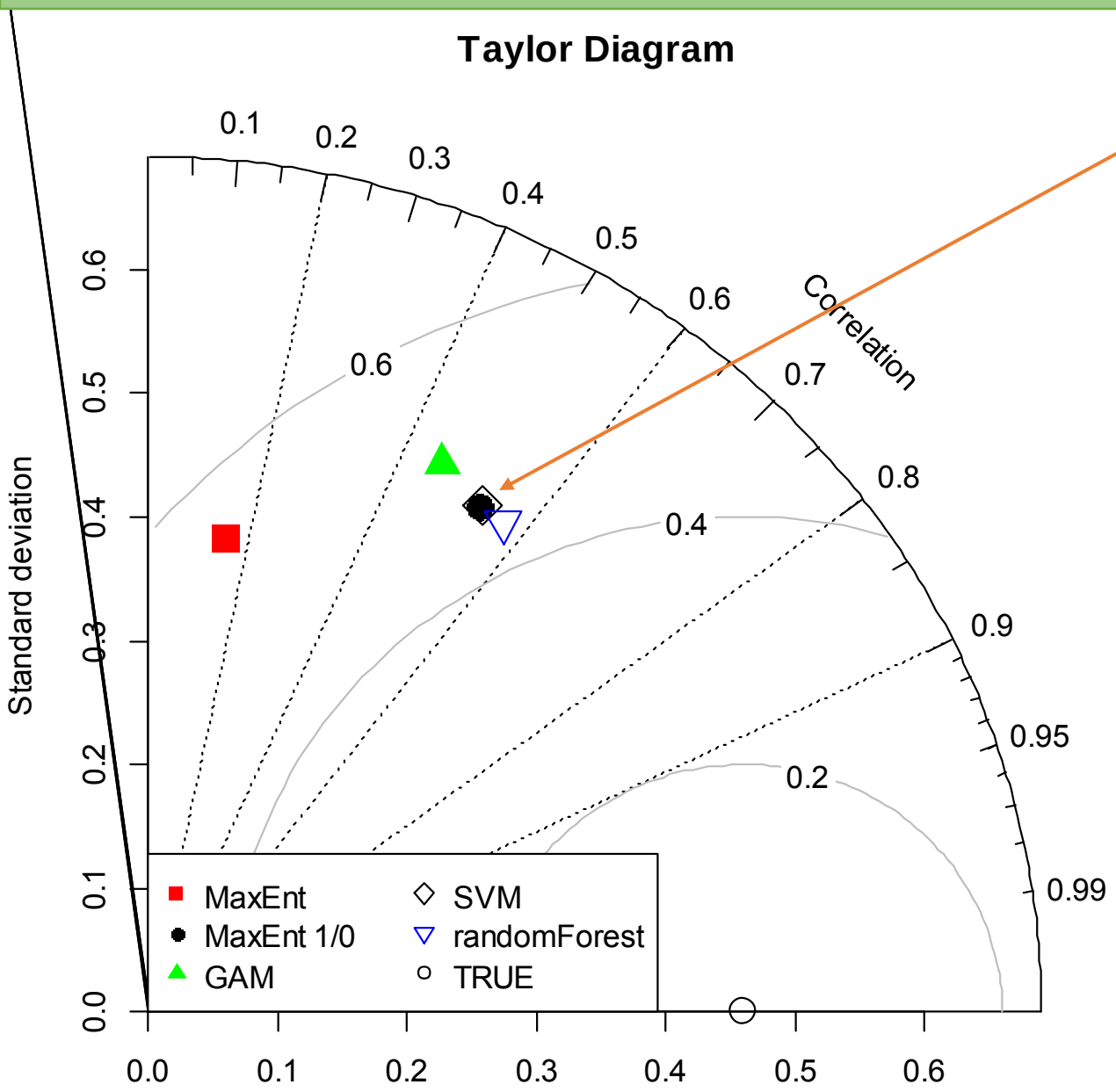
Yes, we could. When we provided MaxEnt the exact locations of absences '-', we got much better results.

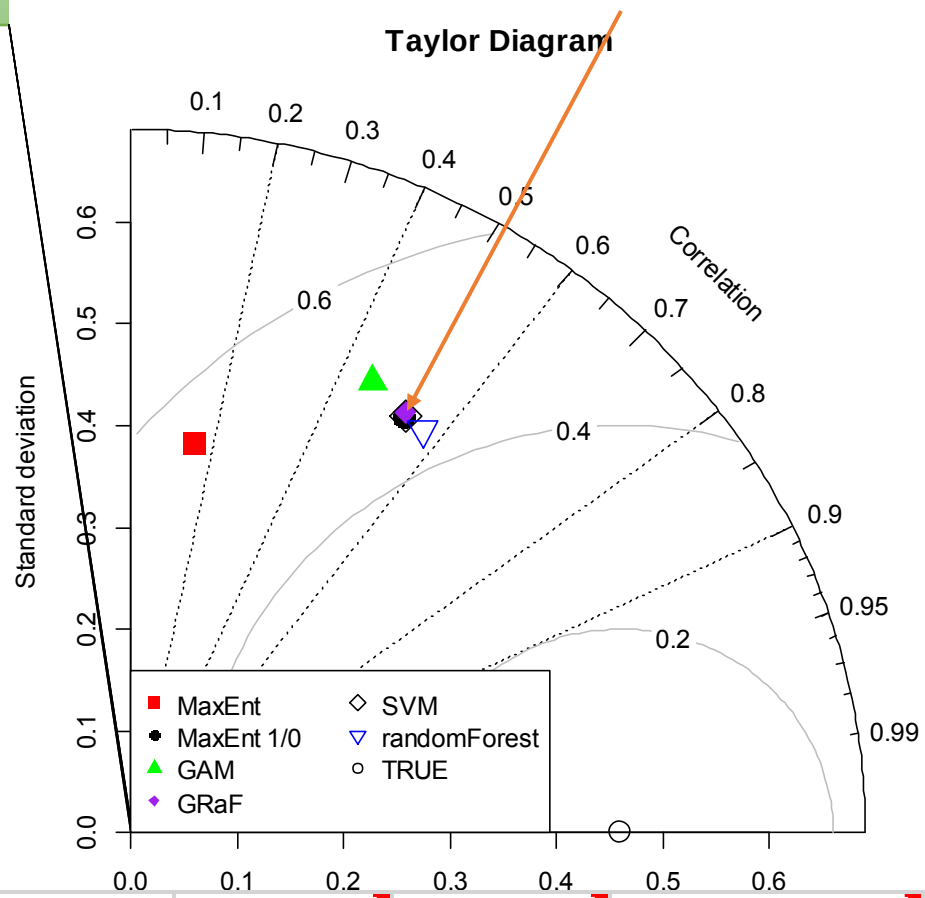




False positives in the JES! GAM, SVM and RandomForest didn't make such errors.

And new results from MaxEnt 1/0 were statistically as good as from SVM

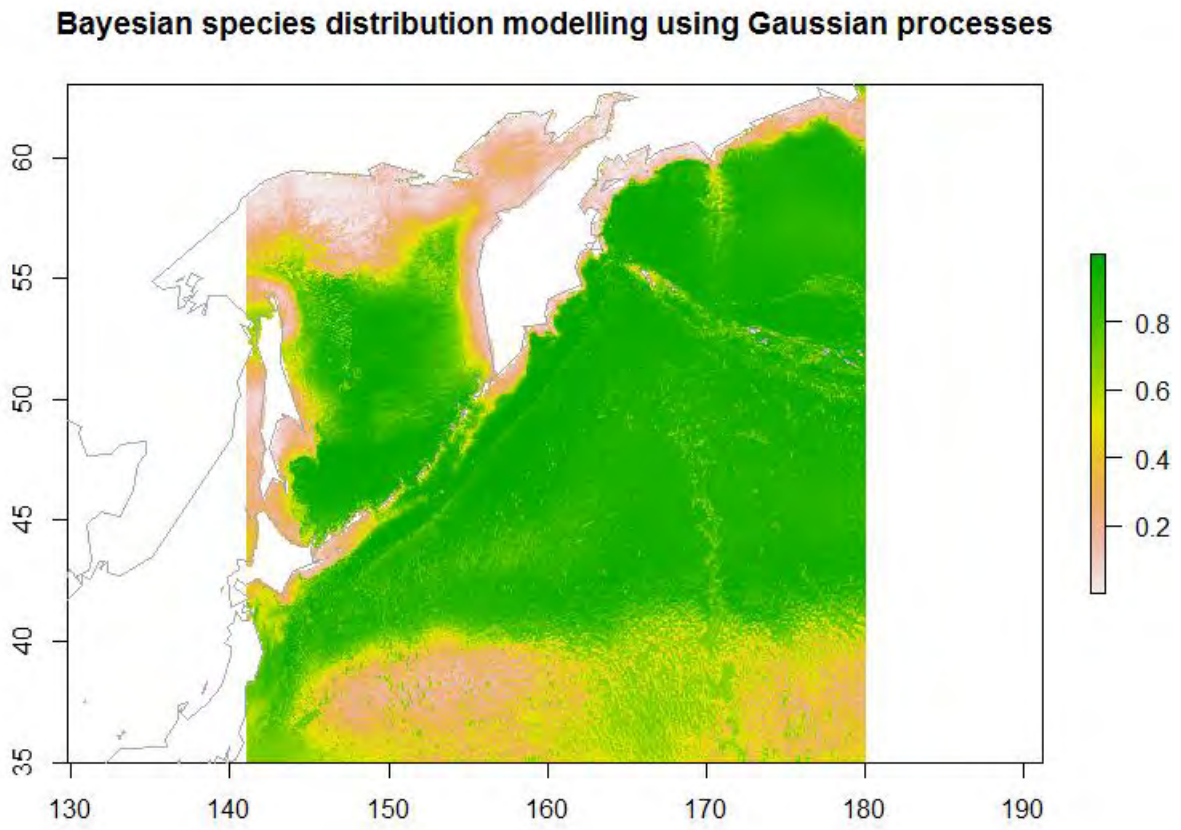


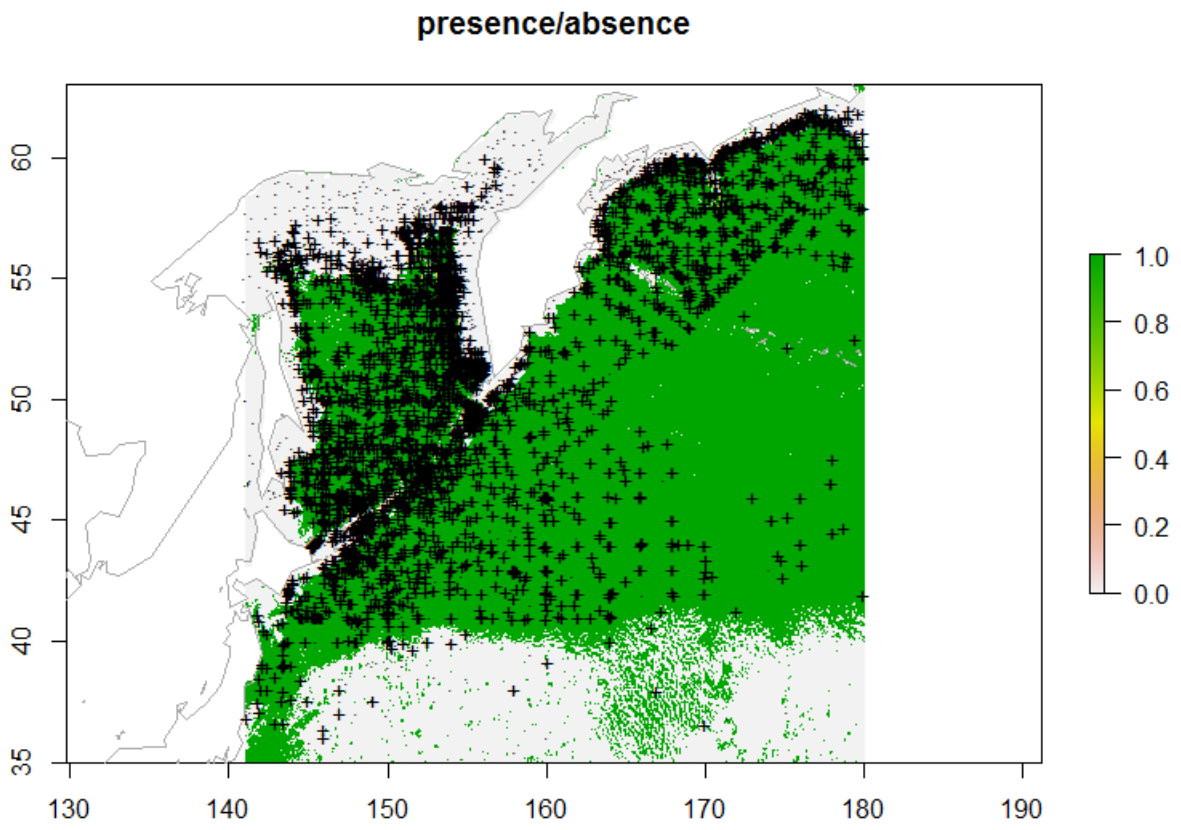


model	*AUC*	*cor*	*max TPR+TNR*
GAM	0.839	0.554	0.833
SVMs	0.831	0.574	0.794
randomForest	0.862	0.619	0.687
Bayes GRaF	0.815	0.527	0.653
MaxEnt	0.528	0.104	0.356
MaxEnt 1/0	0.845	0.575	0.454

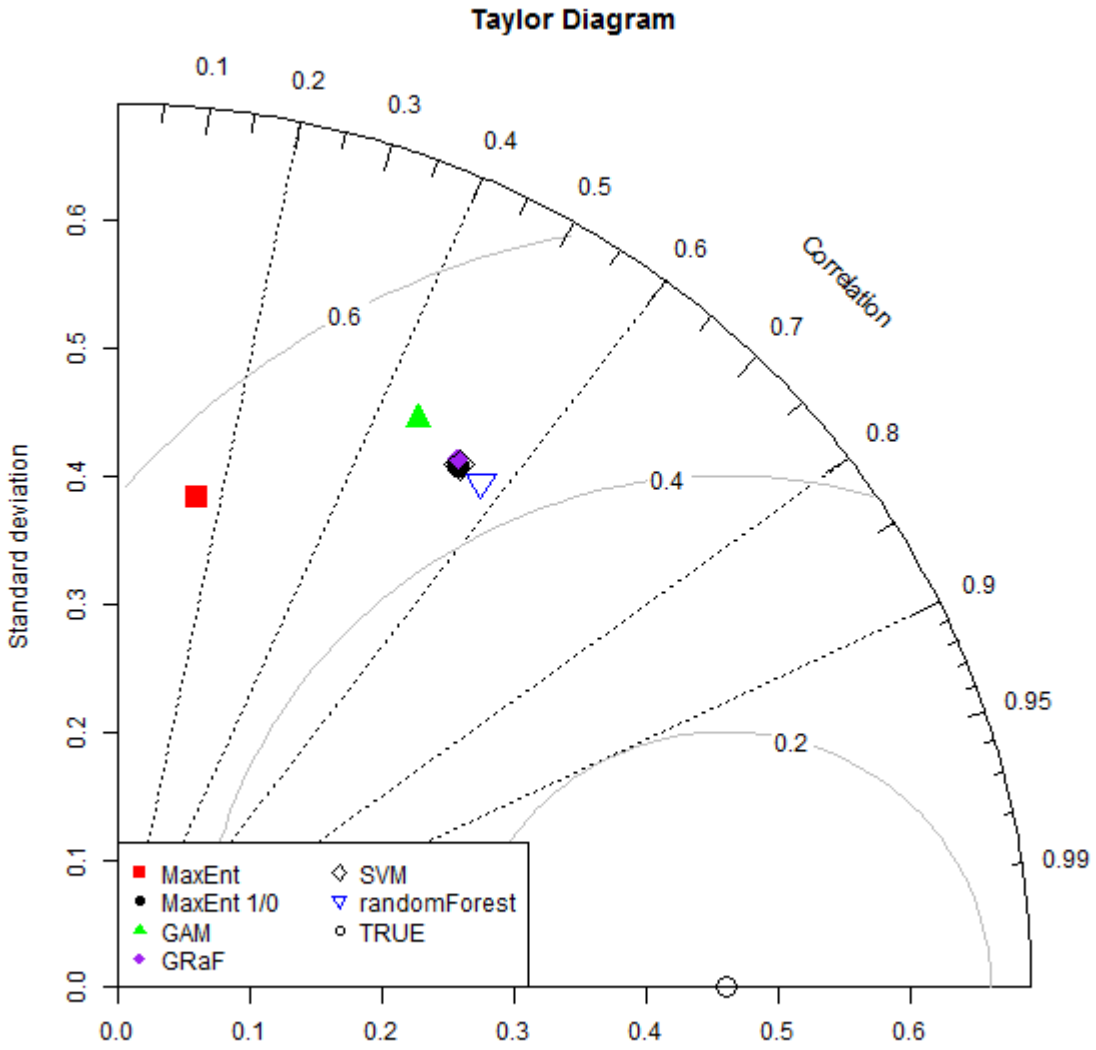
Golding & Purse (2013) **GRaF**: Fast and Flexible Bayesian Species Distribution Modelling Using Gaussian Random Fields. Manuscript in preparation. (preprint as Chapter 5 in http://dx.doi.org/10.6084/m9.figshare.767289)

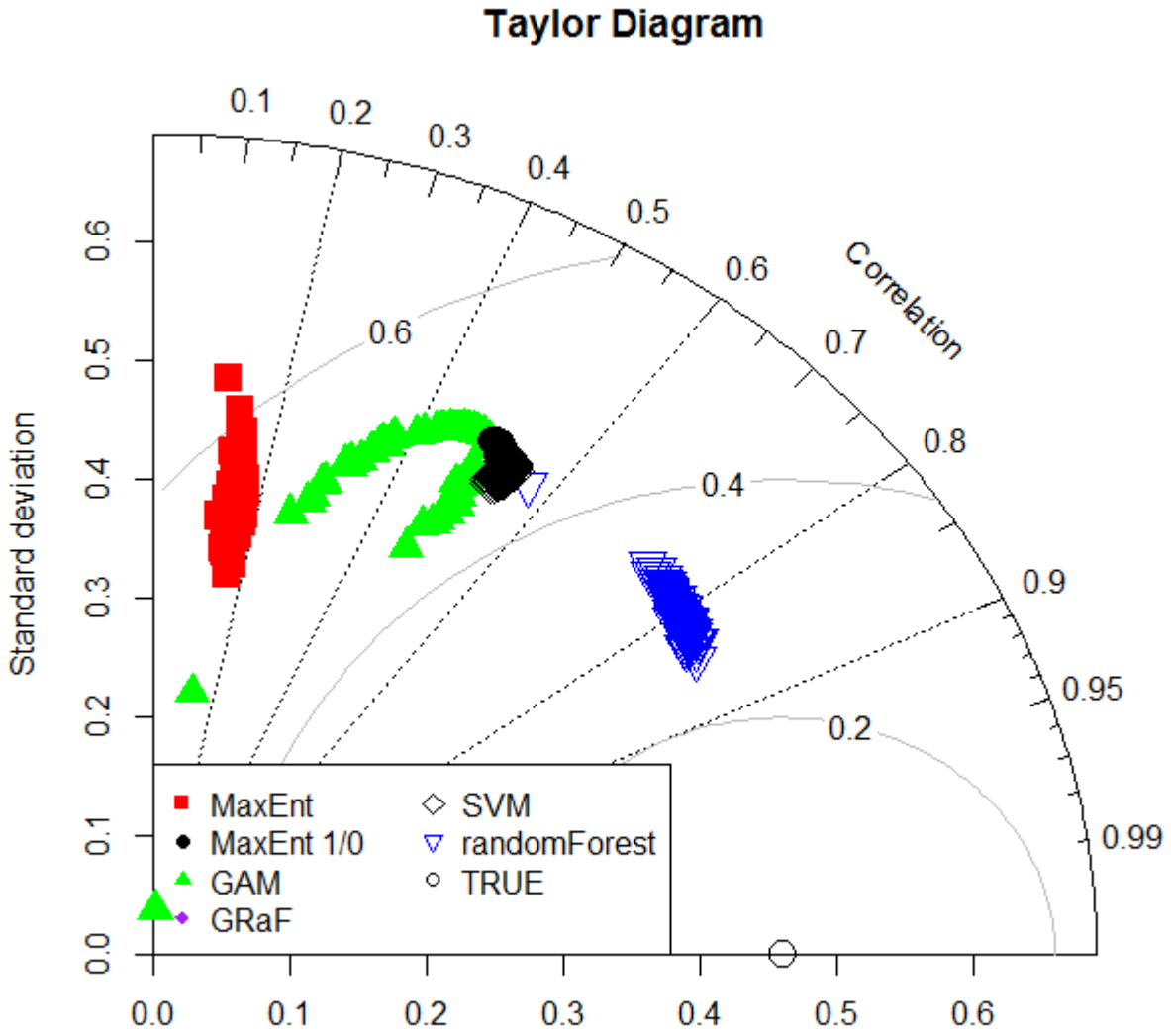
It is good but time consuming: it required more than 19 minutes for the only 1 run. Therefore we didn't try to estimate its sensitivity to subsamples.





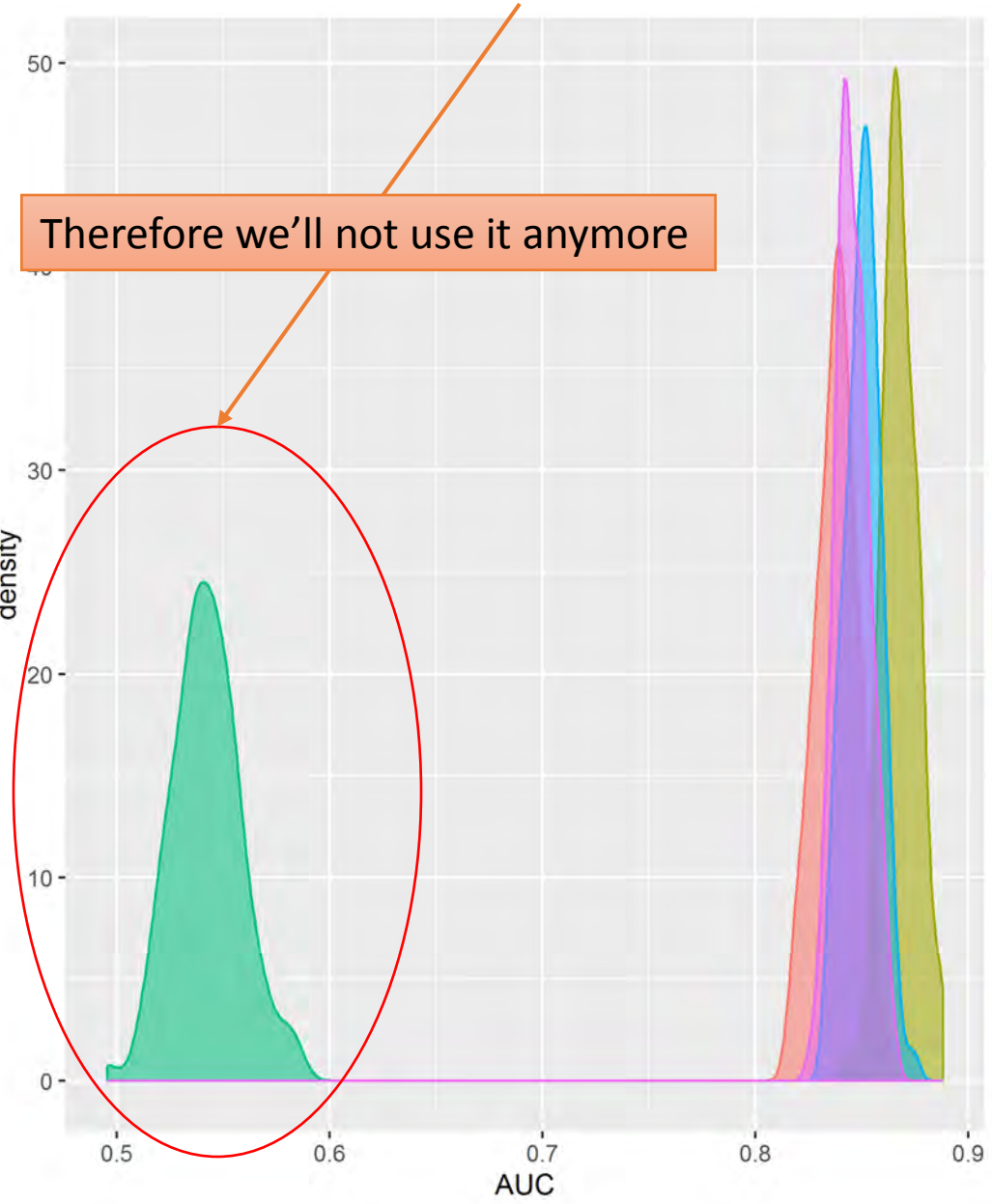
We repeated the process of random sampling (75%) and testing the results against out of sample data 100 times.
Thus we can say that GAM results were not stable and good, at the same time Random Forest were the best.
SVM and MaxEnt 1/0 had the most stable and comparable pefomance.

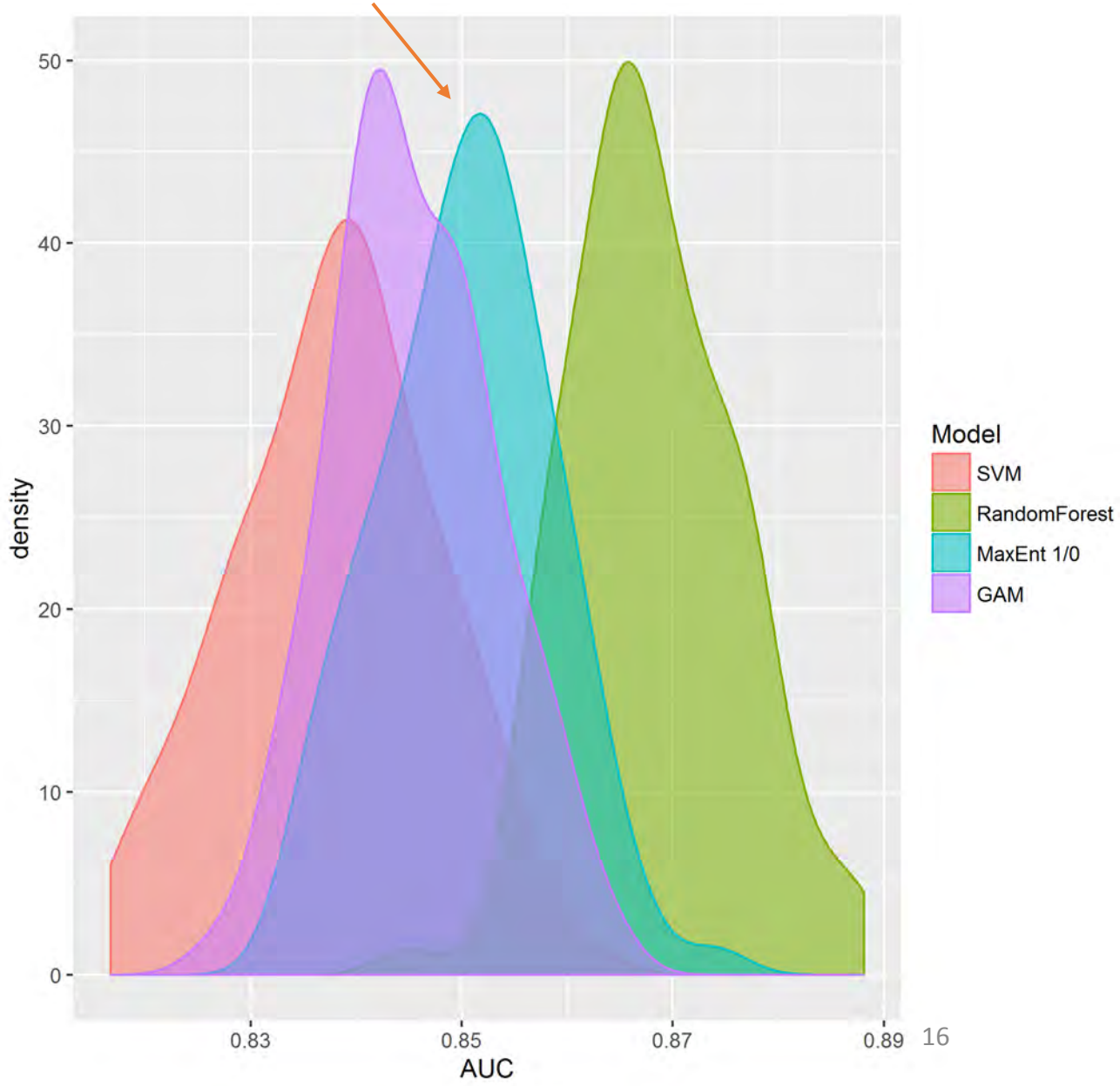




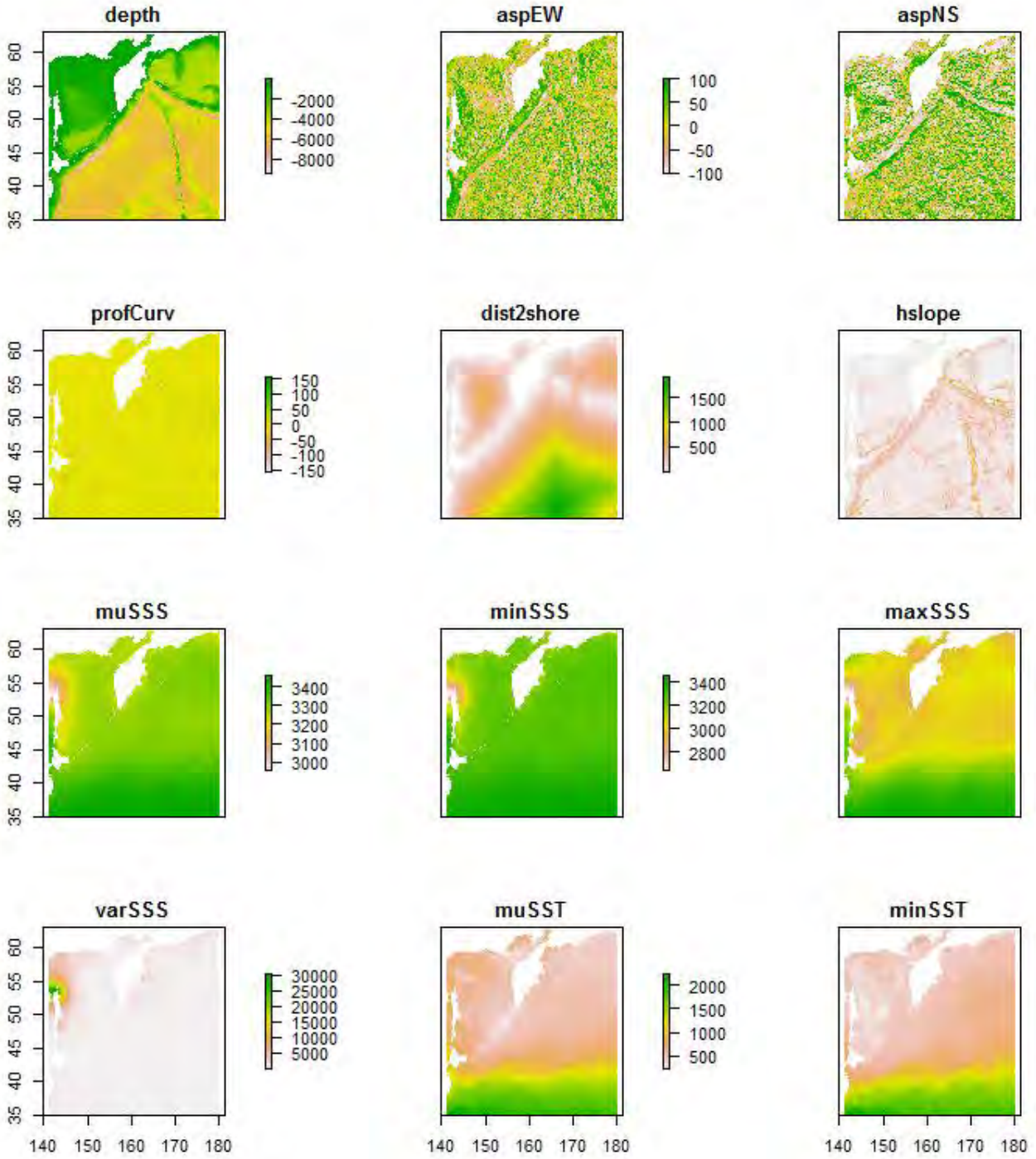
This is to show how important real absences are to get good results like below.

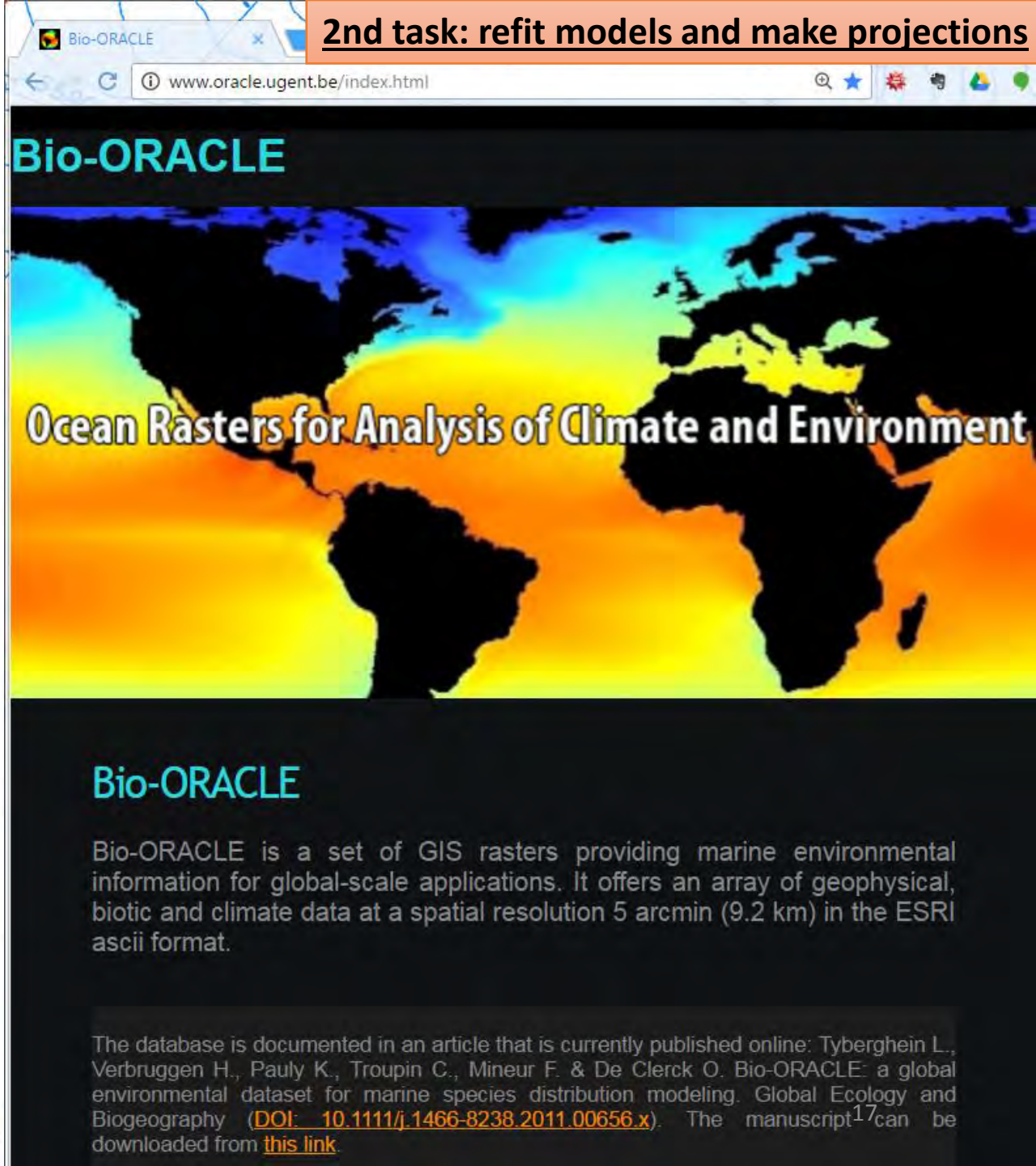
Therefore we'll not use it anymore





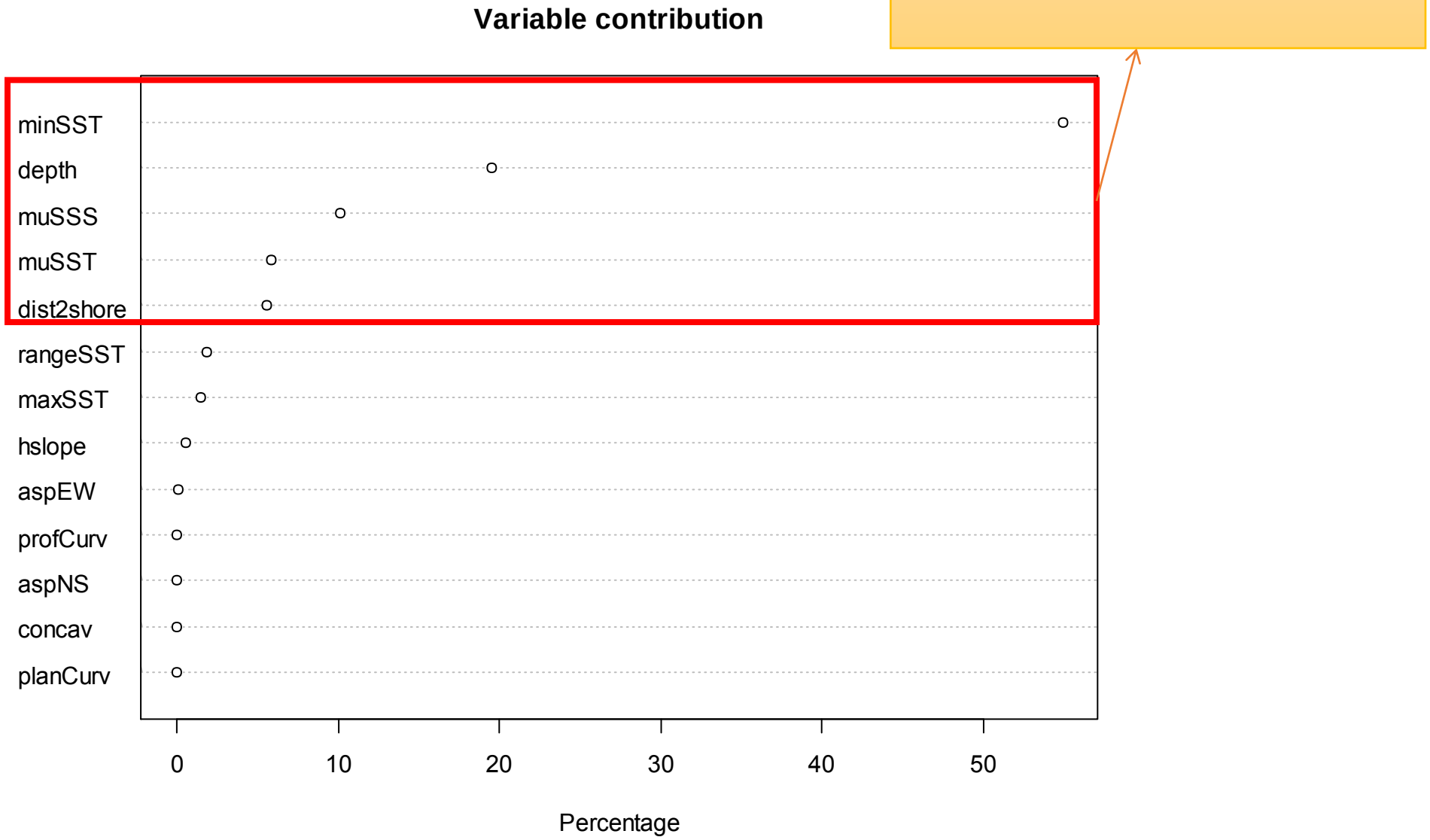
2nd task: refit models and make projections

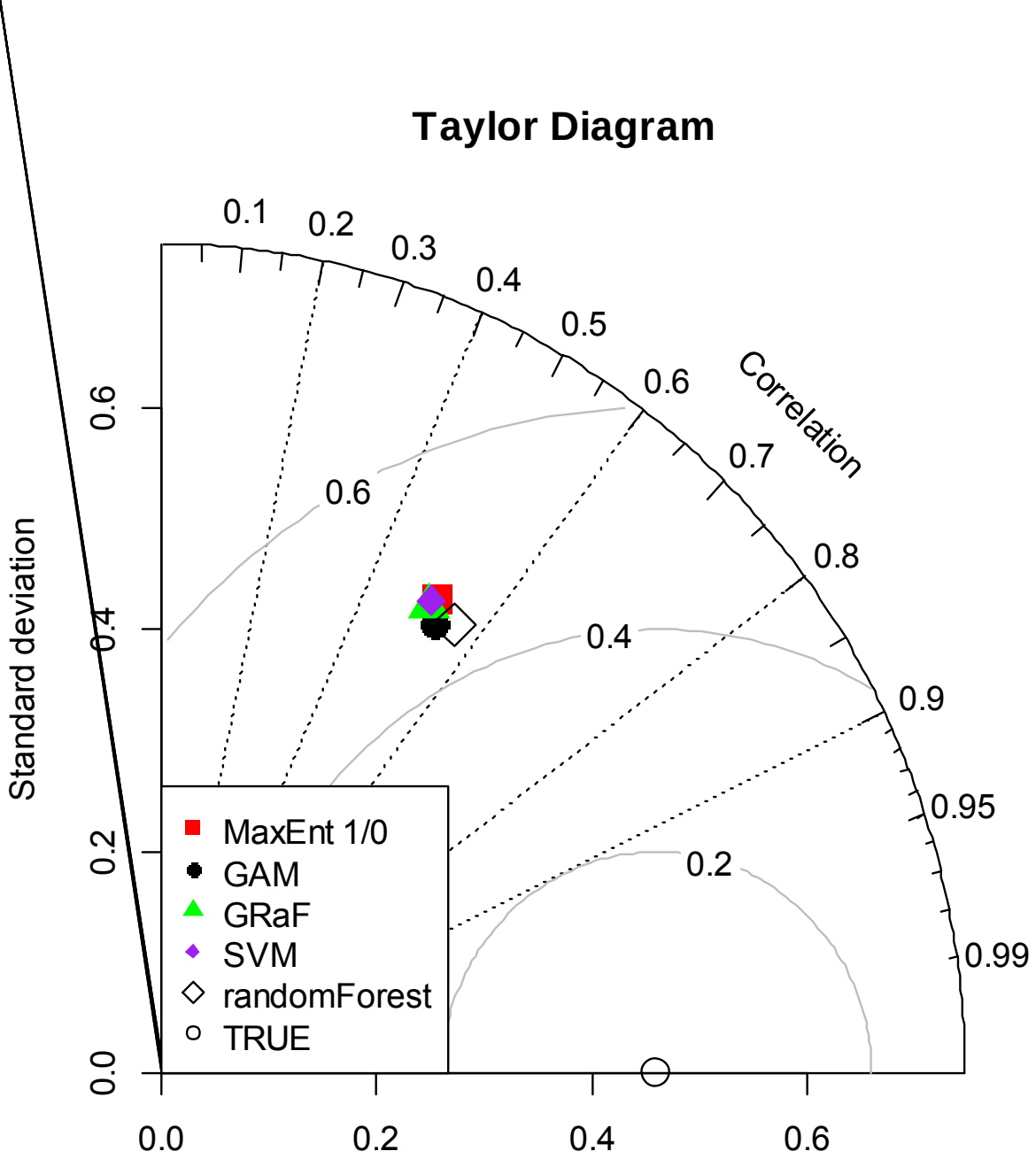




MaxEnt

These variables were used in GAM and SVM





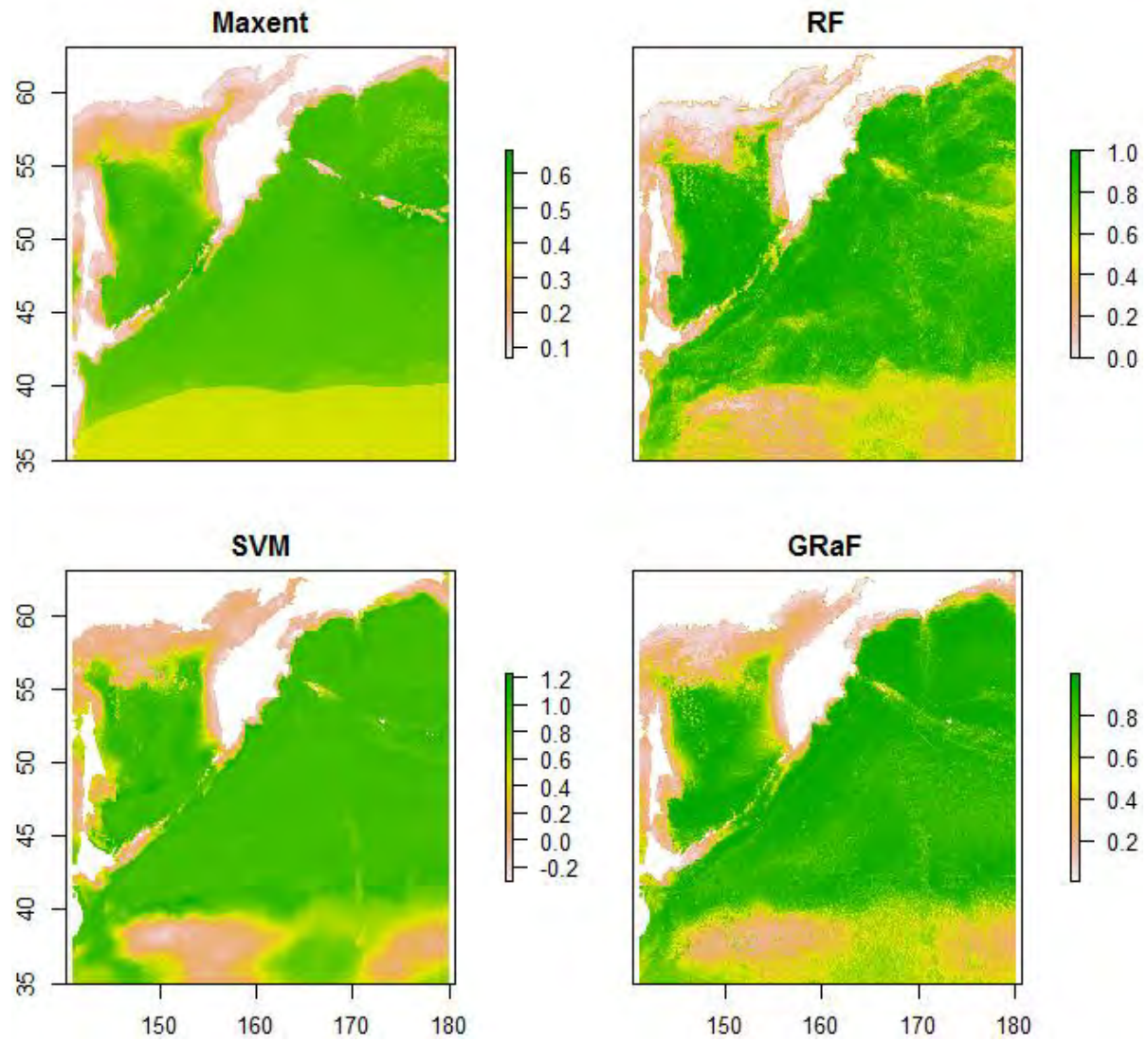
Model perfomance

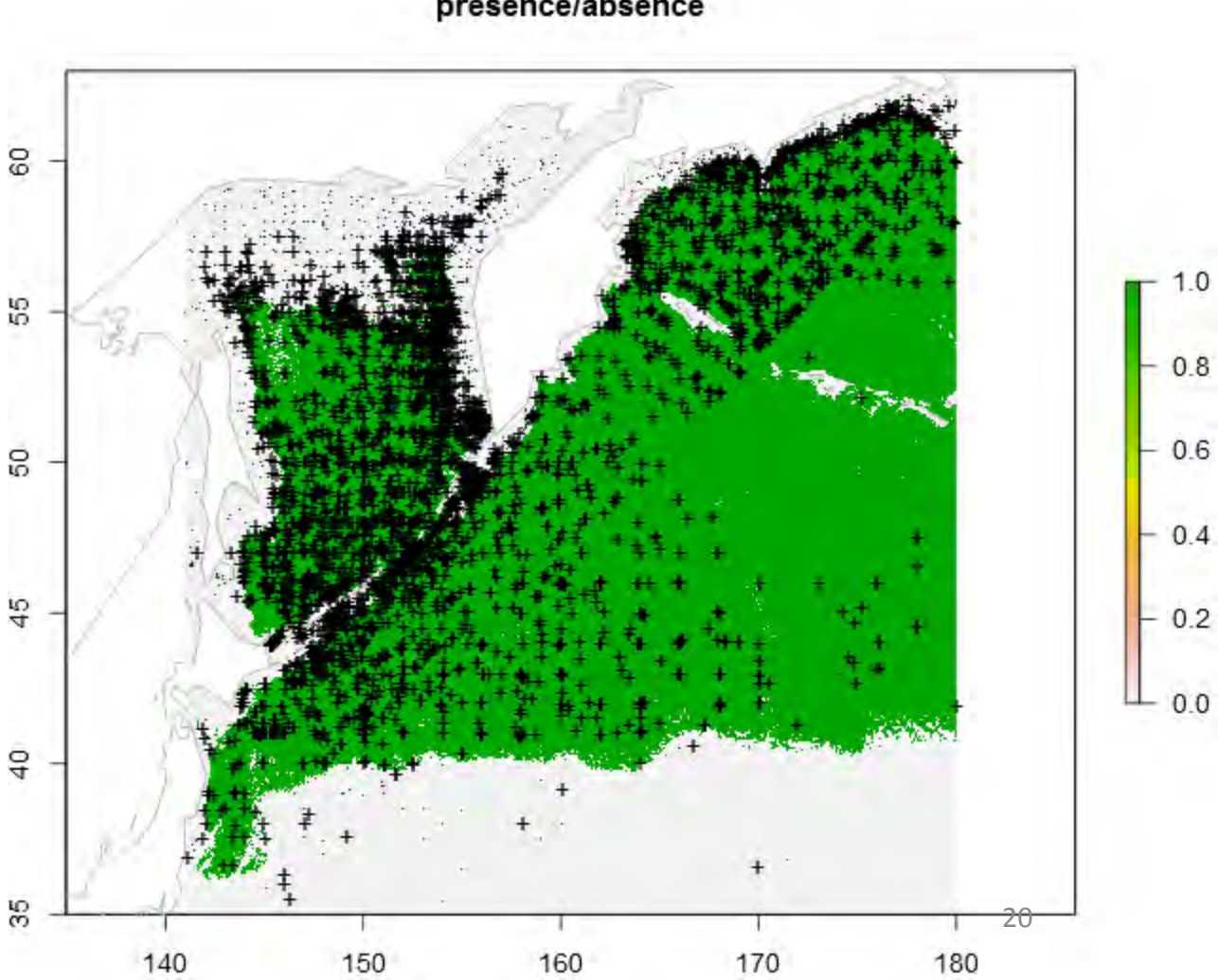
model	*AUC*	*cor*	*max TPR+TNR*
Bayes GRaF	0.815	0.527	0.653
MaxEnt	0.528	0.104	0.356
GAM	0.839	0.554	0.833
randomForest	0.862	0.619	0.687
SVMs	0.831	0.574	0.794
MaxEnt 1/0	0.845	0.570	0.490

Current state SDM (Case 1)

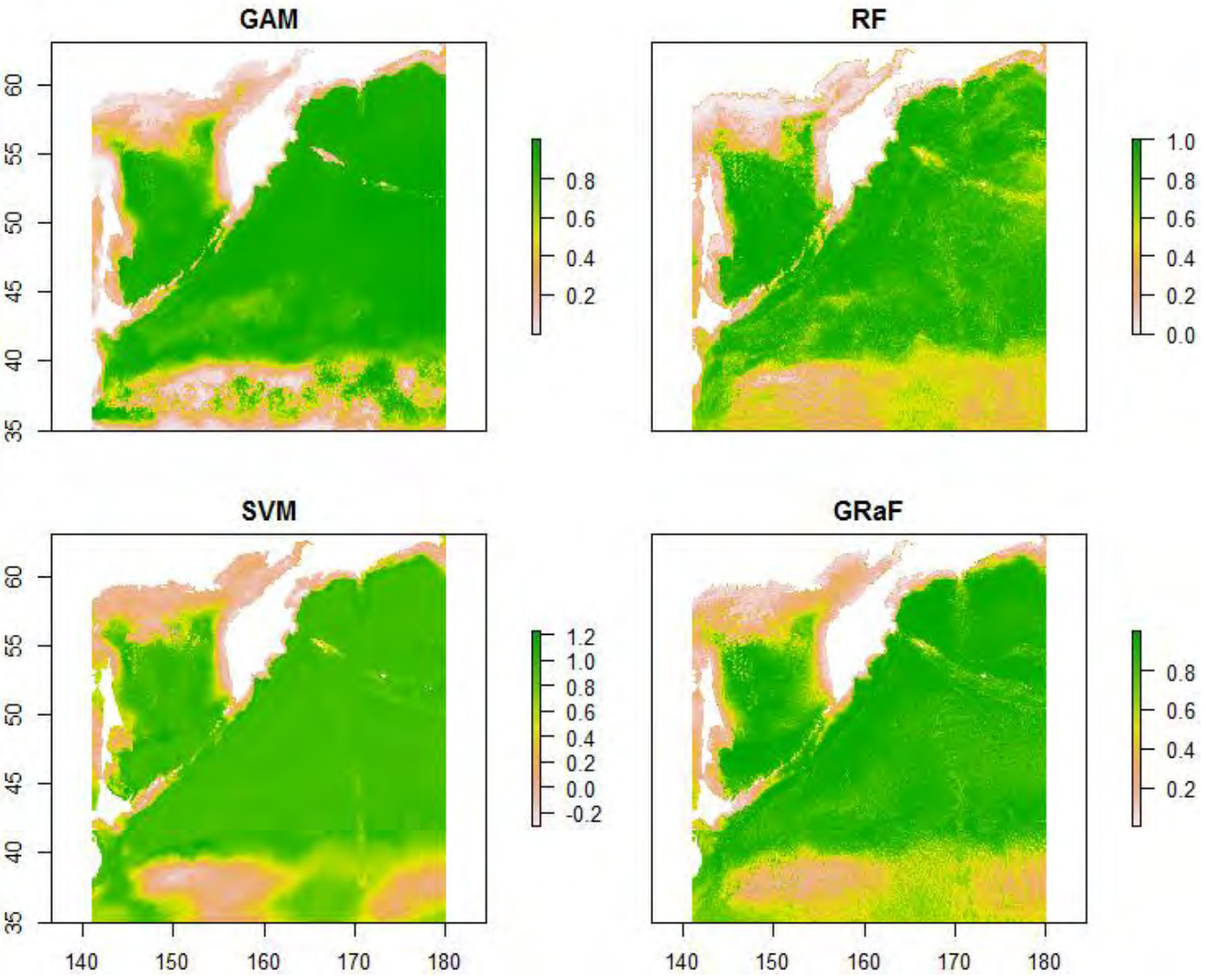
weighted by AUC mean of 4 SDMs of **MaxEnt, RF, SVM, GRaF**

332522 positive cells = 100%





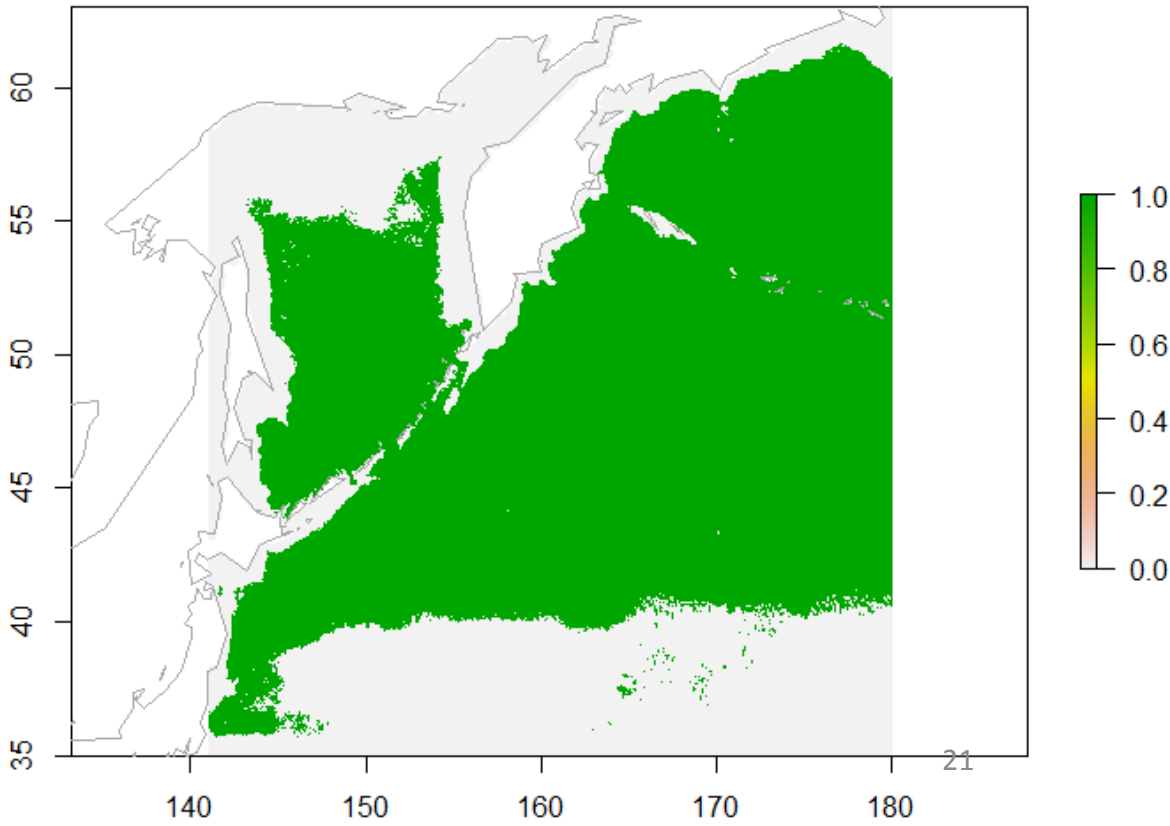
Current state SDM (Case 2)



weighted by AUC mean of 4 SDMs of **GAM, RF, SVM, GRaF**

344447 positive cells = 100%

presence/absence



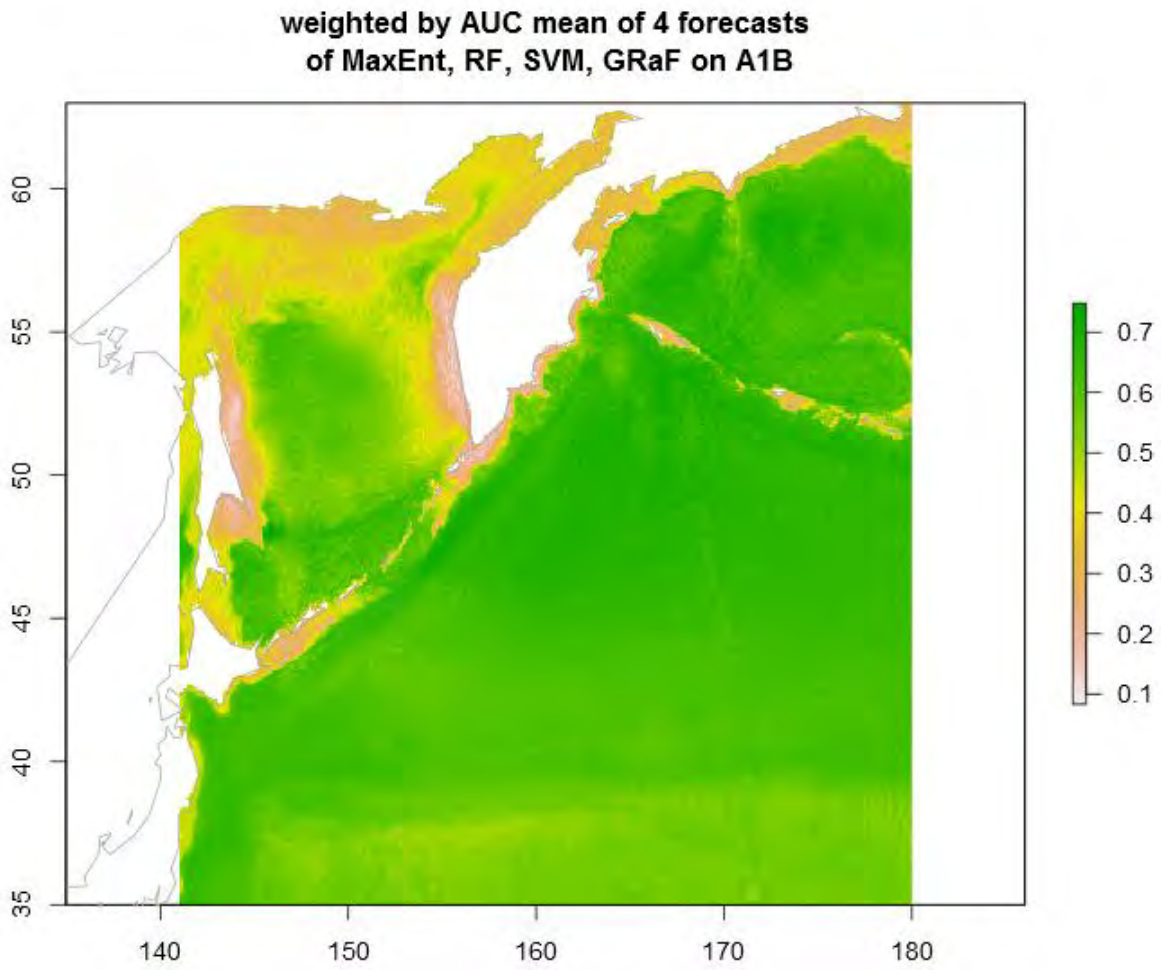
The Emissions Scenarios of the Special Report on Emissions Scenarios (SRES)

A1. The A1 storyline and scenario family describes a future world of very rapid economic growth, global population that peaks in mid-century and declines thereafter, and rapid introduction of new and more efficient technologies. Major underlying themes are convergence among regions, capacity building and increased cultural and social interactions, with a substantial reduction in regional differences in per capita income.

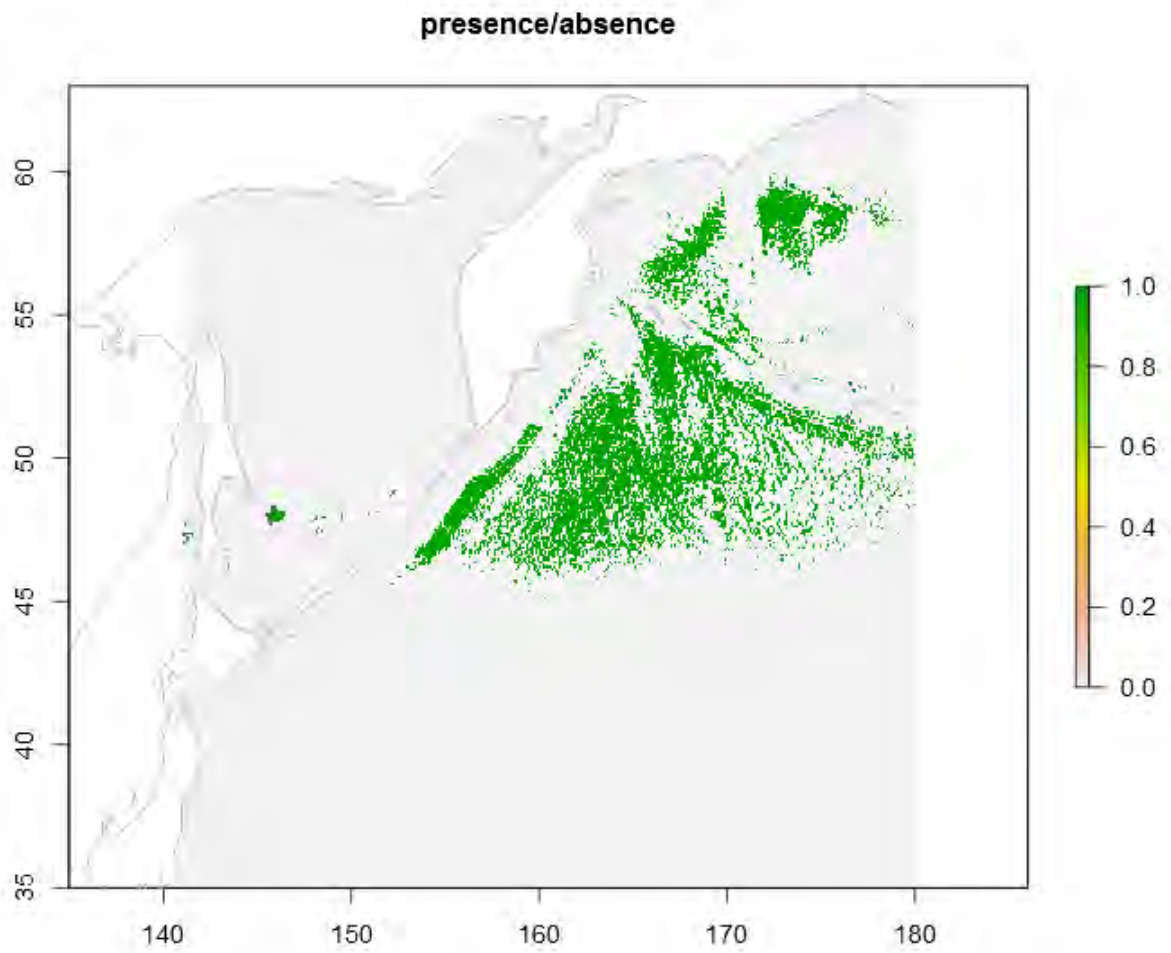
The A1 scenario family develops into three groups that describe alternative directions of technological change in the energy system. The three A1 groups are distinguished by their technological emphasis: fossil intensive (A1FI), non-fossil energy sources (A1T), or a **balance across all sources** (**A1B**) (where balance is defined as not relying too heavily on one particular energy source, but on the assumption that similar improvement rates apply to all energy supply and end-use technologies).

https://www.ipcc.ch/ipccreports/tar/wg1/029.htm#storya1

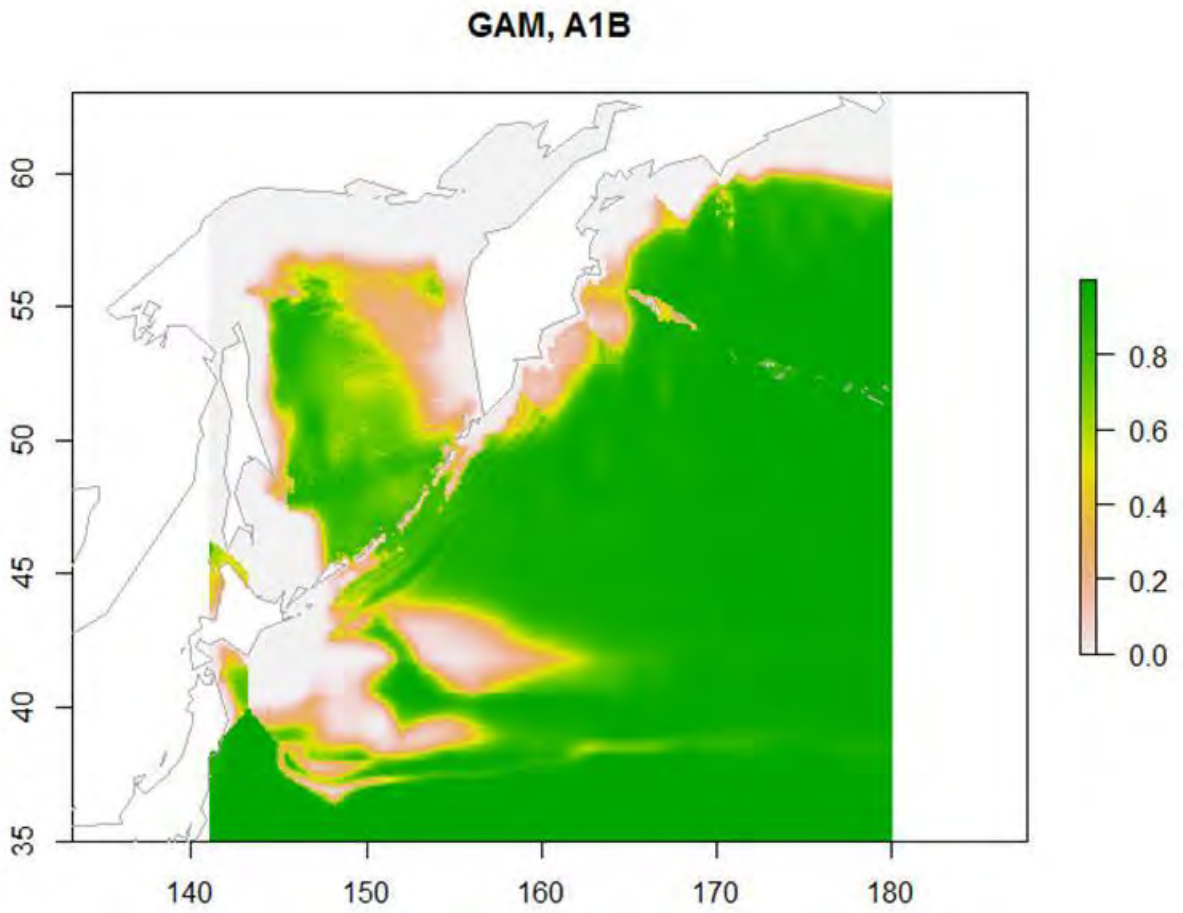
<u>Balance across all sources</u> (**A1B**) (where balanced is defined as not relying too heavily on one particular energy source, but on the assumption that similar improvement rates apply to all energy supply and end-use technologies)

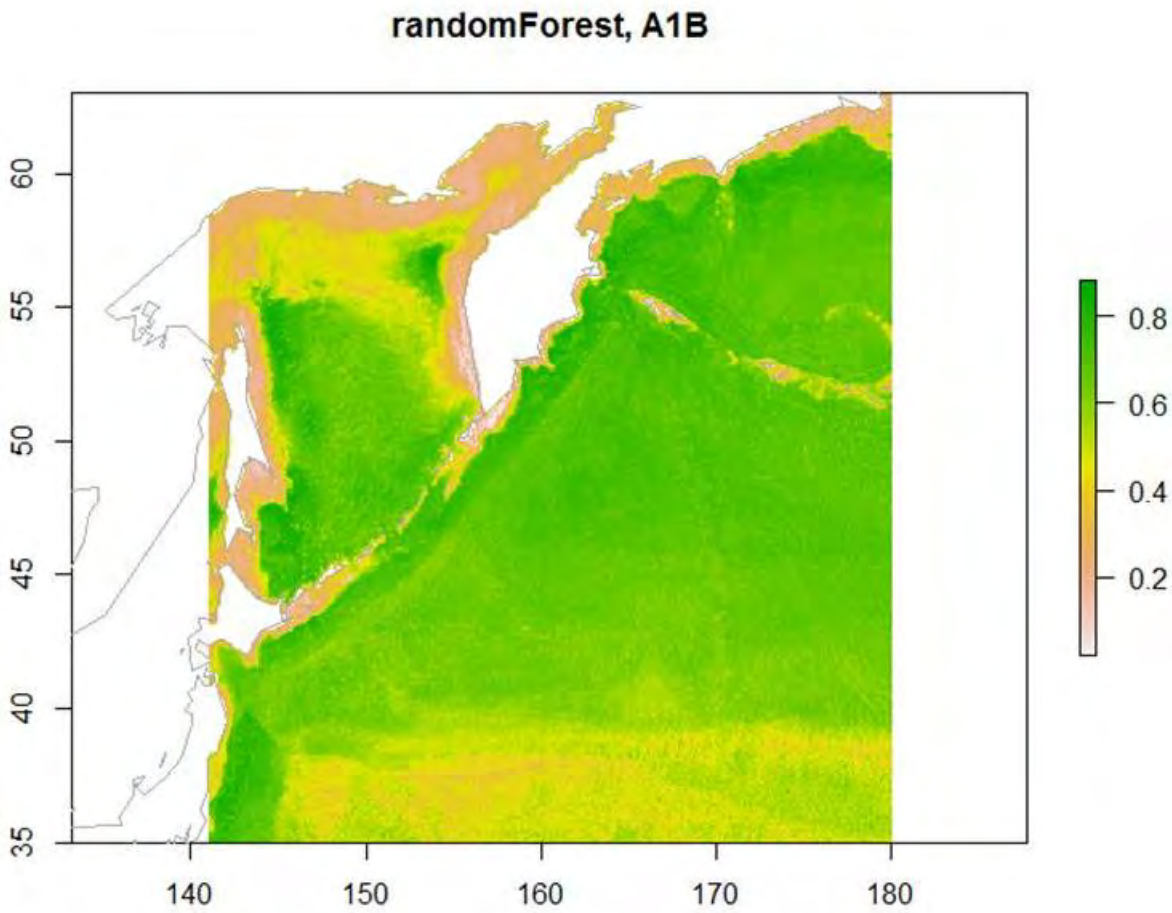


11228 positive cells (3.4%)

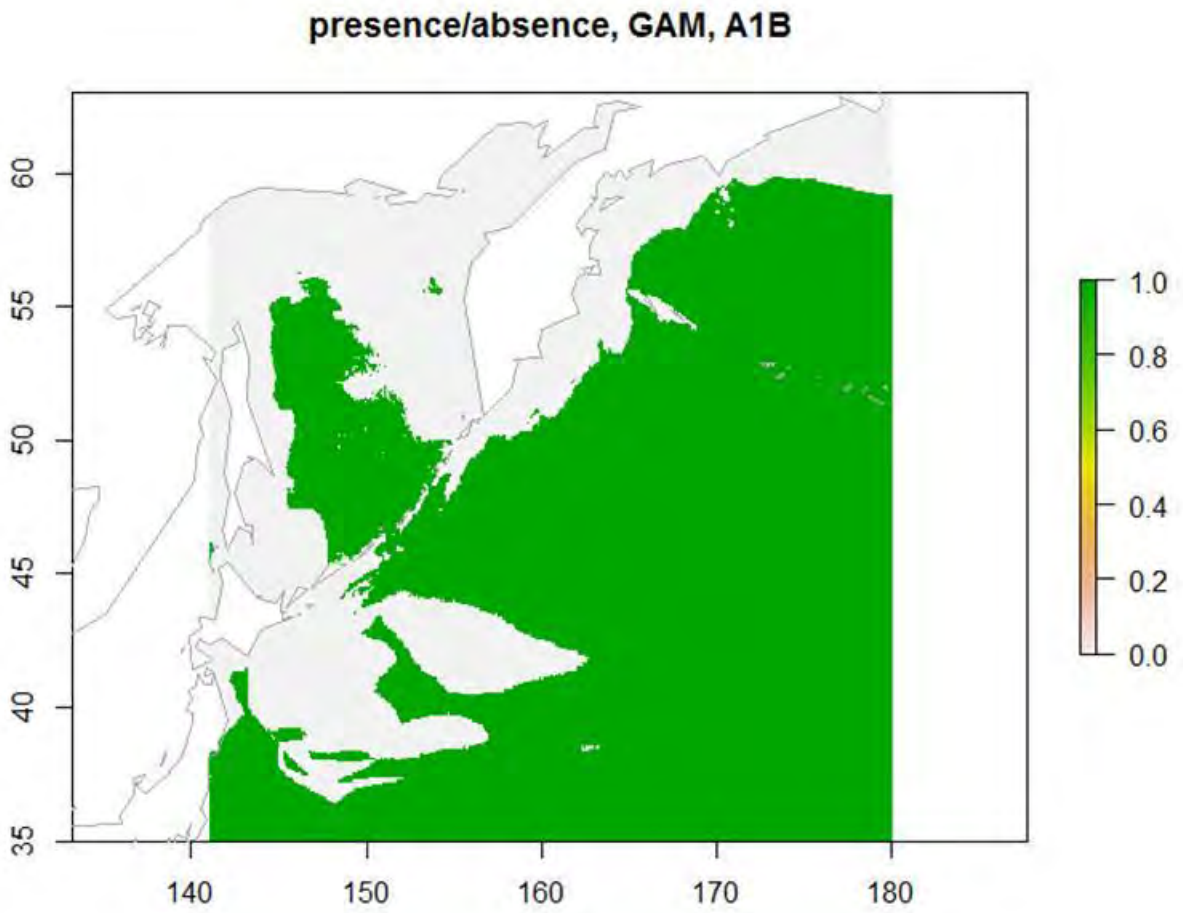


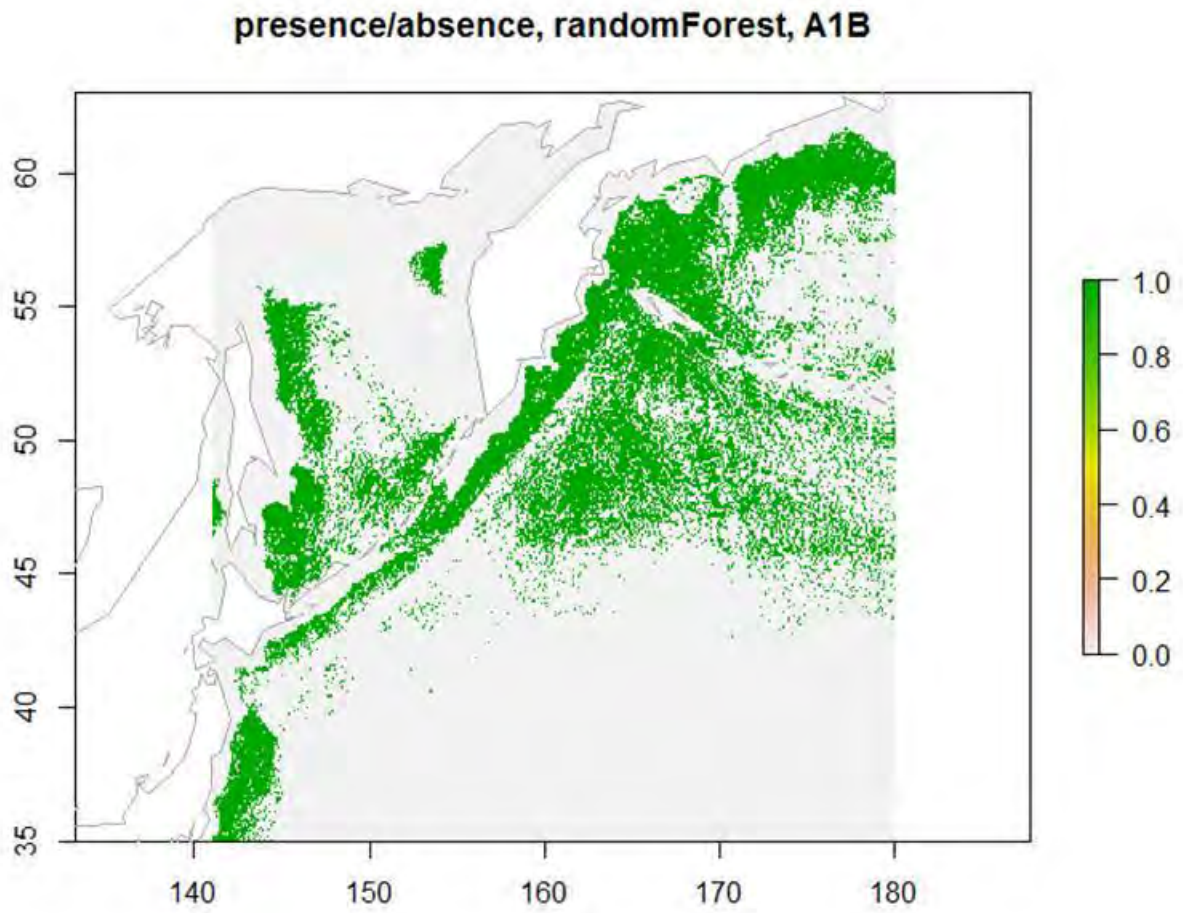
However, separately SDMs showed very different results, e.g.:





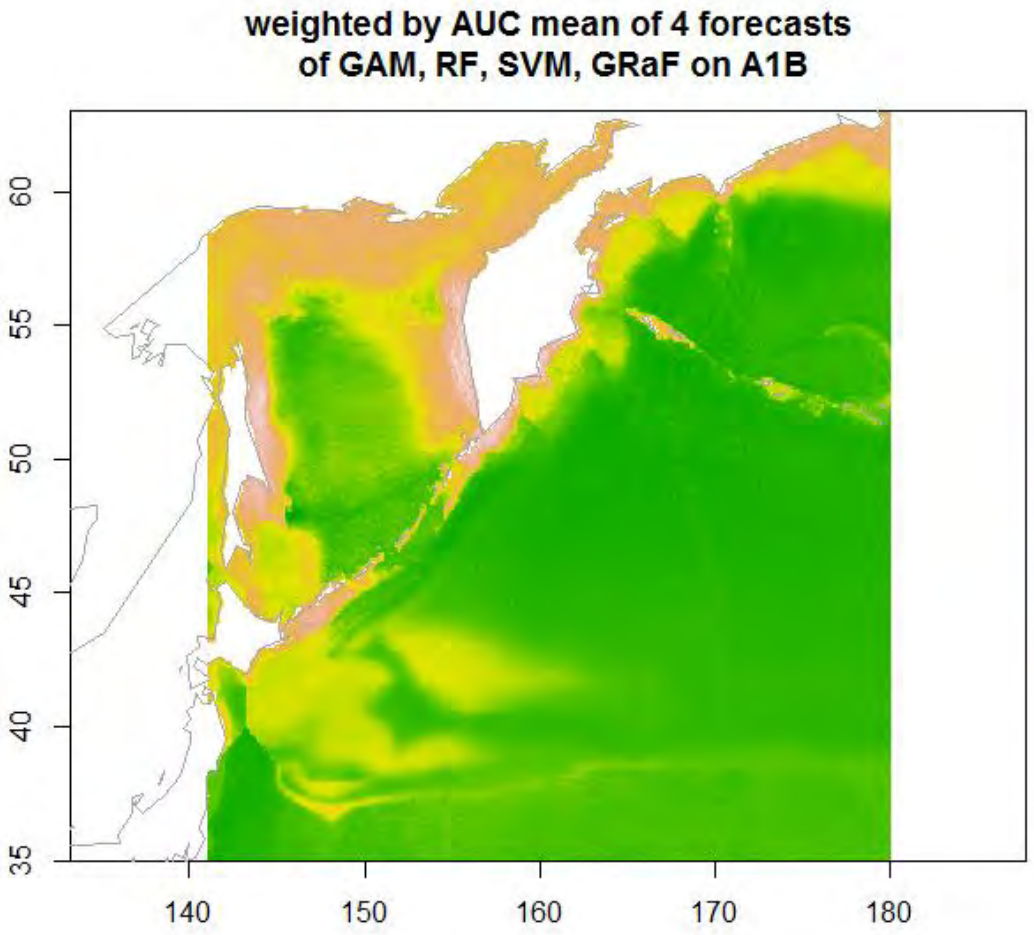
But separately SDMs showed very different results, e.g.:

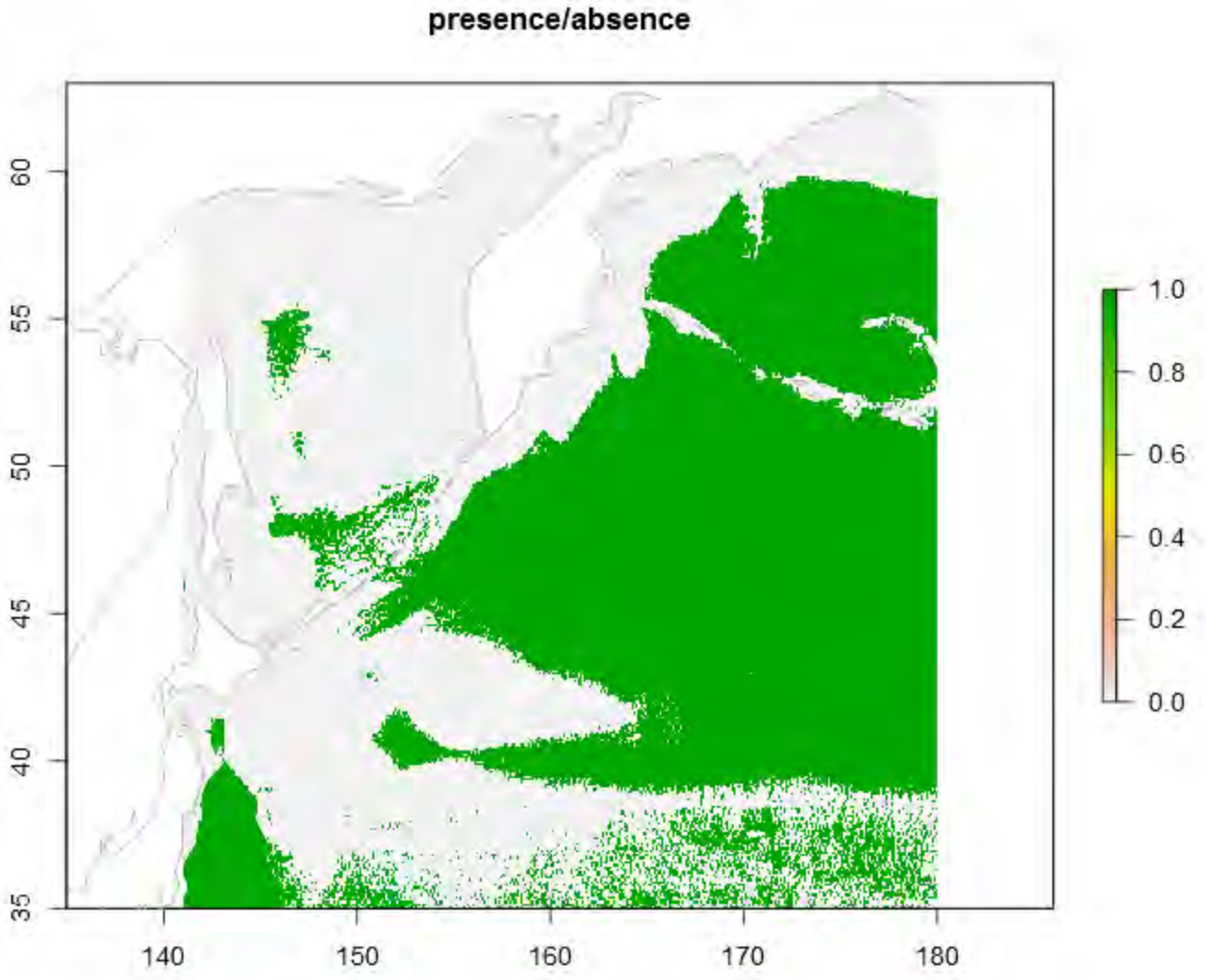




Balance across all sources (**A1B**) (where balanced is defined as not relying too heavily on one particular energy source, but on the assumption that similar improvement rates apply to all energy supply and end-use technologies) **GAM instead of MaxEnt**

66487 positive cells (19.3%)

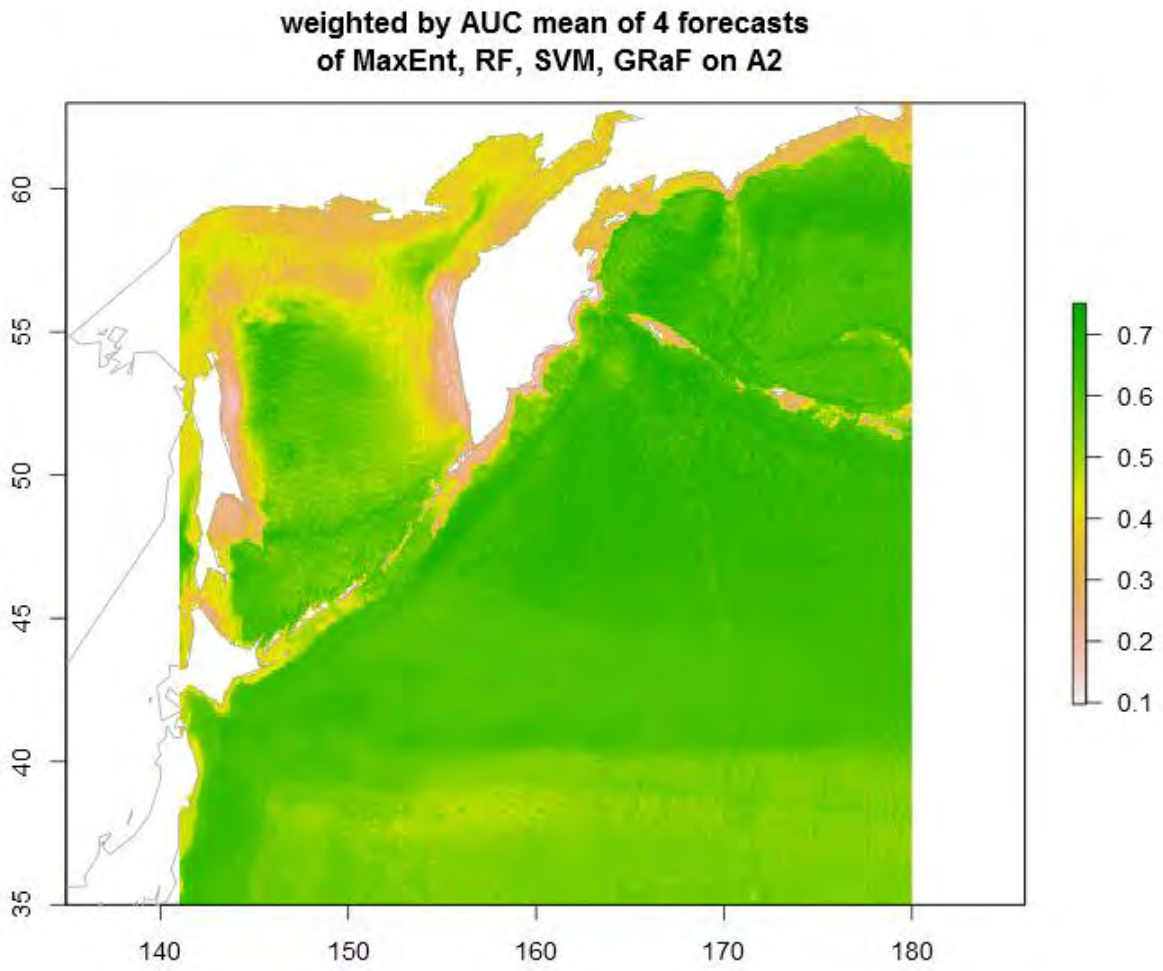


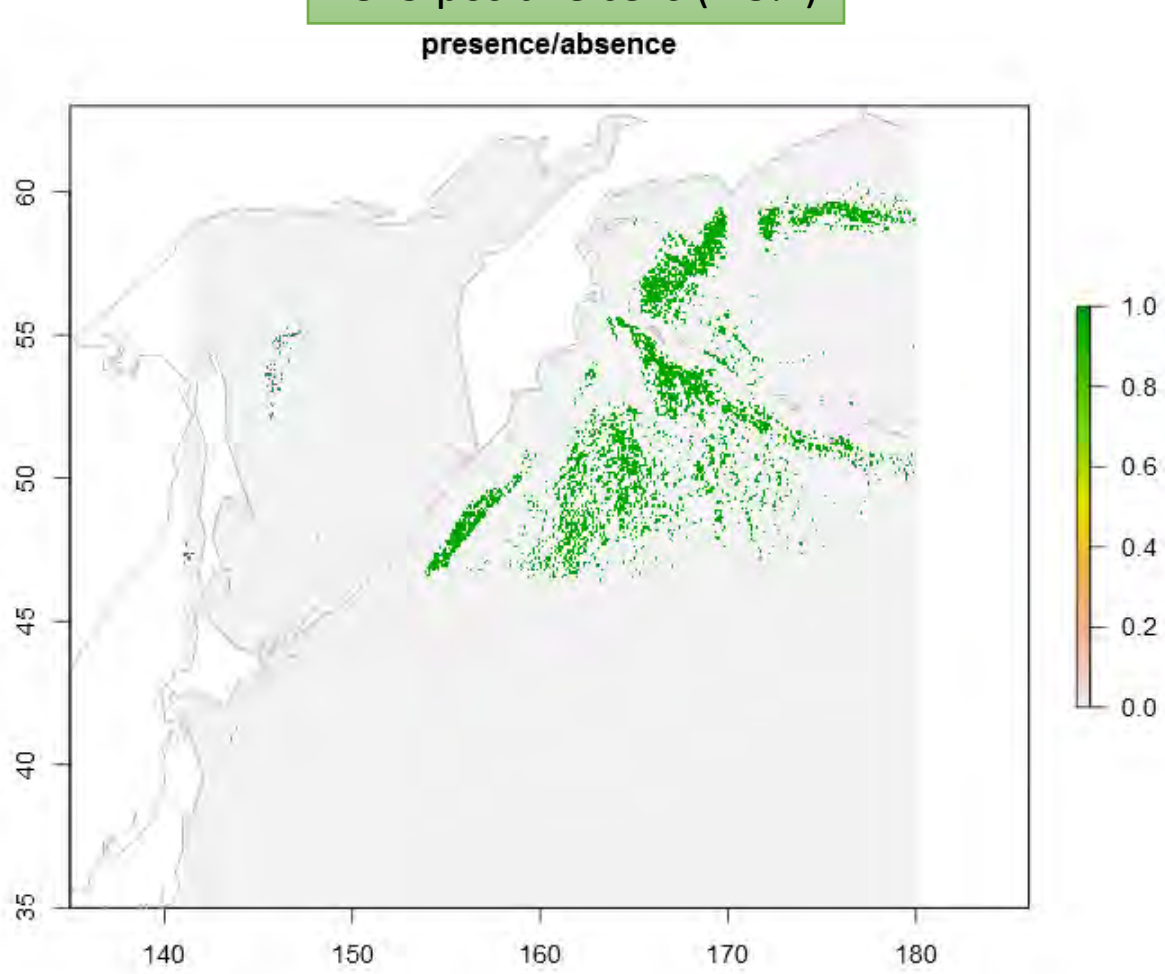


Therefore replacement of the only 1 SDM in the average would lead to different projection (more optimistic in the case #2).

A2. The A2 storyline - a very **heterogeneous world**. The underlying theme is self-reliance and preservation of local identities. Fertility patterns across regions converge very slowly, which results in continuously increasing population. Economic development is primarily regionally oriented and per capita economic growth and technological change more fragmented and slower than other storylines.

4879 positive cells (1.5%)



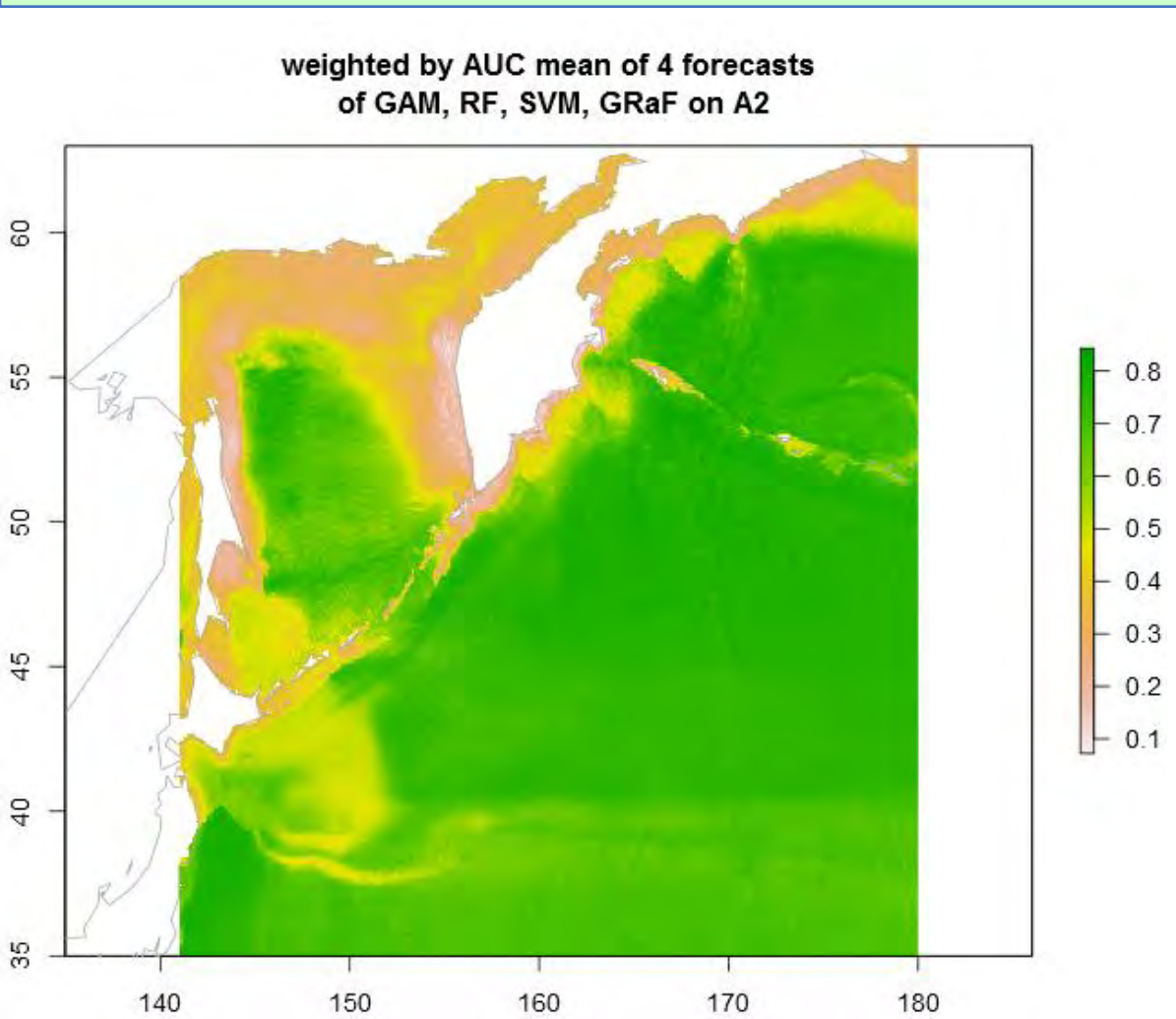


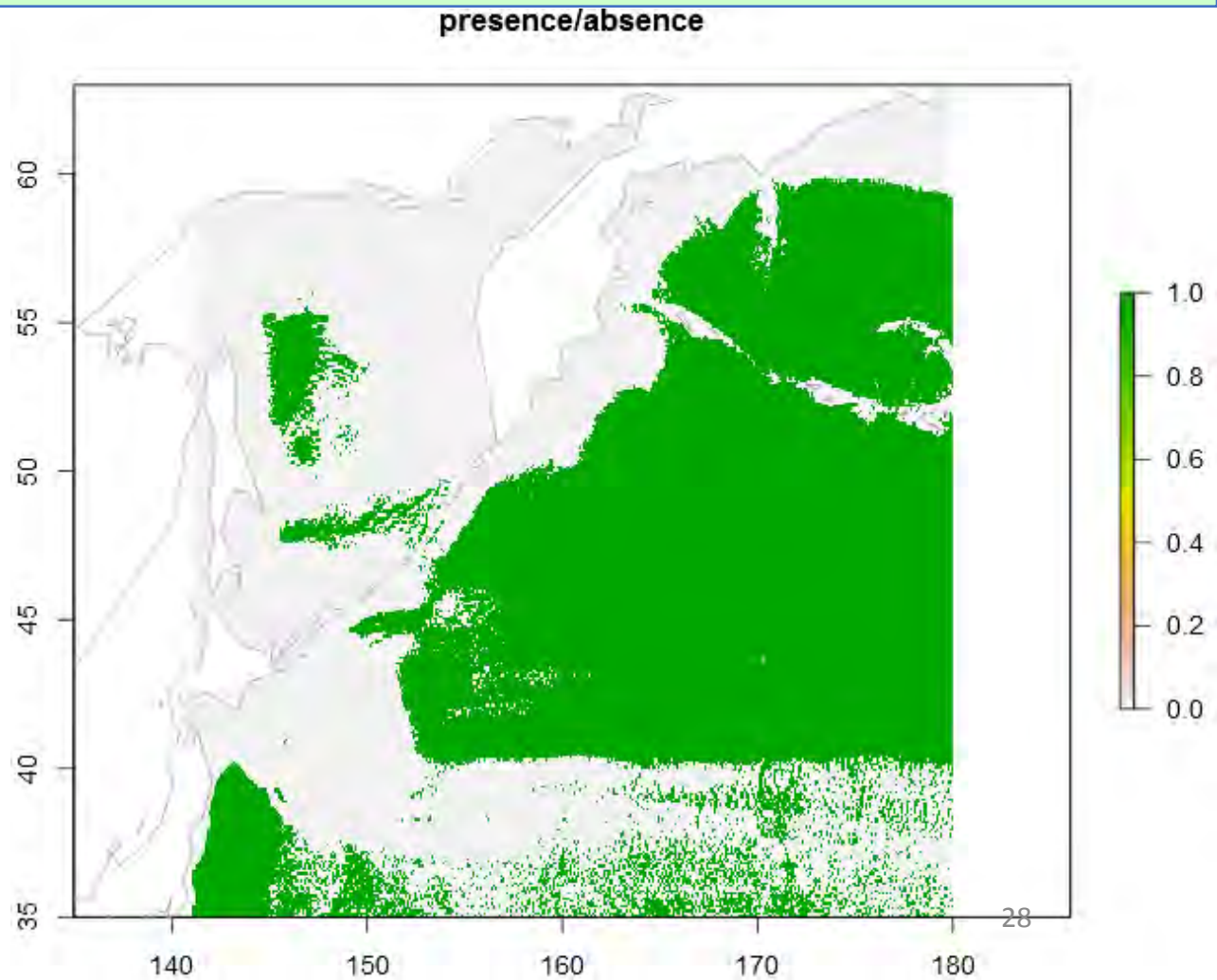
SVM projections didn't include any cell above the threshold!

A2. The A2 storyline - a very **heterogeneous world**. The underlying theme is self-reliance and preservation of local identities. Fertility patterns across regions converge very slowly, which results in continuously increasing population. Economic development is primarily regionally oriented and per capita economic growth and technological change more fragmented and slower than other storylines.

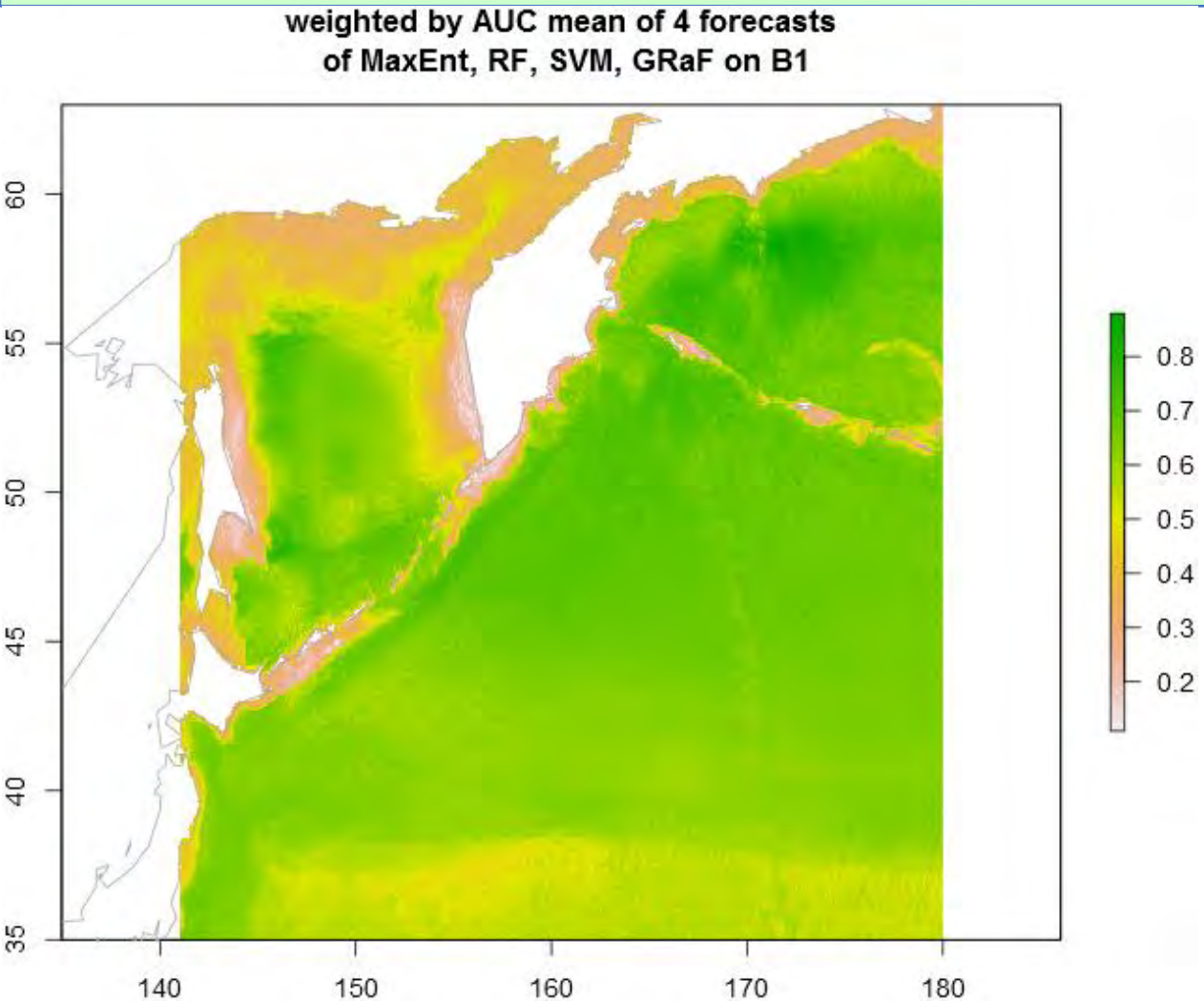
GAM instead of MaxEnt

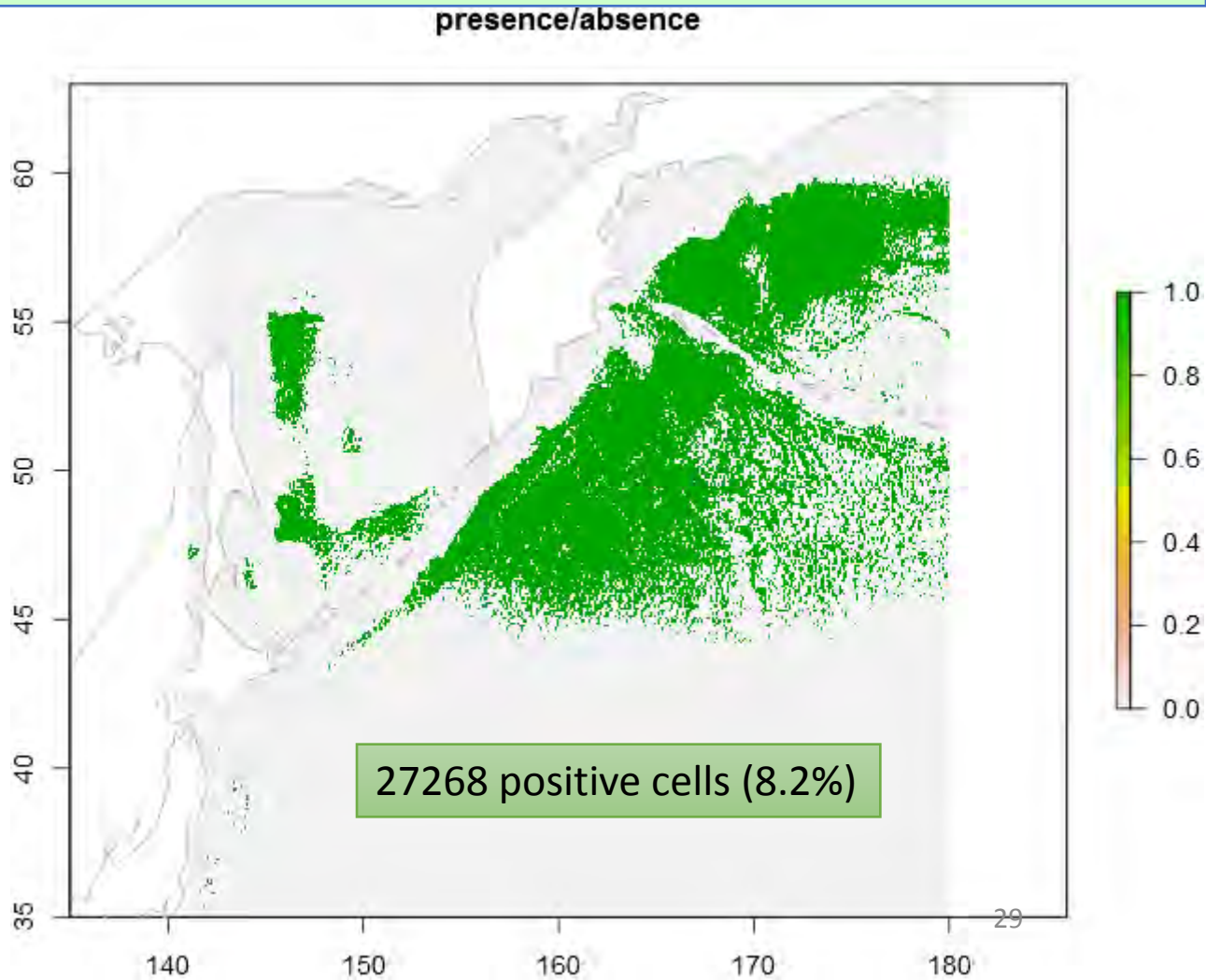
68135 positive cells (19.8%)





The **B1** storyline - **a convergent world** with the same global population, that peaks in mid-century and declines thereafter, as in the A1 storyline, but with rapid change in economic structures toward a service and information economy, with reductions in material intensity and the introduction of clean and resource-efficient technologies. The emphasis is on global solutions to economic, social and environmental sustainability, including improved equity, but without additional climate initiatives.

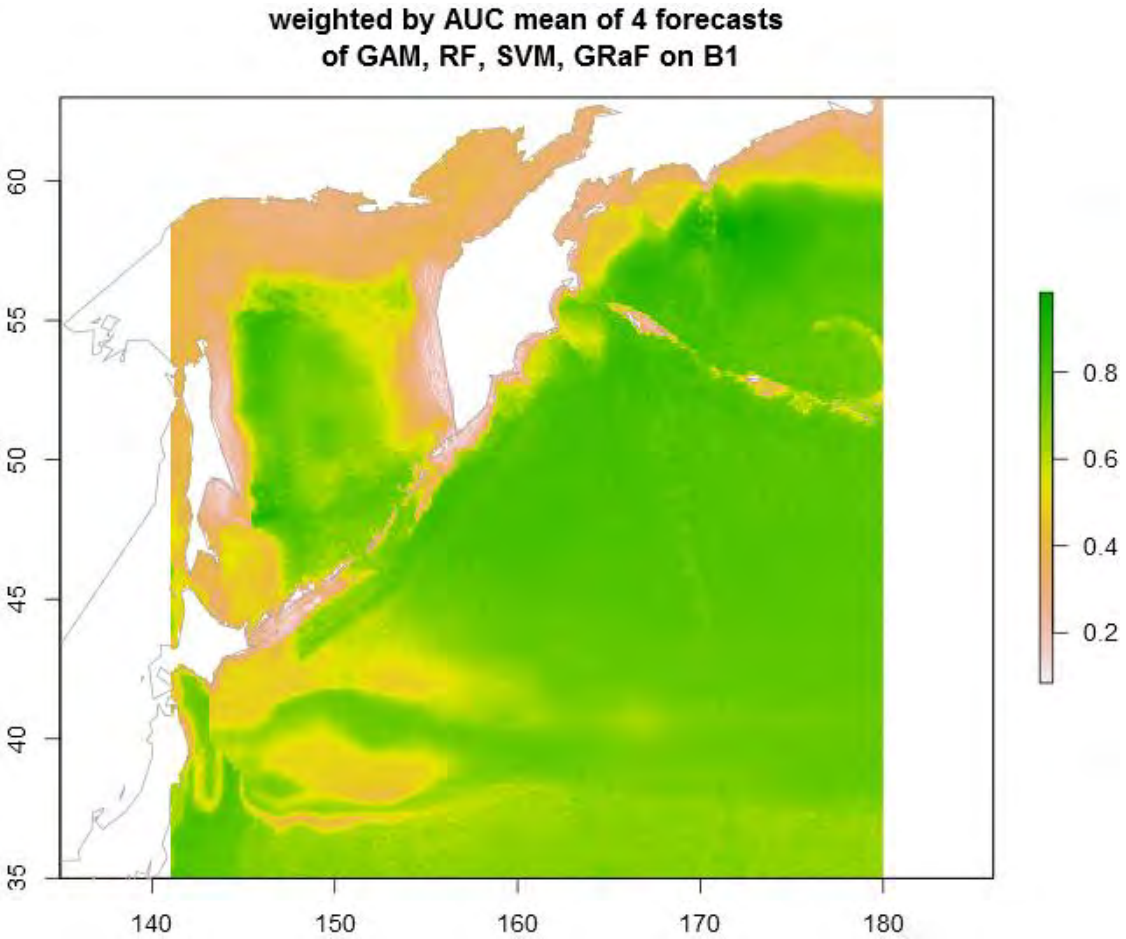


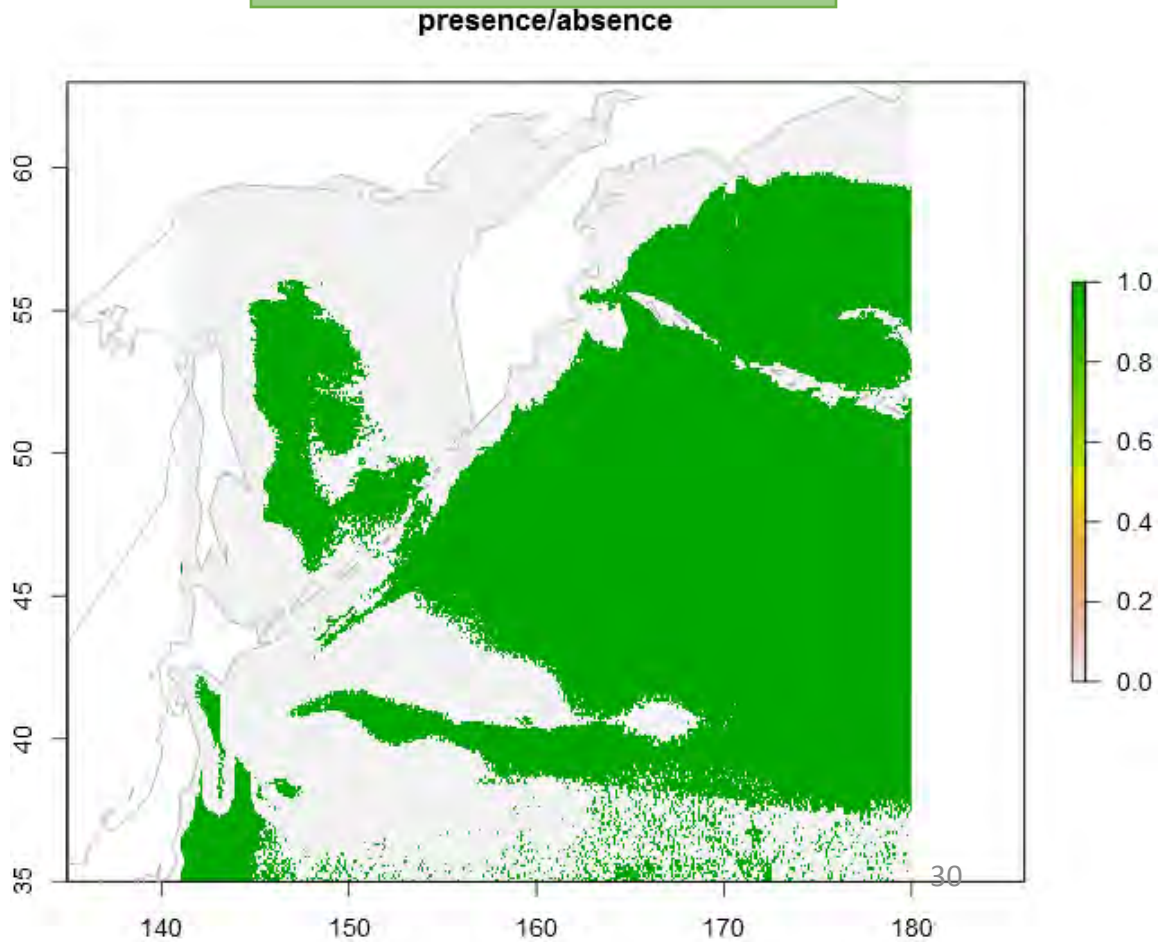


The **B1** storyline - **a convergent world** with the same global population, that peaks in mid-century and declines thereafter, as in the A1 storyline, but with rapid change in economic structures toward a service and information economy, with reductions in material intensity and the introduction of clean and resource-efficient technologies. The emphasis is on global solutions to economic, social and environmental sustainability, including improved equity, but without additional climate initiatives.

GAM instead of MaxEnt

73432 positive cells (21.3%)





Basic Conclusions

1. MaxEnt performance was good only with exact locations of negative observations and it was comparable to SVM and GRaF. In all test cases, MaxEnt showed the lowest AUC around 0.5 when it was not provided with 0 observations.
2. RandomForest outperformed other methods with many covariates obtained from **MARSPEC** database, nevertheless lower number of covariates from **Bio-ORACLE** database made RandomForest to perform as good as other methods.
3. All projections showed a significant decrease (shrinkage) of the distribution area of boreopacific gonate squid at least 5 times until the end of 21st century:
   - A2 – down to 1.5%-19.8%,
   - A1B – down to 3.4%-19.3%, and
   - B1 – down to 8.4%-21.3%
4. Shrinkage of the squid distribution area may influence the distribution of such commercially important fish predators as coho, cherry (masu) and chinook salmon, as well as other abundant pelagic fish that feeds on the squid; therefore, it should be included in future calculations made from SDM at least for some particular North Pacific pelagic species, such as the Pacific salmon.

Thank you for your attention!