

**Developing and applying ensemble species distribution modeling framework  
to assess the spatio-temporal distribution of Atlantic scallop  
in the New York Bight**

**Final report**

Submitted to

The New York Department of Environment Conservation

by

Michael Torre, Ph.D

Yong Chen, Ph.D

School of Marine Sciences

University of Maine, Orono, ME 04469

Email: [michael.torre@maine.edu](mailto:michael.torre@maine.edu), [ychen@maine.edu](mailto:ychen@maine.edu)

<b>Table of Contents</b>	<b>Page</b>
• List of abbreviations and acronyms	3
• List of figures	4
• Executive Summary	5
• Introduction	7
• Materials and Methods	8
○ Survey and environmental data used in model projection	9
○ Development of a habitat suitability index	11
○ Ensemble species distribution modeling algorithm	12
○ Projection of future habitat	14
○ Assessment of uncertainty in the ensemble projections	14
• Results	15
○ Habitat suitability index	15
○ Scallop distribution in the New York Bight	18
• Discussion	21
• Conclusions	23
• References	24
• Appendix A	27

## List of figures

- Figure 1. Delineation of survey areas during a typical NEFSC Scallop Dredge Survey. Figure credit: Virginia Institute of Marine Sciences (VIMS)
- Figure 2. Environmental variables used in modeling framework. Dynamic environmental variables (bottom temperature, bottom salinity, and bottom stress) are annual mean values to represent the range of seasonal exposure for this species.
- Figure 3. Schematic diagram describing the development of ensemble species distribution models.
- Figure 4. Scallop-Environment response curves developed from the random forest model. These curves represent suitability indices, where highest scallop abundance is structured along the peaks of each respective plot.
- Figure 5. Relative importance of each environmental variable with respect to its predictive ability in the random forest model for scallop habitat suitability in the MAB-NYB.
- Figure 6. Spatial distribution of Habitat Suitability Index (HSI) values over the Mid-Atlantic Bight to the New York Bight. HSI values are shown here averaged across three distinct time periods spanning the study period (1980-2014). The HSI value indicates the change of habitat quality with 1 being the highest quality and 0 being unsuitable habitat.
- Figure 7. Temporal changes in the spatial distribution of Habitat Suitability Index (HSI) values for scallops within the Mid-Atlantic Bight to the New York Bight over the study period (1980-2014). Red (warm) color has a positive temporal trend, indicating an improved habitat quality over time, while blue (cold) color has a negative temporal trend, indicating a reduced habitat quality over time from 1980 to 2014.
- Figure 8. Scallop distribution and abundance based on data from the NEFSC Scallop Dredge Survey. Figures are broken up into 5 to 10 year periods from 1980 - 2015 in order to demonstrate temporal changes over the study period. Figures on the left show individual survey samples and figures on the right show a smoothed interpolated surface developed using ordinary kriging.
- Figure 9. Performance criteria (TSS = True Skill Statistic; ROC = Receiver Operator Curve) for each SDM algorithm included in the ensemble modeling framework for scallops in the MAB-NYB.
- Figure 10. Spatial distribution of probability of presence values for scallops over the Mid-Atlantic Bight to the New York Bight. Probability of presence values in this figure represent a median value for hindcasted oceanographic conditions from 2015-2000 in order to portray a “current” distribution upon which projected values can be compared against. The scale (0 - 1000) corresponds to the probability (0 low probability, 1000 high probability) of encountering the modeled species within a given cell.
- Figure 11. Climate-driven change in the spatial distribution of probability of occupancy values for scallops over the Mid-Atlantic Bight to the New York Bight. Changes in probability of occupancy values in this figure represent the degree of change at each raster cell over the 80 year projection with red (warm) color indicating increasing trends (positive trend) and blue (cold) color indicating decreasing trends over the 80 years of projection time period.

## List of abbreviations and acronyms

ANN	Artificial Neural Network
BRT	Boosted Regression Tree
CM	Climate Model
CTA	Classification Tree Analysis
FDA	Flexible Discriminant Analysis
FEP	Final Ensemble Projection
GFDL	Geophysical Fluid Dynamics Laboratory
GAM	Generalized Additive Model
GLM	Generalized Linear Model
GOM-GB	Gulf of Maine and Georges Bank
HSI	Habitat suitability index
IPCC	Intergovernmental Panel on Climate Change
KS Test	Kolmogorov-Smirnov
MAB-NYB	Mid Atlantic Bight to the New York Bight
MARS	Multiple Adaptive Regression Splines
Maxent	Maximum Entropy
NEFSC	Northeast Fisheries Science Center
NEUS-LME	Northeast United States Large Marine Ecosystem
NY DEC	New York Department of Environment Conservation
NMFS	National Marine Fisheries Service
OOB	Out of bag
RCP	Representative Concentration Pathway (for greenhouse)
RF	Random Forest
RMSE	Root Mean Squared Error
ROC	Receiver Operator Curve
SDM	Species Distribution Model
SNE	Southern New England
SRE	Surface range envelope
TSS	True Skill Statistics
VIF	Variance Inflation Factor

## Executive Summary

Altered species biogeography poses a myriad of conservation challenges because changes in species distributions can move stocks into and out of fixed boundaries. Therefore, predicting responses of important fish stocks to future climatic conditions is critical to the implementation of adaptive conservation planning measures. However, likely shifts in biogeography for many species moving into the future remain largely unknown. Correlative species distribution models (SDMs), which rely on quantified species-environment relationships to explain and predict spatial distributions of species, have become important tools in adaptive natural resource management. In the Mid Atlantic Bight to the New York Bight (MAB-NYB), the biogeography of many marine species is currently shifting in response to climate-related changes which will influence large scale marine spatial planning projects going into this area. Understanding these shifts can help the State of New York to better address concerns over impacts from the future offshore wind energy areas to the ecologically and economically important fisheries.

An emerging consensus recommends that uncertainty associated with climate-driven changes in species distribution can be better addressed through an ensemble species distribution modeling approach. This project develops and parameterizes a habitat suitability index (HSI) model and SDM for Atlantic Scallop (*Placopecten magellanicus*) off the New York coast. In this study, we combined an ensemble SDM platform (BIOMOD 2) and a high-resolution global climate model (NOAA GFDL CM2.6) to quantify spatiotemporal changes in the distribution for Atlantic scallop in the MAB-NYB. The ensemble model was calibrated using data from the National Marine Fisheries Service (NMFS) bottom trawl survey as well as NEFSC Scallop Dredge survey to develop the models. Additionally, both the HSI model and SDM are evaluated for performance in a cross validation study. The final model is used to hindcast the suitable scallop habitat and distribution in the past and then forecast the distributions under different climate change scenarios.

The HSI model, built using a random forest regression approach, showed bottom stress as the most important predictor for suitability of habitat for scallop in the MAB-NYB, followed by depth, bottom salinity, and temperature. Sediment grain size was of relatively little importance in predicting scallop habitat quality. The results from the random forest model cross validation study (out-of-bag error evaluation for each tree) showed low error (2.67) and high variance explained (61.16%) suggesting good model performance. General trends for HSI-based bioclimate envelopes for scallop showed a gradient of increasing habitat quality moving from inshore to offshore areas. Warming across the MAB-NYB over the hindcasted study period (1980-2014) resulted in spatial changes in habitat suitability for scallop, with shallower, inshore waters in general undergoing a projected decline in habitat suitability and deeper areas, further offshore, increasing in habitat suitability.

The ensemble modeling framework developed for the Atlantic sea scallop had most (7 out of 10) of the tuned SDM algorithms meet the predefined performance thresholds (True Skill Statistic TSS > 0.5 and Receiver Operator Curve ROC > 0.8). The response curves from the individual SDM algorithms included in final ensemble models indicate that non-linear relationships between the probability of presence and environmental variables for this species. The projections produced from the final ensemble model show a spatial distribution that

corresponds well to the known general, broad scale distribution of scallop over the MAB-NYB with higher probability of presence being generally along offshore areas up to ~ 100m in depth. Changing oceanographic conditions across the MAB-NYB, based on CM2.6 projections, resulted in large spatial changes in probability of presence for scallop. These changes occur along a north-south gradient, with southern areas experiencing a sharper decline in probability of occupancy relative to northern areas. However, selected areas within southern New England (SNE) showed a projected increase in probability of occupancy for scallop.

While the effects of climate change are complex and diverse, the impacts on fisheries can be grouped into two general categories: changes to stock biomass or productivity, and changes to stock distribution, each of which poses different management challenges. This study focused on the latter, and proposed an ensemble means to infer potential future habitats, based on high-resolution climate data, of an economically important fisheries resource in the MAB-NYB. The modeling framework developed in this study adds quality projections of spatiotemporal trends in the distribution of scallop, in response to changing oceanographic conditions over the next 80 years and constitutes a step both towards improving conservation planning within the MAB-NYB as well as increasing the capacity of New York State to generate models needed to inform managers to make decisions based on shifts in species distribution.

## Introduction

Finfish, shellfish and crustacean species, distributed along the coast of northeastern United States, support some of the most valuable fisheries in the USA. Many coastal communities in the northeastern USA heavily depend on the fisheries, and such dependence on the fisheries leaves the coupled natural and human system vulnerable to environmental (e.g., climate changes) and regulatory changes and human activities (e.g., development of offshore wind power). Offshore wind power farm development is considered to have a great potential to provide a more sustainable source of energy. However, before developing an offshore wind power project, it is important to identify and understand potential impacts that the development and operation of offshore wind farms may have on marine environments and the organisms which dwell there. Thus, *it is important to identify possible spatial distributions of key suitable habitat for ecologically and economically important fish species to inform placement of offshore wind power projects, and evaluate how spatial distributions of species may change as a result of climate changes.*

Offshore wind power in the New York Bight has been a topic of much interest. Unfortunately, much of the baseline information that is needed to be able to predict and assess the possible impacts of offshore wind energy development on commercial fisheries is still lacking. Site-specific data such as in-depth sampling of hard bottom habitats along the Northeast U.S. shelf are absent, due to most of stock assessment work focusing on relative abundance and not identification of essential habitat (Petruny-Parker et al. 2015). There is also general recognition that these wind energy areas on the east coast are located in a region of the continental shelf that appears to be undergoing change as a result of climate change, stock depletion and rebuilding, and changes in the ecological dynamics (Petruny-Parker et al. 2015).

To address these concerns it is critical to develop modeling tools to hind-, now-, and forecast key habitat areas for the distribution of key fish species in the New York Bight. These modeling tools allow us to understand spatial dynamics of critical habitats and fish distributions in the past and present, and possibly in the future with different climate change scenarios. Such information has become extremely important to government agencies to properly manage fish quotas, protect important habitat, and to site offshore projects in an environmentally responsible manner. Inability to understand these shifts could result in economic and ecological harm to New York State's valuable marine resources. Currently, the State does not have the capacity to generate the models needed to inform managers to make decisions based on these shifts. Dr. Yong Chen's lab at the University of Maine has been asked to create a distribution modeling framework to assess the spatio-temporal dynamics of Atlantic scallop in the New York Bight.

Dr. Yong Chen's lab at the University of Maine's School of Marine Sciences has been able to develop habitat suitability index (HSI) models and Species Distributional Models (SDMs) that can be used as a tool for the State's marine resource managers. These developed HSI Models and SDMs have been peer reviewed and accepted by the scientific community. These models have been successfully applied across the Gulf of Maine to hindcast and predict the spatio-temporal distributions of suitable habitat and abundance of American lobster (*Homarus americanus*), Atlantic Scallop (*Placopecten magellanicus*), Cusk (*Brosme brosme*), Northern Shrimp (*Pandalus borealis*), and Atlantic Cod (*Gadus morhua*) under present conditions and future climatic conditions (Chang et al. 2010; Tanaka and Chen 2016; Tanaka et al. 2018; Torre et al. 2019; Guan et al. 2017; Li et al. 2018; Runnebaum et al. 2019; Tanaka et al. 2019). The

HSI and SDM models have performed well and the results have identified spatiotemporal blocks where each species is likely to gain, lose, or maintain suitable climatic niches.

Other products produced by the Chen Lab have also been used by the New England Fisheries Management Council and the Atlantic States Marine Fisheries Commission. By utilizing these additional vetted tools, New York marine resource managers can make more informed decisions about our valuable resources. The objectives of this study are to (1) develop and parameterize the HSI models and SDMs for the scallop in the New York Bight; (2) use the defined models to hindcast the distributions of suitable habitat and abundance distributions of scallop in the New York Bight; and (3) forecast the scallop abundance distributions under defined climate change scenarios.

## **Materials and methods**

We used previously developed SDMs for the Gulf of Maine and adapt them as needed for the New York Bight. This modeling framework constitutes a culmination of knowledge acquired by the Chen Lab from numerous peer-reviewed studies on modeling species distribution (Chang et al. 2010; Tanaka and Chen 2016; Tanaka et al. 2018; Torre et al. 2018; Torre et al. 2019). It includes utilizing previously available National Marine Fisheries Service (NMFS) bottom trawl data as well as more recent bottom trawl data if it is available. Additional environmental data specific to the New York Bight are also included in the model. These SDMs are able to show current distribution and/or abundance of a species, as well as shifts in species distribution under future climatic conditions. Thus, the output of these models are flexible and can be tailored to the needs of the New York Department of Environment Conservation (DEC).

New York DEC has requested that Atlantic Scallop be the target species for the HSI model and SDM. This species has previously been modeled for the Gulf of Maine and the model compared extremely well to real world observations (Torre et al. 2018, 2019). Atlantic Scallop are also an extremely economically valuable species found in the waters of the New York Bight. Previous success with this species coupled with large amounts of readily available data, and economic importance make this a good test species for developing an HSI model and SDM for the New York Bight.

Furthermore, the spatial resolution of the models was improved to meet the needs of New York DEC. Currently, the Gulf of Maine SDM for Atlantic Scallop uses a resolution of 0.05 decimal degrees. While this level of resolution is high and acceptable for understanding species shifts, a finer resolution could help resource managers site offshore projects in areas that reduce conflicts and protect valuable habitat. Changing the resolution on the model is a minor task in the overall model development. However, as the resolution becomes finer the required computing power increases, thus increasing the overall time needed to run each model iteration. Balancing resolution needs with model run time is optimized after an initial model is developed using the original model resolution.

A New York Bight specific HSI model and SDM can forecast Atlantic scallop distribution and abundance over different time series (5 years, 10 years, 20 years, etc.) in response to changing climate and environmental conditions on a spatial scale of 0.025 and 0.01 decimal degrees. It is

expected that this New York Bight SDM could be applied to other species of importance with relative ease as long as survey data and appropriate environmental data are available.

### Survey and Environmental data used in model projections

The data available for these analyses were a multi-decade scientific scallop dredge survey as well as a bottom trawl survey dataset and associated environmental (bottom temperature, bottom salinity, and depth) variables. The fishery-independent survey data for Atlantic sea scallop were collected by the Northeast Fisheries Science Center (NEFSC: 1984-2016) covering offshore continental shelf waters (Figure 1). In addition, external data in the form of bottom stress and sediment grain size were used to add important environmental variables relating to scallop distribution on abundance (Figure 2).

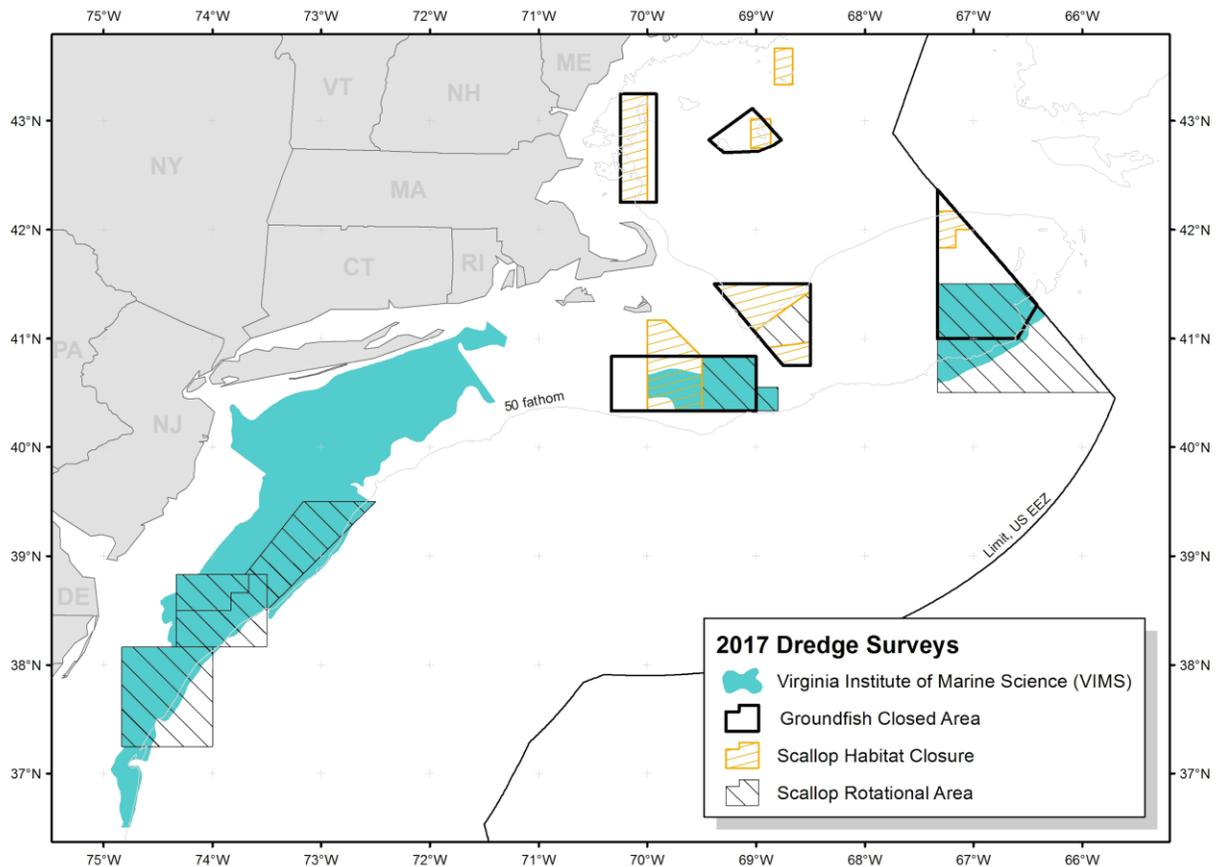


Figure 1. Delineation of survey areas during a typical NEFSC Scallop Dredge Survey. Figure credit: Virginia Institute of Marine Sciences (VIMS)

Historical bottom temperature and salinity climatology within the MAB-NYB were developed using high-resolution, quality-controlled monthly means from the Northwest Atlantic

regional bottom temperature and salinity climatology for 1955-2012 (1/10°; Seidov et al. 2016). Bathymetry data was obtained from the U.S. Coastal Relief Model (NGDC 1999).

Projected oceanographic conditions used in this study were developed using the delta-method (Fogarty et al. 2008; Hare et al. 2012). The delta method is commonly used for future climate projection, which relies on the difference between future climate anomalies and a baseline regional climatology (historical climate condition). The delta-method can remove the climate model projection biases (e.g. drift) and provide a simple and robust projection of future climate conditions (Hare et al. 2012).

Projected bottom temperature and salinity conditions used in this modeling framework are from a high-resolution global climate model developed at the NOAA Geophysical Fluid Dynamics Laboratory (GFDL CM2.6; Delworth et al. 2012; Saba et al. 2016). CM2.6 is a coupled atmosphere-ocean-land-sea ice global climate model, with a 0.1° average horizontal resolution for its ocean component (Saba et al. 2016). CM2.6 resolves mesoscale oceanographic processes and fine-scale bathymetry within the NEUS-LME leading to a better simulation of regional ocean and shelf circulation when compared to global climate models with coarser ocean components (Saba et al. 2016). The bottom temperature and salinity projections from CM2.6 are based on (1) the standard model initialization procedure where global atmospheric CO<sub>2</sub> is fixed at a 1860 pre-industrial concentration to bring the climate system into near-equilibrium, and (2) a transient climate response to simulated 1% year<sup>-1</sup> increase in global atmospheric CO<sub>2</sub> run (i.e. 2xCO<sub>2</sub> simulation) up to 70-years and is then fixed for an additional 10-years. The CM2.6 2xCO<sub>2</sub> simulation can be roughly compared to the IPCC highest greenhouse gas emissions scenario (IPCC-RCP8.5). Under the IPCC RCP8.5 scenario, the global mean surface temperature increases by 2 °C by 2060-70 relative to the 1986-2005 climatology (Winton et al. 2014). The CM2.6 2xCO<sub>2</sub> simulation projects the global mean temperature increases by 2 °C by year 60-80.

The regional 1955-2012 bottom temperature and salinity climatologies were used in conjunction with CM2.6 delta values to depict change in bottom temperature and salinity over the future 80-years. This results in bottom temperature warming on the scale of 1.8-2.9 °C over the simulated 80-years across the NEUS-LMES (Figure 2).

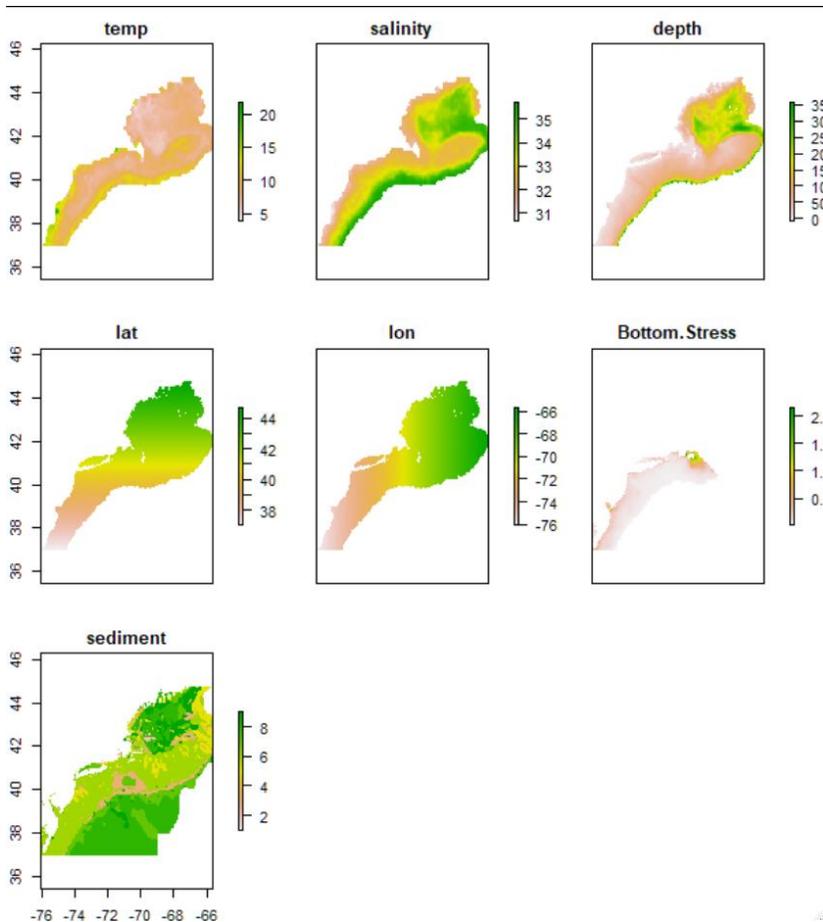


Figure 2. Environmental variables used in modeling framework. Dynamic environmental variables (bottom temperature, bottom salinity, and bottom stress) are annual mean values to represent the range of seasonal exposure for this species.

### Development of a Habitat Suitability Index (HSI)

Suitability indices (SIs) and Habitat Suitability Index scores (HSI) were developed using a machine learning technique, Random Forest models, from the R package randomForest (Liaw and Wiener 2002; R Core Team Development 2016) to regress presence of *P. magellanicus* with a suite of environmental covariates. Random forest regression was used because of its robustness to overfitting and flexibility of implementation (Liaw and Wiener 2002; Prasad et al. 2006). Random Forest regression is an ensemble learning technique based on combinations of sets of decision trees, where each tree is trained through the selection of a random bootstrap subset of the original dataset alongside a random set of predictor variables. To assess the importance of each predictor variable to the presence/absence of scallops in the study area as well as for the calculation of root mean squared error (RMSE), the random forest model withholds data (OOB, out-of-bag) to validate each tree for the classification of OOB error (Liaw and Wiener 2002; Prasad et al. 2006). This commonly used evaluation criterion shows the average error of a model when testing against the independent OOB data (Chai and Draxler 2014).

## Ensemble species distribution modeling algorithm

Ensemble species distribution models (SDMs) for scallop were calibrated using presence-absence data from the NEFSC scallop dredge survey. The environmental variables used for the ensemble scallop habitat modelling were directly obtained from the survey dataset (1982-2016). Bottom temperature, salinity and depth that were available at each tow location were used (Tanaka and Chen 2016; Torre et al. 2018). Furthermore, the location variables (i.e., latitude and longitude) were used to capture localized effects of a suite of proxy habitat factors (Winton et al. 2014; Rooper et al. 2014; Tanaka et al. 2017, 2018). Dynamic environmental variables (bottom temperature and salinity) collected from the bottom trawl survey were averaged across all seasons. Temporal aggregation of these variables was carried out in order to incorporate the largely sedentary nature of this species and reflect an annual range of conditions from a given location (Torre et al. 2018). Potential collinearity among environmental variables used to predict species occurrence were assessed by calculating variance inflation factors (VIFs; Zuur et al. 2007). No predictor variable showed VIF value  $>3$ , therefore the multicollinearity was assumed to be negligible.

Our ensemble SDM platform is based on the BIOMOD2 package developed in the R Programming environment (Alabía et al. 2016, 2018; Thuiller et al. 2016). The BIOMOD2-based ensemble SDM algorithm can incorporate species presence-absence data and combine multiple SDM algorithms to provide robust species occurrence-environment relationships. The following 10 SDM algorithms were explored to predict spatiotemporal changes in presence probability of scallop across the study area under current (average 1980-2015 climatology) and future climatic conditions (CM2.6; Figure 3): generalized linear model (GLM), boosted regression tree (GBM), generalized additive model (GAM), classification tree analysis (CTA), artificial neural network (ANN), surface range envelope (SRE), flexible discriminant analysis (FDA), multiple adaptive regression splines (MARS), random forest (RF), and maximum entropy (Maxent). In order to optimize the fit to species' response curves and increase predictive performance from all SDM algorithms used in the ensemble modeling framework, a tuning process was developed where individual SDM parameters were varied by means of a 10-fold cross-validation procedure (Breiner et al. 2018). For each algorithm, the SDM parameter setting yielding highest receiver operator curve (ROC) value during the cross-validation procedure was kept. For model tuning, the BIOMOD tuning function was used, which uses tuning functions from the CARET R-package to tune GBM, ANN, GAM, MARS, GLM, and CTA (Kuhn 2008), and ENMEVAL R-package to tune Maxent (Muscarella et al. 2014).

Appendix A describes the results from the parameter tuning process for each algorithm used in the ensemble species distribution model (SDM) developed for scallops in the Mid-Atlantic to New York Bight study area. Tuned parameters for each algorithm were set based on these results (Appendix A) for the final ensemble SDM to optimize model performance.

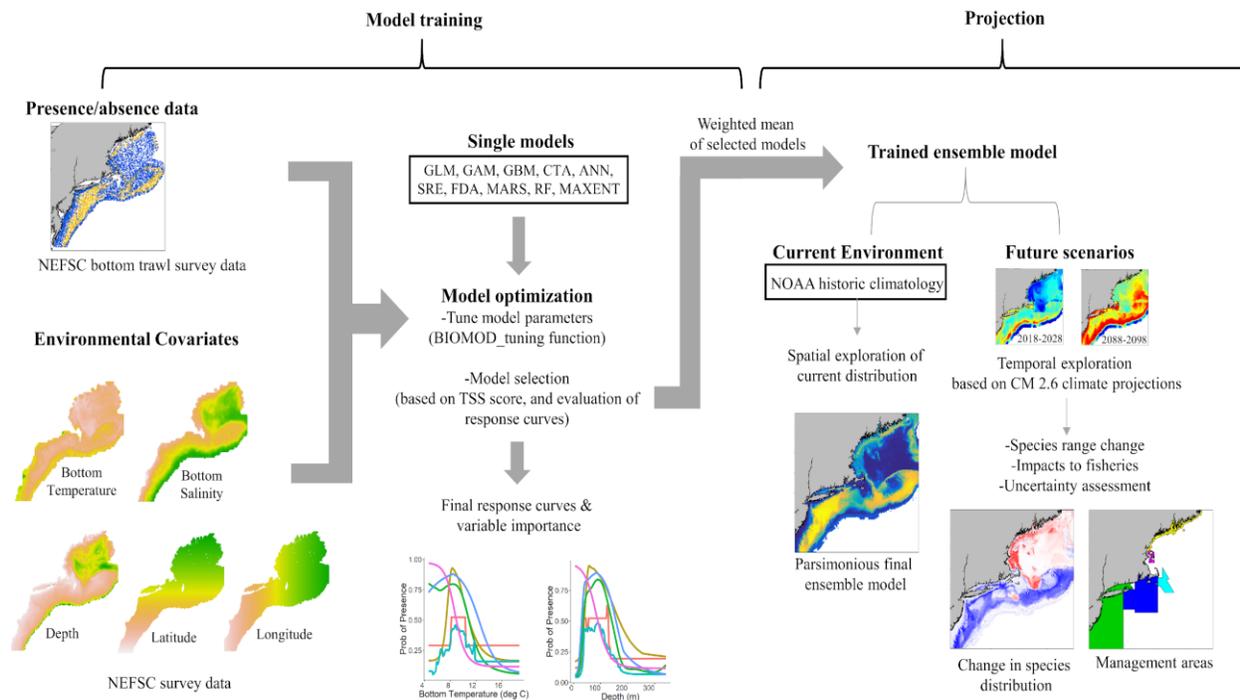


Figure 3. Schematic diagram describing the development of ensemble species distribution models.

Once SDMs were fitted with optimized parameters, all SDMs were run three times each using a randomly chosen 80% of the presence-absence data, with the remaining 20% of the data being used to cross-validate model results. A balance of three runs per each SDM was struck to limit computational demands while still achieving stable results (Thuiller et al. 2016). Two SDM evaluation criteria, ROC and the true skill statistic (TSS), were calculated through cross validation and used to assess the performance of each algorithm, with higher values for each metric being an indication of higher model skill (Hill et al. 2017; Mi et al. 2017; Figure 3). The best-fitting SDM performance were evaluated against predetermined thresholds (TSS > 0.5 and ROC > 0.8; Hill et al. 2017; Mi et al. 2017). The built-in BIOMOD2 function was used to compute the relative variable importance derived from the SDM runs that meet TSS and ROC thresholds.

An ensemble SDM was built for scallop using a composite of the SDM runs that meet predetermined TSS and ROC thresholds. Additionally, the response curves for all selected SDM outputs (response curves and spatial predictions) were visually assessed for validity. Top performing algorithms (excluding those with implausible results) were ranked by the TSS score and combined, using a weighted average of TSS scores, to produce the final ensemble model, which predicts the probability of presence for scallop at 0.1° resolution across the study area. The weighted average of individual SDMs to form the final ensemble projection (FEP) was calculated as follows:

$$FEP = \frac{\sum_{i=1}^n A_i \times TSS_i}{\sum_{i=1}^n TSS_i} \quad [1.]$$

where  $A_i$  denotes the habitat suitability (probability of presence) a single run of one of the 10 SDM algorithms;  $TSS_i$  denotes the true skill statistic score received by that run; and  $n$  is the total number of all runs of all algorithms to be included into the final ensemble model.

### **Projections of future habitat**

Weighted species-specific ensemble SDM was used to project the probability of occupancy of scallop across the NEUS-LME under historic bottom temperature and salinity conditions and over the future 80-years based on the transient climate-driven bottom temperature and salinity changes in response to the doubling of 1%  $CO_2$  per year increase scenario applied in the NOAA GFDL CM2.6 (Saba et al. 2016). The mean regional bottom temperature and salinity climatology for 1955-2012 was used to project species-specific habitat suitability under the “current” conditions. Species-specific habitat suitability change under future bottom temperature and salinity conditions were generated by combining the weighted ensemble SDM with the CM2.6 bottom temperature and salinity fields at each time step.

The magnitude of temperature and salinity-driven changes in species-specific habitat suitability across the study area were analyzed. A two-sample Kolmogorov-Smirnov test (KS-test) test at the 95% level was used to quantify the difference in the modeled probability of presence at a given location between first and last 10-years of the projected 80-years (Conover 1971; DeGroot and Schervish 2002). A two-sample KS-test was conducted at every ensemble projection output grid ( $0.1^\circ$ ), and all the “local” hypothesis tests ( $p < 0.05$  indicating that the distribution of two samples compared are significantly different) were aggregated and summarized using both D-statistic and p-values. In order to further evaluate the spatial changes in probability of occupancy of scallop across the NEUS-LME, ensemble projections were further aggregated for each management zone (see Figure 1) to highlight changes in species habitat suitability within those economically relevant areas (GOM-GB and SNE nearshore areas) and also to isolate distinct areas spanning a gradient of oceanographic conditions (e.g. GOM-GB and SNE).

### **Assessment of uncertainty in the ensemble projections**

Finally, unanimity of agreement among individual ensemble members (i.e., every SDM run with  $TSS > 0.5$  and  $ROC > 0.8$ ) was analyzed to assess the magnitude of uncertainty in ensemble projections (Collins et al. 2013). A linear trend from every individual SDM projection was classified as either an increase (positive slope with  $p < 0.05$ ), decrease (negative slope coefficient with  $p < 0.05$ ), or no change ( $p \geq 0.05$ ). Agreement ratio among individual ensemble members was then calculated for each management area, where a ratio of 1 indicates that all individual SDM projections exhibited the same trend (increase, decrease or no change). Agreement ratio was calculated from both weighted and unweighted individual SDM projections. Unweighted agreement ratio indicated that all individual SDM projections were considered equally, while weighted agreement ratio indicated that individual SDMs with higher

skills were given more weight on agreement ratio. For a given management area, a projected trend was considered “likely” (agreement ratio above 66%) or “unlikely” (agreement ratio less than 33%). Furthermore, agreement ratio from both weighted and unweighted individual ensemble members was compared to assess the robustness of projected habitat suitability changes within each management area.

## Results

### Habitat Suitability Index (HSI)

#### Suitability indices and variable importance

Peak Suitability Index values for each Scallop-Environment relationship were as follows: ideal yearly bottom temperature  $\sim 9^\circ\text{C}$  inshore, ideal depth range  $< 100\text{m}$ , bottom stress  $< 0.10$ ; sediment grain size between 4 and 6.5 (Figure 4). Salinity did not appear to have large contrast over the study area. However, a slightly decreasing trend was observed likely reflecting the offshore distribution of this species in the MAB-NYB.

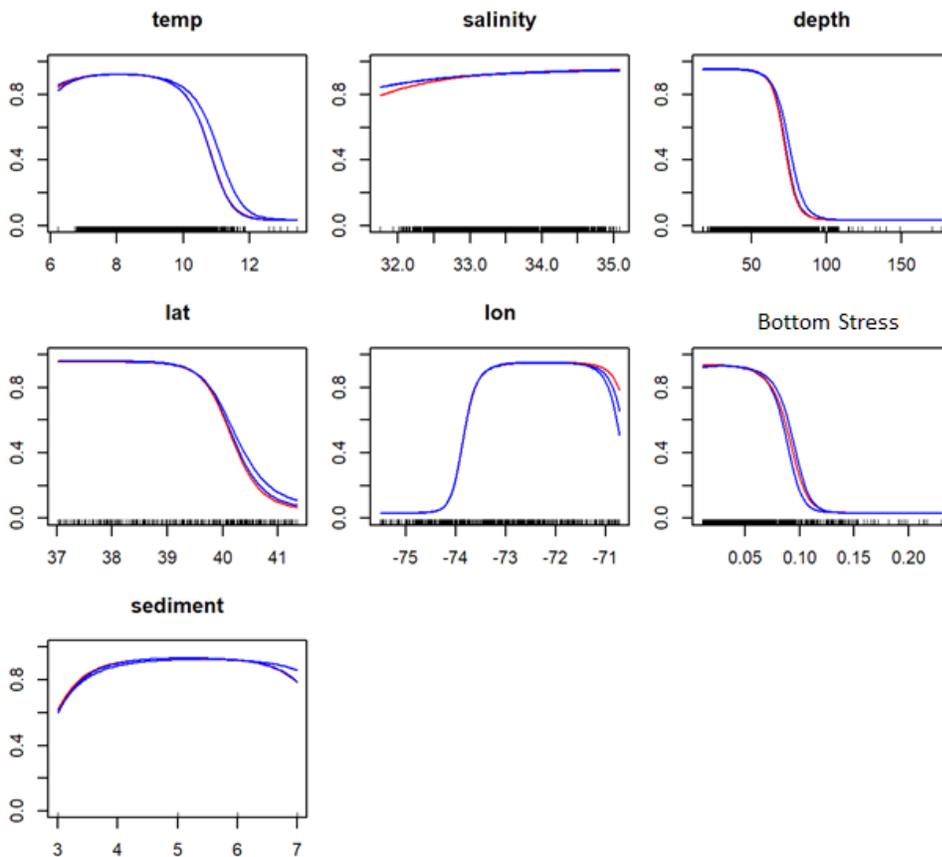


Figure 4. Scallop-Environment response curves developed from the random forest model. These curves represent suitability indices, where highest scallop abundance is structured along the peaks of each respective plot.

The results from the random forest model cross validation study (out-of-bag error evaluation for each tree) show the relative importance of each environmental predictor on habitat quality for scallops in the MAB-NYB. In this model, bottom stress was considered to be the most important predictor, followed by depth, bottom salinity, and temperature (Figure 5). Sediment grain size was of relatively little importance in predicting scallop habitat quality, likely due to the largely homogeneous distribution of sediment type throughout the region. The results from the random forest model cross validation study also show high performance of this approach with mean of squared residuals = 2.67 and percent of explained variance = 61.1.

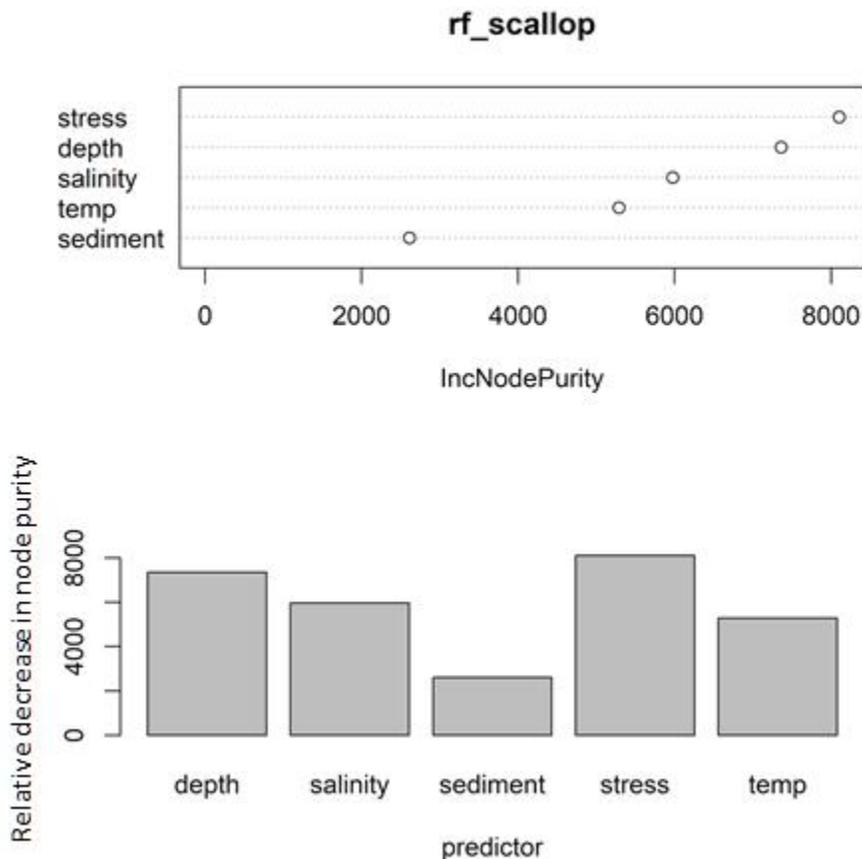


Figure 5. Relative importance of each environmental variable with respect to its predictive ability in the random forest model for scallop habitat suitability in the MAB-NYB.

## Spatial HSI evaluation

The HSI-based bioclimate envelopes for scallop showed higher habitat suitability in offshore ( $> \sim 30\text{nm}$ ) areas relative to inshore ( $< \sim 30\text{nm}$ ) areas along most of the MAB-NYB (Figure 6). However, in SNE along southern Long Island suitable habitat ( $\text{HSI} > 0.8$ ) extends closer to inshore areas. Additionally, the area along southern Georges Bank showed high habitat suitability.

The proportion of total habitat with at least moderate habitat suitability ( $\text{HSI} > 0.5$ ) for scallop along the MAB-NYB shelf was 64.5%. The proportion of total habitat with high habitat suitability ( $\text{HSI value} > 0.8$ ) was 56.3%.

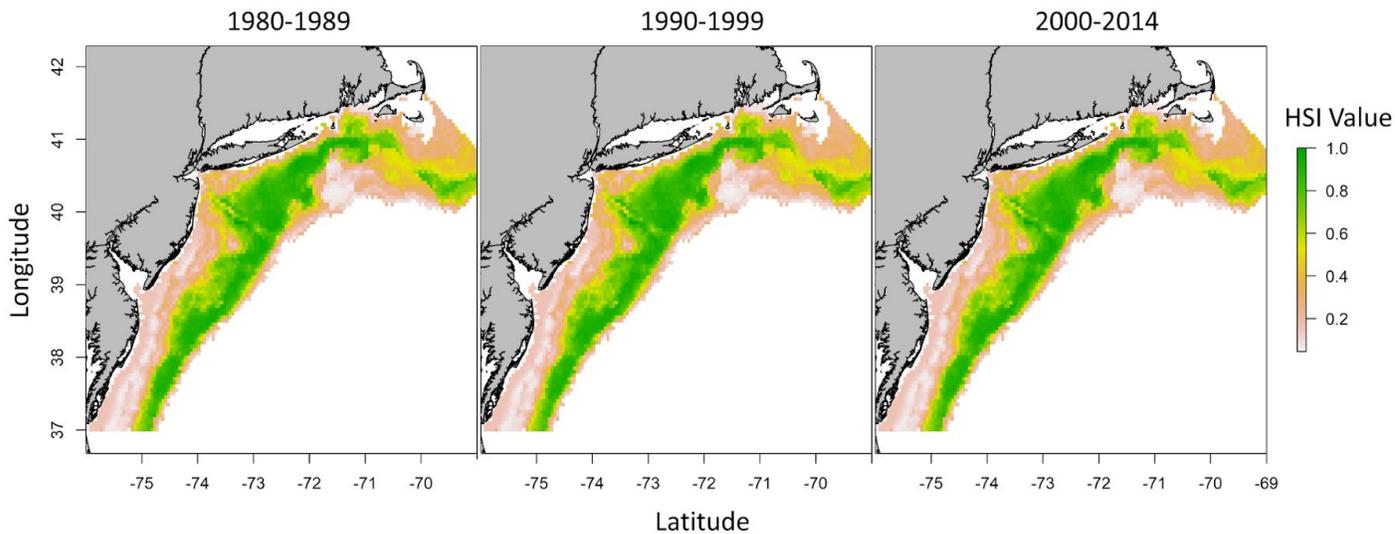


Figure 6. Spatial distribution of Habitat Suitability Index (HSI) values over the Mid-Atlantic Bight to the New York Bight. HSI values are shown here averaged across three distinct time periods spanning the study period (1980-2014). The HSI value indicates the change of habitat quality with 1 being the highest quality and 0 being unsuitable habitat.

Warming across the MAB-NYB over the hindcasted study period (1980-2014) resulted in spatial changes in habitat suitability for scallop (Figure 6). General changes in the distribution of available habitat for scallop shows an inshore-offshore trend where shallower waters have declined in habitat suitability and deeper areas, further offshore, have experienced an increase in habitat suitability. In addition, select areas within southern New England (SNE) have shown an increase in habitat suitability for scallop (Figure 7).

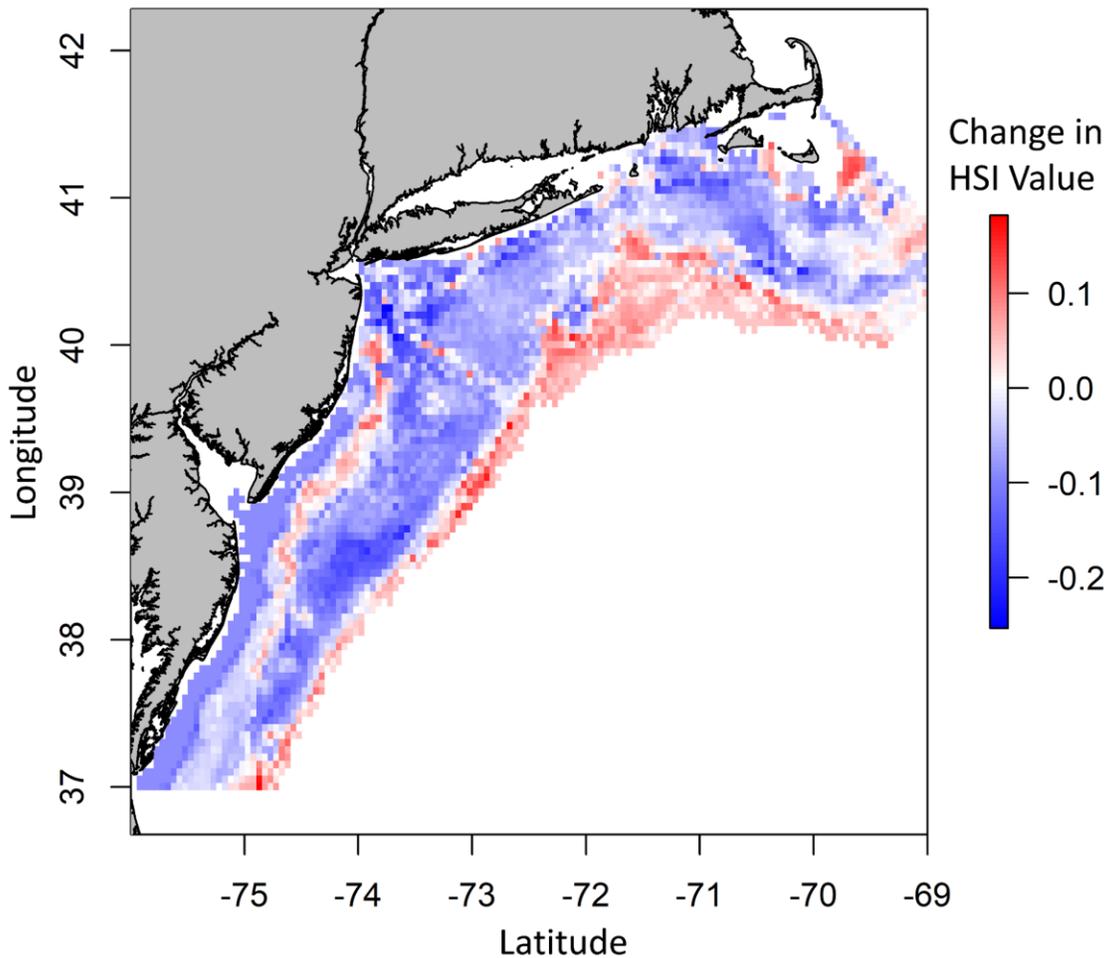


Figure 7. Temporal changes in the spatial distribution of Habitat Suitability Index (HSI) values for scallops within the Mid-Atlantic Bight to the New York Bight over the study period (1980-2014). Red (warm) color has a positive temporal trend, indicating an improved habitat quality over time, while blue (cold) color has a negative temporal trend, indicating a reduced habitat quality over time from 1980 to 2014.

### Scallop distribution in the New York Bight

Mapping of the NEFSC scallop dredge survey data gives an initial look at general scallop distribution in the MAB-NYB and provides a means to visualize the current distribution of scallop biomass in these areas. In general, scallops are primarily found in offshore waters shallower than 100m with very few individuals found deeper than 100m (Figure 8). Highest densities of scallops are shown between 45-75m (Figure 8). Scallops have undergone large decadal variations in distribution and biomass as the fishery has undergone boom and bust cycles over the past 30-40 years.

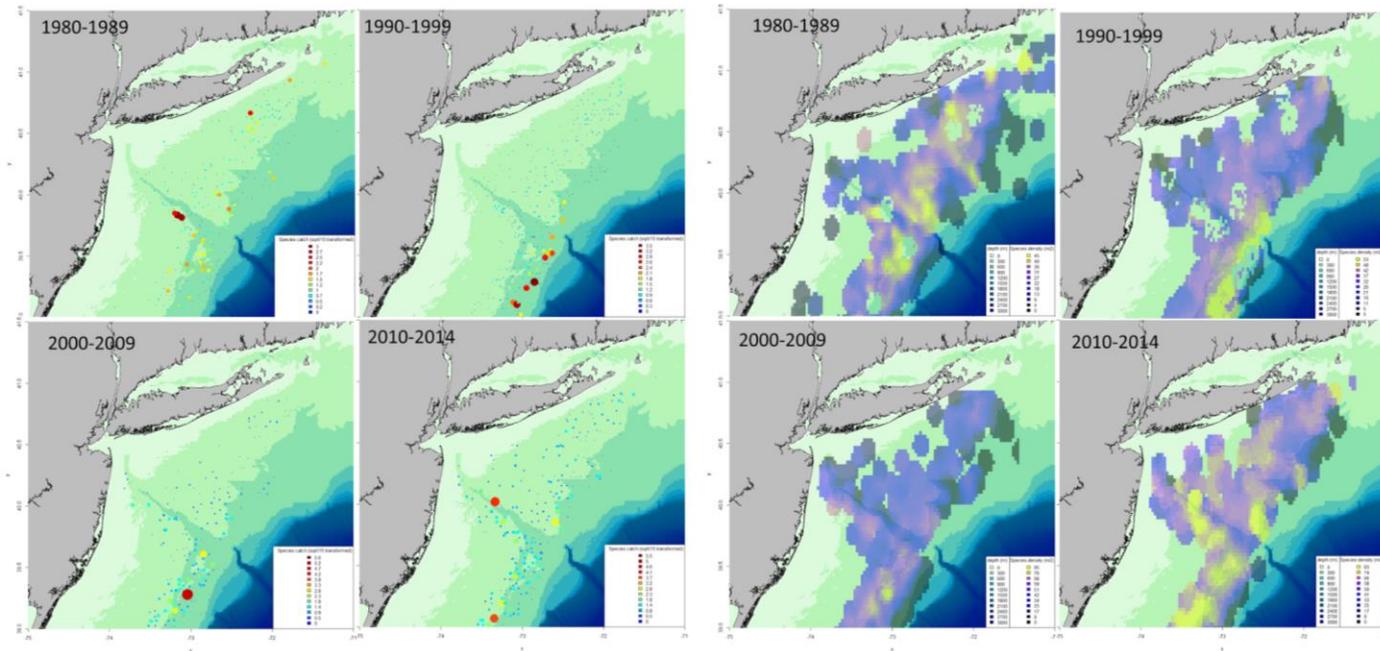


Figure 8. Scallop distribution and abundance based on data from the NEFSC Scallop Dredge Survey. Figures are broken up into 5 to 10 year periods from 1980 - 2015 in order to demonstrate temporal changes over the study period. Figures on the left show individual survey samples and figures on the right show a smoothed interpolated surface developed using ordinary kriging.

### Ensemble species distribution model for scallop

The ensemble modeling framework developed for the Atlantic sea scallop had 8 out of 11 tuned SDM algorithms meet the performance thresholds (TSS > 0.5 and ROC > 0.8; Figure 9). SRE, Maxent *Phillips*, and GLM did not meet the performance thresholds and were rejected from the final model. The prediction accuracies of the selected SDM runs were considered acceptable. Of the selected SDM runs, the best performing and highest weighted on the final ensemble model for scallop was GBM (Figure 9). Maxent *Tsuruoka* had the lowest acceptable performance and thus had the lowest weight on the final ensemble model (Figure 9).

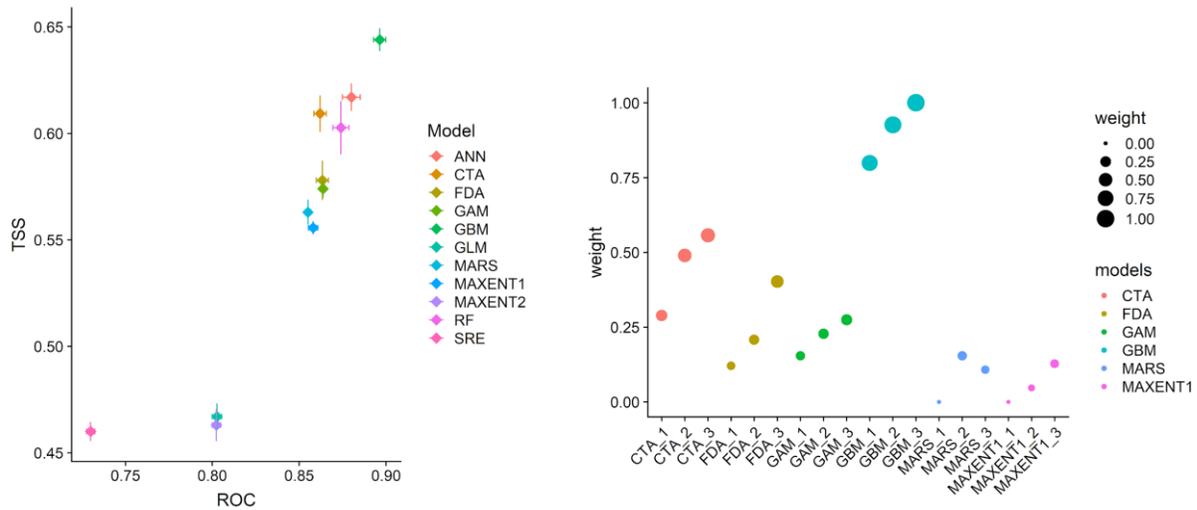


Figure 9. Performance criteria (TSS = True Skill Statistic; ROC = Receiver Operator Curve) for each SDM algorithm included in the ensemble modeling framework for scallops in the MAB-NYB.

Response curves from the individual SDM algorithms included in final ensemble models indicate non-linear relationships between the probability of presence and environmental variables for this species. The final ensemble model produced a spatial distribution that corresponds well to consensus of the general, broad scale distribution of scallop over the MAB-NYB (Figure 10). In general, higher probability of presence is found along offshore areas up to ~ 100m in depth (Figure 10).

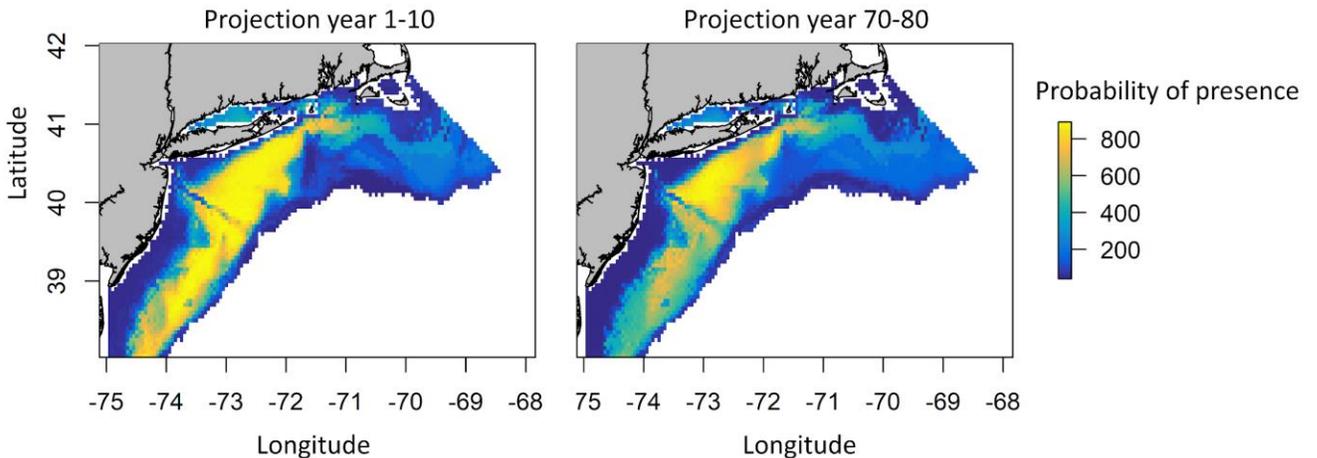


Figure 10. Spatial distribution of probability of presence values for scallops over the Mid-Atlantic Bight to the New York Bight. Probability of presence values in this figure represent a median value for hindcasted oceanographic conditions from 2015-2000 in order to portray a “current” distribution upon which projected values can be compared against. The scale (0 - 1000)

corresponds to the probability (0 low probability, 1000 high probability) of encountering the modeled species within a given cell.

Projected warming across the MAB-NYB over the next 80-years resulted in large spatial changes in probability of presence for scallop (Figure 11). The general distribution of changes in scallop probability of presence over the study period showed a general north-south trend, with southern areas experiencing a projected more sharp decline in probability of occupancy relative to northern areas (Figure 11). Select areas within southern New England (SNE) showed a projected increase in probability of occupancy for scallop (Figure 11).

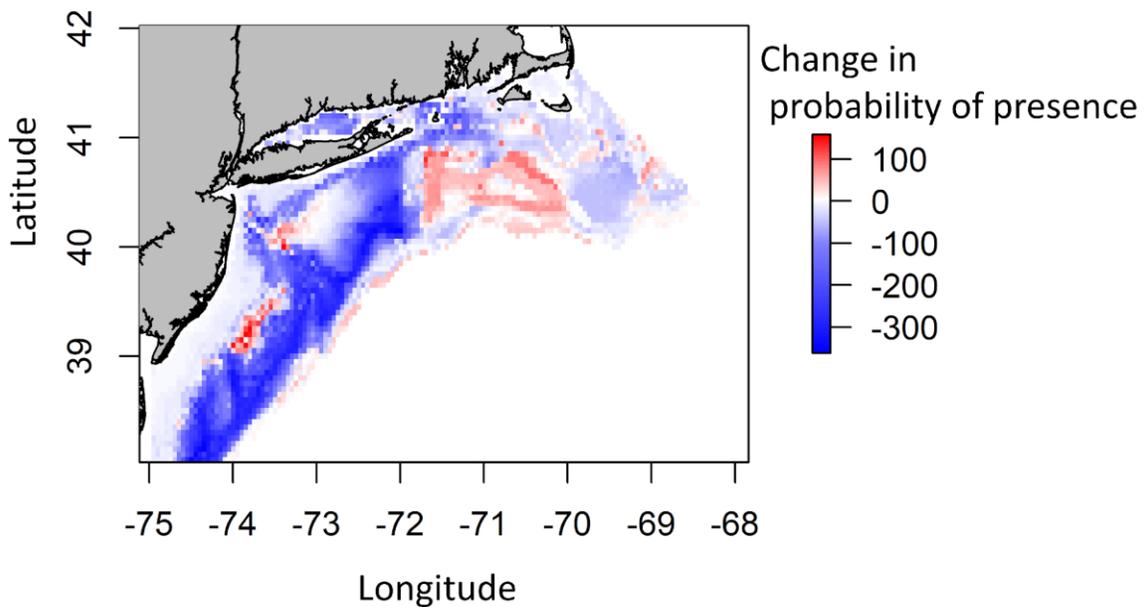


Figure 11. Climate-driven change in the spatial distribution of probability of occupancy values for scallops over the Mid-Atlantic Bight to the New York Bight. Changes in probability of occupancy values in this figure represent the degree of change at each raster cell over the 80 year projection with red (warm) color indicating increasing trends (positive trend) and blue (cold) color indicating decreasing trends over the 80 years of projection time period.

## Discussion

Altered oceanographic conditions driven by climate change are likely to result in shifts in species distribution, which may influence large scale marine spatial planning projects in the New York Bight. The ability to better understand these shifts can help the State to address concerns over impacts from future offshore wind energy areas to ecologically and economically important fisheries. Specifically, this project develops and parameterizes a species distribution model (SDM) for Atlantic Scallop in the MAB-NYB and establishes a means to infer the potential future habitats, based on the high-resolution climate data. The modeling effort also developed the

ability to visualize the current distribution of scallops within this area as well as projected changes to this distribution. The suite of modeling and visualization tools made available by this project will allow the state to better interface with wind energy developers and help to inform responsible placement of wind turbines within lease sites.

### **Justification for using ensemble SDM approach**

This modeling framework uses an ensemble approach in the development of the species distribution model (SDM) for scallop. Ensemble SDMs are increasingly being used in ecology as they address a major challenge associated with the choice of SDM algorithm that can have a large impact on projections (Araújo and New 2007; Buisson et al. 2010; Forester et al. 2013). A major challenge in the development of SDMs deals with the selection of an appropriate modeling methodology among a range of alternatives (Araújo and New 2007). Studies have shown that projections based on a single SDM, out of the myriad biostatistical approaches currently available, can have enough variability to cause misinterpretation of even a simple application (Pearson et al. 2006; Araújo and Luoto 2007). This challenge constitutes a large source of uncertainty in SDM outputs, however, a growing body of literature suggests that the combination of individual algorithms yield lower mean error than any individual constituent part (Araújo and New 2007; Buisson et al. 2010). In order to avoid issues associated with model selection bias, an ensemble modeling approach was used where multiple individual SDM algorithms are selected and weighted based on performance metrics along with visual assessment to ensure that model results fall within known physiological understanding of scallop. Through providing a method to alleviate issues associated with variability in ecological predictions across a wide range of currently available statistical methods, ensemble SDMs provide the distinct advantages of offering a more straightforward approach to model selection.

### **Limitations to this modeling approach**

It is critical to acknowledge that that modeled habitat suitability in this study should be interpreted as a proxy for probability of presence (occupancy) as opposed to actual lobster and scallop habitat suitability, given that measured catch was affected by some niche dimensions and processes not explicitly included in the predictors (e.g., territorial occupancy occurring at smaller scales) (Tanaka et al. 2018; Torre et al. 2018). Furthermore, projected species habitat suitability changes in this study should be viewed as a potential change in occupancy of a species due to changes in bottom temperature and salinity while holding all other top-down and bottom up variables constant (e.g. fishing pressure and species interaction). For example, environment-presence relationships for sea scallop were defined upon only five predictor variables when in reality, regional predator-prey interactions such as sea star predation on sea scallops can have a significant effect on the distribution of the species at smaller scales (Hart and Chute 2004; Hart 2006). Furthermore, we used latitude and longitude as proxy variables to capture a wide range of covarying bio-climate factors such as spatiotemporally variable fishing pressure and larval supply (Guernier et al. 2004; Wikgren et al. 2014; Shumway and Parsons 2016; Tanaka et al. 2018). As a result, certain areas with high quality habitats may have lower probability of detected presence. Another potential limitation of our modeling approach is that the interpolation of survey-derived environmental data masks the scale at which fine-scale habitat selection (active or passive) is occurring for each species. While these are important points to consider in

future studies, our study focused on evaluating changing probability of presence over a large spatial scale. The large spatiotemporal extent of observations used in this study reflect species presence-environment relationships reasonably well for both lobster and scallop. As more comprehensive environmental data becomes available in the future, a further detailed ensemble SDM approach could include additional variables such as pH, dissolved oxygen, predator-prey, and other food-web interactions to capture a more comprehensive representation of the biogeography of lobster and scallop (e.g., Bio-ORACLE <http://www.bio-oracle.org/>).

### **Spatial trends in projected scallop distribution**

The dominant temporal trend for scallop over the simulated 80-years is a climate-driven habitat reduction in the MAB-NYB. A significant decreasing trend in scallop habitat quality was projected in many parts of the study area. These results signify a likely contraction of suitable scallop habitat in offshore waters of the MAB-NYB. These findings parallel other studies on the spatial distribution of scallop in response to oceanographic change in the Northwest Atlantic (Howell 2012; Steneck and Wahle 2013; Wahle et al. 2013; Tanaka and Chen 2016; Torre et al. 2018). However, select northern areas within the MAB-NYB (south of Nantucket sound) show a projected increase in habitat quality which suggests a north-south gradient of habitat change over the course of the model projection. This modeling effort suggests that climate change will act as a likely stressor to the southern scallop fisheries and continues to drive the further contraction of scallop and habitats into northern areas. Similar findings were also reported for other major commercial stocks in the regions Nye et al. (2009).

### **Conclusions**

This study provides ensemble estimates of climate-driven changes and associated uncertainties in the biogeography of the Atlantic sea scallop, in the area from the Mid-Atlantic to the New York Bight (MAB-NYB), and has developed a modeling framework capable of being extended to other species of interest. Large spatial heterogeneity is present in the habitat suitability and distributions for the scallop in the MAB-NYB. Such spatial variabilities change over time, with some areas having improved habitat and distribution and other areas having reduced habitat quality and distribution for scallop over time in the MAB-NYB. Projected climate change in the MAB-NYB will pose management challenges for the scallop, and these ensemble projections provide useful information for climate-ready management of commercial fisheries.

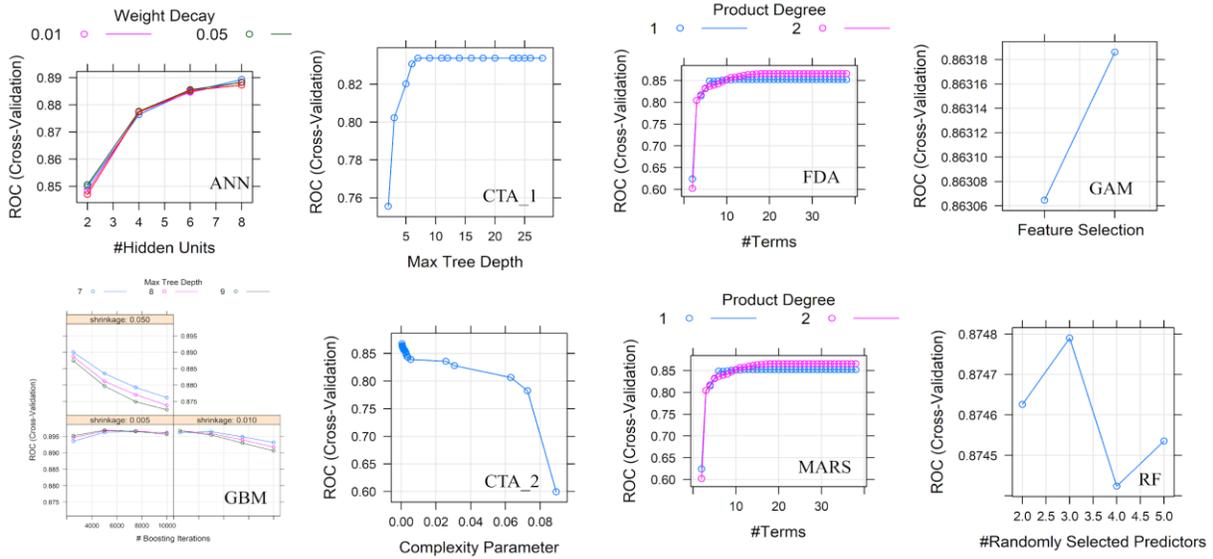
## References

- Araújo, M.B., and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Glob. Ecol. Biogeogr.* **16**: 743–753. doi:10.1111/j.1466-8238.2007.00359.x.
- Araújo, M.B., and New, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**(1): 42–47. doi:10.1016/j.tree.2006.09.010.
- Araujo, M.B., and Peterson, A.T. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**(7): 1527–1539.
- Boulangeat, I., D. Gravel, and W. Thuiller. 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters* 15(6):584–593.
- Chang, J.-H., Y. Chen, D. Holland, and J. Grabowski. 2010. Estimating spatial distribution of American lobster *Homarus americanus* using habitat variables. *Marine Ecology Progress Series* 420:145–156.
- Fogarty, Michael, Lewis Incze, Katherine Hayhoe, David Mountain, and James Manning. 2008. “Potential Climate Change Impacts on Atlantic Cod (*Gadus Morhua*) off the Northeastern USA.” *The North Atlantic Oscillation: Climatic Significance and Environmental Impact* 13 (5–6): 453–66. doi:10.1007/s11027-007-9131-4.
- Guan, L., Y. Chen, K. Staples, J. Cao, and B. Li. 2017. Spatial dynamics of Atlantic cod (*Gadus morhua*) in the Gulf of Maine. *ICES Journal of Marine Science* fsx064, <https://doi.org/10.1093/icesjms/fsx064>
- Hare, Jonathan A., John P. Manderson, Janet A. Nye, Michael A. Alexander, Peter J. Auster, Diane L. Borggaard, Antonietta M. Capotondi, et al. 2012. “Cusk (*Brosme Brosme*) and Climate Change: Assessing the Threat to a Candidate Marine Fish Species under the US Endangered Species Act.” *ICES Journal of Marine Science* 69 (10): 1753–68.
- Hill, L., A. Hector, G. Hemery, S. Smart, M. Tanadini, and N. Brown. 2017. Abundance distributions for tree species in Great Britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. *Ecology and Evolution* 7(4):1043–1056.
- Howell, P., and Auster, P.J. 2012. Phase Shift in an Estuarine Finfish Community Associated with Warming Temperatures. *Mar. Coast. Fish.* **4**(1): 481–495. doi:10.1080/19425120.2012.685144.

- Intergovernmental Panel on Climate Change (IPCC). 2013. “Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.” Edited by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi:10.1017/CBO9781107415324.
- Kuhn, M. 2008. Building predictive models in r using the caret package. *J. Stat. Softw.* **28**: 1–26.
- Li, B., J. Cao, L. Guan, M. Mazur, Y. Chen, R.A. Wahle. 2018. Estimating spatial non-stationary distribution of species: a case study from American lobster in the Gulf of Maine. *ICES Journal of Marine Sciences* 75(4):1473–1482, <https://doi.org/10.1093/icesjms/fsy024>
- Mi, C., Huettmann, F., Sun, R., and Guo, Y. 2017. Combining occurrence and abundance distribution models for the conservation of the Great Bustard. *PeerJ* **5**: e4160. doi:10.7717/peerj.4160
- Miller, J. 2010. Species Distribution Modeling. *Geography Compass* 6:490–509.
- Muscarella, R., Galante, P.J., Soley-Gurdia, M., Boria, R.A., Kass, J.M., and Anderson, R.P. 2014. An r package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods Ecol. Evol.* **5**: 1198–1205.
- Nye, J.A., Link, J.S., Hare, J.A., and Overholtz, W.J. 2009. Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States.
- Pearson, R.G., Dawson, T.P., Berry, P.M., and Harrison, P.A. 2002. SPECIES : A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecol. Modell.* **154**: 289–300.
- Petruny-Parker, M.A. Malek, M. Long, D. Spencer, F. Mattera, E. Hasbrouck, J. Scotti, K. Gerbino, J. Wilson. 2015. Identifying Information Needs and Approaches for Assessing Potential Impacts of Offshore Wind Farm Development on Fisheries Resources in the Northeast Region. US Dept. of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs, Herndon, VA. OCS Study BOEM 2015-037 pp.
- Runnebaum, J., L. Guan, J. Cao, L. O’Brien, and Y. Chen 2018. Habitat suitability modeling based on a spatio-temporal model: an example for Cusk in the Gulf of Maine. *Canadian Journal of Fisheries and Aquatic Sciences* 75(11): 1784-1797, <https://doi.org/10.1139/cjfas-2017-0316>

- Saba, Vincent S., Charles A. Stock, James R. Spotila, Frank V. Paladino, and Pilar Santidrián Tomillo. 2012. "Projected Response of an Endangered Marine Turtle Population to Climate Change." *Nature Climate Change* 2 (11). Nature Publishing Group: 814–20. doi:10.1038/nclimate1582.
- Seidov, D., O.K. Baranova, D.R. Johnson, T.P. Boyer, A.V. Mishonov, and A.R. Parsons. 2016. "Northwest Atlantic Regional Climatology, Regional Climatology Team, NOAA/NCEI (www.nodc.noaa.gov/OC5/regional\_climate/nwa-Climate), Dataset doi:10.7289/V5RF5S2Q." [https://www.nodc.noaa.gov/OC5/regional\\_climate/nwa-climate/](https://www.nodc.noaa.gov/OC5/regional_climate/nwa-climate/).
- Tanaka, K., and Y. Chen. 2016. Modeling spatiotemporal variability of the bioclimate envelope of *Homarus americanus* in the coastal waters of Maine and New Hampshire. *Fisheries Research* 177:137–152. Elsevier B.V.
- Tanaka, K. R., J.-H. Chang, Y. Xue, Z. Li, L. Jacobson, and Y. Chen. 2018. Mesoscale climatic impacts on the distribution of *Homarus americanus* in the US inshore Gulf of Maine. *Fisheries Oceanography*.
- Torre, M. P., K. R. Tanaka, and Y. Chen. 2018. A spatiotemporal evaluation of *Placopecten magellanicus* habitat in the Gulf of Maine using a bioclimate envelope model. *Marine and Coastal Fisheries* 10(2):224–235.
- Torre, M.P., Tanaka, K.R., and Chen, Y. 2019. Development of a climate-niche model to evaluate spatiotemporal trends in *Placopecten magellanicus* distribution in the Gulf of Maine, USA. *J. Northwest Atl. Fish. Sci.* 50: 37–50. [10.2960/J.v50.m721](https://doi.org/10.2960/J.v50.m721).
- Torre, M. P., K. R. Tanaka, and Y. Chen. In Review. Using local ecological knowledge for evaluating statistical species distribution models: A case study with *Placopecten magellanicus* in the Gulf of Maine, USA. *Mar. Eco. Prog. Ser.*
- Zuur, A. F., E. N. Ieno & G.M. Smith. 2007. *Analysing ecological data*. Vol. 680. New York, NY: Springer. 672 pp.

# Appendix A



Appendix A. Each figure represents the results from the parameter tuning process for each algorithm used in the ensemble species distribution model (SDM) developed for scallops in the Mid-Atlantic to New York Bight study area. Tuned parameters for each algorithm were set based on these results for the final ensemble SDM to optimize model performance.