SPECIES DISTRIBUTION MODELING FOR KING MACKEREL (*SCOMBEROMORUS CAVALLA*) AND ITS PREY SPECIES IN THE GULF OF MEXICO

A Thesis

by

XIAOPENG CAI

BS, Xiamen University, China, 2010
MS, Xiamen University, China, 2013

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

COASTAL AND MARINE SYSTEM SCIENCE

Texas A&M University-Corpus Christi
Corpus Christi, Texas

December 2016

December 2016

SPECIES DISTRIBUTION MODELING FOR KING MACKEREL (*SCOMBEROMORUS CAVALLA*) AND ITS PREY SPECIES IN THE GULF OF MEXICO

A Thesis

by

XIAOPENG CAI

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Alexey Sadovski, PhD
Chair

James Simons, PhD
Co-Chair

Blair Sterba-Boatwright, PhD
Committee Member

December 2016

ABSTRACT

Ecosystem based fisheries management (EBFM) has been broadly recognized throughout the world as a way to achieve better conservation. Therefore, as an important part of EBFM, mapping multi-species interactions or spatial distributions has been strongly needed. Species distribution models are widely applied since information regarding the presence of species is usually only available for limited locations due to the high cost of field surveys. Furthermore, a large proportion of the fisheries survey data have only presence records instead of regular presence and absence records, Thus, presence-only species distribution models are needed.

In this study, four presence-only species distribution algorithms (Bioclim, Domain, Mahal and Maxent) were applied using 12 environmental parameters as predictors to model the distribution of king mackerel (*Scomberomorus cavalla*) and 31 of its prey species in the Gulf of Mexico. Based on the results, 10 major distribution patterns were proposed to describe the distribution of the 32 species. Post hoc with Tukey's test shows that area under curve (AUC) for the Maxent-based models were significantly ($p<0.05$) higher than those for Bioclim and Domain based models, but insignificantly different from those for Mahal-based models ($p=0.955$); while correlation coefficients (r) for the Maxent-based models were significantly higher than those for all the other three types of models ($p<0.05$). Thus, Maxent-based models were concluded to have the best performance.

Generalized linear models (GLM), generalized additive models (GAM) and random forest models (RF) were applied to model the abundance distribution of three shrimp species throughout the Gulf. Results show that abundance distributions predicted were quite close to the species distribution predicted by the presence-only models, which validated the good performance of the presence-only models. Evaluation of the models by correlation shows that the GAM models had the best performance for brown shrimp abundance modeling, while the RF models had the best performance for the other two shrimp species.

Good performance of the species distribution/abundance models shows that interesting distribution patterns, especially the special zones (eg. the dead zone), can provide some insights for scientists or government managers to better manage fisheries resources in the Gulf of Mexico.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Alexey Sadovski, my committee members, Dr. James Simons and Dr. Blair Sterba-Boatwright, and Dr. Cristina Carollo for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University-Corpus Christi an enjoyable experience. I also want to extend my gratitude to my wife, who always supported me while I was finishing my thesis.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES PAGE

LIST OF TABLES

TABLES PAGE

CHAPTER I: INTRODUCTION

1.1 Ecosystem modeling

Marine and coastal ecosystems are extremely important for human beings, not only for providing food, energy, and natural products, but also for playing important roles in nutrient cycling, storm protection and climate regulation. Unfortunately, they are under various stages of impairment (US Commission on Ocean Policy, 2004). In recent years, NOAA has proposed using integrated ecosystem assessment (IEA) as a tool combining biological, chemical, physical and socioeconomic factors to keep the marine ecosystem healthy and sustainable under human and environmental stresses.

Fisheries management, directly linked to socioeconomics, has been a key part of the IEA. The need for ecosystem-based fisheries management (EBFM) has been widely recognized throughout the world. Instead of considering single issues, species, or functions, EBFM takes the complexity of the interactions between them into account, and considers the inherent links between ecosystem condition and human activity.

Ecosystem modeling is one of the ways that help us better understand, assess, and manage marine ecosystems to reach the goals of IEA and EBFM (Link et al., 2010). There have been numerous approaches using different methods to model marine ecosystems. Sophisticated models such as Atlantis, Ecopath with Ecosim (EwE), and OSMOSE have been broadly used all over the world to explore ecosystem dynamics.

The Atlantis model is an ecosystem box model primary aimed at management strategy assessment. The key of the Atlantis model is a three-dimensional, spatially-resolved, deterministic biophysical sub model using a map consisting of boxes and slab-like layers. The model tracks the nutrient (usually silica and nitrogen) flows through the major biological groups of interest in the marine ecosystem. The key ecological processes taken into account in the model are consumption,

production, waste cycling, migration, predation, mortality, recruitment and habitat dependency. Currently, 13 Atlantis models have been developed or are under development for marine ecosystems in Australia and the United States (Fulton et al., 2011; Link et al., 2010).

EwE combines software for trophic mass analysis (Ecopath) with a dynamic modeling capability (Ecosim) for modeling past and future impacts of fishery and environmental disturbances. Over 300 EwE models have been developed worldwide, of which 15 are from the Gulf of Mexico (Christensen and Walters, 2004; Christensen et al., 2008; Geers et al., 2016).

OSMOSE is an individual-based and multispecies two-dimensional model mainly focused on the major processes in the life cycle of invertebrate species and high trophic level (HTL) groups of fish (Shin and Cury, 2001; Shin et al., 2004). OSMOSE has been used to model trophic dynamics and the impacts of fishing management strategies in a variety of ecosystems, including the west Florida shelf (Grüss et al., 2015; Marzloff et al., 2009; Travers et al., 2009; Travers et al., 2010).

However, those well-developed models are highly dependent on high quality trophic data that are usually difficult to obtain. There are always gaps in trophic data, which include gaps in species distribution, abundance and interaction data. A gap analysis to identify and fill those gaps will be a very important step to improve the results from these ecosystem models (Masi et al., 2014). Species interactions are based on species distribution and abundance. Once the gaps of distribution and abundance of target species has been identified and filled, we will have a better understanding to further fill the gaps in species interactions. Therefore, in this study, I will build species distribution models and species abundance models from incomplete information to identify and fill the distribution and abundance gaps for king mackerel and its prey species in the Gulf of Mexico.

1.2 Species distribution modeling

1.2.1 Presence-absence vs. presence-only models

Presence and absence models are commonly applied in ecological species distribution studies. When researchers sample at a location, if the target species is observed and sampled, it is recorded

as present; if the target is not observed or sampled, it is recorded as absent. However, usually the absence here is not "true absence", it can be biased and incomplete, because even if a species not observed or sampled at a given time, it may not be absent from that location at all times. It may appear here at other times, or it may be already here, but the researchers did not sample it due to other reasons (Kéry et al., 2010). For example, researchers use trawls to sample fishes, including king mackerel; king mackerel usually appear at the surface of the water and swim very fast, so it is quite easy for the king mackerel to escape from the trawl, which is mainly focused on catching bottom fishes. In this case, the trawl does not harvest any king mackerel; however, it is not true absence.

For better conservation and management purpose, there is a strong need for mapping of a species' geographical distribution or use of habitats. However, due to high cost of wildlife species distribution surveys, species presence or absence at every location on the map are rarely available. Therefore, prediction models using environmental parameters or remote sensing data as covariates are applied to footprint the species distribution-interpolate, or extrapolate beyond the locations where target species presence is known (Pearce and Boyce, 2006). Presence-absence classification is a special multiclass (binary) classification while presence-only classification is a one-class classification approaches. If the survey data contains presence and absence, then it is a presence-absence classification of traditional multiclass question, which is easier to process. However, a large proportion of the fish survey data has only presence records; therefore, the modeling of those species distribution belongs to presence-only classification category.

1.2.2 Pseudo absence and background points

One way to make the presence only classification is to add so-called "Pseudo absences" into the dataset to make it a regular presence-absence classification (Pearce and Boyce, 2006; VanDerWal et al., 2009). For example, if in regular presence and absence sampling, a sample of 100 observations contain 30 observations of the target species, the probability of occurrence is 0.3 ($=\frac{30}{(30+70)}$). However, in presence-only data, the presence locations were sampled independently and then pseudo-absence locations were selected by the researchers. In this case, the proportion of occurrences does not represent the true presence of the target species in the population, but represent the relative proportion selected by the researcher. For instance, the above presence-

absence example could be reproduced by taking 30 presence records and adding a set of 70 'pseudo-absence' records to them in a statistically independent way. In this manner, the probability of occurrence is also 0.3. However, if we select 300 pseudo-absence locations, the probability of occurrence would then become 0.09 ($=\frac{30}{(30+300)}$). Therefore, while pseudo absence makes the classification problem easier to solve, determining the appropriate number of pseudo absences is still a big challenge (VanDerWal et al., 2009).

Background points (Phillips et al., 2009), similar to pseudo-absence, are also used for creating a non-presence class. Unlike pseudo-absences, background points are not trying to guess at absence points. They are independent of where the species is known to be present, and are used to establish the environmental domain of the study region. The method of background points requires fewer assumptions than pseudo-absences, which is an advantage of background points. A second advantage is that there are some existing statistical methods for solving the "overlap" problems between background points and presence (Ward et al., 2009).

1.2.3 Existing presence-only species distribution modeling methods

A number of biologists, collaborating with statisticians, have published a variety of algorithms to build species distribution models using presence-only data. The models can be classified into three categories: profile, regression, and machine learning models (Hijmans and Elith, 2015). Profile methods only consider presence-only data, not absence or background data, while regression and machine learning methods (except Maxent, which can deal with presence only data) are applied to both presence and absence or background data. In this study, four presence-only species distribution models will be discussed: three profile models (Bioclim, Domain and Mahal) and one machine learning model (Maxent).

The Bioclim algorithm is a classic 'climate-envelope-model' which has been widely used for species distribution modeling (Booth et al., 2014). The Bioclim algorithm calculates the likelihood of occurrence at a location by comparing the similarity of the values of environmental parameters at this location to a percentile distribution of values for those parameters at known locations of presences (training sites). The closer to the 50th percentile (median), the more likely a presence

will be at the location. There is no difference for the tails of the distribution, that is, 80th percentile is considered as equivalent to 20th percentile.

In the Domain algorithm (Hijmans and Graham, 2006), to evaluate the suitability of a location, the Gower distance (G) is computed for that location. Let x be a vector of values of the environmental parameters at a location, μ a vector of the means of these parameters at training sites, and t a vector of the ranges of these parameters at training sites. Then G is defined as the mean of the vector $\frac{|x-\mu|}{t}$, where the calculations are made component-wise. Therefore, G will range from 0 to 1. The Domain similarity statistic (D) is then calculated as $100\times(1-\mathrm{G})$, thus, D will range from 0 to 100, and a high value (e.g. 90) indicates a high likelihood of the species being present at the location.

The Mahal algorithm is similar to the Domain algorithm, except that it computes the Mahalanobis distance (De Maesschalck et al., 2000; Tsoar et al., 2007) between environmental parameters at a location and those parameters at training sites instead of the Gower distance. Mahalanobis distance consider the correlations of the parameters in the data set, and it is independent of the scale of measurements. The smaller the Mahalanobis distance, the greater the likelihood of a presence.

Maximum entropy (Maxent), is a machine learning method that consists of two components, a constraint component and an entropy component (Elith et al., 2011; Phillips et al., 2006). The constraint component is similar to the profile methods: it defines a set of constraints that represent the incomplete information for the probability distribution of the environmental parameters at the training sites. The entropy component estimates the unknown probability distribution, ensures that the estimation satisfies any constraints on the target distribution, and determines the best probability distribution among many distributions (e.g. uniform distribution) by maximizing the entropy. The Maxent method was proven to have good performance among the presence-only species distribution modes (Phillips et al., 2009; Vierod et al., 2014; Yackulic et al., 2013).

1.3 Species abundance modeling

Many regression and machine learning methods have been applied to build species abundance modeling (Guisan et al., 2002; Leathwick et al., 2006; Potts and Elith, 2006). In this study, two regression models (generalized linear models (GLM) and general additive models (GAM)) and one machine learning model (random forest (RF)) will be discussed.

GLM are an extension of linear models. They differ from regular linear models in that they use a probability distribution on the actual dependent variable to optimize the relationship between a transformation of the mean of that dependent variable and a linear combination of the independent parameters. The transformation, called a link function, is used to avoid forcing data into unnatural scales by describing the probability distribution of the data, such as normal, Poisson, binomial, negative binomial distribution. By considering the probability distribution, GLM fit better for non-constant and non-linearity variance structures of most ecological data than regular linear models, and they are more robust for ecological modeling (Guisan et al., 2002; Hastie and Tibshirani, 1990).

GAM are semi-parametric extensions of GLM with an additive smooth function. This makes GAMs strong at dealing with highly non-monotonic and non-linear data (Hastie and Tibshirani, 1990). Like GLM, GAM have also been used successfully for ecological modeling. For example, a GAM including factors such as depth, chlorophyll a, temperature, oxygen, and sediment type was used by Drexler and Ainsworth (2013) to predict pink shrimp abundance in the Gulf of Mexico.

The RF model is a machine learning method based on classification and regression trees (CART). Strengths of random forest models include out of bag (OOB) cross-validation and variable importance evaluation. The RF model are generally good for regression, clustering and classification problems (Breiman, 2001). An RF model was proven to perform better than GLM and GAM models when predicting bottlenose dolphin distribution in the northwestern Mediterranean Sea (Marini et al., 2015).

1.4 King mackerel (*Scomeromorus cavalla*) diet data modeling

Advancing the scientific foundation for fisheries and other marine resources management is important to improve our understanding for the Gulf of Mexico Large Marine Ecosystem. We need to strengthen our current capabilities of collecting, cataloguing, archiving, assessing, and understanding predator-prey relationships, which would help stock assessments for federally managed fish species, and improve ecosystem models being built for the purposes of getting better ecosystem-based fisheries management. To achieve those goals, the Gulf of Mexico Species Interaction database (GoMexSI) has been established by Simons et al. (2013a) to focus on trophic interactions of fishes. However, there are still numerous information gaps in the database. Accurate information for the spatial and temporal distributions of the target species, and anthropogenic impacts of these species are essential to develop management strategies for the conservation. Any lack of data for a specific species creates an information gap that impedes proper management actions. Thus, building models that identify and fill spatial gaps in species diet in the GOM will be critical to the success of the GoMexSi project. Spatial distribution modeling of king mackerel and its prey species is a first step to reach that goal.

1.4.1 King mackerel and its prey species distribution modeling

King mackerel (*Scomeromorus cavalla*) was selected as the first species for data compilation and diet data gap analysis because it is a top predator with high commercial and recreational value distributed nearly throughout the Gulf of Mexico. In addition, king mackerel is a species with fairly high Hg concentrations. Consumption of estuarine and marine fish is the primary way that humans are exposed to monomethylmercury (MMHg) (Sunderland et al., 2012). Many of these fishes (eg. mackerel, tunas) are high order carnivores at or near the top of the food chain. These fishes contain high concentrations of mercury (Hg) because of the bio-magnification process. Thus, it is important to research the food webs that include these species. With better knowledge of the ways Hg gets into the food chain and is transferred to predators, we can take measures to reduce the Hg at the source.

There have been a number of research projects focused on king mackerel diet data, and six areas in GOM have been recorded with detailed diet data (Simons et al., 2013b), but the majority of the GOM lacks diet data. Therefore, a gap analysis footprint of the diet data to fully understand the

diet for king mackerel throughout the GOM will be developed. Once the king mackerel diet data is successfully foot printed, other researchers will be able to map the overall trophic data of other species in the GOM, or other ecosystems. This will also provide solid information for Atlantis, EwE, and OSMOSE models that will ultimately inform the fisheries managers to better manage the GOM ecosystem. To reach the goal of king mackerel diet data modeling, prey species distribution and abundance will need to be known first. However, only a small amount of data for those prey species. has been recorded. Therefore, gap analysis modeling of the king mackerel and its prey species' distribution and abundance in the GOM will be of great importance, and this is what this study will focus on.

1.4.2 Three shrimp species abundance modeling

Besides the importance of king mackerel and its prey species distribution modeling, modeling their abundance will also be of great importance. However, good quality abundance data would require a great quantity of fisheries surveys, and of course result in a correspondingly high cost. Therefore, it is difficult to access good quality abundance data. Fortunately, the Southeast Monitoring and Assessment Program (SEAMAP) has annual fisheries survey across the northern Gulf of Mexico, and because they are species of great economic value, detailed abundance data for three common shrimps: *Farfantepenaeus aztecus* (brown shrimp), *F. duorarum* (pink shrimp) and *Litopenaeus setiferus* (white shrimp). For this study, we will use these three shrimp species as examples for building species abundance models in the Gulf of Mexico.

CHAPTER II: METHODS AND MATERIALS

2.1 Data collection and preprocessing

2.1.1 King mackerel and prey species distribution data

Overall 13626 occurrence records with specific coordinates for the target species (king mackerel and 31 prey species) were retrieved from Ocean Biogeographic Information System (OBIS, http://www.iobis.org/) and Global Biodiversity Information Facility (GBIF, http://www.gbif.org/). These data were used to build the species distribution models. King mackerel preys on more than 31 species; the 31 prey species used in this study are the most common prey species, with a good number of occurrence records in the Gulf of Mexico stored in the two databases. Data from the two resources were compiled into one .csv file for later analysis.

2.1.2 Species abundance data for three shrimps

In total, 46,372 geo-referenced abundance records of three shrimp species, *Farfantepenaeus aztecus* (brown shrimp), *F. duorarum* (pink shrimp) and *Litopenaeus setiferus* (white shrimp), were collected from Southeast Monitoring and Assessment Program (SEAMAP, http://www.gsmfc.org/default.php?p=sm_ov.htm). The unit of abundance is catch per unit effort (CPUE). The data were used to build species abundance models. To gain more insight whether the three shrimp species combined together as a functional group would have a different abundance distribution pattern from individual species, the sum of the abundance for the three shrimp species at each location was also analyzed (in the results below, the sum is named as "all shrimps").

2.1.3 Environmental parameters

Twelve predictors were collected from National Oceanographic Data Center (NODC, http://www.nodc.noaa.gov/OC5/GOMclimatology/). The predictors were: depth (m), four bottom types: sand, mud, rock and gravel concentrations (%) and eight annual average water quality parameters: temperature (°C), salinity (‰), silicate (mg/L), phosphate (mg/L), nitrate (mg/L), dissolved oxygen (mg/L), and oxygen saturation (%). For locations where the 12 predictors were not directly measured, values were interpolated using inverse distance weighting (IDW) method

with the resolution in 0.064° × 0.064° by using ArcGIS10.2. The interpolated predictors were compiled into one single file by using the "dismo" package in R 3.3.2.

2.2 Species distribution modeling

2.2.1 Model setup

Training and test dataset for occurrence data: For each species, to get a reasonable training and test dataset, five-fold cross validation was applied, i.e., 4/5 of the presence data were randomly selected to be training dataset, and the rest 1/5 of the presence data were saved to be test dataset. The training dataset was used to for fitting models, while the test dataset was used to evaluate the predictive capability of the models. The goal of this cross validation is to limit problems like overfitting: if there is a substantial difference in how the model performs on the training set versus the test set, then the model has probably fit random noise in the training set and is too complex.

Background points: For each species, 1000 background points were randomly created, five-fold cross validation is also applied to separate the background points into training and test dataset, i.e., 1/5 of the background points (200 points) were selected as test dataset to evaluating the model performance. Thus, the test set for a given model combines 1/5 of the actual presence data with 1/5 = 200 background points, while the training set is the combination of the remaining 4/5 of each set.

2.2.2 Model fitting on training datasets

For each species, the Bioclim, Domain, Mahal and Maxent algorithm were applied to build species distribution models. For each model, the 12 predictors and the training dataset of occurrence were used to build the model. In total, 128 models were built: four algorithms for each of the 32 species. The "dismo" package in R (3.3.2) software were used to fit the models. Details of the model fitting process may be found in the package tutorial (Hijmans and Elith, 2015).

2.2.3 Model evaluation on test datasets

To evaluate the performance for each model, the models were used to predict presence in the occurrence test dataset and the background test dataset. Area under the receiver operator curve (AUC) and correlation coefficient (r) were the parameters used to quantify the performance of each model.

Receiver operator curve (ROC) is a curve plotting true positive rate (TPR) against false positive rate (FPR), and AUC is the area under the ROC. AUC can be range from 0 to 1; higher AUC indicates better predictive performance of a model, while AUC = 0.5, indicates the model is just as good as a random guess.

Correlation coefficient calculating the correlation between predicting values (range from 0 to 1) and the presence and absence in the test dataset (absence comes from the background points, for example, 181 out of 200 background points were treated as absence). Correlation coefficient ranges from -1 to 1, higher positive correlation coefficient indicates the better performance of the model.

2.2.4 Model prediction

The presence-only species distribution models are threshold dependent. That indicates a threshold must be set first (e.g., 0.2). Predicted values (range from 0 to 1) greater than threshold indicate a prediction of presence, while values less than the threshold mean absence.

The threshold for each model was calculated as the threshold resulting in the maximum value for the sum of true positive rate and true negative rate (TPR+TNR). Each model would have raw predicted values (using predictors to predict presence and absence, so the resolution of the predicted value is the same as the predictors, i.e., $0.064° \times 0.064°$). After filtering by the threshold, the model would have the final output, that is presence (raw predicted value greater than threshold) and absence (raw predicted value less than threshold).

2.2.5 Performance comparison for the four algorithms

One-way analysis of variance (ANOVA) with Tukey post hoc testing was applied to compare the AUC and r for each of the algorithm. Each algorithm had 32 models for the 32 species. The result showed whether one algorithm had significantly better AUC or r than the other algorithms.

2.3 Abundance modeling for three shrimp species

2.3.1 Model setup

Predictors for model fitting: unlike the species distribution modeling above, instead of being used as a single file, the 12 predictors would be used as 12 individual predictors here, and each of the predictors would be extracted to the points (46372) where abundance were recorded, thus, the new data set would include species name, abundance, latitude, longitude and the corresponding predictors for each of the 46372 records (i.e., dimension: 46372×16).

Predictors for model prediction: like the species distribution modeling above, the single raster stack of the 12 predictors would be used to predict the abundance of each shrimp species across the Gulf of Mexico. The raster stack of the predictors was used because it is evenly distributed in the Gulf with high resolution. In this case, the model prediction outputs would cover everywhere in Gulf, and is not just limit to the locations where the abundance data were recorded.

Training and test dataset: similar to species distribution modeling above, three-fold cross validation would also be used to separate the data into training and test dataset, i.e., 2/3 of the data would be training dataset while the rest 1/3 as test dataset.

2.3.2 Model fitting on training datasets

For each species, GLM, GAM and RF algorithms were applied to build the species abundance models, respectively. In each model, the training datasets were used applying the 12 predictors to fit the abundance. There are three algorithms for four species, so in total 12 models would be created. Since the abundance is count related data, the negative binomial was used as the underlying model in the GLM, with a log link function.

2.3.3 Model evaluation on test datasets

Correlation coefficient calculating the correlation between predicted abundance and the abundance values in the test datasets would be used to evaluate the performance for each model. Correlation coefficient ranges from -1 to 1, higher positive correlation coefficient indicates the better performance of the model.

2.3.4 Model prediction

The raster stack of the 12 predictors would be used applying each of the model built to predict the abundance distribution for each shrimp species across the Gulf of Mexico.

2.3.5 Performance comparison for the three algorithms

One-way analysis of variance (ANOVA) with Tukey post hoc test was applied to compare the correlation coefficient for each of the algorithm, each algorithm would have four models for the four species (three shrimp species + all shrimps), the result would show whether one algorithm has significantly better correlation coefficient than the other algorithm or not.

CHAPTER III: RESULTS AND DISCUSSIONS

3.1 Species distribution modeling results and discussions

3.1.1 Species list and corresponding presence data collected

Detail on the number of occurrences for each species (32 species from 15 families) are listed in Table 1. Ballyhoo halfbeak had the lowest number of occurrences obtained (121), while Atlantic croaker had the highest (1256). Totally, 13625 occurrences data were collected for the 32 species, and the average was 426 for each species.

Table 1: Species list and corresponding number of occurrences

Family name	Species name	Common name	Number of occurrences
Carangidae	*Decapterus punctatus*	Round scad	456
Carangidae	*Chloroscombrus chrysurus*	Atlantic bumper	231
Carangidae	*Caranx crysos*	Blue runner	363
Clupeidae	*Sardinella aurita*	Spanish sardine	248
Clupeidae	*Brevoortia patronus*	Gulf menhaden	613
Clupeidae	*Opisthonema oglinum*	Atlantic thread herring	457
Engraulidae	*Anchoa hepsetus*	Striped anchovy	857
Haemulidae	*Orthopristis chrysoptera*	Pigfish	417
Hemiramphidae	*Hemiramphus brasiliensis*	Ballyhoo halfbeak	121
Hemiramphidae	*Hyporhamphosus unifasciatus*	Common halfbeak	126
Loliginidae	*Lolliguncula brevis*	Atlantic brief squid	442
Loliginidae	*Loligo pealei*	Longfin inshore squid	154

Lutjanidae	*Lutjanus campechanus*	Red snapper	502
Lutjanidae	*Rhomboplites aurorubens*	Vermilion snapper	260
Lutjanidae	*Lutjanus synagris*	Lane snapper	439
Mugilidae	*Mugil cephalus*	Striped Mullet	453
Mugilidae	*Mugil curema*	White mullet	246
Mullidae	*Upeneus parvus*	Dwarf goatfish	424
Penaeidae	*Litopenaeus setiferus*	White shrimp	220
Penaeidae	*Farfantepenaeus aztecus*	Brown shrimp	635
Penaeidae	*Farfantepenaeus duorarum*	Pink shrimp	330
Penaeidae	*Rimapenaeus similis*	Roughback shrimp	479
Scaridae	*Nicholsina usta*	Emerald parrotfish	149
Sciaenidae	*Cynoscion arenarius*	Sand seatrout	183
Sciaenidae	*Micropogonias undulatus*	Atlantic croaker	1256
Sciaenidae	*Leiostomus xanthurus*	Spot croaker	939
Scombridae	*Euthynnus alletteratus*	Little tunny	464
Scombridae	*Sarda sarda*	Atlantic bonito	461
Scombridae	*Scomberomorus maculatus*	Spanish mackerel	345
Scombridae	*Scomberomorus cavalla*	King mackerel*	715
Serranidae	*Diplectrum bivittatum*	Dwarf sand perch	457
Trichiuirdae	*Trichiurus lepturus*	Atlantic cutlassfish	183

Notes: "*" indicates the predator (king mackerel); the remaining 31 species are all prey.

3.1.2 Spatial distribution of interpolated environmental parameters

Figure 1 shows the spatial patterns of the annual average for the 12 predictors. The following are brief description for the spatial pattern for each of the 12 predictors in the Gulf of Mexico: **Depth**: range from 0 to 7900 m, being classified into three categories: continental shelf (depth less than 200 m), continental slope (depth between 200 m to 2000 m) and open sea (depth greater than 2000 m); **Temperature**: range from about 18 to 31 °C, the western Gulf has relatively low temperatures compared to the eastern Gulf; **Salinity**: range from 5 to 37, evenly distributed in the Gulf with salinity 35, with low salinity off Louisiana-Mississippi estuaries; **Oxygen**: range from 4 to 7 mg/L, evenly distributed with 5 mg/L in the Gulf, while the while relatively higher (7 mg/L) off Louisiana-Mississippi estuaries; **Oxygen saturation**: range from 90 to 120 %, evenly distributed; **Nitrate** (0.01 to 4.6 mg/L) and **Silicate** (0.2 to 8.4 mg/L) has similar patterns, both has higher level off Louisiana-Mississippi estuaries; **Phosphate**: range from 0.02 to 0.67 mg/L, highest in the Louisiana-Mississippi estuaries, the western Gulf has the medium level, while the eastern Gulf has the lowest level; **Bottom type-Clay** and **Mud** has similar pattern, higher percentage in the open sea, while lower percentage in the coastal area. Higher clay percentage in the western Gulf, while low clay percentage in the eastern Gulf; **Gravel** has overall low percentage in the Gulf, the southern Gulf has higher percentage than the northern and medium Gulf; **Sand** shows opposite patterns to the clay and mud: the eastern Gulf has the highest percentage, followed by the western Gulf, lowest in the open sea. I note that the sum of concentration for the four bottom types at one location should be less or equal to 100%, however, the sum of concentration of clay and mud in the open sea is higher than 100%, this should be because of the interpolation of each bottom type was done separately. The problem might be solved by setting the sum of bottom type in one location equal to 100% when doing the interpolation, however, if the setting processed, the collinearity problem which two or more predictor parameters are highly correlated would happen. Therefore, to avoid collinearity, there is no setting of the sum of bottom types concentration when processing the spatial interpolation.

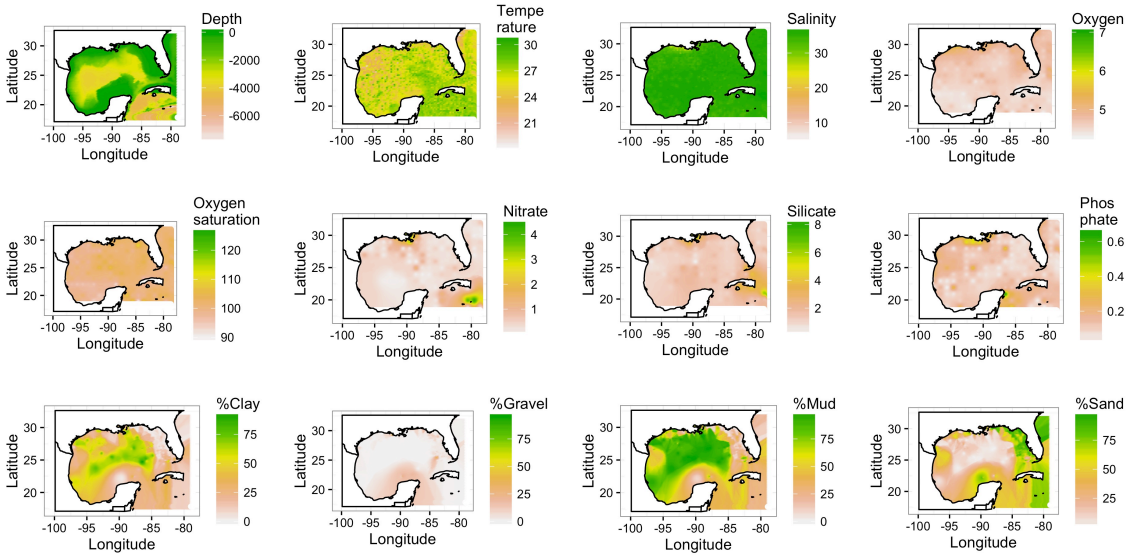

Figure 1: Spatial distribution of the 12 predictors in the Gulf of Mexico

3.1.3 Species distribution model visualization for selected species

The model visualization for all the 32 species are not included in the main text, only nine selected species results are listed, the visualization for the remaining of 23 species are included in the Appendix A. **King mackerel** is selected because it is the predator; **White shrimp, brown shrimp and pink shrimp** are selected because they are the three shrimp species used in the species abundance modeling; **Atlantic bonito** is selected because it is the species with most different distribution pattern; Clupeidae, Hemiramphidae, Sciaenidae, Trichiuirdae were reported to be the families that king mackerel eat the most in the Gulf of Mexico (Simons et al., 2013b). **Gulf menhaden** (Clupeidae), **common halfbeak** (Hemiramphidae), **Atlantic croaker** (Sciaenidae) **and Atlantic cutlassfish** (Trichiuirdae) are selected because they are the species with the highest number of occurrences (see Table 1) for the corresponding families in this study.

Figure 2 shows all four models predicted that king mackerel was mainly distributed at shallower depths (most within continental slope while only some in open sea, refer to Figure 1). However, it is unlikely to be occur in the dead zone-hypoxic area at the mouth of Mississippi River, the north-east or the south Gulf. Domain model predicted the widest distribution followed by the Bioclim,

Maxent and then Mahal; Mahal model had the best AUC (0.96), followed by Maxent, Bioclim and then Domain; while Maxent had the best r (0.66), followed by Domain, Bioclim, and then Mahal.

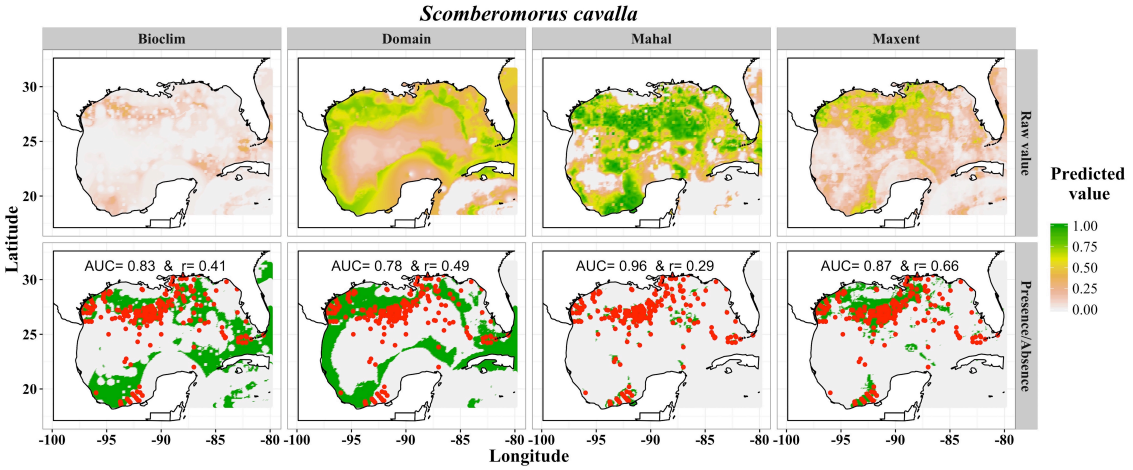

Figure 2: Visualization of king mackerel (*Scomberomorus cavalla*) distribution models.

Notes: "Raw value" indicates predicted raw values for each algorithm based model; "Predicted/Absence" indicates presence/absence predicted by filtering the raw values using a threshold as described above. The values in the color bar range from 0 to 1, 0 (gray color) indicates absence while 1 (green color) indicates presence. Red dots represent the king mackerel presence records used to build the models.

Figure 3 shows all four models predicted that white shrimp was primarily distributed within the continental shelf. Domain model predicted the widest distribution followed by the Mahal, Maxent and then Bioclim; Domain model predicted white shrimp distributed along the western, northern and eastern Gulf, but not in the south zone- coastal area off Merida, Mexico; while Maxent and Bioclim model predicted white shrimp to be primarily distributed in the northern Gulf. Maxent model had the best AUC (0.96), followed by Mahal, Bioclim and then Domain; for correlation coefficient, Maxent also the best r (0.66), followed by Bioclim, Domain, and then Mahal.

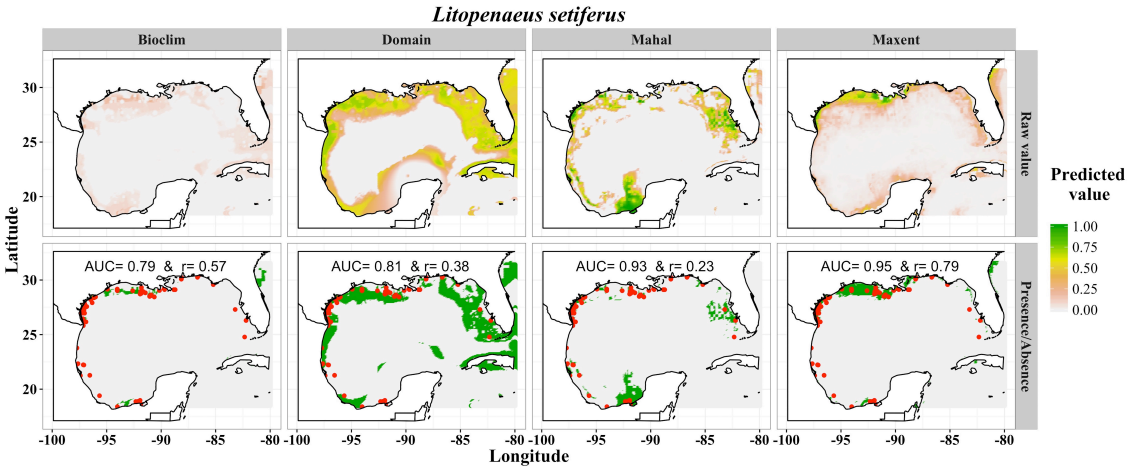

Figure 3: Visualization of white shrimp (*Litopenaeus setiferus*) distribution models.

Similar to white shrimp, Figure 4 shows all four models predicted that brown shrimp was mainly distributed within continental shelf. Domain model predicted the widest distribution followed by the Bioclim, Maxent and then Mahal. Domain model predicted brown shrimp distributed all along the coast of the Gulf (i.e., western, northern, eastern and southern Gulf), but absent from the south zone and northern east coastal shelf of the Gulf; all the rest three model predicted brown shrimp primarily distributed only in the western and northern Gulf, but absent from southern and eastern Gulf. Mahal model had the best AUC (0.97), followed by Maxent, Bioclim and then Domain; while Maxent had the best r (0.85), followed by Bioclim, Domain and then Mahal.

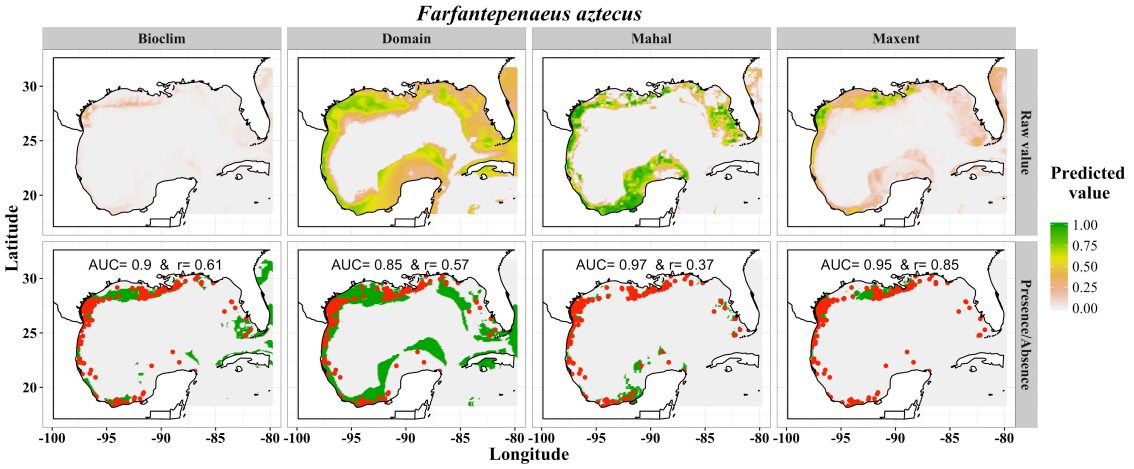

Figure 4: Visualization of brown shrimp (*Farfantepenaeus aztecus*) distribution models.

Figure 5 shows all four models predicted that pink shrimp was mainly distributed within continental shelf and slope, which was relative wider than white and brown shrimp. Domain model predicted the widest distribution followed by the Bioclim, Maxent and then Mahal. All the models predicted pink shrimp to be distributed all along the Gulf but absent from the dead zone at the mouth of Mississippi River. Mahal model had the best AUC (0.94), followed by Maxent, Bioclim and then Domain; while Maxent had the best r (0.74), followed by Domain, Bioclim, and then Mahal.

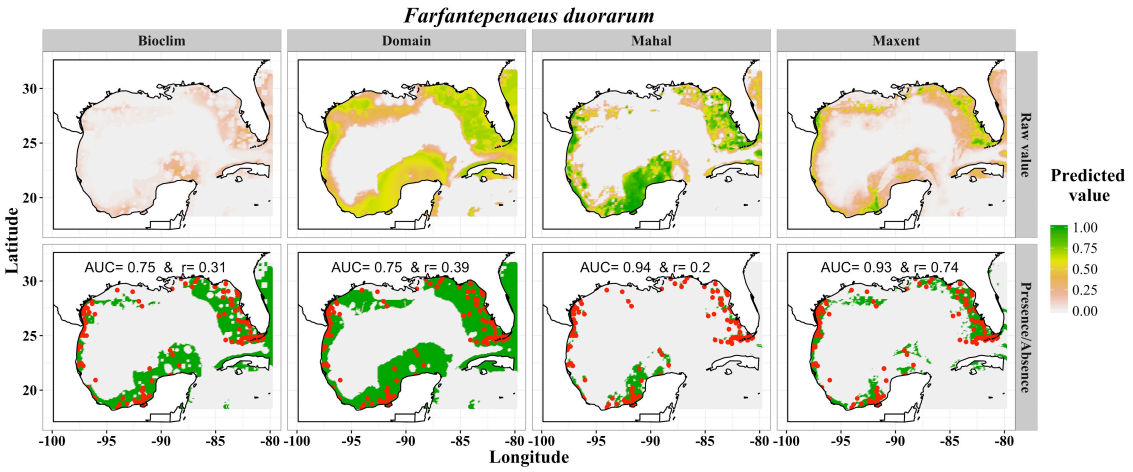

Figure 5: Visualization of pink shrimp (*Farfantepenaeus duorarum*) distribution models.

Figure 6 shows all four models predicted that Atlantic bonito was mainly distributed in the open sea, which was quite different from the rest of the 31 species (king mackerel had the closest distribution patterns). Mahal model had the best AUC (0.93), followed by Maxent, Bioclim and then Domain; while Maxent had the best r (0.72), followed by Bioclim, Domain and then Mahal.

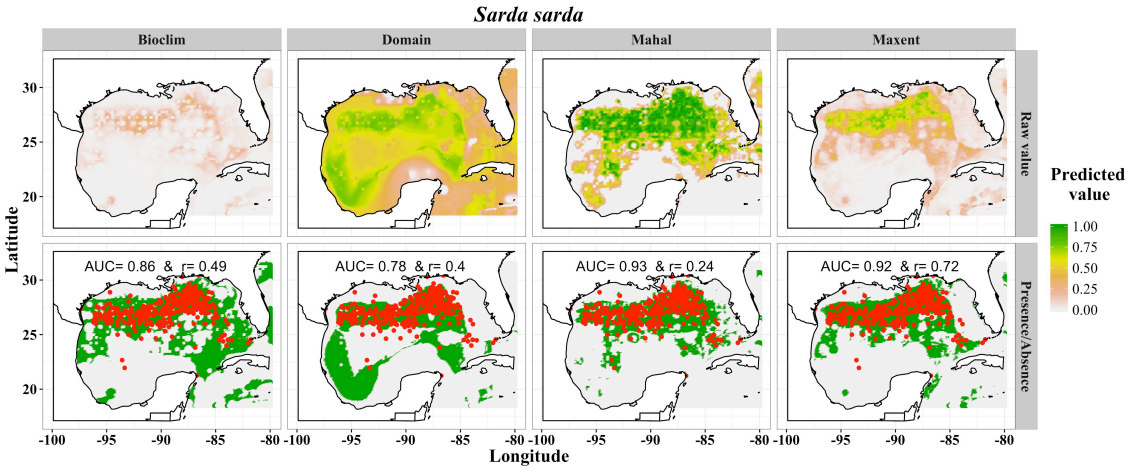

Figure 6: Visualization of Atlantic bonito (*Sarda sarda*) distribution models.

Figure 7 shows all four models predicted that gulf menhaden was mainly distributed in the continental shelf. Domain model predicted the widest distribution that gulf menhaden distributed all along the Gulf, while the rest three models all predicted gulf menhaden only occur in the northern Gulf (absent from the southern Gulf). The modeling results fit the gut content analysis results (king mackerel eats Clupeidae primarily in the northern Gulf) by Simons et al. (2013b). Mahal model had the best AUC (1), followed by Maxent, Bioclim and then Domain; while Maxent had the best r (0.89), followed by Bioclim, Domain and then Mahal.

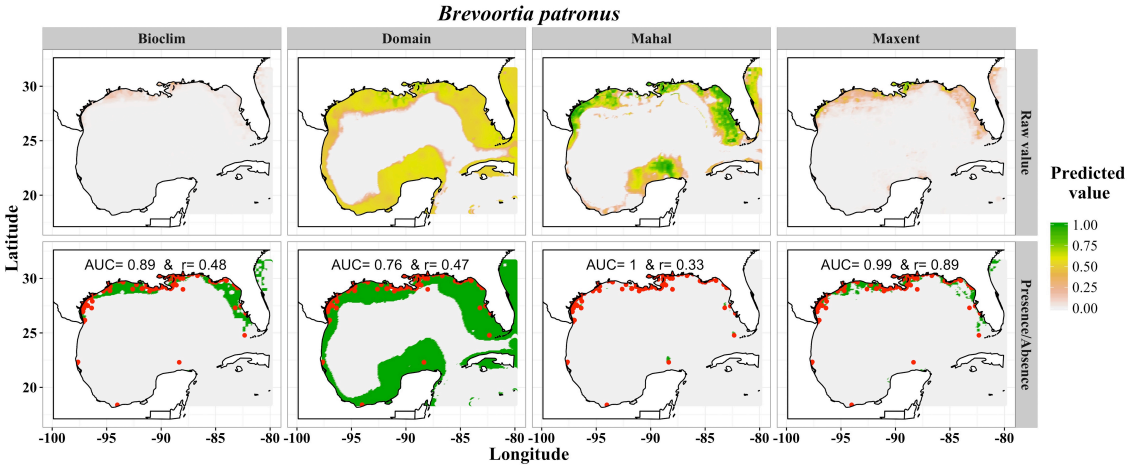

Figure 7: Visualization of gulf menhaden (*Brevoortia patronus*) distribution models.

Figure 8 shows all four models predicted that common halfbeak was mainly distributed in the continental shelf. Bioclim, Mahal and Maxent models all predicted that common halfbeak primarily distributed in the eastern and southern Gulf, but absent from western Gulf and the dead zone, while Domain model predicted similar but wider distribution that common halfbeak also distributed in the western Gulf. The modeling results fit the gut content analysis results (king mackerel eats Hemiramphidae primarily in the eastern Gulf, no data from southern Gulf was included in their analysis) by Simons et al. (2013b). Mahal and Maxent models had the best AUC (0.89), followed by Bioclim and then Domain; while Maxent had the best r (0.51), followed by Bioclim, Domain and then Mahal.

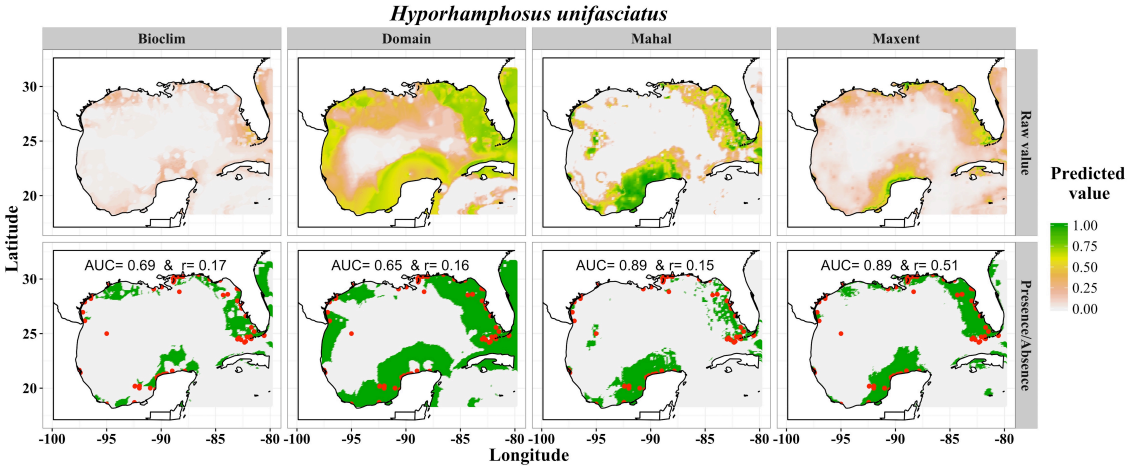

Figure 8: Visualization of common halfbeak (*Hyporhamphosus unifasciatus*) distribution models.

Figure 9 shows all four models predicted that Atlantic croaker was mainly distributed in the northern and western continental shelf. Bioclim, Mahal also predicted the distribution covers eastern and southern Gulf (but absent from south zone), while Mahal and Maxent predicted Atlantic croaker was absent from eastern and southern Gulf. The modeling results fit the gut content analysis results (king mackerel eats Sciaenidae primarily in the northern Gulf, no data from southern Gulf was included in their analysis) by Simons et al. (2013b). Mahal model had the best AUC (0.99), followed by Maxent, Bioclim and then Domain; while Maxent had the best r (0.81), followed by Domain, Bioclim and then Mahal.

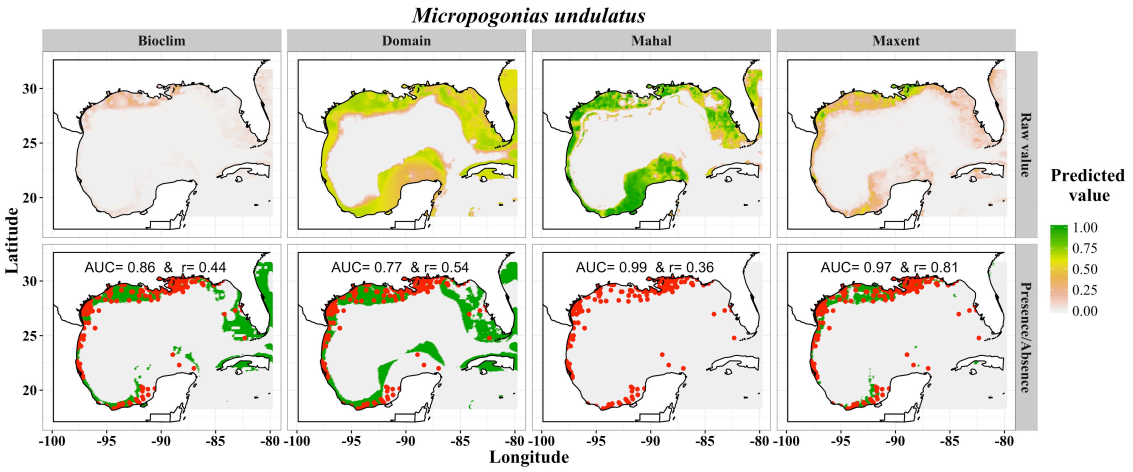

Figure 9: Visualization of Atlantic croaker (*Micropogonias undulates*) distribution models.

Figure 10 shows all four models predicted that Atlantic cutlassfish was mainly distributed within the continental shelf. Domain and Mahal predicted that Atlantic cutlassfish occurs all along the Gulf, while Bioclim and Maxent predicted that Atlantic cutlassfish primarily occurs in the northern Gulf. The modeling results fit the gut content analysis results (king mackerel eats Trichiuridae primarily in the northern Gulf) by Simons et al. (2013b). Maxent model had the best AUC (0.96), followed by Mahal, Domain and then Bioclim; while Maxent had the best r (0.81), followed by Bioclim, Domain and then Mahal.

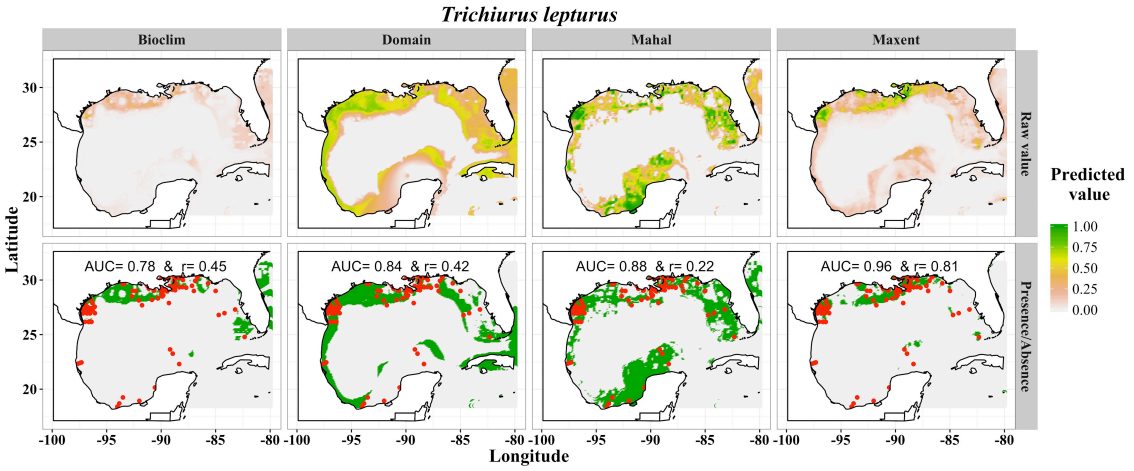

Figure 10: Visualization of Atlantic cutlassfish (*Trichiurus lepturus*) distribution models.

3.1.4 Predicted distribution patterns summary

Figures similar to Figure 2-10 for the remaining 23 prey species can be found in Appendix A. I here summarize the results of the distribution models for all the 32 species.

For each species, the four models had similar predicted distribution, but Domain models always predicted wider distribution ranges, while Bioclim and Maxent models predicted medium ranges, and Mahal always predicted the most narrow ranges.

Different distributions were predicted for the 32 species, most of the species distributed in continental shelf (depth less than 200 m) or slope (depth less than 2000 m), while only three species (Atlantic bonito, blue runner and king mackerel) were predicted to live in the open sea (depth greater than 2000 m). Some species were predicted to occur all along the coastal area of the Gulf, while some only distributed in the eastern coastal area, and others only along the Northern coast. Interestingly, the three "special zones" where king mackeral was not predicted to be found—the dead zone at the mouth of the Mississippi River, the north-east corner of the Gulf, and the southern Gulf—recur as areas of absence for many prey species.

In total, 10 major distribution patterns based on three categories (depth, west-east-north-south Gulf, and special zones) were proposed to describe the distribution of the 32 species (details are listed in Table 2). For example, the sediments in the western Gulf have high mud concentration while the eastern Gulf is very sandy. Some species like brown shrimp prefer muddy bottoms, so they were predicted to appear in the western Gulf rather than the eastern Gulf. The dead zone has high nitrate, phosphate and silicate concentration while the salinity is low, and some of the species don't thrive in the conditions there.

Since the four models had different predictions, some models predicted one species to live in one area, while some other models predicted not. Thus, five levels (0, 1/4, 1/2, 3/4, 1) were used to show how many models out of the total four models predicted one species to be present in one specific area. For example, 1/4 indicates that only one out of the four models predicted a given species to be present in the area. The overall summary of the distribution patterns are displayed in Figure 11.

Table 2: Distribution patterns predicted for the 32 species

Species name	Common name	C.Shelf	C.Slope	Open. Sea	West. Gulf	East. Gulf	North. Gulf	South .Gulf	Dead. Zone	North.East. Zone	South. Zone
Decapterus punctatus	Round scad	1	1	0	1/2	1	1	1	0	1	0
Chloroscombrus chrysurus	Atlantic bumper	1	1	0	1/4	1/2	1	1/2	1	1/2	0
Caranx crysos	Blue runner	1	1	1	1/2	1	1	1	0	3/4	1/2
Sardinella aurita	Spanish sardine	1	1	0	1	1	1	1	0	3/4	1
Brevoortia patronus	Gulf menhaden	1	1/2	0	1/4	1	1	1/4	1	1	1/4
Opisthonema oglinum	Atlantic thread herring	1	1	0	1/2	1	1	1	1	1/2	1/2
Anchoa hepsetus	Striped anchovy	1	1/2	0	1/4	1	1	1	1	1/4	1
Orthopristis chrysoptera	Pigfish	1	1	0	1/2	1	1	1	3/4	1	1/2
Hemiramphus brasiliensis	Ballyhoo halfbeak	1	1	0	1/4	1	1	1	0	1/2	1/4
Hyporhamphosus unifasciatus	Common halfbeak	1	1	0	1/4	1	1	1	0	3/4	1
Lolliguncula brevis	Atlantic brief squid	1	1/4	0	0	1	1	0	0	1/4	0
Loligo pealei	Longfin inshore squid	1	1	0	3/4	1	1	1	0	1	1/4
Lutjanus campechanus	Red snapper	1	1	0	1	1	1	1	1	1/4	0
Rhomboplites aurorubens	Vermilion snapper	1	1	0	1	1	1	1	0	3/4	1

Lutjanus synagris	Lane snapper	1	1	0	1	1	1	1	1/2	1/4	1/2
Mugil cephalus	Striped Mullet	1	1	0	1/4	1	1	1/4	1/2	3/4	0
Mugil curema	White mullet	1	1	0	1	1	3/4	1	1/2	1/4	0
Upeneus parvus	Dwarf goatfish	1	1	0	1	1	1	1	0	0	0
Litopenaeus setiferus	White shrimp	1	1	0	1/2	1/4	1	1/4	1	0	0
Farfantepenaeus aztecus	Brown shrimp	1	1	0	1	1/4	1	1/2	1	0	0
Farfantepenaeus duorarum	Pink shrimp	1	1	0	1	1	3/4	1	0	1	1
Rimapenaeus similis	Roughback shrimp	1	1	0	3/4	1/2	1	1/2	1/2	0	1/4
Nicholsina usta	Emerald parrotfish	1	1	0	1/4	1	0	1	0	1/4	1/2
Cynoscion arenarius	Sand seatrout	1	1	0	1/4	1/2	1	0	1/2	3/4	0
Micropogonias undulatus	Atlantic croaker	1	1	0	1	1	1	3/4	3/4	0	0
Leiostomus xanthurus	Spot croaker	1	1	0	1	1	1	1/4	1	1	1/4
Euthynnus alletteratus	Little tunny	1	1	0	1/2	1	1	1	1/4	0	0
Sarda sarda	Atlantic bonito	0	1	1	1/2	1/2	1/2	0	0	0	0
Scomberomorus maculatus	Spanish mackerel	1	1	0	1/4	1/2	1	1	3/4	1/4	0
Scomberomorus cavalla	King mackerel	1	1	1	1/4	1/2	1	1/2	0	0	0

Diplectrum bivittatum	Dwarf sand perch	1	1	0	1	1	1	1	1	1/4	0
Trichiurus lepturus	Atlantic cutlassfish	1	1	0	1/2	1	1	1/2	1/2	0	1/4

Notes: 10 distribution patterns were listed to summarize the distribution of the 32 species predicted by the models. “C.Shelf ” indicates the continental shelf with depth less than 200 m; “C.Slope” indicates the continental slop with depth between 200 m and 2000 m; “Open.Sea” indicates the open sea area with depth higher then 2000 m; “West.Gulf” indicates the western part of the Gulf of Mexico; “East.Gulf” indicates the eastern part of the Gulf of Mexico; “North.Gulf” indicates the northern part of the Gulf of Mexico; “South.Gulf” indicates the southern part of the Gulf of Mexico; “Dead.Zone” indicates the Gulf of Mexico hypoxic zone at the mouth of Mississippi river; “North.East.Zone” indicates the northern east coastal shelf of the Gulf; “South.Zone” indicates the coastal area off Merida, Mexico. The numbers in the table indicates the frequency of the presence in the 10 listed zones out of the four models prediction, “0” indicates zero out of the four models predicted that the species is being presented in the zone; while “1/4” indicates one out of four, and “1/2” for two out of four, and “3/4” for three out of four, and “1” for four out of four.

Figure 11 shows that most of the species were present in continental shelf or slope, except Atlantic Bonito, which presented mostly in the open sea area, and sometimes in continental slope. King mackerel and blue runner were both present mostly within continental slope, and some time in the open sea. Most of the species were present in the northern and eastern Gulf, while some of them not in the western or southern Gulf, e.g, sand seatrout was predicted to be absent in the southern Gulf. There were also some species were predicted to be absent in either one or two or all of the special zone, e.g. dwarf goatfish and king mackerel were predicted to be absent in all the three special zones, while white shrimp and brown shrimp predicted to be absent in north east zone and south zone, but appeared in the dead zone.

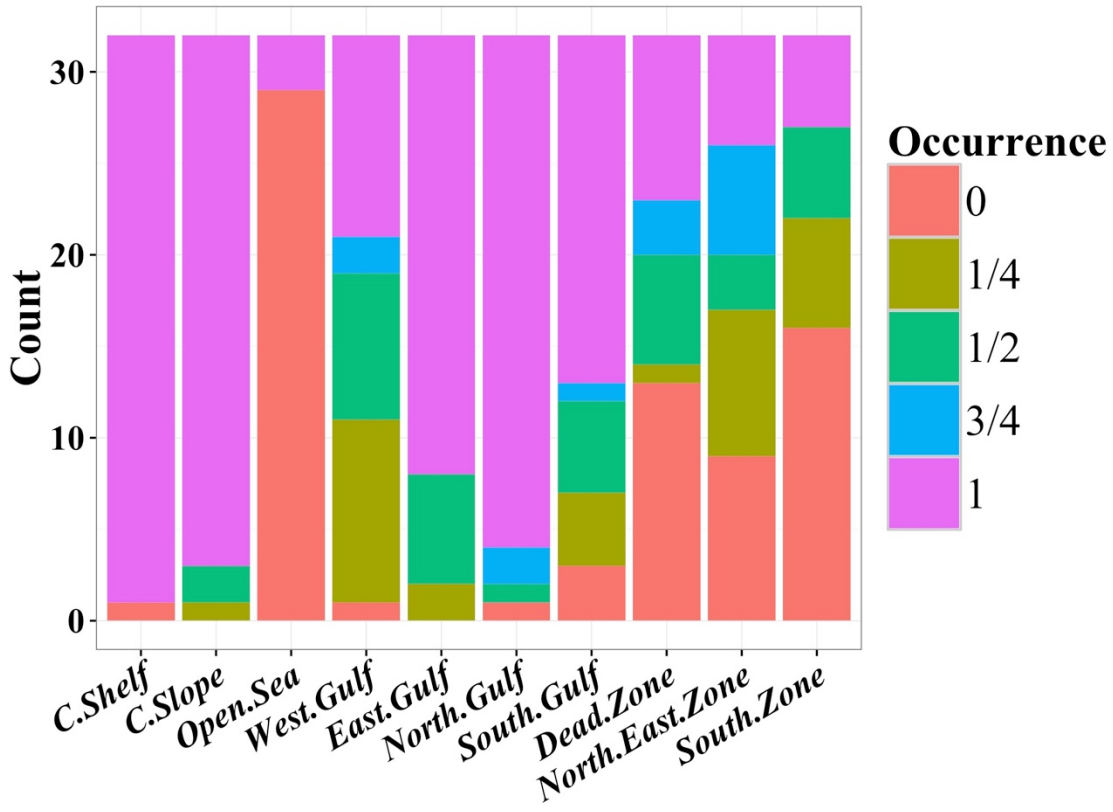

Figure 11: Summary of the predicted distribution patterns

3.1.5 Model evaluation and comparison for each algorithm

Area under curve (AUC) and correlation coefficient (r) were used to evaluate the model performances. Figure 12 shows the AUC of the models based on these four algorithms were 0.83±0.07, 0.77±0.08, 0.94±0.03 and 0.94±0.03, respectively; while Figure 13 shows the r for the models were 0.47±0.13, 0.43±0.12, 0.27±0.08 and 0.76±0.08, respectively. Figure 14 shows the Post hoc with Tukey's test comparing the performance of each models, it indicates that AUC for the Maxent-based models were significantly ($p<0.05$) higher than those for Bioclim and Domain based models, but insignificantly different from those for Mahal-based models ($p=0.955$); while Figure 15 shows that r for the Maxent-based models were significantly higher than those for all the other three types of models ($p<0.05$). Thus, we conclude that the Maxent-based models had the best performance for this data.

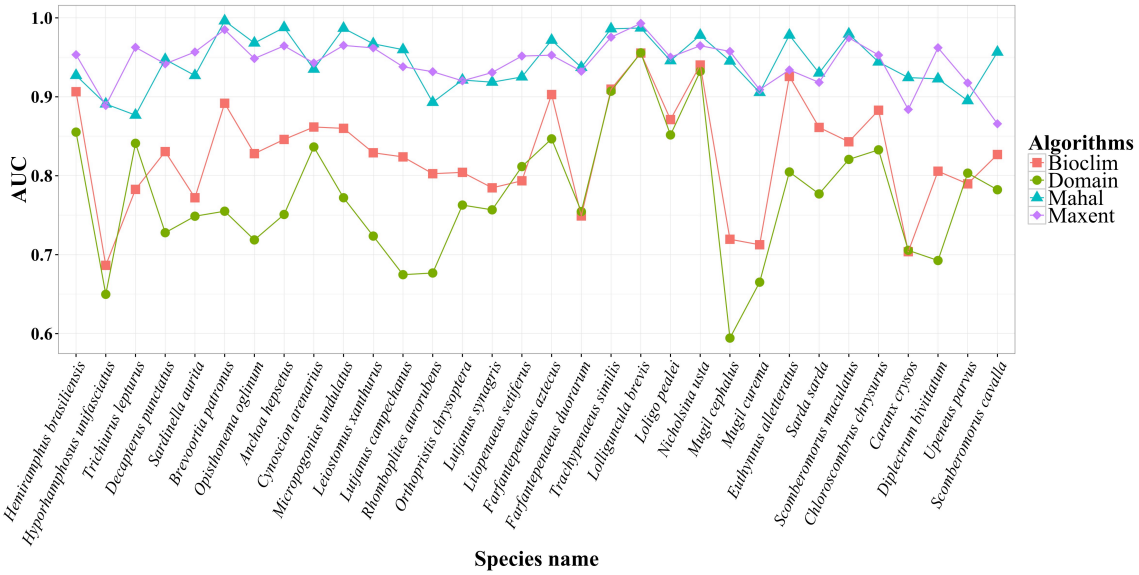

Figure 12: Model performances evaluated by AUC for each of the 128 (four Algorithms × 32 species) models.

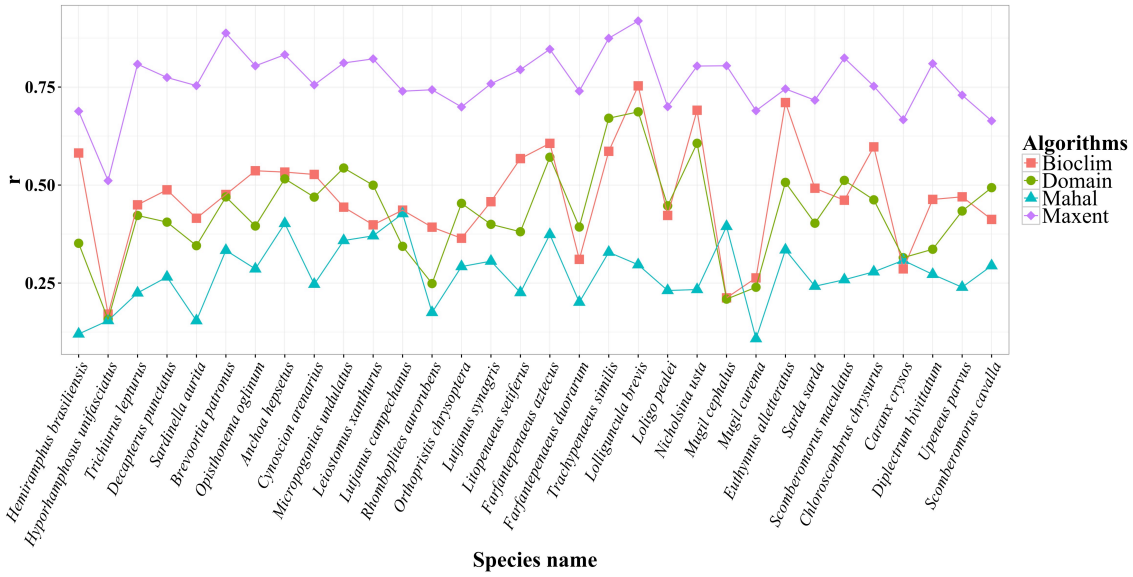

Figure 13: Model performances evaluated by r for each of the 128 (four Algorithms × 32 species) models.

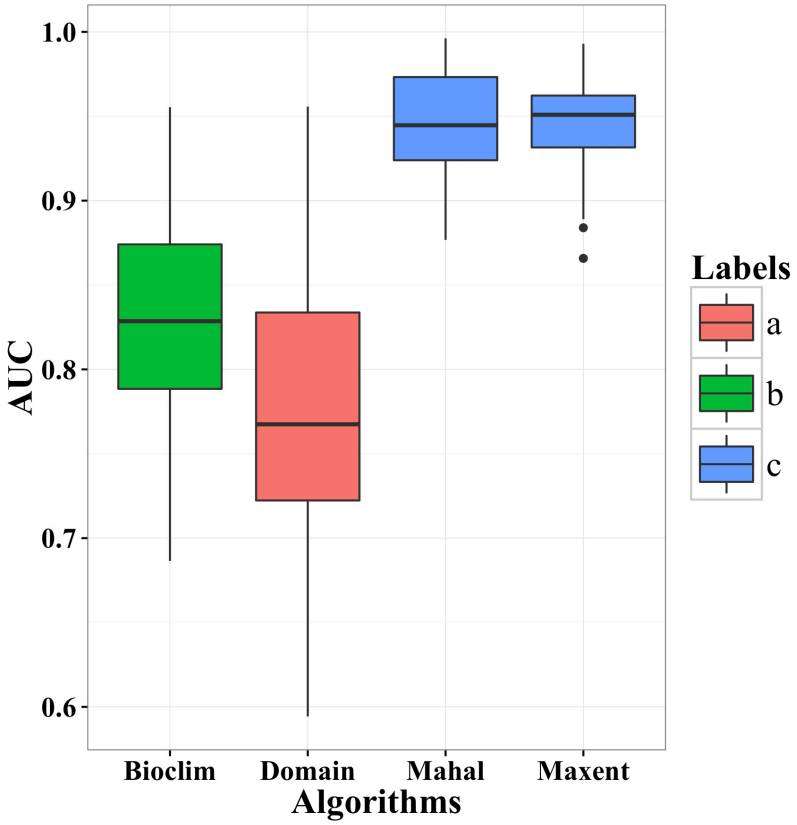

Figure 14: Boxplot results of one-way ANOVA comparing AUC from each of the four algorithms.

Notes: Algorithms were compared by Westfall's extension of the Tukey HSD test using alpha=0.05. Algorithms labeled with the same letter were not significantly different.

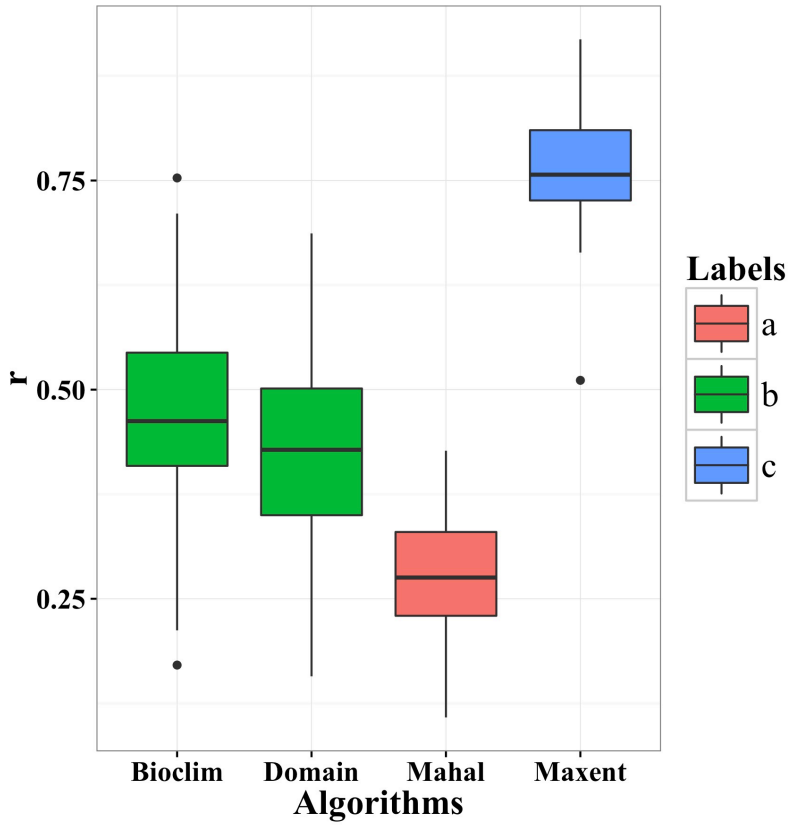

Figure 15: Boxplot results of one-way ANOVA comparing r from each of the four algorithms.

Notes: Algorithms were compared by Westfall's extension of the Tukey HSD test using alpha=0.05. Algorithms labeled with the same letter/color were not significantly different.

3.2 Shrimp species abundance modeling results and discussions

Figure 16 shows the abundance distribution predicted for the three shrimps (and all three shrimps combined) by GLM, GAM and RF algorithms. The original data without modeling is also included as "No model". Some general patterns can be concluded for the three shrimps' abundance distribution: brown shrimp was mostly distributed in the western Gulf with nearly zero abundance for eastern Gulf; pink shrimp shows the opposite distribution to brown shrimp, it was mostly

distributed in the eastern Gulf, while nearly no abundance in the western Gulf; white shrimp shows similar patterns to brown shrimp, but it was only distributed in the northern Gulf, with no distribution in the rest of the Gulf. We note that the GAM model predicted pink shrimp to have high abundance (300 to 400 CPUE) in the continental slope and open sea, however, it seemed to be over predicting, when compared to other modeling results and the original data before modeling, the abundance in the continental slope and open sea should be very low or even zero.

Figure 17 shows that for brown shrimp, the GAM model has the best r, while for other shrimps, the RF model has the best performance.

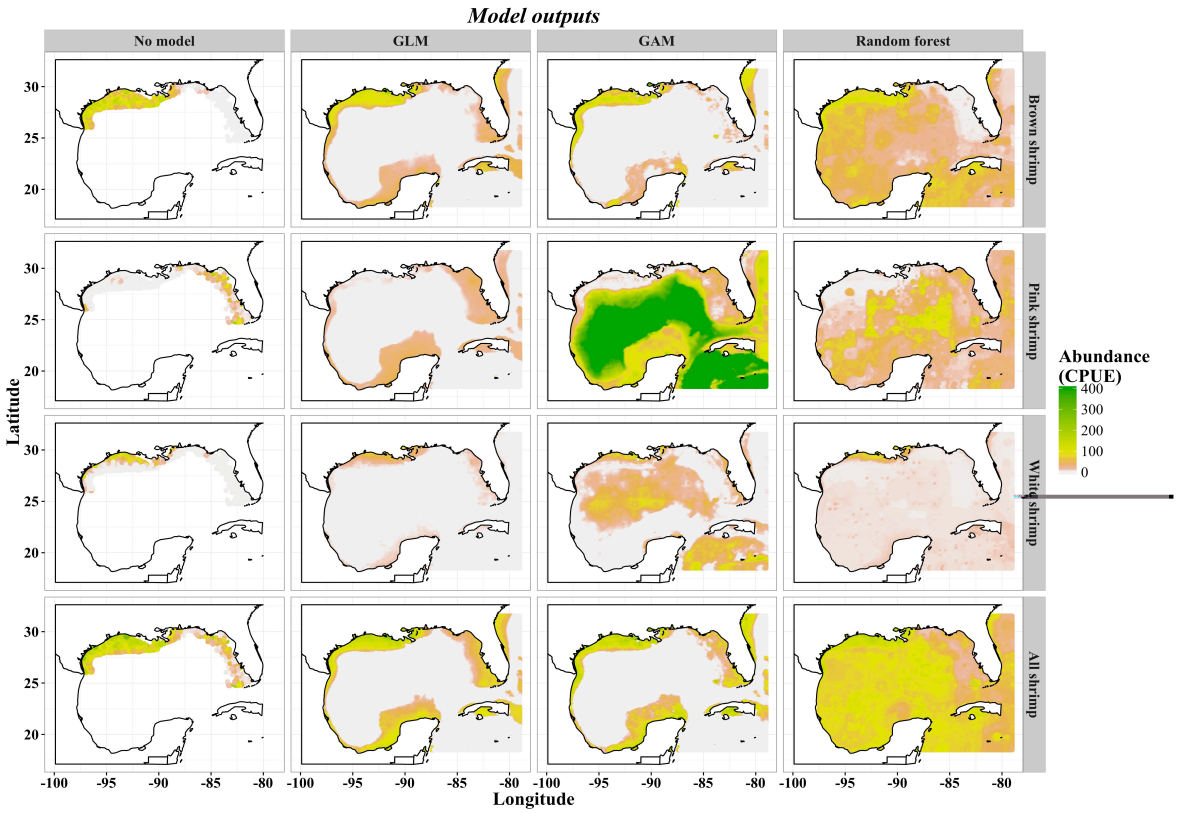

Figure 16: Model outputs visualization for the three shrimp species based on the three algorithms

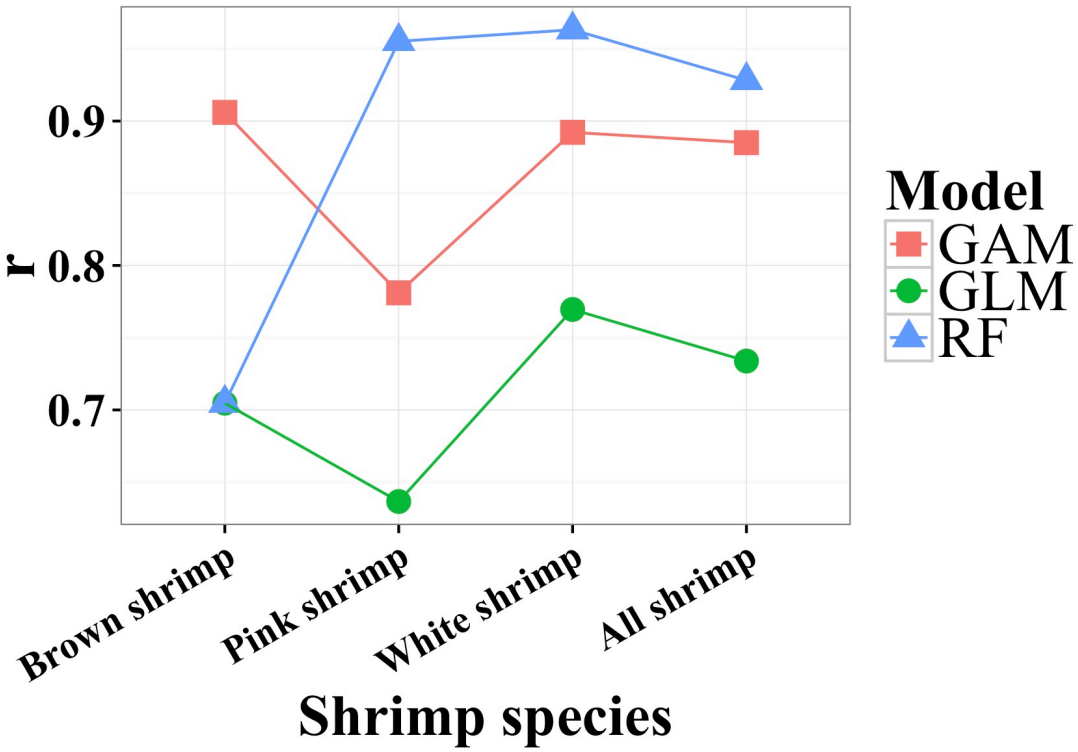

Figure 17: Model performance comparison by coefficient correlations (r) for the three shrimp species based on the four algorithms

CHAPTER IV: CONCLUSIONS

This study introduced four presence-only species distribution algorithms (Bioclim, Domain, Mahal and Maxent) to do spatial gap analysis for king mackerel and its prey species in the Gulf of Mexico. Overall, the four models predicted similar distribution for each species, the Domain-based model predicted the widest distribution ranges, followed by the Bioclim and Maxent-based models, while the Mahal-based model predicted the narrowest ranges. Post hoc analysis on AUC and r indicates that Maxent-based model has the best performance.

Ten distribution patterns based on three categories (depth, west-east-north-south Gulf, and special zones) were proposed to describe the distribution of the 32 species studied. Most of the species were distributed in continental shelf or slope, while only 3 species (Atlantic bonito, blue runner and king mackerel) were predicted to live in the open sea. Some species were present all along the Gulf, while some were only distributed in eastern Gulf, and some only in northern Gulf; some species like king mackerel were predicted not to be present in the "dead zone" at the mouth of Mississippi River. Each species has its own ecological niche and preferecne of habitat type, and this is why the distribution of each species is more or less different. For example, brown shrimp prefers muddy bottom, so they were predicted to be primarily appeared in the western Gulf where the mud concentration is high; and were predicted to be absent from eastern Gulf where the mud concentration is low. Additionally, the distribution patterns predicted for gulf menhaden, common halfbeak, Atlantic croaker and Atlantic cutlassfish fitted well with the king mackerel gut content analysis results done by Simons et al. (2013b), this also showed good performance of the models.

Abundance modeling results for the three shrimp species predicted that brown shrimp was mostly distributed in the western Gulf while nearly zero abundance for eastern Gulf; pink shrimp shows opposite distribution to brown shrimp, it was mostly distributed in the eastern Gulf, while nearly no abundance in the western Gulf; white shrimp shows similar patterns to brown shrimp, but it was only distributed in the northern Gulf, with no presence in the rest of the Gulf. The abundance distribution patterns were quite close to the distribution pattern predicted by the presence-only models, this consistency supports the validity of the two different types of models (distribution

and abundance). Evaluation of the models shows that the GAM models has the best performance for brown shrimp abundance modeling, while the RF models has the best performance for the rest of the shrimp species.

The distribution patterns, especially the special zones, can be further validated by doing some more research in those special zones. The distribution patterns can also provide some hints for scientist or government mangers to better manage the fisheries in the Gulf of Mexico.

CHAPTER V: REFERENCES

Booth, T. H., Nix, H. A., Busby, J. R., Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and distributions, 20*(1), 1-9.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Christensen, V., Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities and limitations. *Ecological Modelling, 172*(2), 109-139.

Christensen, V., Walters, C. J., Pauly, D., Forrest, R. (2008). Ecopath with Ecosim version 6. User Guide. *Fisheries Centre, University of British Columbia, Vancouver, 235*.

De Maesschalck, R., Jouan-Rimbaud, D., Massart, D. L. (2000). The mahalanobis distance. *Chemometrics and intelligent laboratory systems, 50*(1), 1-18.

Drexler, M., Ainsworth, C. H. (2013). Generalized additive models used to predict species abundance in the Gulf of Mexico: an ecosystem modeling tool. *PloS one, 8*(5), e64458.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions, 17*(1), 43-57.

Fulton, E. A., Link, J. S., Kaplan, I. C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R. J., Smith, A. D. (2011). Lessons in modelling and management of marine ecosystems: the Atlantis experience. *Fish and Fisheries, 12*(2), 171-188.

Geers, T., Pikitch, E., Frisk, M. (2016). An original model of the northern Gulf of Mexico using Ecopath with Ecosim and its implications for the effects of fishing on ecosystem structure and maturity. *Deep Sea Research Part II: Topical Studies in Oceanography, 129*(2016), 319-331.

Grüss, A., Schirripa, M. J., Chagaris, D., Drexler, M., Simons, J., Verley, P., Shin, Y.-J., Karnauskas, M., Oliveros-Ramos, R., Ainsworth, C. H. (2015). Evaluation of the trophic structure of the West Florida Shelf in the 2000s using the ecosystem model OSMOSE. *Journal of Marine Systems, 144*, 30-47.

Guisan, A., Edwards, T. C., Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling, 157*(2), 89-100.

Hastie, T. J., Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43): CRC Press.

Hijmans, R. J., Graham, C. H. (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Global change biology, 12*(12), 2272-2281.

Hijmans, R. J., Elith, J. (2015). Species distribution modeling with R.

Kéry, M., Gardner, B., Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography, 37*(10), 1851-1862.

Leathwick, J., Elith, J., Hastie, T. (2006). Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling, 199*(2), 188-196.

Link, J. S., Fulton, E. A., Gamble, R. J. (2010). The northeast US application of ATLANTIS: a full system model exploring marine ecosystem dynamics in a living marine resource management context. *Progress in Oceanography, 87*(1), 214-234.

Marini, C., Fossa, F., Paoli, C., Bellingeri, M., Gnone, G., Vassallo, P. (2015). Predicting bottlenose dolphin distribution along Liguria coast (northwestern Mediterranean Sea) through different modeling techniques and indirect predictors. *Journal of environmental management, 150*, 9-20.

Marzloff, M., Shin, Y.-J., Tam, J., Travers, M., Bertrand, A. (2009). Trophic structure of the Peruvian marine ecosystem in 2000–2006: insights on the effects of management scenarios for the hake fishery using the IBM trophic model Osmose. *Journal of Marine Systems, 75*(1), 290-304.

Masi, M., Ainsworth, C., Chagaris, D. (2014). A probabilistic representation of fish diet compositions from multiple data sources: A Gulf of Mexico case study. *Ecological Modelling, 284*, 60-74.

Pearce, J. L., Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology, 43*(3), 405-412.

Phillips, S. J., Anderson, R. P., Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*(3), 231-259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications, 19*(1), 181-197.

Potts, J. M., Elith, J. (2006). Comparing species abundance models. *Ecological Modelling, 199*(2), 153-163.

Shin, Y.-J., Cury, P. (2001). Exploring fish community dynamics through size-dependent trophic interactions using a spatialized individual-based model. *Aquatic Living Resources, 14*(02), 65-80.

Shin, Y.-J., Shannon, L., Cury, P. (2004). Simulations of fishing effects on the southern Benguela fish community using an individual-based model: learning from a comparison with ECOSIM. *African Journal of Marine Science, 26*(1), 95-114.

Simons, J., Yuan, M., Carollo, C., Vega-Cendejas, M., Shirley, T., Palomares, M. L., Roopnarine, P., Gerardo Abarca Arenas, L., Ibañez, A., Holmes, J. (2013a). Building a fisheries trophic interaction database for management and modeling research in the Gulf of Mexico large marine ecosystem. *Bulletin of Marine Science, 89*(1), 135-160.

Simons, J., Poelen, J., Hewitt, R., Miller, T., Gonzalez, B. (2013b). Food habits and food webs of high Hg content commercially and recreationally harvested fishes in the Gulf of Mexico. *Final Report for WQ-003. Submitted to Florida Department of Environmental Quality. 29p + 3 App.*

Sunderland, E. M., Kriens, D., Von Stackelberg, K. (2012). Pilot Analysis of Gulf of Mexico State Residents' Methylmercury Exposures from Commercial and Locally Caught Fish. *Report to: Florida Department of Environmental Protection*, 1-39.

Travers, M., Shin, Y.-J., Jennings, S., Machu, E., Huggett, J., Field, J., Cury, P. (2009). Two-way coupling versus one-way forcing of plankton and fish models to predict ecosystem changes in the Benguela. *Ecological Modelling, 220*(21), 3089-3099.

Travers, M., Watermeyer, K., Shannon, L., Shin, Y.-J. (2010). Changes in food web structure under scenarios of overfishing in the southern Benguela: comparison of the Ecosim and OSMOSE modelling approaches. *Journal of Marine Systems, 79*(1), 101-111.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R. (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and distributions, 13*(4), 397-405.

US Commission on Ocean Policy. (2004). An Ocean Blueprint for the 21st Century. Washington, DC: US Commission on Ocean Policy. .

VanDerWal, J., Shoo, L. P., Graham, C., Williams, S. E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling, 220*(4), 589-594.

Vierod, A. D., Guinotte, J. M., Davies, A. J. (2014). Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. *Deep Sea Research Part II: Topical Studies in Oceanography, 99*, 6-18.

Ward, G., Hastie, T., Barry, S., Elith, J., Leathwick, J. R. (2009). Presence-only data and the EM algorithm. *Biometrics, 65*(2), 554-563.

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Campbell Grant, E. H., Veran, S. (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution, 4*(3), 236-243.

APPENDIX A

The following appendix (Figure A1-A23) are the visualizations of the distribution models for the 23 remaining species.

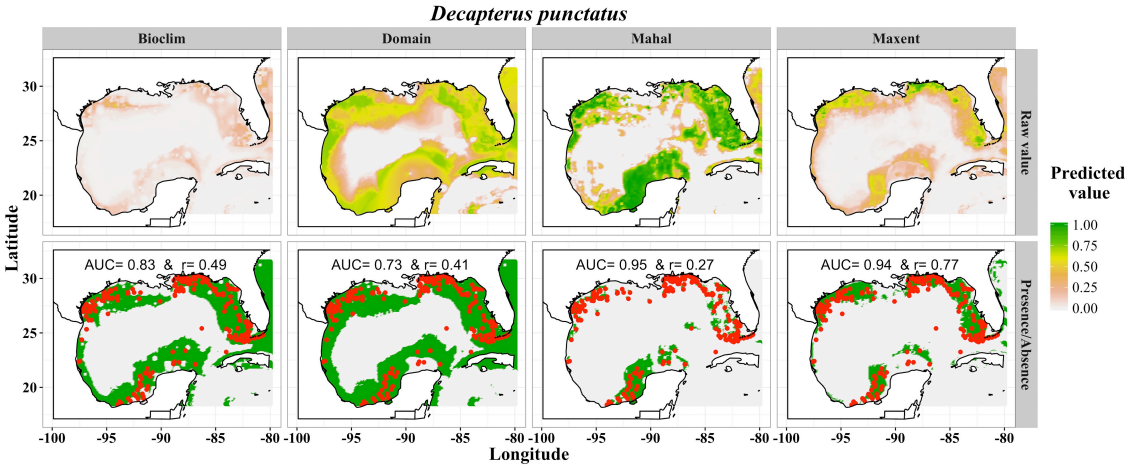

Figure A1: Visualization of round scad (*Decapterus punctatus*) distribution models

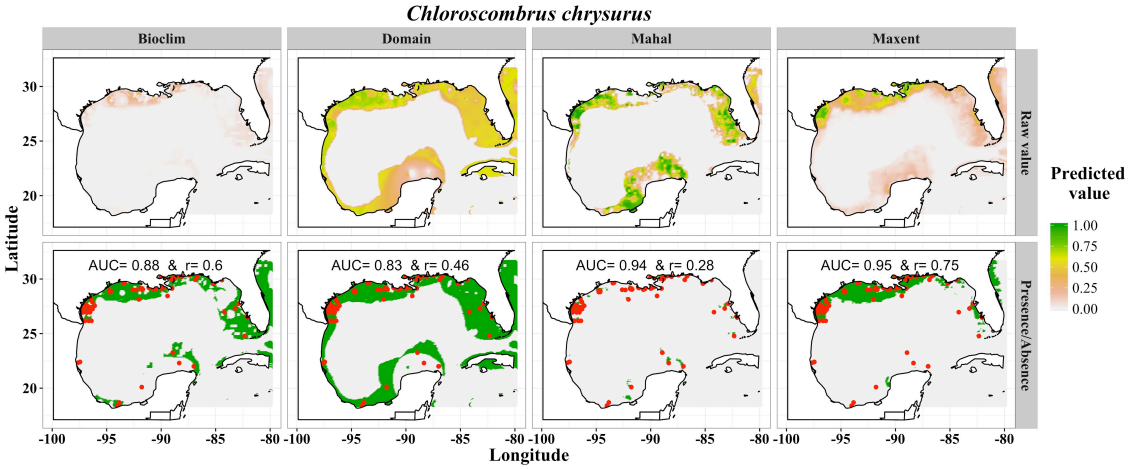

Figure A2: Visualization of Atlantic bumper (*Chloroscombrus chrysurus*) distribution models.

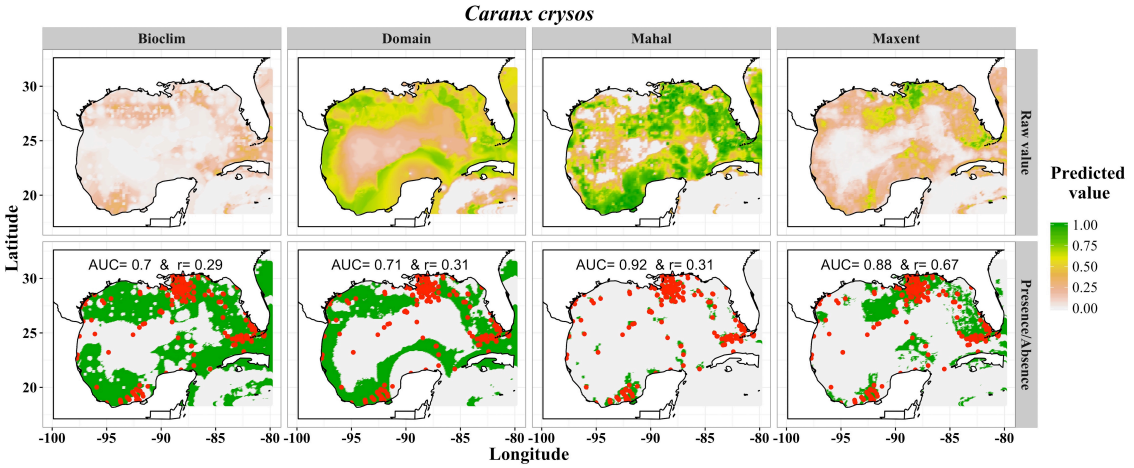

Figure A3: Visualization of blue runner (*Caranx crysos*) distribution models.

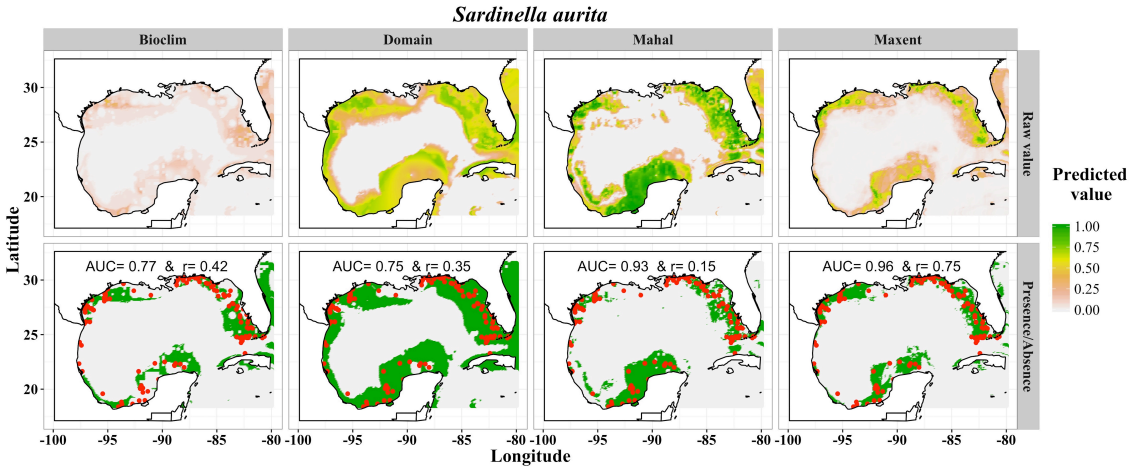

Figure A4: Visualization of Spanish sardine (*Sardinella aurita*) distribution models.

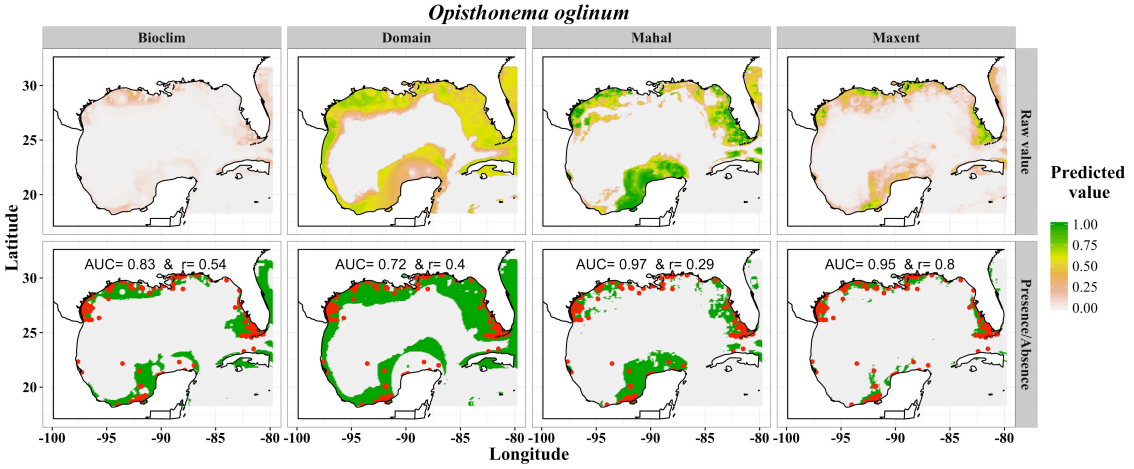

Figure A5: Visualization of Atlantic thread herring (*Opisthonema oglinum*) distribution models.

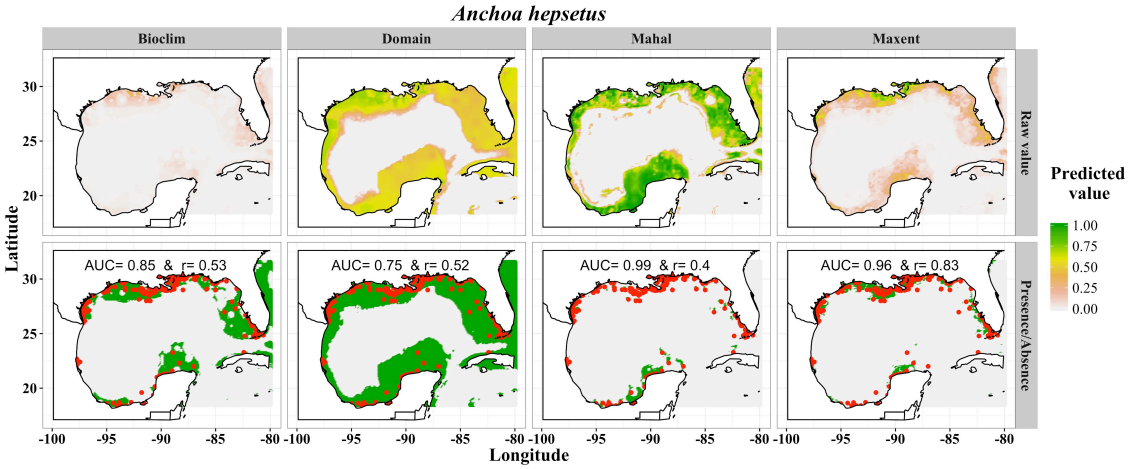

Figure A6: Visualization of striped anchovy (*Anchoa hepsetus*) distribution models.

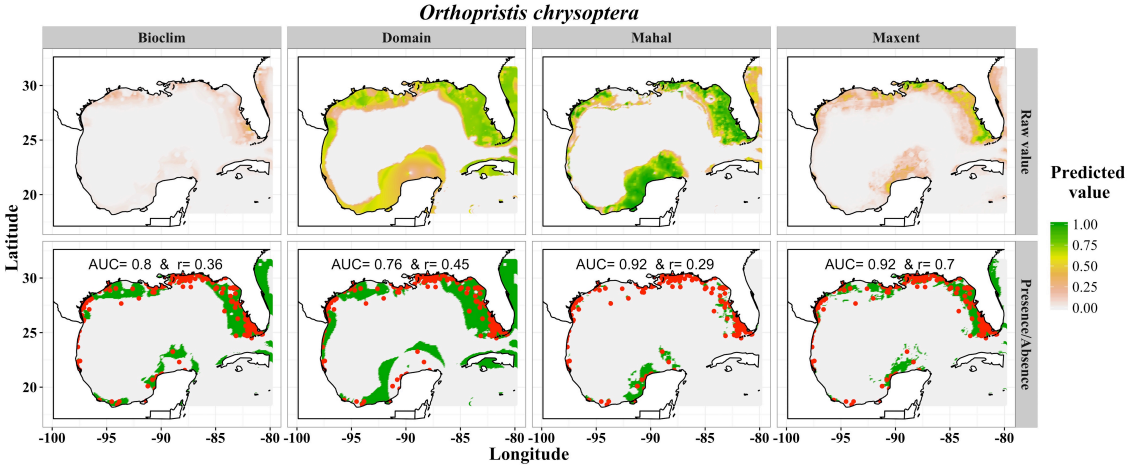

Figure A7: Visualization of pigfish (*Orthopristis chrysoptera*) distribution models.

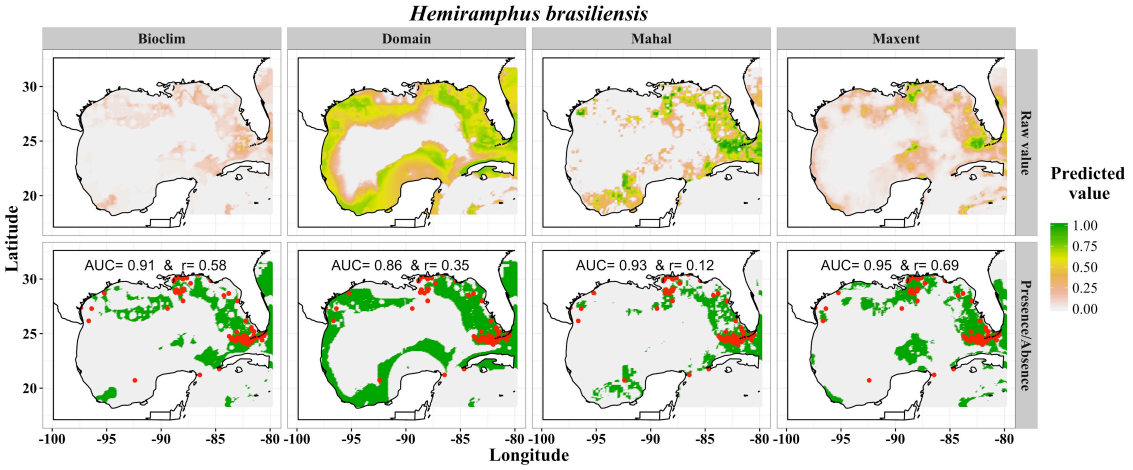

Figure A8: Visualization of ballyhoo halfbeak (*Hemiramphus brasiliensis*) distribution models.

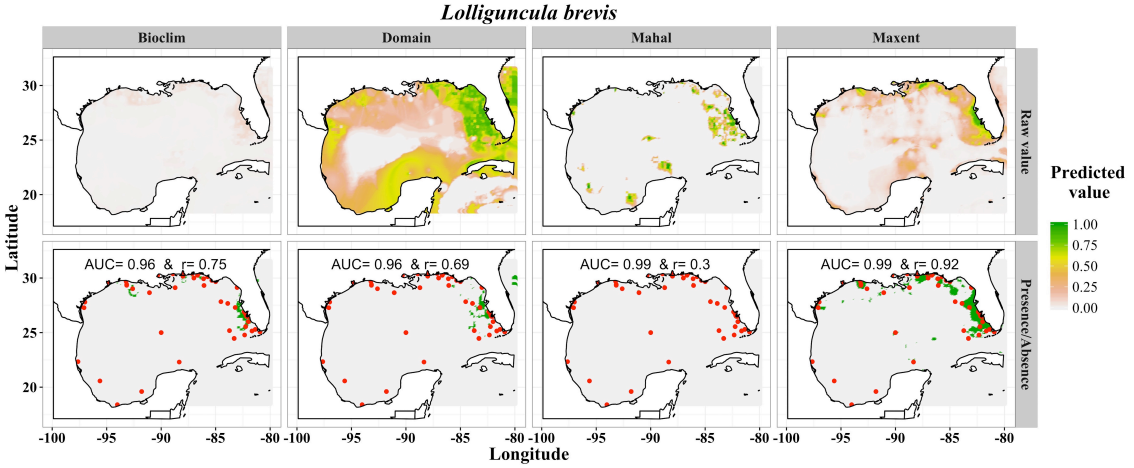

Figure A9: Visualization of Atlantic brief squid (*Lolliguncula brevis*) distribution models.

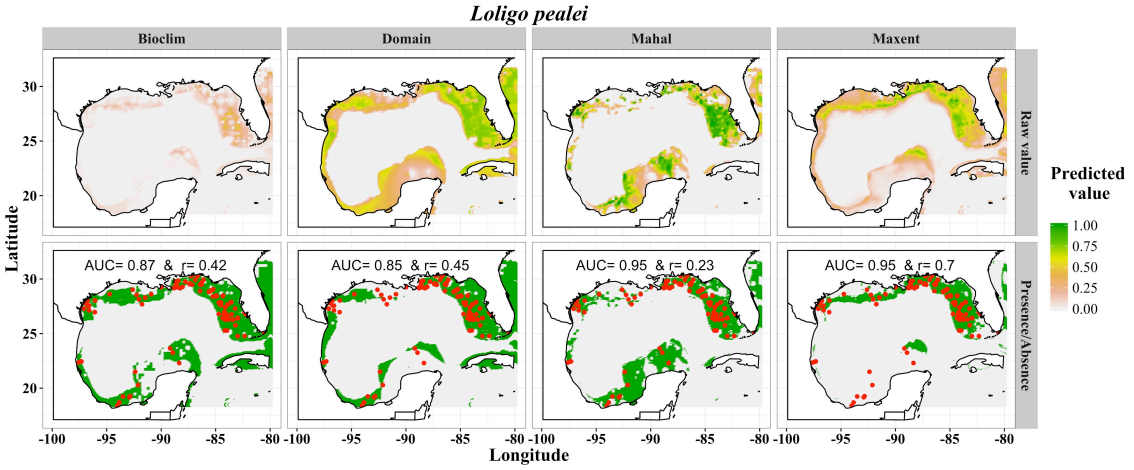

Figure A10: Visualization of longfin inshore squid (*Loligo pealei*) distribution models.

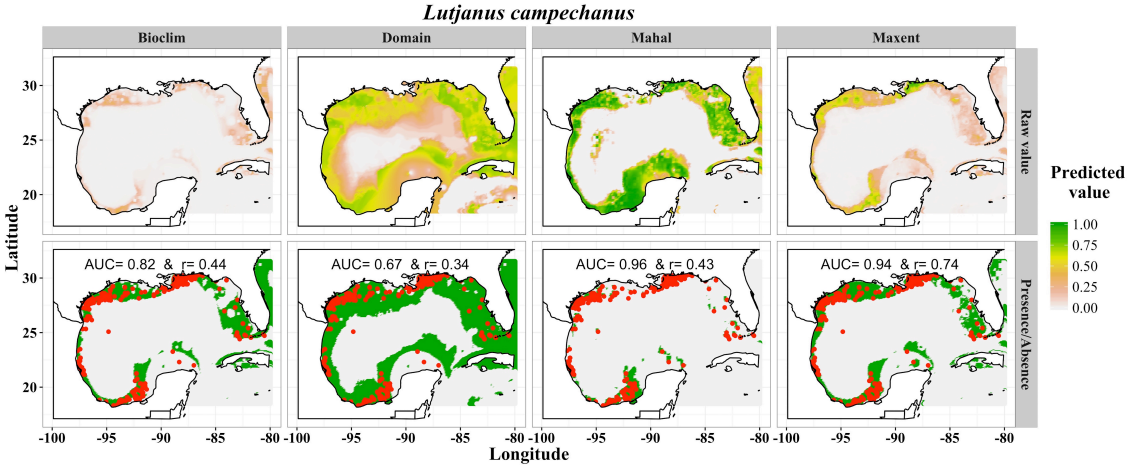

Figure A11: Visualization of red snapper (*Lutjanus campechanus*) distribution models.

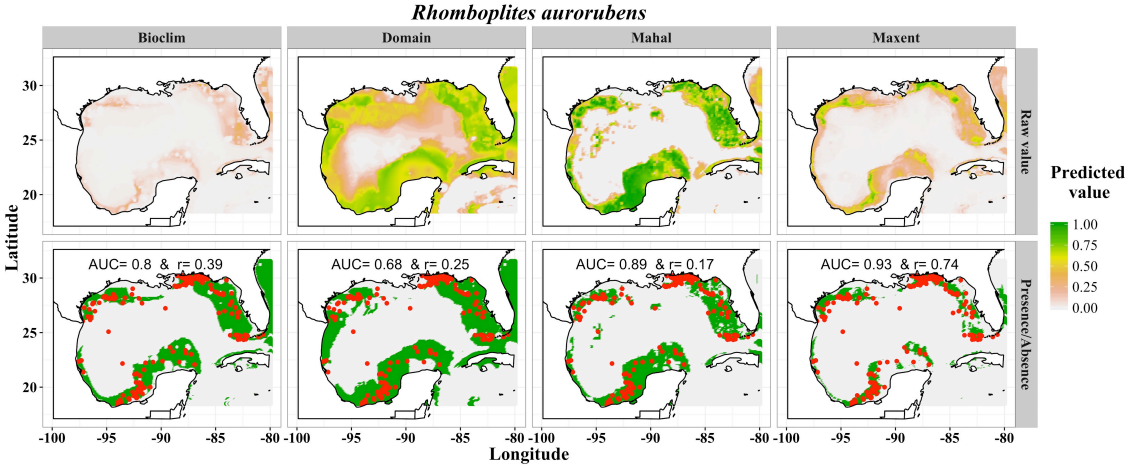

Figure A12: Visualization of vermilion snapper (*Rhomboplites aurorubens*) distribution models.

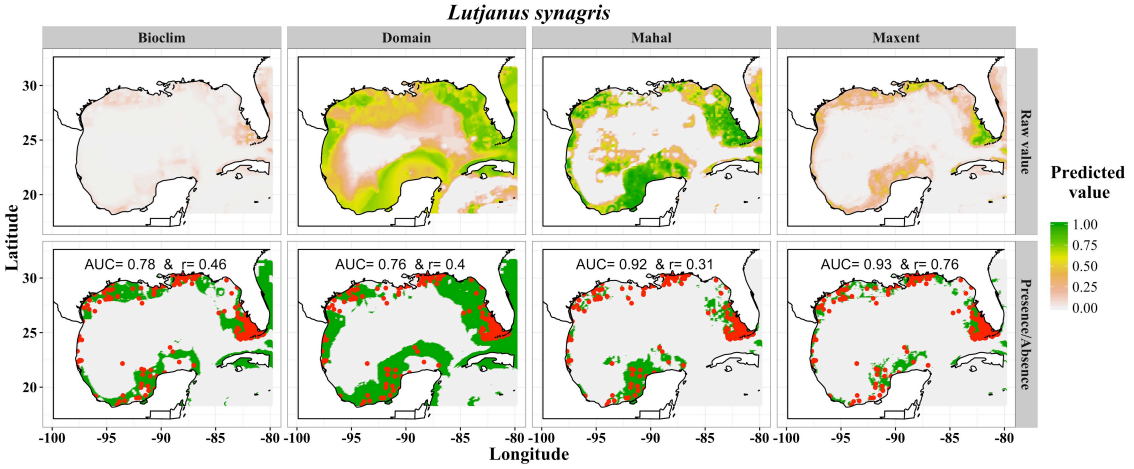

Figure A13: Visualization of lane snapper (*Lutjanus synagris*) distribution models.

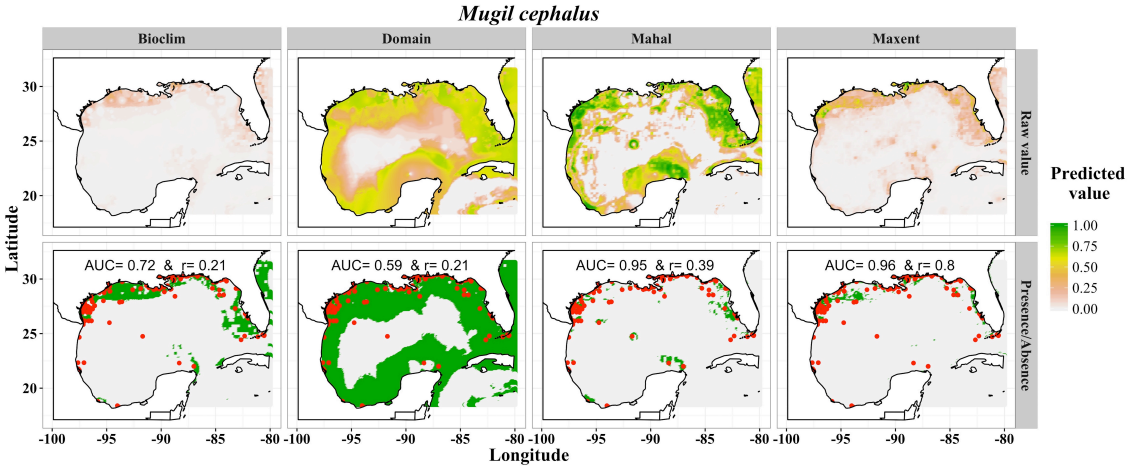

Figure A14: Visualization of striped Mullet (*Mugil cephalus*) distribution models.

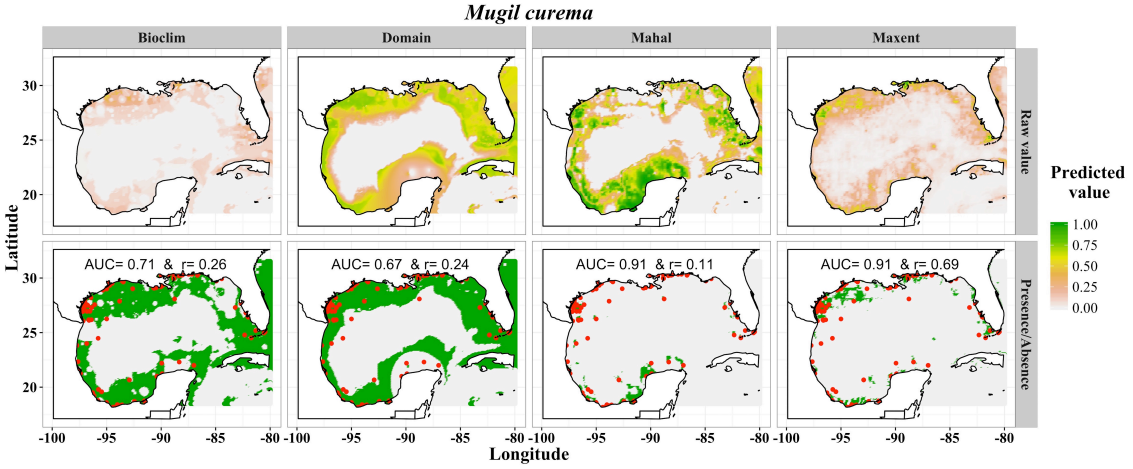

Figure A15: Visualization of white mullet (*Mugil curema*) distribution models.

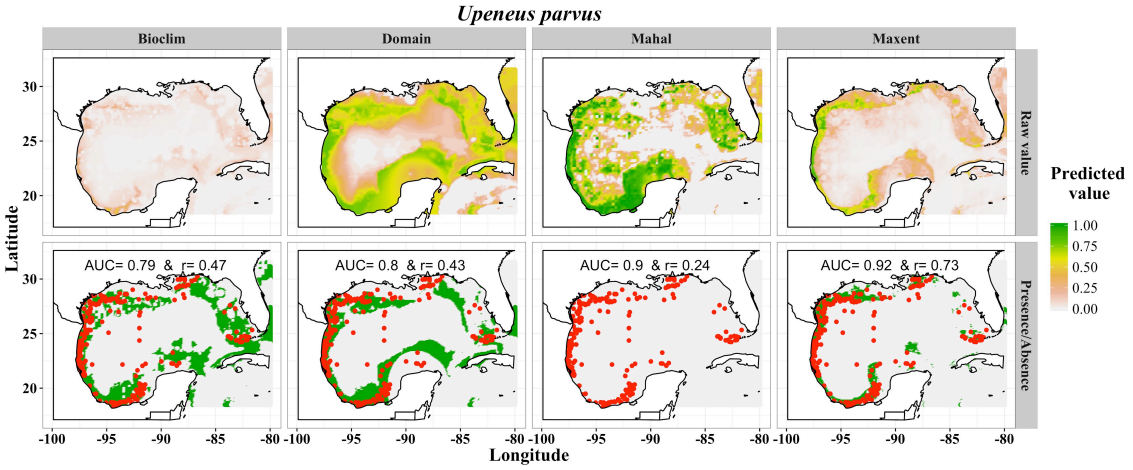

Figure A16: Visualization of dwarf goatfish (*Upeneus parvus*) distribution models.

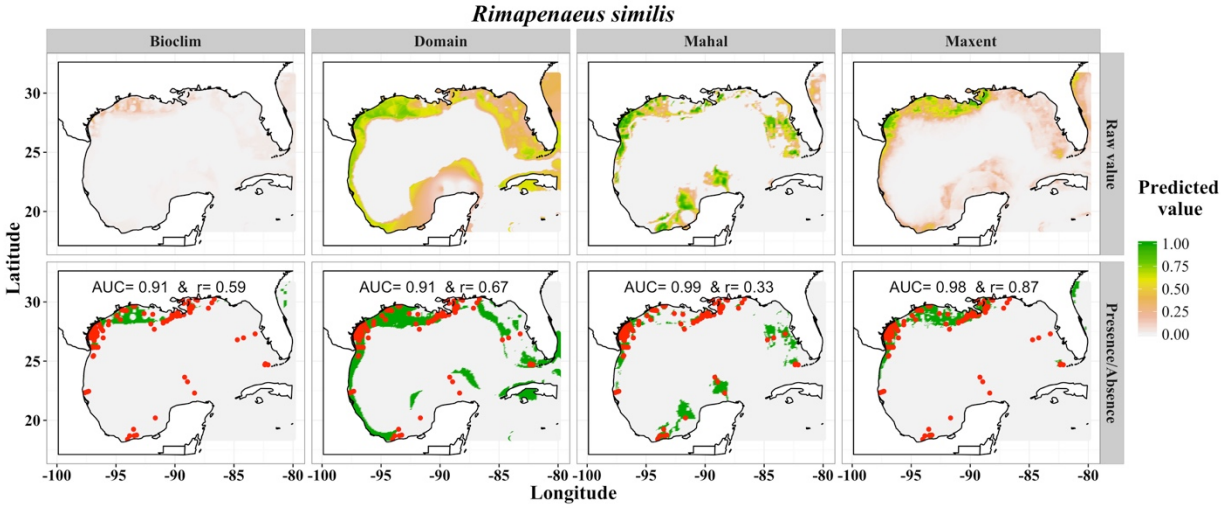

Figure A17: Visualization of roughback shrimp (*Rimapenaeus similis*) distribution models.

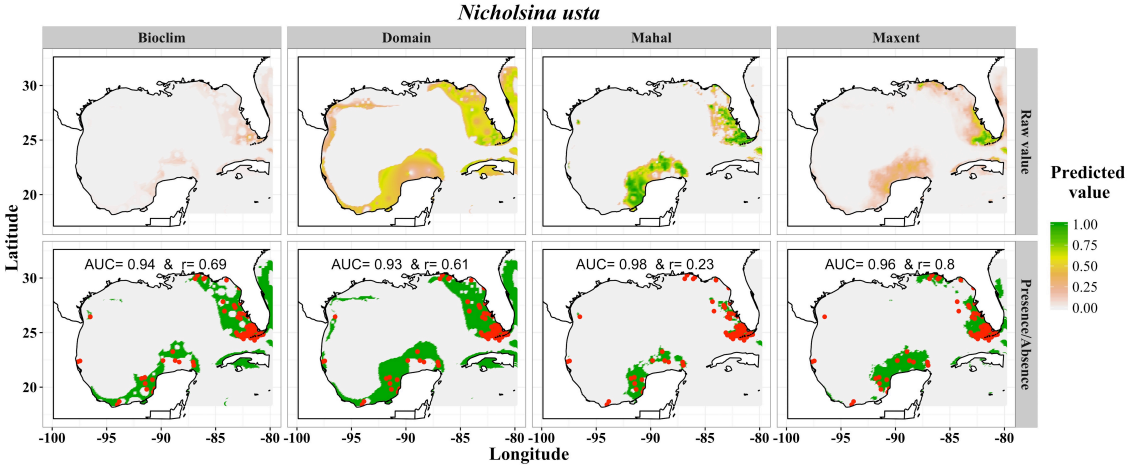

Figure A18: Visualization of emerald parrotfish (*Nicholsina usta*) distribution models.

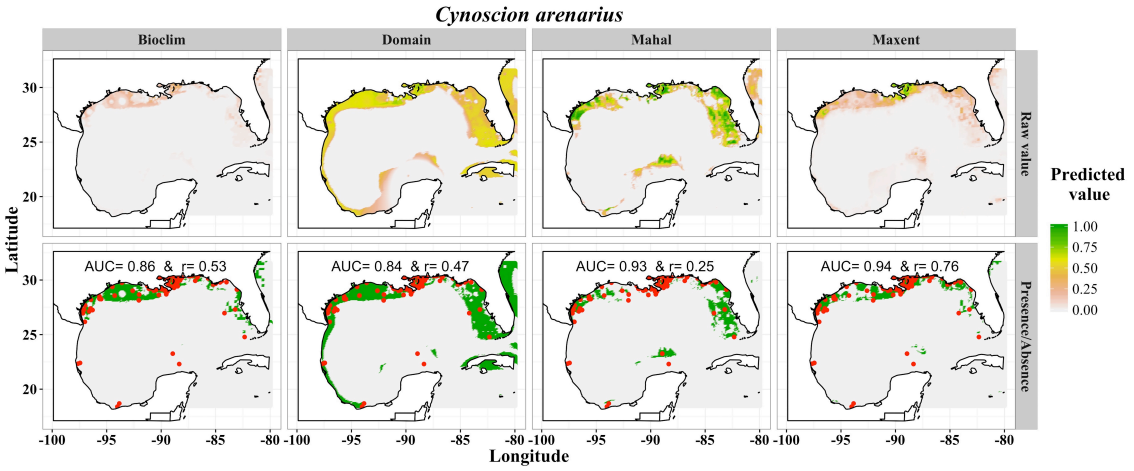

Figure A19: Visualization of sand seatrout (*Cynoscion arenarius*) distribution models.

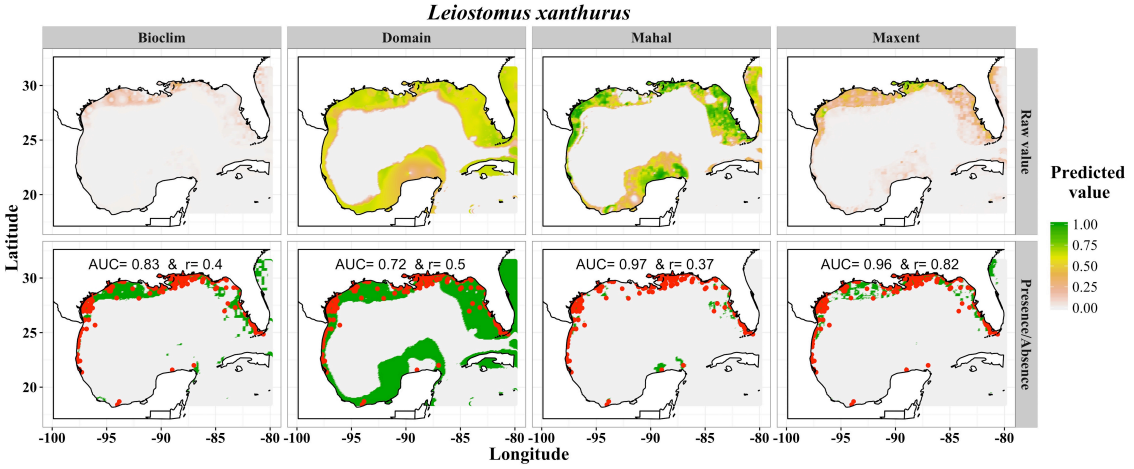

Figure A20: Visualization of spot croaker (*Leiostomus xanthurus*) distribution models.

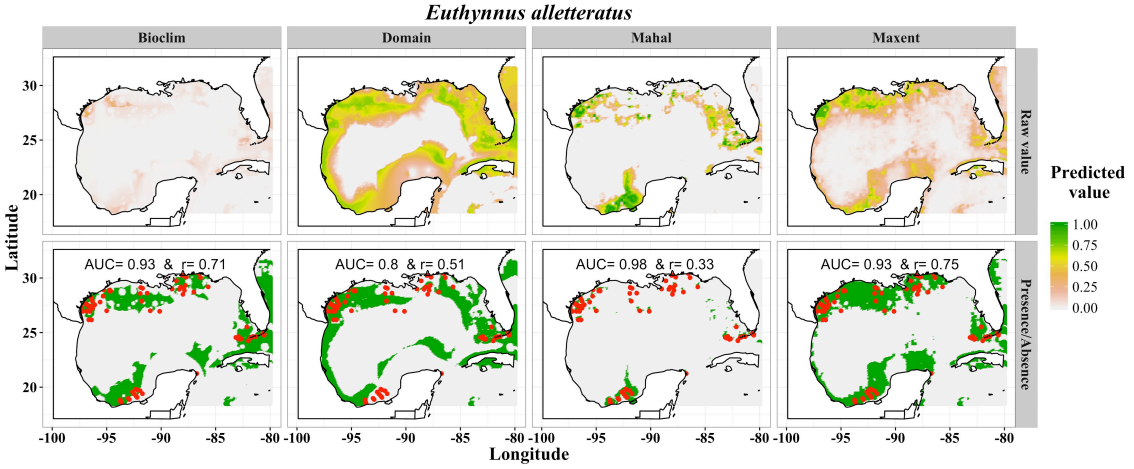

Figure A21: Visualization of little tunny (*Euthynnus alletteratus*) distribution models.

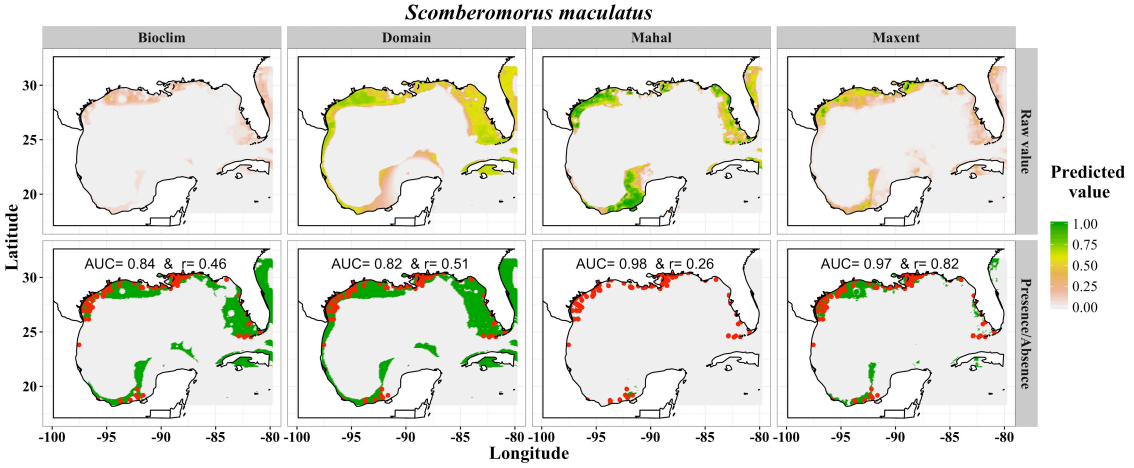

Figure A22: Visualization of Spanish mackerel (*Scomberomorus maculatus*) distribution models.

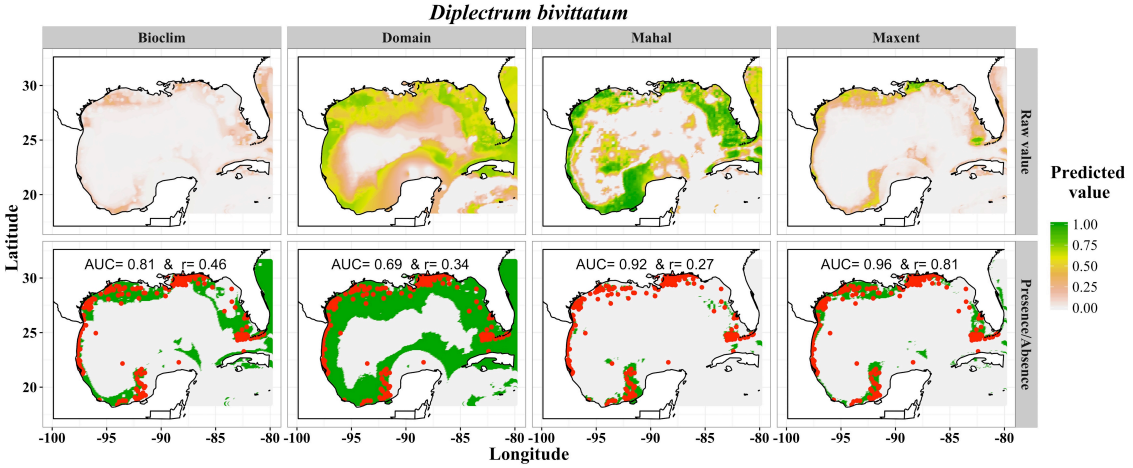

Figure A23: Visualization of dwarf sand perch (*Diplectrum bivittatum*) distribution models.