Spring 5-10-2019

Linking Environmental Variability to the Biogeography of Placopecten magellanicus in the Gulf of Maine

Michael Paul Torre
University of Maine, Michael.torre@maine.edu

Recommended Citation

Torre, Michael Paul, "Linking Environmental Variability to the Biogeography of Placopecten magellanicus in the Gulf of Maine" (2019). *Electronic Theses and Dissertations*. 3113.
https://digitalcommons.library.umaine.edu/etd/3113

LINKING ENVIRONMENTAL VARIABILITY TO THE BIOGEOGRAPHY OF *PLACOPECTEN MAGELLANICUS* IN THE GULF OF MAINE

By

Michael Paul Torre

B.A. Elizabethtown College, 2010

M.S. University of Delaware, 2014

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Ecology and Environmental Sciences)

The Graduate School

The University of Maine

May 2019

Advisory Committee:

Yong Chen, Professor of Fisheries Sciences (Advisor)

Huijie Xue, Professor of Oceanography

David Townsend, Professor of Oceanography

David Hiebeler, Professor of Mathematics

Deborah Hart, Operations Research Analyst, Northeast Fisheries Science Center

LINKING ENVIRONMENTAL VARIABILITY TO THE BIOGEOGRAPHY OF *PLACOPECTEN MAGELLANICUS* IN THE GULF OF MAINE

By Michael Paul Torre

Dissertation Advisor: Dr. Yong Chen

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Ecology and Environmental Science)
May 2019

The Atlantic sea scallop (*Placopecten magellanicus*) supports a highly valuable fishery in the United States over its range on the Northwest Atlantic Shelf. Scallop distribution has been shown to be highly affected by changes in climactic variables. Therefore, long-term changes in the thermal regime of the Gulf of Maine are expected to greatly impact scallop ecology; however, these projected changes have rarely been quantified. The modeling framework developed for my dissertation research will improve our understanding of the distribution of scallop habitat as well as the biogeography for this species. Additionally, this modeling capacity will provide several tangible tools to visualize species distribution over space and time as well as to evaluate potential impacts of a changing Gulf of Maine ecosystem.

The framework for my dissertation research is comprised of 1) a bioclimate envelope covering the Gulf of Maine to quantify spatiotemporal variability in scallop habitat; 2) a statistical species distribution model to predict spatiotemporal changes in scallop distribution in the Gulf of Maine; 3) the design of a dredge survey in the Northern Gulf of Maine to obtain scallop biomass estimates; and 4) a two-stage modeling and computer simulation framework to refine fisheries surveys.

Due to changing oceanographic conditions within the Gulf of Maine ecosystem it is becoming increasingly important to view resource management from within the context of climate change. Effective management of marine resources requires knowledge of population distribution and dynamics, however; fisheries managers must frequently base decisions on limited information. The modeling framework developed in my dissertation establishes the ability to better visualize sea scallop distribution as well as to evaluate the potential impacts of a changing ecosystem on this species. The results provided by this research increase the extent of knowledge about sea scallop ecology and have the potential to contribute to the conservation of this species. Additionally, the modeling approaches developed throughout my dissertation are highly generalizable to a variety of commercially important species and may be useful in advising conservation efforts for other fisheries in the Northwest Atlantic to help ensure the implementation of adaptive management strategies under uncertain climate conditions.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Yong Chen for his support and encouragement as well as providing me with several great opportunities to apply my research to various projects throughout my time at the University of Maine. I would also like to thank the members of my advisory committee, Huijie Xue, David Townsend, David Hiebeler, and Deborah Hart for their guidance, support, and helpful feedback during my dissertation research.

Kevin Kelly of Maine Department of Marine Resources provided survey and interview data along with model feedback. Credit is due to Mike Kersula of Maine Department of Marine Resources for playing a large part in the design and implementation of the 2016 NGOM scallop survey. Many Chen lab members have helped me throughout my degree, especially Kisei Tanaka.

This research was supported by the US National Science Foundation Adaptation to Abrupt Climate Change IGERT program grant DGE-1144423, NOAA Scallop RSA fund, NSF Coastal SEES program and Maine Sea Grant College Program. This work uses the FVCOM developed by Dr. Changsheng Chen’s research team at the Marine Ecosystem Dynamics Modeling Laboratory, University of Massachusetts Dartmouth as well as NOAA Global Historical Climatology data from the National Climate Data Center and CM2.6 developed at the NOAA Geophysical Fluid Dynamics Laboratory.

I would like to thank my friends and family who have always supported me. Finally, I would like to thank Sarah and Mackiecito (dog-burrito) Ebel who have been a source of inspiration and endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1- BACKGROUND ON SEA SCALLOP ECOLOGY AND THE FISHERY IN THE UNITED STATES

1.1 Sea Scallop Ecology

The Atlantic sea scallop (*Placopecten magellanicus*) is a bottom-dwelling bivalve mollusk of the family *Pectinidae* which occurs on the continental shelf and has a distribution extending from the Gulf of St. Lawrence to Cape Hatteras, North Carolina, USA (Shumway and Parsons 2006).

Like most benthic species, sea scallop abundance and distribution are influenced by a complex array of interacting environmental variables such as depth, bottom composition, currents, temperature, and salinity (Wildish and Saulnier 1992; Stokesbury and Himmelman 1995; Hart and Chute 2004). Throughout their geographic range, sea scallops occur mainly at depths ranging from 15 to 110 m, but can be found in shallower water in the northern part of its range, where they have been reported at depths up to 2 m (Naidu and Anderson 1984; Carsen et al. 1995). Juveniles and adults are most abundant on gravel substrate, but are also commonly found on sand, shells, or small rocks (Thouzeau et al. 1991). Temperature is an important environmental factor influencing growth rates in this species with adult scallops showing optimal growth at temperature between 10-15 °C and temperatures above 21 °C being lethal (Stewart and Arnold 1994). This species prefers full strength seawater (~35 ppt), with salinities of 16.5 ppt or lower being lethal (Stewart and Arnold 1994).

Sea scallops are suspension filter feeders that use cilia-generated currents to filter water and suspended particulate material (Hart and Chute 2004). Phytoplankton and microzooplankton make up the majority of nutrition for this species, however, detritus particles and dissolved organic matter can be ingested as well and make up a minor portion of the diet for this species (Hart and Chute 2004). Sea scallops are usually found in environments with strong currents (Hart and Chute 2004), and flow velocity has been shown to be a key factor controlling waste removal, oxygen uptake, feeding, and growth rates (Stewart and Arnold 1994; Shumway and Parsons 2006). Optimal growth for this species is around 0.1 m s^{-1} (Wildish and Saulnier 1992) and feeding inhibition has been observed to start at ~ 0.25 m s^{-1} Pilditch and Grant (1999).

Reproduction in sea scallops is dependent on a number of local ecological conditions, but late summer to early fall spawning is typical in most populations (Shumway and Parsons 2006). Once fertilized, eggs generally reside on the sediment before developing into the larval stage (Hart and Chute 2004). The larval stage consists of a transition from the trocophore stage through the veliger stage and lasts for approximately 40 days (Stewart and Arnold 1994; Shumway and Parsons 2006). During this time the larvae are planktonic and can only minimally control their position via vertical migration (Hart and Chute 2004).

Sea scallops are largely sedentary, especially during the adult phase of their life history (Shumway and Parsons 2006). Thus, this species does not actively move to select for optimal habitat. Rather, the abundance and distribution of scallops are influenced by a multitude of habitat characteristics and ocean currents that interact to control larval settlement and survival into the adult stage (MacDonald and Thompson 1985a, 1985b; Macdonald et al. 1987; Thouzeau et al. 1991; Wildish and Saulnier 1992; Stokesbury and Himmelman 1995; Hart and Chute 2004).

1.2 Status of Sea Scallop fisheries in the United States

The federal scallop fishery has undergone a large recovery since the mid-1990s and now supports the most valuable fishery in the U.S. with total ex-vessel values exceeding $486 million in 2016 (NMFS 2016). Conversely, the Maine scallop fishery is relatively depleted, having bottomed out in 2005 when only 33,000 pounds of scallop were reported landed. Historically, the state fishery has undergone large, irregular fluctuations in abundance with high fishing pressure following peaks in scallop populations (Kelly 2012). While landings from this fishery are still relatively low, data from 2017 show that fisherman brought in 550,000 pounds, a marked increase from 2005.

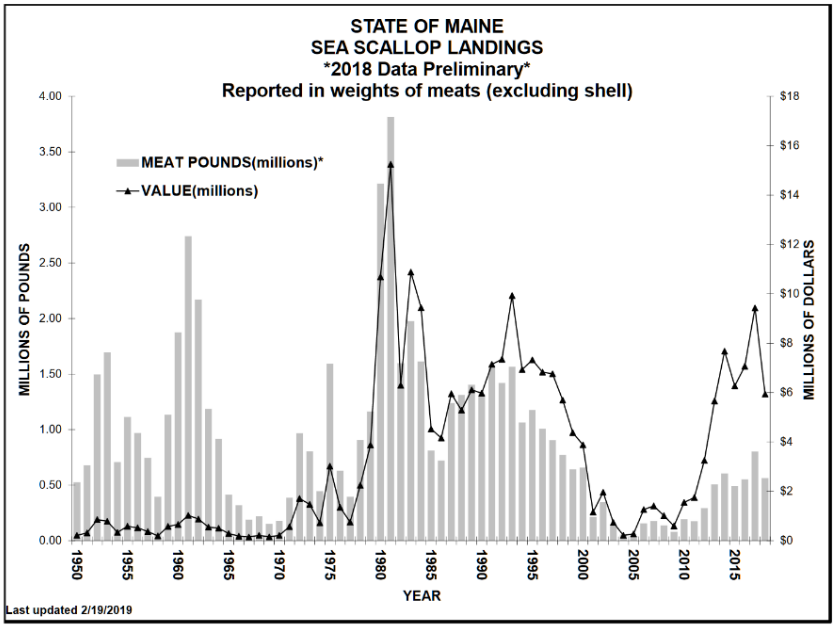


Figure 1.1: Maine landings from 1950 to 2018. Maine Division of Marine Resources

The value of scallops has increased from $8.23 per pound in 2005 to over $10 per pound in 2017. This fishery has been very valuable historically during periods of high abundance and has the potential to build back to this point, supporting local Maine fisherman and other stakeholders. Thus, improving stock assessment and management to maximize sustainable yield is highly important to the economies of coastal Maine areas where scallops are landed. Recent survey data shows that the increasing trend in scallop abundance is continuing, which provides an ideal opportunity to establish a sustainable, highly valuable fishery through properly informed management (Kelly 2012).

1.3 Impacts of a climactically altered Gulf of Maine ecosystem on sea scallops

There is compelling evidence that anthropogenic climate change is already causing long-term changes in oceanographic conditions that are impacting marine ecosystems (Townsend et al. 2004; Sumaila et al. 2011; Mills et al. 2013). The Northwest Atlantic in particular is undergoing rapid biophysical changes with water temperature increasing an average of 0.03 °C yr^{-1} since 1982 along with an accompanying increase in salinity (Mills et al. 2013; Pershing et al. 2015; Saba et al. 2015). This enhanced warming exceeds that which could be accounted for by the magnitude of natural decadal variability, and is associated with a northerly shift in the Gulf Stream, resulting in an increase of Warm-Temperate Slope Water entering the Northwest Atlantic Shelf (Pershing et al. 2015; Saba et al. 2015). Based on the transient climate response, which is the global mean temperature rise at the time of CO_2 doubling, warming on the scale of 3-4 °C is expected in the Northwest Atlantic Ocean over the next 70 years (Saba et al. 2015). Thus, climate change will continue to restructure ecosystems and have escalating effects on both marine resources and ecosystem services (Pershing et al. 2015).

The distribution of sea scallops has been shown to be keenly affected by changes in climactic variables (Kurihara 2008; Caputi et al. 2013; Tanaka et al. 2018; Torre et al. 2018). Dickie (1955) and Caddy (1979) showed that fluctuations in the abundance of sea scallops in the Bay of Fundy was correlated with bottom water temperature, with higher temperatures leading to increases in abundance. The influence of increased temperature is thought to result in rapid larval development and improved survival of juveniles and adults (Dickie 1955; Caddy 1979).Therefore, long-term changes in the thermal regime in the Northwest Atlantic are expected to greatly impact sea scallop larval settlement, survival, and resulting adult distribution (Thouzeau et al. 1991; Wildish and Saulnier 1992; Crossin et al. 1998; Caputi et al. 2013). In light of a changing GOM ecosystem (Mills et al. 2013; Pershing et al. 2015) it is becoming increasingly important to document the importance and synergistic effects of climate forcing on the dynamics of species abundance and distribution.

1.4 Structure of dissertation

The distribution and abundance of species are central concepts to ecological research and vital components of both conservation planning and fisheries management (Franklin 2010). The modeling framework developed for my dissertation research will improve our understanding of the distribution of habitat for sea scallops as well as the biogeography for this species. Additionally, this modeling capacity will establish the ability to quantify relationships between a commercially important fish stock and its surrounding environment, while providing several tangible tools to visualize species distribution over space and time as well as to evaluate potential impacts of a changing Gulf of Maine ecosystem. These chapters are arranged in order of the level of sophistication of the modeling approach and represent a progression of my skills in biostatistical modeling and computer programming.

Chapter 2 presents a bioclimate envelope model to evaluate the impacts of climate variability on spatiotemporal availability of suitable habitat for sea scallops in the Gulf of Maine (Torre et al. 2018). This modeling approach constitutes the qualitative component of my dissertation research. Bioclimate envelopes were established through Habitat Suitability Indices (HSI) based on bottom temperature, bottom salinity, current velocity, depth, and bottom composition. This model was coupled with a regional circulation model to establish the ability to hindcast spatiotemporal dynamics of suitable habitat for sea scallops in coastal and offshore waters of the Gulf of Maine from 1978 to 2013. The relationship between *sea scallop* abundance and each environmental variable was quantified using Suitability Indices (SIs) which were generated based on standardized scallop abundance sampled over 10 years of dredge survey data.

Boosted regression tree (BRT) models were used to determine the relative importance of each environmental variable to scallop abundance, which established a weighting scheme within the HSI. Higher habitat suitability was found along inshore areas compared with offshore. Model predictions indicate an increasing trend in habitat suitability in inshore waters since 1978 and decreasing habitat suitability in offshore waters.

Chapter 3 presents a climate-niche species distribution model to evaluate spatiotemporal trends in sea scallop distribution along the coastal waters of the Gulf of Maine (Torre et al. in press). A Tweedie-generalized additive model (GAM) was used to quantify the relationships between scallop abundance and key environmental variables. A boosted regression tree was used to identify significant interactions among environmental variables to integrate within the Tweedie GAM and a regional circulation model was incorporated with the Tweedie GAM to hindcast projections of scallop distribution and assess the impacts of environmental change on this species. Additionally, within this chapter, two common model fitting and variable selection methods for GAMs were evaluated to ensure high model performance. A classic backward variable selection procedure was compared to penalized thin plate regression splines. Projections from the climate-niche species distribution model show higher scallop density along inshore areas relative to those farther offshore. An increasing temporal trend in scallop density was observed along inshore areas and a decreasing temporal trend was observed in areas farther offshore. Additionally, the GAM incorporating thin plate regression splines was found to outperform the widely used backwards stepwise procedure.

Chapters 4 and 5 present an application of this modeling framework to improve fishery management for sea scallops in the Gulf of Maine. Chapter 3 comprises a technical report that details the design of the 2016 Northern Gulf of Maine (NGOM) sea scallop dredge survey and analysis of resulting data. This survey took place during May- June of 2016 and covered seven key areas within the NGOM management area. The goal of this survey was to provide reliable biomass estimates of harvestable size sea scallops within the area in order to inform management actions for the 2017 fishing season. This survey is an ongoing process, so it remains important to use each iteration of survey information to optimize the design with regards to accuracy and precision of sea scallop biomass estimates. To this end, chapter 5 presents a coupled two-stage species distribution model (SDM) and computer simulation framework to evaluate and optimize this dredge survey (Torre and Chen, in review). Model-predicted distribution of sea scallop biomass from the 2016 survey was used as the underlying population upon which multiple sampling effort allocation schemes were simulated to evaluate the bias and precision of biomass estimates from survey design implemented in 2016 and to test alternative approaches. The SDM developed in this chapter performed well predicting both the occurrence and biomass distribution of sea scallops. There was considerable difference, in terms of accuracy and precision, in the reliability of biomass estimates across survey designs. A marked improvement was observed in the relative root mean squared error (RMSE%) of biomass estimates from the best performing survey design (11.4) relative to the implemented in the 2016 dredge survey (26.0), justifying changes be made for future surveying efforts. The framework developed in this study offers a dynamic and adaptive approach for evaluating and improving fishery survey design.

In light of changing oceanographic conditions within the Gulf of Maine ecosystem it is becoming increasingly important to view resource management from within the context of climate change (Mills et al. 2013; Pershing et al. 2015; Saba et al. 2015). Effective management of marine resources requires knowledge of population distribution and dynamics (Langton et al. 1995), however; fisheries managers must frequently base decisions on limited information

(Brown et al. 2000). Additionally, many management plans fail to incorporate environmental variability (NMFS, 2010). The modeling framework developed in this dissertation provides several tools to visualize sea scallop biogeography as well as to evaluate the potential impacts of a changing ecosystem on this species. Additionally, the modeling approaches developed throughout my dissertation are highly generalizable to a variety of commercially important species and can be useful in advising conservation efforts for the scallop fishery in the Northwest Atlantic to help ensure the implementation of adaptive management strategies under uncertain climate conditions.

CHAPTER 2 - A SPATIOTEMPORAL EVALUATION OF *PLACOPECTEN MAGELLANICUS* HABITAT IN THE GULF OF MAINE USING A BIOCLIMATE ENVELOPE MODEL

2.1 Abstract

A bioclimate envelope model was developed to evaluate the impacts of climate variability on spatiotemporal availability of suitable habitat for Atlantic sea scallop (*Placopecten magellanicus*) in the Gulf of Maine. Bioclimate envelopes were established through Habitat Suitability Indices (HSI) based on bottom temperature, bottom salinity, current velocity, depth, and bottom composition. The relationship between *P. magellanicus* abundance and each environmental variable was quantified using Suitability Indices (SIs) which were generated based on standardized scallop abundance sampled over 10 years of dredge survey data. Boosted regression tree (BRT) models were used to determine the relative importance of each environmental variable to scallop abundance, which established a weighting scheme within the HSI. A regional circulation model was coupled with the weighted HSI to hindcast spatiotemporal dynamics of suitable habitat for *P. magellanicus* in coastal and offshore waters of the Gulf of Maine from 1978 to 2013. Higher habitat suitability was found along inshore areas compared with offshore. Model predictions indicate an increasing trend in habitat suitability in inshore waters since 1978 and decreasing habitat suitability in offshore waters. This research provides a novel modeling framework for *P. magellanicus* to enhance research and management of commercially valuable stocks over broad spatiotemporal scales in the climatically altered Gulf of Maine.

2.2 Introduction

The Atlantic sea scallop (*Placopecten magellanicus*) is a bivalve mollusk of the family Pectinidae. The species occurs on the continental shelf and its distribution extends from the Gulf of St. Lawrence to Cape Hatteras, North Carolina, USA(Shumway and Parsons 2006). *P. magellanicus* supports a highly valuable fishery overall in the United States. However, the scallop fishery in the Gulf of Maine is depleted, having bottomed out in 2005 when only 33,000 pounds were reported landed, compared with over 3 million pounds landed in the early 1990's (Kelly 2012). Recent data show an increasing trend in *P. magellanicus* abundance, which provides an ideal opportunity to establish a persistent, valuable fishery through focused research and management efforts (Kelly 2012).

The distribution and abundance of many benthic species are closely tied to their surrounding environment, which fluctuates over space and time (Dickie 1955; Slacum et al. 2010). Abundance and distribution of *P. magellanicus* are influenced by a multitude of habitat characteristics such as depth, bottom composition, currents, temperature, and salinity (MacDonald and Thompson 1985a, 1985b; Macdonald et al. 1987; Thouzeau et al. 1991; Wildish and Saulnier 1992; Stokesbury and Himmelman 1995; Hart and Chute 2004). *P. magellanicus* occur mainly at depths ranging from 15 to 110 m throughout its geographic range, but can be found in shallower water in the northern part of its range, where it has been reported at depths up to 2 m (Naidu and Anderson 1984). Both juveniles and adults of this species are generally found on sand, gravel, shells, or small rock substrate with gravel typically holding the highest abundances (Thouzeau et al. 1991). Adult *P. magellanicus* show optimal growth at temperature between 10-15 °C with temperatures above 21 °C being lethal and prefer full strength seawater (~35 ppt), with salinities of 16.5 ppt or lower being lethal (Stewart and Arnold 1994). *P. magellanicus* are usually found in environments with strong currents (Hart and Chute

2004), and flow velocity has been shown to be a key factor controlling waste removal, oxygen uptake, feeding, and growth rates (Stewart and Arnold 1994; Shumway and Parsons 2006). The feeding response of most suspension-feeding bivalves to increases in unidirectional flow velocity is expected to follow a unimodal function where, at low flow rates, increases in velocity enhance filtration up to an optimal rate upon which further increases result in feeding inhibition (Wildish and Kristmason 1993; Pilditch and Grant 1999). Optimal growth for this species is around 0.1 m s^{-1} (Wildish and Saulnier 1992). Pilditch and Grant (1999) observed inhibition of feeding at 0.25 m s^{-1}. Despite the clear influence of habitat quality on bivalve ecology, evaluation of *P. magellanicus*-habitat relationships and spatiotemporal trends of suitable habitats remain scarce (Shumway & Parsons, 2006; Mendo et al., 2014).

The Gulf of Maine (GOM) has warmed faster than the majority of the world's oceans with temperatures increasing an average of 0.026 °C yr^{-1} since 1982 (Mills et al. 2013; Pershing et al. 2015). Both bottom temperature and bottom salinity are also increasing (Saba et al. 2015). Rapidly increasing temperatures are impacting the abundance and distribution of numerous marine species (Overholtz et al. 2011; Howell and Auster 2012; Hollowed et al. 2013) with many fish stocks exhibiting a poleward shift in their center of biomass and/or an increase in depth (Nye et al. 2009). *P. magellanicus* distribution and abundance have been shown to be impacted by climatic variability (Frank et al. 1990; Kurihara 2008). Dickie (1955) and Caddy (1979) showed that fluctuations in the abundance of *P. magellanicus* in the Bay of Fundy was correlated with bottom water temperature, with higher temperatures leading to increases in *P. magellanicus* abundance. The influence of increased temperature is thought to result in rapid larval development and improved survival of juveniles and adults (Dickie 1955; Caddy 1979). In light of a changing GOM ecosystem (Mills et al. 2013; Pershing et al. 2015) it is becoming increasingly important to document the importance and synergistic effects of climate forcing on the dynamics of species abundance and distribution.

A bioclimate envelope model was developed in this study through coupling empirical Habitat Suitability Indices (HSI) with a regional ocean circulation model, to evaluate the climate-driven changes in habitat suitability for *P. magellanicus* from 1978–2013. Bioclimate envelopes are generally referred to as a multivariate space defined by a set of key climatic and environmental variables that best describes the physical and/or biological conditions of observed species distribution (Cheung et al. 2009; Araujo and Peterson 2012). HSI is an ecological index that can quantify the relationships between environmental variables and species abundance and predict where a species is likely to persist (Giannoulaki et al. 2011; Araujo and Peterson 2012). Habitat suitability-based bioclimate envelope models are increasingly used to quantify the impact of climatic variation on the spatiotemporal availability of suitable habitat for a given species (Pearson and Dawson 2003; Araujo and Peterson 2012; Tanaka and Chen 2016).

Bioclimate envelopes for *P. magellanicus* were defined by HSI derived from bottom temperature, depth, bottom composition, flow velocity, and bottom salinity. This HSI-based bioclimate envelope model has been adapted to incorporate the largely sedentary nature of adult *P. magellanicus* through temporal aggregation of dynamic environmental variables (temperature, salinity, and flow velocity) to reflect an annual range of conditions in a given location. The modelling effort developed in this study establishes the ability to evaluate spatiotemporal changes in bioclimate envelopes due to the incorporation of a regional ocean circulation model. Spatiotemporal trends in bioclimate envelopes are discussed with relation to potential climate-driven changes in *P. magellanicus* abundance and distribution. This research is novel as it’s the first bioclimate envelope model developed for *P. magellanicus* and provides a framework that can facilitate ecosystem-based management of commercially valuable GOM stocks.

2.3 Methods

2.3.1 Study Area

This modeling effort covers the inshore (<3 nm from shore) and offshore (>3 nm from shore) waters within the GOM from Cape Cod, Massachusetts to Downeast Maine, USA (Figure 2.1). The GOM is characterized by a mixture of oceanic influences directly affected by the Labrador Current, the Gulf Stream, and the freshwater discharge from the St. Lawrence River (Sutcliffe et al. 1976). As such, water temperature follows a gradient moving up the coast and also offshore. Within the study area, *P. magellanicus* were found in areas with maximum yearly temperature ranging from 8 °C in deeper areas to 16 °C in shallow areas. Salinity ranged from 26 ppt in inshore areas subject to freshwater inputs to full seawater (35 ppt) in offshore locations. This study covered depths to ~ 60 m since *P. magellanicus* are known to be uncommon deeper than this depth range (Hart and Chute 2004).

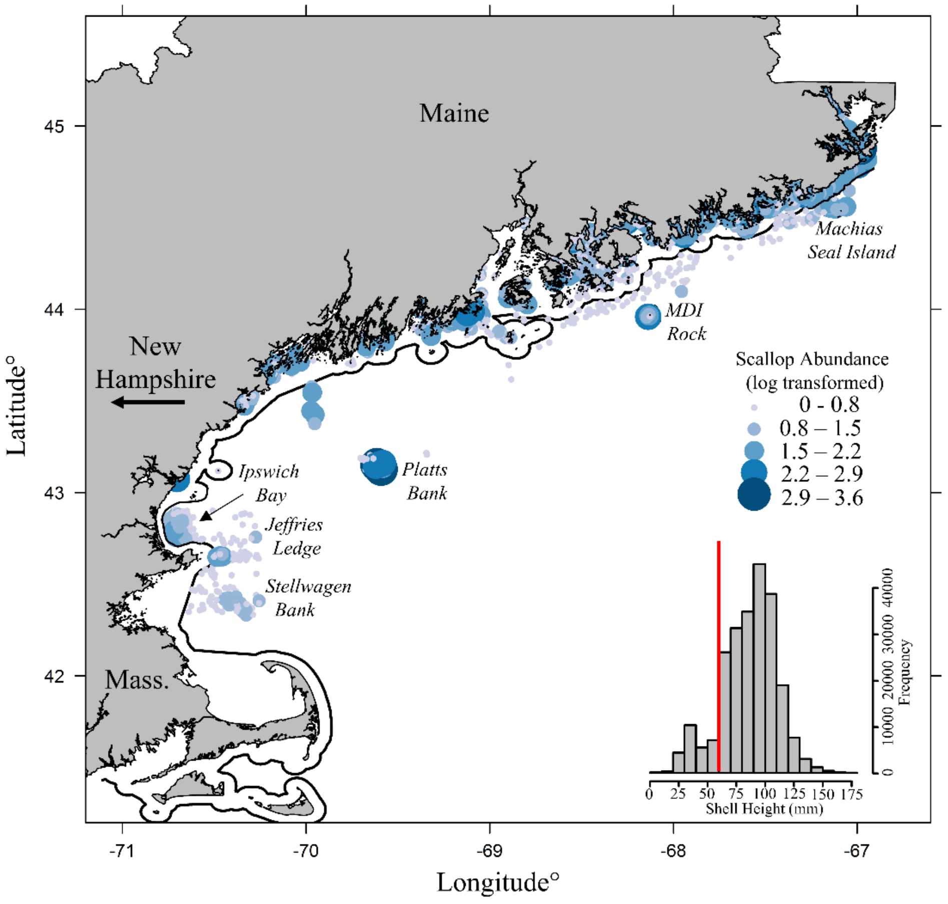


Figure 2.1: Spatial distribution of natural log standardized sea scallop abundance. The study area is in the Gulf of Maine from Cape Cod, Massachusetts to the Maine-Canada border. Size frequency is also shown with the vertical red line indicating the cutoff at 65mm shell height below which scallops were not included in the model.

2.3.2 Survey data

This study incorporates 10 years of dredge-based fishery-independent surveys of *P. magellanicus* along the inshore and offshore waters of Maine and Massachusetts conducted by the Maine Department of Marine Resources (2005-2014: Kelly, 2012; Figure 2.1). Sections of this total area were sampled intermittently throughout this time period. An annual spring survey covers alternating portions of inshore Maine waters and follows a stratified random design. An annual systematic survey covering select coastal areas occurs each fall. Offshore areas were surveyed in 2009 and 2012 using an adaptive two-stage random stratified design. The gear used for all surveys is an unlined, 7 ft New Bedford style drag with 2 in rings, 1.75 in head bale, 3.5 in twine top, 10 in pressure plate and rock chains. Since *P. magellanicus* < 65 mm in shell height were not efficiently sampled with the 2 in rings (Kelly 2012), these were excluded from all analyses. Tows were conducted at 3.5-4 knots and lasted from 2.5 to 5 minutes depending on the location, bottom type, and amount of fixed fishing gear in the area. All combined surveys yielded a total of 2,469 tows and captured 235,111 samples (Figure 2.1).

2.3.3 Environmental data

The unstructured-grid Finite-Volume Community Ocean Model (FVCOM) configured in the Northwest Atlantic Shelf region was used to simulate monthly estimates of bottom temperature, salinity, and current velocity from 1978 to 2013 throughout the entire study area. The FVCOM is a regional ocean circulation model developed by the University of Massachusetts-Dartmouth and the Woods Hole Oceanographic Institution (Chen et al. 2006). It has a horizontal resolution ranging from 0.02 km to 10 km (Chen et al. 2011).The unstructured FVCOM grid captures complex and irregular coastal geometry, making FVCOM suitable for physical and biological studies in coastal regions and estuaries (Chen et al. 2011). Since the survey did not measure temperature, salinity, or current velocity, FVCOM predictions were matched to survey tows from the nearest neighboring FVCOM point during time of sample (Figure 2.2). The absolute value of two-dimensional current velocity was taken to approximate the magnitude of water flow at a given location. Current velocity C was estimated at station i, and year y from FVCOM predictions using the following equation:

$$C_{i,y} = \sqrt{u_{i,y}^2 + y_{i,y}^2}$$

where *C* is the magnitude of the predicted current velocity; *u* and *y* are the x and y vector components of the velocity (Chen et al. 2011).

Bathymetry and substrate data were obtained from the U.S. Coastal Relief Model (CRM) and the Continental Margin Mapping (CONMAP) GIS database, respectively (NGDC, 1999; Poppe et al., 2005). Substrate type in the study area included gravel (pebbles defined as 2.00 - 64.00 mm, cobbles defined as 64 - 256 mm, boulder defined as above 256 mm), gravel-sand (0.62 - 2.00 mm), sand-clay (0.001 - 0.004 mm), sand- clay/silt (0.004 - 0.062 mm), sand-silt/clay, and sand/silt/clay (Poppe et al. 2005).

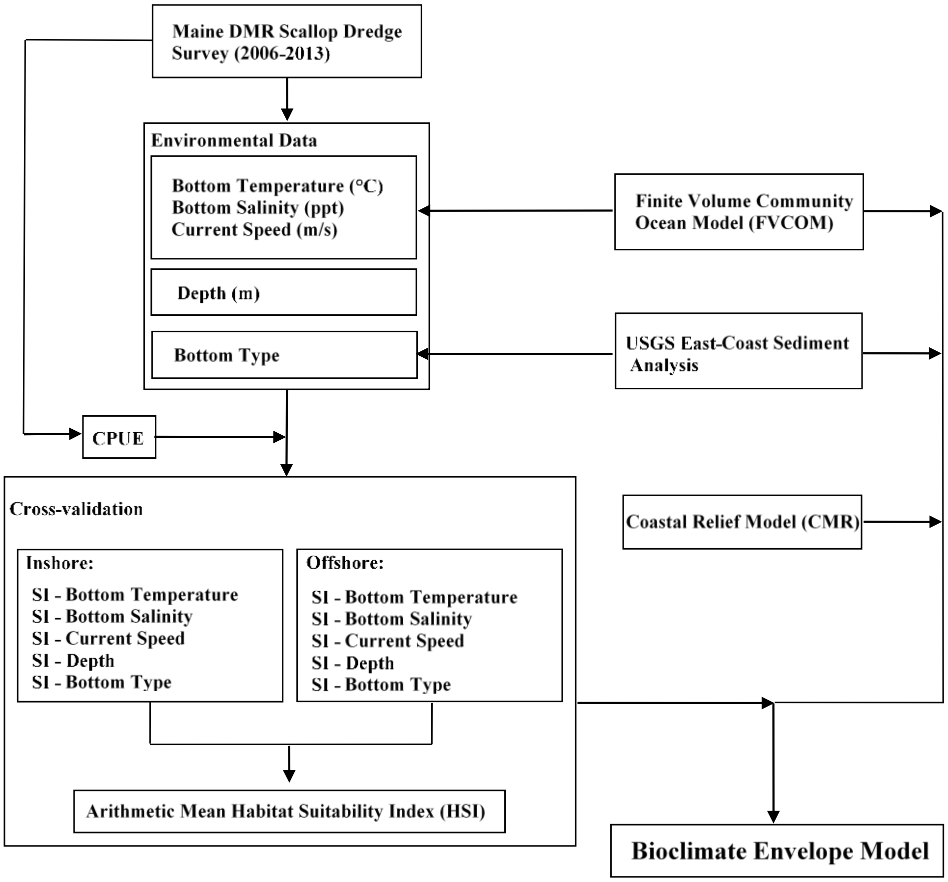


Figure 2.2: Schematic diagram of the bioclimate envelope modeling effort implemented in this study. All data exploration and modeling procedures were conducted in R programming environment.

2.3.4 Model development

The modeling approach used in this study to develop the HSI-based bioclimate envelope model (Figure 2.2) for *P. magellanicus* is an extension of previous modelling efforts for *Homarus americanus* in Long Island Sound and coastal New Hampshire to Maine (Tanaka and Chen 2015, 2016). Adult *P. magellanicus* are known to be largely sedentary, with little to no movement reported from previous studies (Posgay 1981; Carsen et al. 1995). Additionally, any movement that does happen is thought to be random in distance and direction, and any net movement over time is likely the result of tidal currents as opposed to active habitat selection (Posgay 1981). Given this, distribution of *P. magellanicus* corresponds to both successful settlement in an area and survival until capture. Consequently, temporally dynamic environmental variables (bottom temperature, bottom salinity, current velocity) in this model were incorporated as yearly aggregates (mean bottom temperature, mean salinity, and mean current speed across the 12 months prior to capture) in order to better reflect the range of conditions that an individual would experience in its location over time.

The standardized *P. magellanicus* abundance derived from the dredge surveys was used to develop suitability indices (SIs) for each environmental variable. The nominal abundance index was calculated as a survey catch per unit of sampling effort (CPUE) at station *i*, and year *y* (Chang et al. 2012; Tanaka and Chen 2015, 2016);

$$CPUE_{iy} = \left(\frac{Count_{iy}}{Tow\ Duration_{iy}}\right) * 2.5$$

where *Count* represents the total abundance of all captured *P. magellanicus* > 65 mm in shell height. *Tow duration*, measured in minutes, varied from 2.5 to 7 minutes depending on location, bottom type, and amount of fixed fishing gear in the area and was standardized to 2.5 minutes at each station. In this study, *P. magellanicus* inhabiting offshore waters (>3 nm from shore) were found to exhibit different habitat preferences relative to those inhabiting inshore waters (<3 nm from shore). Thus, a separate set of SIs were developed for inshore and offshore areas to more accurately reflect an optimal (SI >0.8) range for each environmental variable.

All continuous environmental variables were binned using Fisher's natural breaks classification method (Bivand 2013). The number of bins ranged from 6-10 to ensure adequate sample size in each data grouping. Each SI of class *k* for environment variable *i*, $SI_{i,k}$, was calculated on a scale of 0.0 - 1.0 using the following equation (Tanaka and Chen 2015, 2016);

$$SI_{i,k} = \frac{CPUE_{i,k} - CPUE_{i,min}}{CPUE_{i,max} - CPUE_{i,min}}$$

where $CPUE_{i,k}$ represents the average CPUE over all sampling stations falling within the class *k* of environmental variable *i* in each *P. magellanicus* group. $CPUE_{i,min}$ and $CPUE_{i,max}$ represents the minimum and maximum values of the average CPUEs of all the classes for environmental variable *d*, respectively. To analyze the relationships between each environmental variable and *P. magellanicus* abundance, estimated SI was assigned to each class of environmental variables in the form of a linear transfer function, where the most suitable class (SI = 1) and the least suitable class (SI = 0) were identified (Bayer and Porter 1988).

Inshore and offshore Suitability Indices (SIs) were estimated using the histogram method (Tanaka and Chen 2016). Local polynomial regression fitting (LOESS) smoothing was applied to the SIs (R Core Development Team, 2008). Suitable ranges were identified as SI values above 0.8 (Tanaka and Chen 2015, 2016). Boosted regression tree (BRT) models were used to identify relative importance of environmental variables on the response variable (Elith et al. 2008; Xue et al. 2017). Using this method, weights were assigned to each environmental variable corresponding to its relative contribution (%) to the deviance explained in the BRT model (Xue et al. 2017). BRTs were developed with the “gbm.step” function within the “gbm” R package (Ridgeway 2015). The SIs were then combined to form a composite HSI (also on a 0 - 1 scale) using an arithmetic mean model (Xue et al. 2017);

$$HSI = \frac{1}{\sum_{i=1}^{n} w_i} \; x \; \sum_{i=1}^{n} SI_i w_i$$

where SI_i represents a SI value associated with the *i*th environmental variable, and w_i represents the weight of variable *i* based on BRT results and *n* represents the number of environmental variables included in the Arithmetic Mean Model (AMM). In this study, only the AMM was used to generate HSI predictions as previous studies have shown consistently better performance of this model over a geometric mean model (GMM) (Tanaka and Chen 2015, 2016).

2.3.5 Model validation

A cross-validation study was implemented to evaluate performance and predictive ability of weighted HSIs. A randomly selected subset representing 80% of all data (training data) was used for HSI development, while the remaining 20% (testing data) was used for the evaluation of HSI performance (Smith 1994; Zuur et al. 2007; Tanaka and Chen 2015, 2016). The predicted HSI values based on the training data were compared against the observed HSI values, based on testing data, and linear regression analysis was performed to evaluate the predictive performance of the HSI. This cross-validation procedure was repeated 100 times using random data selection in each round to obtain 100 sets of linear regression parameters (intercept, slope, and R^2). Good model performance was indicated by an intercept parameter close to zero, a slope close to one and an R^2 close to one. This process was carried out separately for inshore and offshore stations.

2.3.6 Spatio-temporal HSI-based bioclimate envelope evaluation

The weighted HSI model coupled with FVCOM was used to predict spatiotemporal variability of bioclimatic envelope for *P. magellanicus (*> 65mm in shell height) in inshore and offshore GOM areas between 1978 and 2013. A spatial interpolation technique using ordinary Kriging with a semivariogram function was used to produce continuous model outputs (Bailey & Gatrell, 1995; R Core Development Team, 2008). HSI values were aggregated temporally by obtaining the median HSI value over the 36 years of study period at each FVCOM node. Median HSI values were used as opposed to mean because it provides a clearer interpretation of the tendency over the 36 year study period, i.e. it is not susceptible to skewing in rare cases of outliers. The distribution of median HSI over 36 years was evaluated for the spatial distribution in the quality of bioclimate envelopes. In this study, an area with a HSI value larger than 0.8 was designated as highly suitable habitat, while areas with a HSI value below 0.3 was considered poor habitat (Tanaka and Chen 2015, 2016). Linear regression analysis was performed at every FVCOM node and the derived slope (β) coefficient was used to evaluate temporal change in quality of *P. magellanicus* bioclimate envelopes over 36 years.

2.4 Results

2.4.1 Suitability indices

Highest yearly bottom temperature ranged from ~9 – 16 °C inshore, and ~7 – 12 °C offshore. Lowest yearly bottom salinity ranged from ~26 – 31.5 ppt inshore and ~30.5 – 33ppt offshore. Depth ranged from ~ 2 – 25 m inshore and ~15 – 57 m offshore. Average current speed ranged from close to 0 – 0.1 m s^{-1} for both inshore and offshore (Figure 2.3). Substantial differences in SI curves were found between *P. magellanicus* located within inshore areas relative to offshore (Figure 2.3). Peak SI for each environmental variable were as follows: highest yearly bottom temperature ~ 15 °C inshore, ~10 °C offshore; Lowest yearly bottom salinity ~31 ppt inshore, ~33 ppt offshore; depth ~ 10 m inshore, ~37 m offshore; average current speed ~0.05 m s^{-1} inshore, 0.1 m s^{-1} offshore (Table 2.1, Figure 2.3).

Table 2.1: Summary of location specific suitable ranges (Suitability Index > 0.8) of each environmental variable.

		Suitable Variable Range (Suitability Index > 0.8)			
Location	Mean Bottom Temperature (°C)	Mean Bottom Salinity (ppt)	Mean Flow Velocity (m/s)	Depth (m)	Bottom Composition
Inshore	8.6 – 9.4	30.1 - 31.3	0.04 - 0.07	5 - 12	sand
Offshore	8.1 – 8.7	33.0 +	0.07 - 0.13	35 - 41	gravel-sand

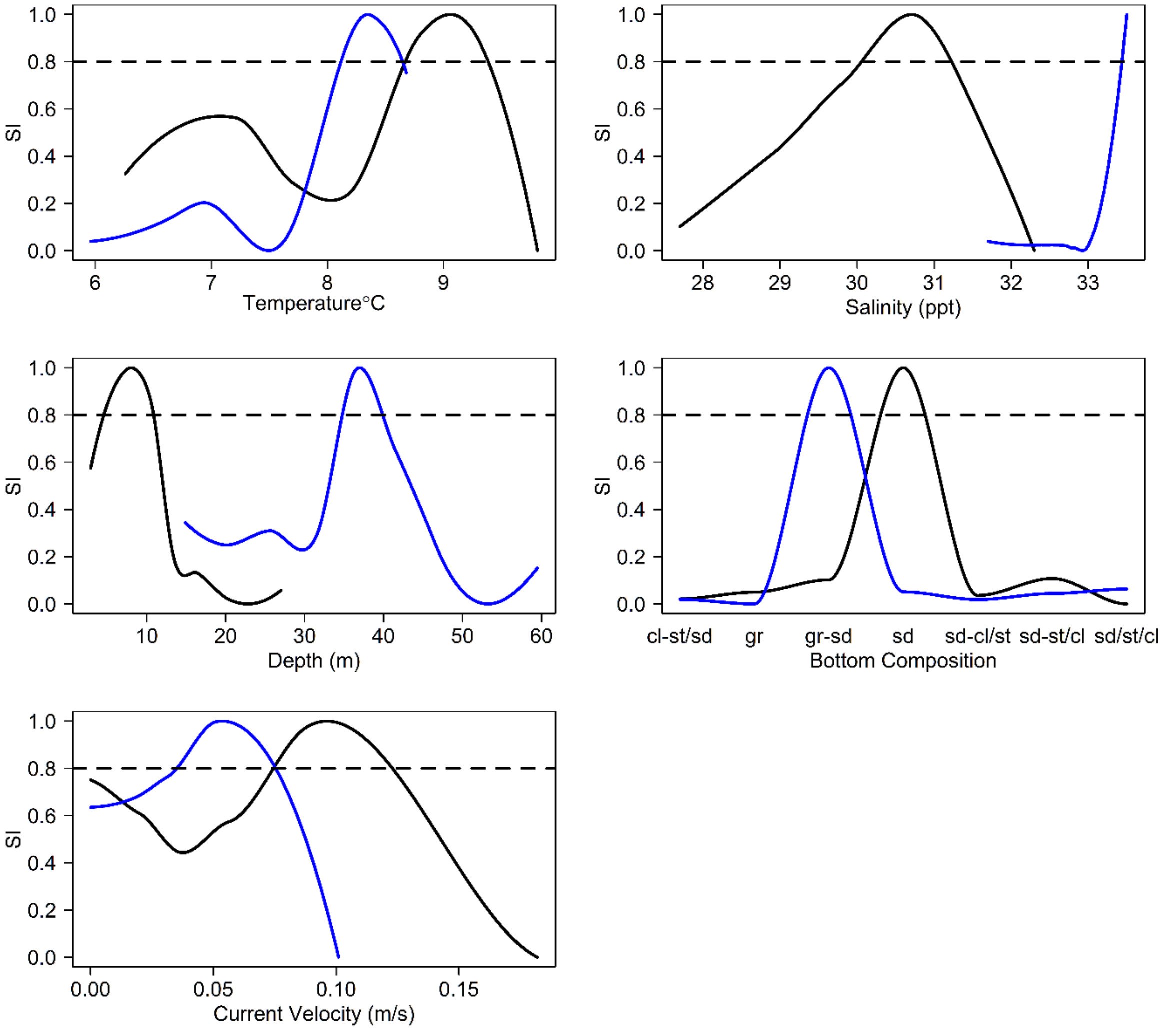


Figure 2.3: Inshore (black line) and offshore (blue line) Suitability Index (SI) curves showing the relationship between *Placopecten magellanicus* (>65 mm shell height) abundance and bottom temperature, depth, bottom salinity, current velocity, and bottom composition. Horizontal dotted lines represent the cutoff above which the suitability of a given habitat variable was considered high.

2.4.2 Variable weighting and model validation

BRT-based variable weighting showed bottom salinity, bottom temperature, and flow velocity were the most important variables inshore, while bottom salinity, bottom temperature, and depth were the most important variables offshore (Table 2.2).

Table 2.2: Relative contribution (%) of all environmental variables used in inshore and offshore Habitat Suitability Index (HSI) models to the deviance explained by the Boosted Regression Tree (BRT)

		Relative Contribution (%)			
Location	Mean Bottom Temperature	Mean Bottom Salinity	Mean Flow Velocity	Depth	Bottom Composition
Inshore	28.18	32.30	16.25	8.39	14.86
Offshore	29.47	34.32	13.12	15.83	7.24

AMM model performance, tested through cross-validation, had better predictive ability within inshore stations relative to those offshore. Median values from linear regression show that inshore stations had a median intercept of 0.14, a slope of 0.88, and r^2 of 0.81. Offshore stations had a median intercept of 0.14, a slope of 0.57, and r^2 of 0.35.

2.4.3 Spatio-temporal HSI-based bioclimate envelope evaluation

Projected HSI-based bioclimate envelopes for *P. magellanicus* showed higher habitat suitability inshore compared with most offshore areas (Figure 2.4). However, offshore shoal areas displayed high habitat suitability on par with inshore areas. Cobskook bay appears to have the highest habitat suitability over the study area (Figure 2.4).

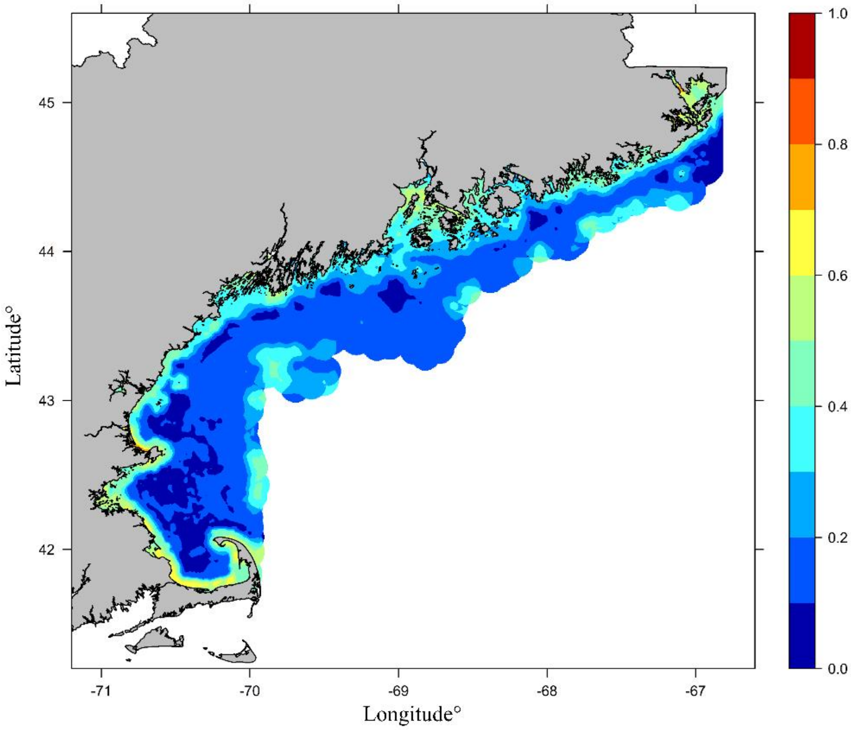


Figure 2.4: Map showing the spatial distribution of median habitat suitability index (HSI) over 1978-2013 for *Placopecten magellanicus* (>65 mm shell height) from Massachusetts to Maine. HSI values larger than 0.8 are designated as good habitat, while HSI values below 0.3 are considered poor habitat. The color ramp corresponds to predicted HSI value, where blue indicates poor habitat and red indicates good habitat.

Changes in climate-driven habitat suitability over the 36 years of study period (1978 – 2013) were apparent throughout the study area. Overall, inshore areas show a trend of increasing habitat suitability (Figure 2.5). Offshore areas show a decreasing trend in general, with the exception of shoal areas that show an increasing trend (Figure 2.5).

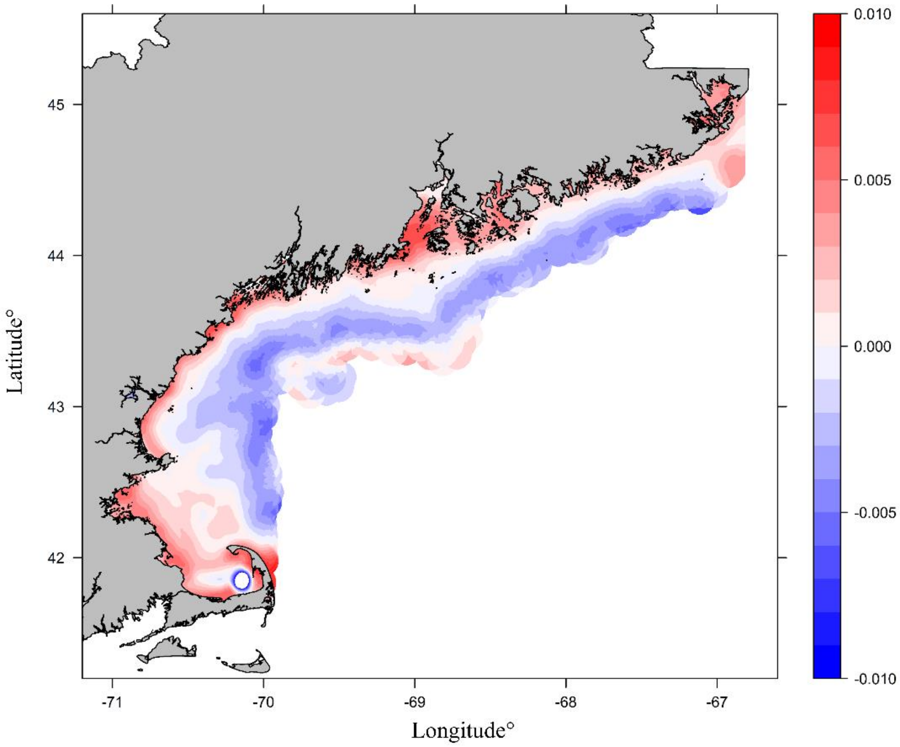


Figure 2.5: Map showing the temporal change in habitat suitability index (HSI) values for *Placopecten magellanicus* (>65mm shell height) from Massachusetts to Maine over 1978 – 2013. The color ramp corresponds to the linear regression slope (*β*) coefficient. Red areas have a positive slope and blue areas have a negative slope.

Proportion of total habitat with at least moderate habitat suitability (HSI > 0.5) for *P. magellanicus* in the Gulf of Maine ranged from 14.49% (1980) to 46.66% (2001) during 1978 to 2013. Total habitat with HSI > 0.5 from the median over the 36-year study period was 26.16%. Proportion of total habitat with high habitat suitability (HSI value > 0.8) for *P. magellanicus* in the Gulf of Maine ranged from 0.03% (1988) to 6.04% (2012) during 1978 to 2013. Total habitat with HSI > 0.8 from the median over the 36-year study period was 0.02%.

2.5 Discussion

A dominant spatial trend made apparent by this modeling approach is a decline in habitat suitability moving from inshore to offshore locations. It is likely that decreased habitat suitability in offshore areas is largely driven by increases in depth and decreases in temperature. This finding corresponds with habitat suitability being higher among offshore shoal areas, relative to adjacent deep areas. Previous studies sampling *P. magellanicus* along a depth gradient (10-30m) have observed decreases across a range of ecological energetics (shell growth, somatic growth, somatic production, gonad production, gonad output reproductive effort, and residual reproductive value) in deeper waters (MacDonald and Thompson 1985a, 1985b; Macdonald et al. 1987). These differences are attributed to deteriorating food availability and temperature conditions with water depth, which is thought to represent a natural gradient of habitat quality (Sarro and Stokesbury 2009; Hennen and Hart 2012).

Modeled nonlinear responses of the suitability indices (SI) reflect larval supply coupled with the species‘ ability to survive environmental variability present in a given area. The SI curves for bottom temperature, depth, bottom salinity, current velocity, and bottom composition fell within known habitat ranges for *P. magellanicus* (Naidu and Anderson 1984; Thouzeau et al. 1991; Wildish and Kristmason 1993; Stewart and Arnold 1994; Pilditch and Grant 1999; Hart and Chute 2004). However, within these broad ranges, this study identified considerable differences in habitat preference between *P. magellanicus* inhabiting inshore and offshore areas. Inshore *P. magellanicus* were most abundant in shallower areas with stronger currents, higher temperatures and lower salinities relative to offshore *P. magellanicus*. Abundance of inshore *P. magellanicus* was highest on sand substrate as opposed to offshore where gravel-sand was preferred. Inshore-offshore differences in dispersal and retention patterns of *P. magellanicus* larvae were found by Tremblay & Sinclair (1991), which may factor into differences in habitat selectivity observed in the present study. Additionally, Beyer et al. (2010) speculate that habitat selection is context-dependent with functional responses in preference resulting from changing availability. While *P. magellanicus* inhabiting both inshore and offshore areas were within known, broad ranges of physiologically suitable environmental conditions, they still subject to a different composite of habitat variables. Thus, it is possible that perceived habitat preference may partially reflect given habitat availability rather than the species' physiologically preferred habitat range. Also, interactive effects of multiple habitat variables could have resulted in the observed difference in inshore and offshore SI curves. For example, higher optimal flow velocity observed in inshore *P. magellanicus* may reflect a required higher feeding rate due to increased metabolic rate resulting from higher temperatures in these areas. Such complex interactions among key habitat variables highlight a benefit of evaluating habitat quality in a holistic manner as opposed to analyzing each variable in isolation.

Total suitable habitat coverage (HSI value of > 0.8) in the model showed large inter-annual variations, ranging from 0.03% to 6.04%, which reflect changes in dynamic environmental variables (bottom temperature, current velocity, bottom salinity). The mean proportion of total suitable habitat (HSI > 0.8) for *P. magellanicus* in the GOM over the 36 year study period was 0.02%. Brown et al., (2000), developed HSI models for eight fish and invertebrate species in Casco Bay and Sheepscot Bay, Maine. Total suitable habitat (HSI value of > 0.84; note difference in "suitable" habitat cutoff between this study and the current) for these species ranged from 6% (*Ammodytes americanus*) to 95% (adult *Homarus americanus*). Coverage of suitable habitat for *P. magellanicus* in the current study is low relative to the species modeled in Brown et al. (2000); however, consideration should be given to the spatial scale at which these models are applied. Brown et al. (2000) confined their model to a much smaller, coastal area while the current model was applied over a large portion of the GOM. Applying the current model over a large spatial scale increased the likelihood that a gradient in habitat quality will be covered, effectively lowering percent of suitable habitat coverage. Environment-biota relationships can include a hierarchy of factors operating at different scales (Willis and Whittaker 2002; Pearson and Dawson 2003; Hattab et al. 2014). This study highlights spatial effects in the relationships between *P. magellanicus* and habitat variables. Thus, bioclimate envelope models may perform differently if applied over different spatial scales. In the current study, this was addressed through the individual development of both inshore and offshore SIs to reflect the perceived change in habitat preference between these areas. Future refinement of this model could include evaluating *P. magellanicus* – habitat relationships over even smaller spatial scales, to further explore the scale at which each environmental variable operates.

An increasing temporal trend in climate-driven (i.e. bottom temperature and salinity) habitat suitability was observed for inshore areas with a decreasing trend in offshore areas (Figure 2.5). Any change in habitat suitability over time resulted from changes in dynamic habitat variables (bottom temperature, bottom salinity, and current velocity), suggesting that the composite of these 3 factors has changed favorably in regard to *P. magellanicus* from 1978-2013 in inshore areas and remained relatively stable in offshore areas. These trends assume that the habitat preference of *P. magellanicus* does not change over the study period (Pearson and Dawson 2003; Crisp et al. 2009; Catullo et al. 2015).

Model validation revealed that offshore areas had lower predictive ability relative to inshore areas. Higher inshore performance likely corresponds to having ~4.5 times more tows conducted in these areas. Future iterations of this model can include more data in both inshore and offshore areas that will have the potential to increase predictive ability in offshore areas. However, it is also possible that it is more difficult to predict the distribution of offshore *P. magellanicus* habitat due to possible complex or de-coupled interactions between *P. magellanicus* and habitat variables in these areas as described above (Beyer et al. 2010).

Modeling methodology in ecological research has historically been largely quantitative (Bradbury et al. 1986). However, qualitative models effectively capture ecological pattern and have the advantage of avoiding data driven biases that quantitative models are subject to (Bradbury et al. 1986; Store and Kangas 2001; Tanaka and Chen 2016). This HSI-based bioclimate envelope modeling approach can be applied to a number of different research areas such as modeling potential species distribution, and evaluating the effects of climate-driven changes in habitat suitability on this distribution through hindcasting/forecasting analyses (Pearson and Dawson 2003; Araujo and Peterson 2012; Tanaka and Chen 2016). However, there are intrinsic limitations to this approach that should be considered when evaluating model results (Pearson and Dawson 2003; Luoto et al. 2005). FVCOM predictions, used in this study as inputs for all dynamic habitat variables, provide the highest resolution and broadest spatial coverage for temperature, salinity, and current velocity data available in the study area; however, *P. magellanicus* beds can frequently occur over relatively small spatial scales. Due to the inherent resolution of this environmental data set, it is unlikely that the modeling approach used here would be able to resolve fine-scale patches of habitat with the potential to support *P. magellanicus* beds. Instead, the bioclimate envelopes developed in this study are more useful for exploring spatio-temporal trends in mesoscale climate-driven habitat suitability.

Development of bioclimate envelope models relies upon environmental data, and as with any environmental data there are several possible sources of error that could cause misrepresentation of model predictions. Since all data obtained through FVCOM are outputs from model simulations as opposed to directly measure values, prediction accuracy needs to be taken into account. To evaluate performance of FVCOM within the study area, Tanaka and Chen (2016) and Li *et al.* (2017) used a collection of observed bottom temperatures collected by environmental monitors on lobster traps (eMOLT) to compare with FVCOM predictions. They found that while some variability occurred in FVCOM outputs relative to eMOLT observations, FVCOM adequately captured general spatial and temporal trends in bottom temperature and salinity. These findings add validity to the quality and accuracy of FVCOM predictions over broader scales. Another important consideration is the bottom composition data obtained from the U.S. Continental Margin Mapping (CONMAP) GIS database (Poppe et al. 2005). This GIS layer provides a relatively coarse resolution of bottom type which may be insufficient to resolve potential fine-scale *P. magellanicus* habitat. However, the CONMAP GIS database provides the most comprehensive coverage of bottom composition and is useful here as a key component to identifying large scale spatial trends in habitat suitability.

In this study, the bioclimate envelopes were defined upon five environmental variables when in reality, a large number of physical, biological, and chemical conditions likely factor into the life history and distribution of *P. magellanicus.* As more comprehensive environmental data becomes available in the future, studies to develop a further detailed bioclimate envelope model could include additional variables such as pH, dissolved oxygen, predator-prey, and other food-web interactions to capture a more comprehensive representation of *P. magellanicus* ecology (Araújo and Luoto 2007). Additionally, environmental predictors in this study were selected based on availability, and assumed correlation with habitat quality. However, as is generally the case with species-environment modeling, variables used to build this model may be operating as surrogates for factors directly controlling species distribution through physiological mechanisms (Austin 2007; Araujo and Peterson 2012). It is from the associations between variables that we can infer the relationship between spatiotemporal variability of environmental factors and habitat quality. For instance, salinity in these models may act as a proxy for broad spatial patterns in *P. magellanicus* distribution coupled with, and driven by, the origin of water mass existing in a given area. An additional example is that temperature and depth likely correspond to gradients in food availability (MacDonald and Thompson 1985a, 1985b; Macdonald et al. 1987.

An important assumption to consider in this modeling framework is that predicted habitat quality is directly related to observed *P. magellanicus* density, when in reality, a number of other factors collectively act on the ecology of this species. For instance, *P. magellanicus* density is highly influenced by larval supply (Shumway and Parsons 2006), and also not accounted for in this modeling approach, is spatiotemporally variable fishing pressure, which acts directly on adult *P. magellanicus* density. Thus, certain areas with quality habitat may still have low *P. magellanicus* density, which could impact modeled nonlinear responses of suitability indices. However, while this is an important point to consider, as a result of the large spatiotemporal extent of *P. magellanicus* observations used to calibrate this model, it is likely that collectively, suitability indices accurately reflect preferred ranges of environmental variability by this species in the given area.

In light of recent abrupt warming events as well as long-term warming trends within the GOM ecosystem it is becoming increasingly important to view resource management from within the context of climate change (Mills et al. 2013; Pershing et al. 2015). Effective management of marine resources requires knowledge of population distribution and dynamics (Langton et al. 1995), however; fisheries managers must frequently base decisions on limited information (Brown et al. 2000). Even when intensive sampling efforts are conducted, they sometimes fail to provide adequate spatial or temporal coverage to capture an entire range of available habitat which can result in misinformed management decisions (Brown et al. 2000). Additionally, many stock assessments fail to incorporate environmental variability (NMFS, 2010). The bioclimate envelope model developed in this study provides a unique tool to visualize the extent of available habitat for *P. magellanicus* as well as to evaluate the potential impacts of a changing ecosystem on the distribution of available habitat in the Gulf of Maine. This study sheds a new light on spatio-temporal trends in habitat suitability that could potentially inform and improve stock assessments and the management of *P. magellanicus.* Other potential management applications for this modeling framework could include the development of habitat maps in poorly sampled areas (Brown et al. 2000), refinement of fisheries independent surveying efforts, and prioritizing areas for conservation actions.

CHAPTER 3 - DEVELOPMENT OF A CLIMATE-NICHE MODEL TO EVALUATE SPATIOTEMPORAL TRENDS IN *PLACOPECTEN MAGELLANICUS* DISTRIBUTION IN THE GULF OF MAINE, USA

3.1 Abstract

We developed a climate-niche species distribution model to evaluate spatiotemporal trends in Atlantic sea scallop (*Placopecten magellanicus*) along the coastal waters of the Gulf of Maine. We used a Tweedie-generalized additive model (GAM) to quantify the relationships between scallop abundance and key environmental variables. A boosted regression tree was used to identify significant interactions among environmental variables to integrate within the Tweedie GAM and a regional circulation model was incorporated with the Tweedie GAM to hindcast projections of scallop distribution and assess the impacts of environmental change on this species. Additionally, we evaluate two common model fitting and variable selection methods for GAMs to ensure high model performance. A classic backward variable selection procedure was compared to penalized thin plate regression splines. Projections from the climate-niche species distribution model show higher scallop density along inshore areas relative to those farther offshore. An increasing temporal trend in scallop density was observed along inshore areas and a decreasing temporal trend was observed in areas farther offshore. Additionally, we found that the GAM incorporating thin plate regression splines outperformed the widely used backwards stepwise procedure. This modeling framework will help to inform adaptive management strategies for the scallop fishery within the context of a changing Gulf of Maine ecosystem.

3.2 Introduction

Atlantic sea scallops (*Placopecten Magellanicus*) support a highly valuable fishery in the United States. However, the scallop fishery within Maine state waters is relatively depleted. The fishery there bottomed out in 2005 when only 33,000 lb was taken, about 1% of what was landed in the early 1990's (Kelly 2012). However, recent years have shown an increasing trend in scallop abundance, with over 800,000 lb landed in 2018, allowing for an opportunity to establish a persistent, valuable fishery through research-informed management and conservation efforts (Kelly 2012).

The distribution and abundance of species are central concepts to ecological research and vital components of conservation planning and fisheries management (Franklin 2010). Species distributions are influenced by many interacting biotic and abiotic processes that can manifest as complex occurrence-environment relationships (Boulangeat et al. 2012; Merow et al. 2014). Thus, a key step in understanding the biogeography of species is identifying environmental factors that regulate the distribution of a species (Merow et al. 2014). Like most benthic species, scallop abundance and distribution are influenced by an array of interacting variables such as depth, current, temperature, and salinity (Stokesbury and Himmelman 1995; Hart and Chute 2004).

Throughout their geographic range, scallops occur mainly at depths of 15 to 110 m, but can be found as shallow as 2 m in the northern part of its range (Naidu and Anderson 1984; Carsen et al. 1995). Temperature is an important environmental factor influencing growth rates of this species with adult scallops showing optimal growth at temperature between 10-15 °C and temperatures above 21 °C being lethal (Stewart and Arnold 1994). This species prefers full strength seawater (~35 ppt), with salinities of 16.5 ppt or lower being lethal (Stewart and Arnold 1994). Scallops are usually found in environments with strong currents (Hart and Chute 2004), and flow velocity has been shown to be a key factor controlling waste removal, oxygen uptake, feeding, and growth rates (Stewart and Arnold 1994; Shumway and Parsons 2006). Optimal growth for this species occurs near 0.1 m s^{-1} (Wildish and Saulnier 1992) and feeding inhibition has been observed to start at ~ 0.25 m s^{-1} (Pilditch and Grant 1999). While the influence of key environmental variables on bivalve ecology is apparent, quantitative evaluations of scallop-environment relationships and spatiotemporal trends of distribution are uncommon (Shumway and Parsons 2006; Mendo et al. 2014).

Water temperatures within the Gulf of Maine (GoM) have increased over the past 35 years at an average rate of 0.026 °C yr^{-1} (Mills et al. 2013; Pershing et al. 2015). Rapidly increasing temperatures are changing the distribution of numerous marine species (Overholtz et al. 2011; Howell and Auster 2012; Hollowed et al. 2013), with many fish stocks undergoing a poleward shift in their center of biomass and/or an increase in depth (Nye et al. 2009). Scallop distribution has also been shown to be impacted by climatic variability (Frank et al. 1990; Kurihara 2008). Dickie (1955) and Caddy (1979) demonstrated that higher temperatures in the Bay of Fundy were correlated with changes in scallop abundance. Additionally, climatic changes may alter larval development as well as the survival of juvenile and adult scallops (Dickie 1955; Caddy 1979). Considering a changing GoM ecosystem (Mills et al. 2013; Pershing et al. 2015), it is important to document the importance and potential synergistic effects of climate forcing on the dynamics of species abundance and distribution.

In this study, we use data describing the distribution of scallops in the GoM to develop a climate-niche species distribution model (SDM). This SDM predicts the spatial distribution of scallops within the inshore Gulf of Maine across unsampled areas and hindcasts spatiotemporal changes in the distribution of scallops from 2005-2013 to evaluate the effects of shifting environmental conditions on this species. We used Tweedie-generalized additive models (GAMs) to quantify the relationships between scallop abundance and key environmental variables. Additionally, we evaluate two prevalent model fitting and variable selection methods for GAMs to ensure high model performance. A classic backward variable selection procedure was compared to penalized thin plate regression splines following Wood (2003, 2006). This modeling framework will help to inform adaptive management strategies for the scallop fishery within the context of a changing GoM ecosystem.

3.3 Methods

3.3.1 Study Area and survey data.

Dredge-based fishery-independent scallop surveys conducted over 15 years, from 2005 to 2017, by the Maine Department of Marine Resources were used for this modeling effort (DMR: Kelly, 2012; Figure 3.1). Survey coverage extends out to 3 nm from shore from southern Maine to the Maine-Canadian border, USA (Figure 3.1).

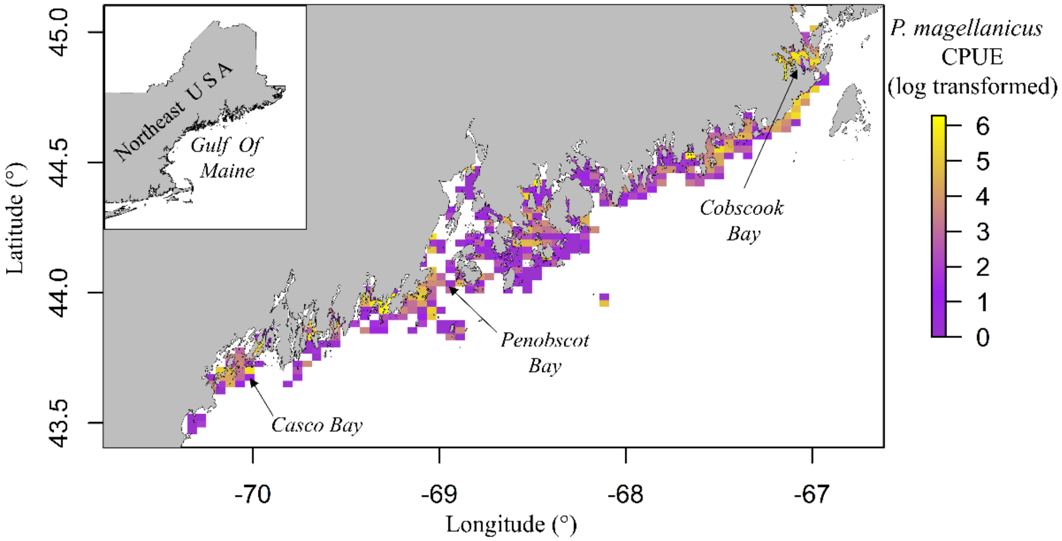


Figure 3.1: Spatial distribution of natural log standardized scallop (*Placopecten magellanicus*) density from the Maine Department of Marine Resources Scallop Dredge Survey from 2005-2017. The study area encompasses nearshore waters of the Gulf of Maine from Casco Bay to the Maine-Canada border.

This dataset comprised two annual random systematic surveys, one in the spring survey covering alternating portions of inshore Maine waters and one in the fall covering select coastal areas. The gear used for both surveys is an unlined, 7 ft New Bedford style drag with 2 in rings, 1.75 in head bale, 3.5 in twine top, 10 in pressure plate and rock chains. Since scallops < 65 mm in shell height were not efficiently sampled with the 2 in rings (Kelly 2012), these were excluded from all analyses. Tows were conducted at 3.5-4 knots and lasted ~ 2.5 minutes. A total of 4,321 tows were made yielding 507,911 total observed scallops in this dataset (Figure 3.2). All tows from the survey were included within the modeling framework. Scallop abundance from each tow was standardized to catch-per-unit-effort (CPUE) over a 2.5-minute tow.

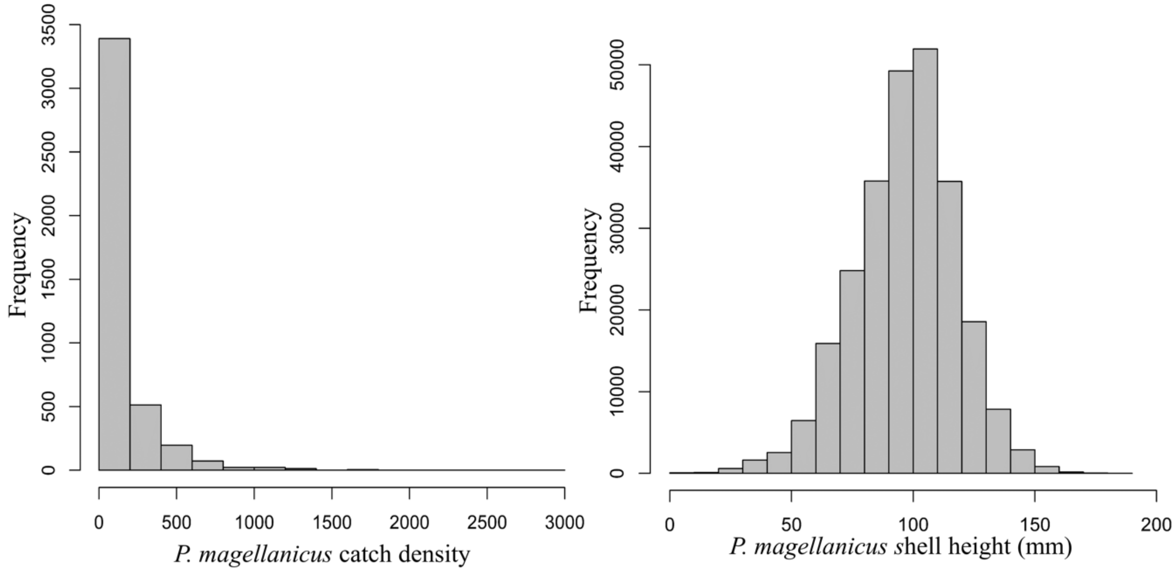


Figure 3.2: Frequency plots of scallop (*Placopecten magellanicus*) density (left) and size frequency (right) in the density data from the Maine Department of Marine Resources Scallop Dredge Survey

The Gulf of Maine is characterized by a mixture of oceanic influences directly affected by the Labrador Current, the Gulf Stream, and the freshwater discharge from the St. Lawrence River (Tremblay 1997; Drinkwater and Gilbert 2004). As such, water temperature follows a gradient moving up the coast and offshore. Within the study area, scallops were found where maximum yearly temperature ranged from 8 °C in deeper areas to 16 °C in shallow areas. Salinity ranged from 26 ppt in areas subject to freshwater inputs to full seawater (35 ppt) in offshore locations. The study area covered depths to ~ 110 m, since scallops are uncommonly found outside of this depth range (Hart and Chute 2004), with the majority of tows occurring in < 60 m.

3.3.2 Environmental data.

Because the DMR surveys did not measure temperature, salinity, or current velocity, the Finite-Volume Community Ocean Model (FVCOM), a regional ocean circulation model developed by the University of Massachusetts-Dartmouth and the Woods Hole Oceanographic Institution (Chen et al. 2006) was used to simulate monthly estimates of bottom temperature, salinity, and current velocity from 2005 to 2013. It has a horizontal resolution ranging from 0.02 km to 10 km and captures complex and irregular coastal geometry, making it suitable for physical and biological studies in coastal regions and estuaries (Chen et al. 2011). FVCOM predictions were matched to survey tows from the nearest neighboring FVCOM node during time of sampling (Figure 3.3). Horizontal current velocity was calculated to approximate the magnitude of water flow at a given FVCOM node. Current velocity C was estimated at station i, and year y from FVCOM predictions using the following equation

$$C_i = \sqrt{u_i^2 + y_i^2}$$

where C is the magnitude of the predicted current velocity and u and y are the x and y vector components of the velocity (Chen et al. 2011; Torre et al. 2018). Bathymetry data were obtained from the U.S. Coastal Relief Model (CRM) (National Geophysical Data Center 1999).

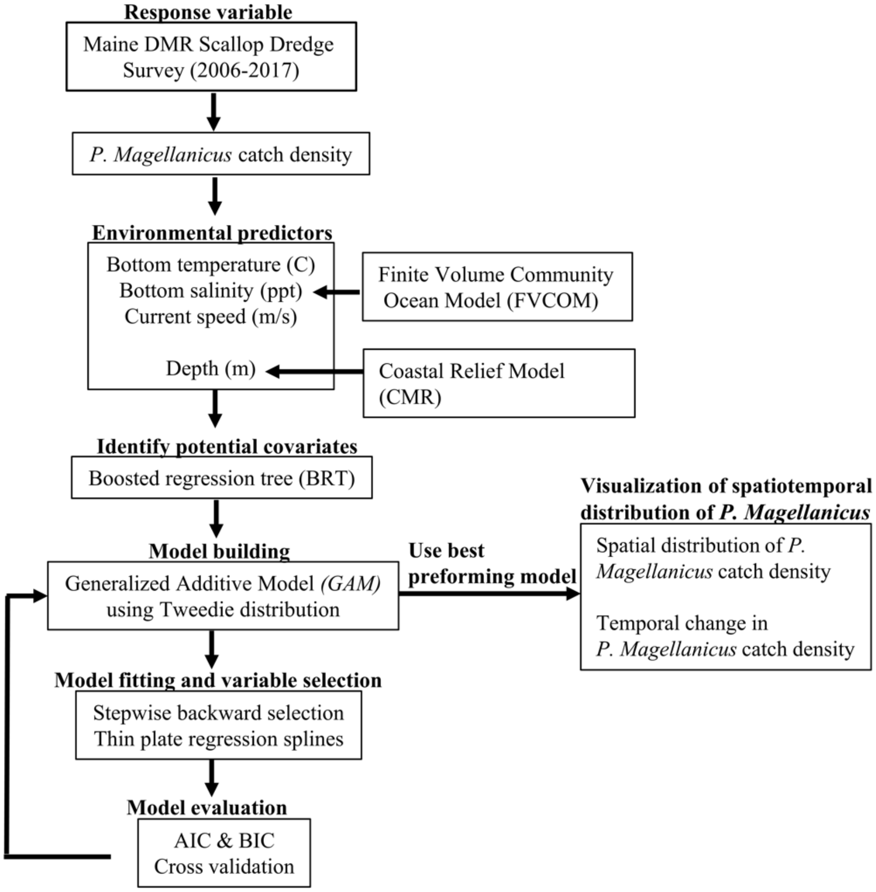


Figure 3.3: Schematic diagram of the modeling framework implemented in this study. All data exploration and modeling procedures were conducted within the R programming environment.

3.3.3 Generalized additive models.

A generalized additive model (GAM) was used to make spatiotemporal predictions of scallop distribution in the inshore GoM (Figure 3.3). Conceptually, GAMs are generalized linear models with a linear predictor upon which smooth functions are applied to covariates (Guisan et al. 2002; Marra and Wood 2011). The strength of GAMs lie in their ability to handle, in a multivariate regression setting, non-linear and non-monotonic relationships between the response and covariates that arise often in nature (Guisan et al. 2002). Resulting from their ability to deal with a variety of distributions that occur in ecological data, the use of GAMs has been extensively applied to species distribution modeling efforts (Guisan et al. 2002; Sagarese et al. 2014; Young and Carr 2015).

Terms included within the full model were selected according to boosted regression tree (BRT) analysis (Elith et al. 2008), and bivariate interaction terms were also identified and included based on this analysis. BRT models were used to determine the relative importance of each environmental variable with relation to scallop density. Bivariate terms were included when interaction between two variables in the BRT was considered high (> 100).

Formulation of the GAM for the estimate of scallop CPUE can be expressed as follows;

$$g(n) = \propto + \sum_{j=1}^{p} S_j(x_j) + \varepsilon$$

where *g(.)* is a log link function between the response variable, *n*, and each additive predictor, x_j; α is the intercept term; s_j are smooth functions of the predictors, represented by either cubic splines, that are linear or nonlinear, or thin plate regression splines with a penalty; ε is the residual error (Wood 2003; Marra and Wood 2011). Predictors comprised either a single variable or interacting pair of variables that are thought to relate to scallop distribution (Table 3.1). Smooth terms were used in conjunction with a pair of variables to model interactions. Where cubic splines were used, the maximum value for degrees of freedom were set at 5 for univariate functions and 30 for bivariate functions. These values were chosen as a balance between over generalization and over-fitting as suggested in the literature (Zuur et al. 2009; Sagarese et al. 2014). Statistical analyses were carried out within the R programming environment (R Core Team Development 2016). Boosted regression tree analysis was done using the *dismo* package (Hijmans et al. 2017) and GAMs were fitted using the *mgcv* package (Wood 2011).

Table 3-1: Variables used in generalized additive modeling of scallop (*Placopecten magellanicus*) density in the nearshore Gulf of Maine.

Variables	Description
Longitude (°)	Measurement of longitude of tow starting location
Latitude (°)	Measurement of latitude of tow starting location
Bottom Temperature (°C)	bottom temperature at tow location (imported from FVCOM)
Bottom Salinity (ppt)	bottom salinity at tow location (imported from FVCOM
Current Velocity (m s^{-1}**)**	current velocity at tow location (imported from FVCOM)
Depth (m)	depth at tow location

FVCOM: Finite Volume Community Ocean Model

A Tweedie distribution was used as the likelihood to measure GAM fit to account for a high proportion of zero-catch tows and skewness in the CPUE data (Figure 3.2). The Tweedie distribution handles zero and positive values simultaneously, and works as a Poisson-Gamma compound distribution when the power parameter p is greater than 1 but less than 2 (Li et al. 2011; Wood 2011). This distribution has been shown to outperform other methods for dealing with zero inflated data (Shono 2008; Li et al. 2011). Tweedie GAMs were fitted through a process of optimizing its profile likelihood. Power parameter p was estimated with the range of $1 < p < 2$ during the fitting process (Shono 2008; Wood 2011; Tanaka et al. in review).

Final GAMs were used to predict the density of scallops at every FVCOM node in the study area during 2005-2013. Predictive fields were interpolated using ordinary kriging with a semivariogram function to produce continuous model outputs (Bailey and Gatrell 1995; R Core

Team Development 2016). Predicted density of scallops was aggregated temporally by obtaining the median density value over the 8-year study period at each FVCOM node. Median values were used as opposed to means because they provide a clearer interpretation of the tendency over the study period, i.e. not susceptible to skewing in rare cases of outliers. Linear regression was performed at every FVCOM node and the derived slope (β) coefficient was used to evaluate temporal change in predicted density of scallops over the 8-year study period.

3.3.4 Model selection.

In this modeling framework, two methods of fitting and variable selection were performed to assess their relative performance (Figure 3.3). Here, performance signifies both a balance between goodness of fit and parsimony, and maximizing prediction accuracy while maintaining model interpretability (Marra and Wood 2011). The first method of model fitting and variable selection was a conventional backwards stepwise procedure (BSP), where terms were removed iteratively from an initial full model using AIC (Akaike 1974). This method is widely used in ecological modeling due to its simplicity and demonstrated effectiveness (Burnham and Anderson 2002; Marra and Wood 2011). During each step, the variable with the highest *p*-value was removed and AIC was recalculated for the reduced model. This iterative process was repeated as long as variable removal lowered AIC.

The second method of model fitting and variable selection was a shrinkage approach where each variable in the full model was fitted with a thin plate regression spline including a thin plate spline penalty (TPRS). These are low rank isotropic smoothers of covariates that include a modification to the smoothing penalty, so that whole terms can be reduced to zero, effectively removing superfluous variables (Wood 2003; Marra and Wood 2011). Unlike the stepwise algorithm from BSP, this procedure is carried out in a single step.

3.3.5 Model Validation.

The performance of final models from BSP & TPRS was evaluated using multiple evaluation criteria (Figure 3.3). Two traditional model evaluation criteria, AIC and BIC (Burnham and Anderson 2002; Wood 2006), were used to compare relative performance of BSP and TPRS. AIC and BIC are widely used as model evaluation criteria. AIC is an estimator of the relative quality of statistical models based on goodness of fit. BIC is closely related to AIC, but has been shown to penalize complexity to a higher degree than AIC (Hastie 2009).

A cross-validation study was implemented to evaluate predictive performance of final GAMs, where a randomly selected subset, training data, (80% of all data) was used for GAM development, while the remaining 20%, testing data was used for the evaluation of performance (Smith 1994; Wood 2006; Zuur et al. 2007; Tanaka and Chen 2015, 2016). The GAM-predicted CPUE values based on training data were compared against observed CPUE values, based on testing data, and linear regression analysis was performed to evaluate the predictive performance of the GAM. The cross-validation procedure was repeated 100 times using random data selection in each round to obtain 100 sets of linear regression parameters (intercept, slope, and R^2). Good model performance was indicated by an intercept parameter close to zero, a slope close to one and an R^2 close to one.

3.4 Results

3.4.1 GAM performance.

Full GAMs included nine terms in total (Tables 2 and 3). Single terms included latitude (°), longitude (°), depth (m), bottom temperature (°C), bottom salinity (ppt), and current velocity (m/s). BRT analysis identified strong two-way interactions (value > 100) between three pairs of variables, depth-latitude (163.13), salinity-temperature (112.79), and longitude-latitude (125.10)

(Table 3.2). All three of these two-dimensional terms were significant and included in both full and final GAMs (Table 3.3, Figure 3.4 and 3.5).

Table 3.2: Results from boosted regression tree analysis. Higher values are associated with stronger interaction between variables. Variable pairs with a value > 100 were considered to have a "strong" interaction and included as terms in the generalized additive models.

	Longitude	**Latitude**	**Bottom Temperature**	**Bottom Salinity**	**Current Velocity**	**Depth**
Longitude	0	125.1	13.33	46.73	5.34	30.65
Latitude	0	0	60.15	12.93	8.55	163.13
Bottom Temperature	0	0	0	112.79	24.78	44.98
Bottom Salinity	0	0	0	0	49.63	27.83
Current Velocity	0	0	0	0	0	4.77
Depth	0	0	0	0	0	0

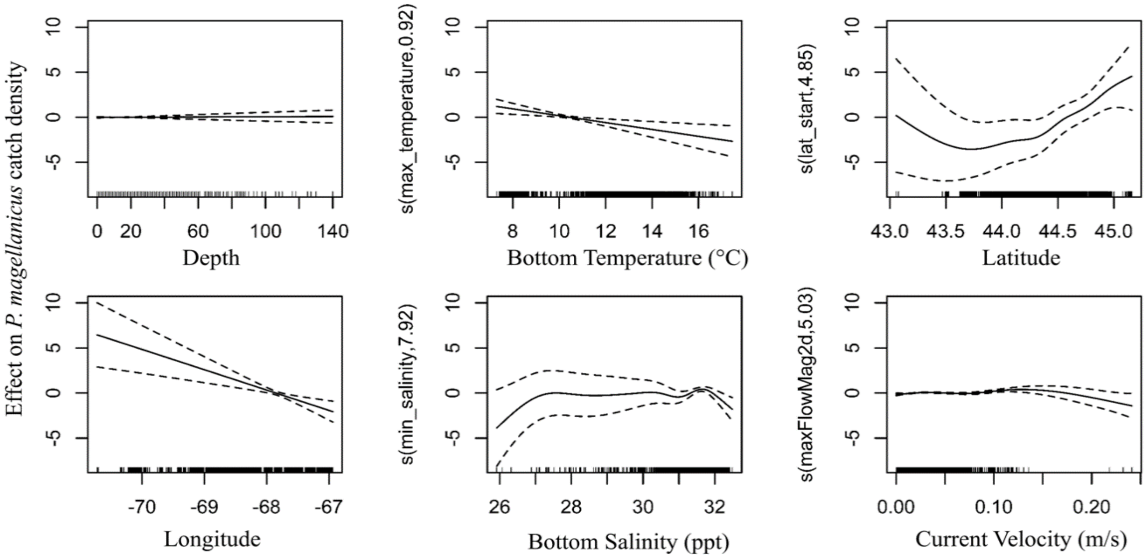


Figure 3.4: Plots describing the partial effect of significant univariate explanatory variables in the thin plate regression spline generalized additive model. The response curves describe the relationship between a given environmental term and smoothed density of scallops (*Placopecten magellanicus*). Tick marks on the x-axis correspond to number of observations.

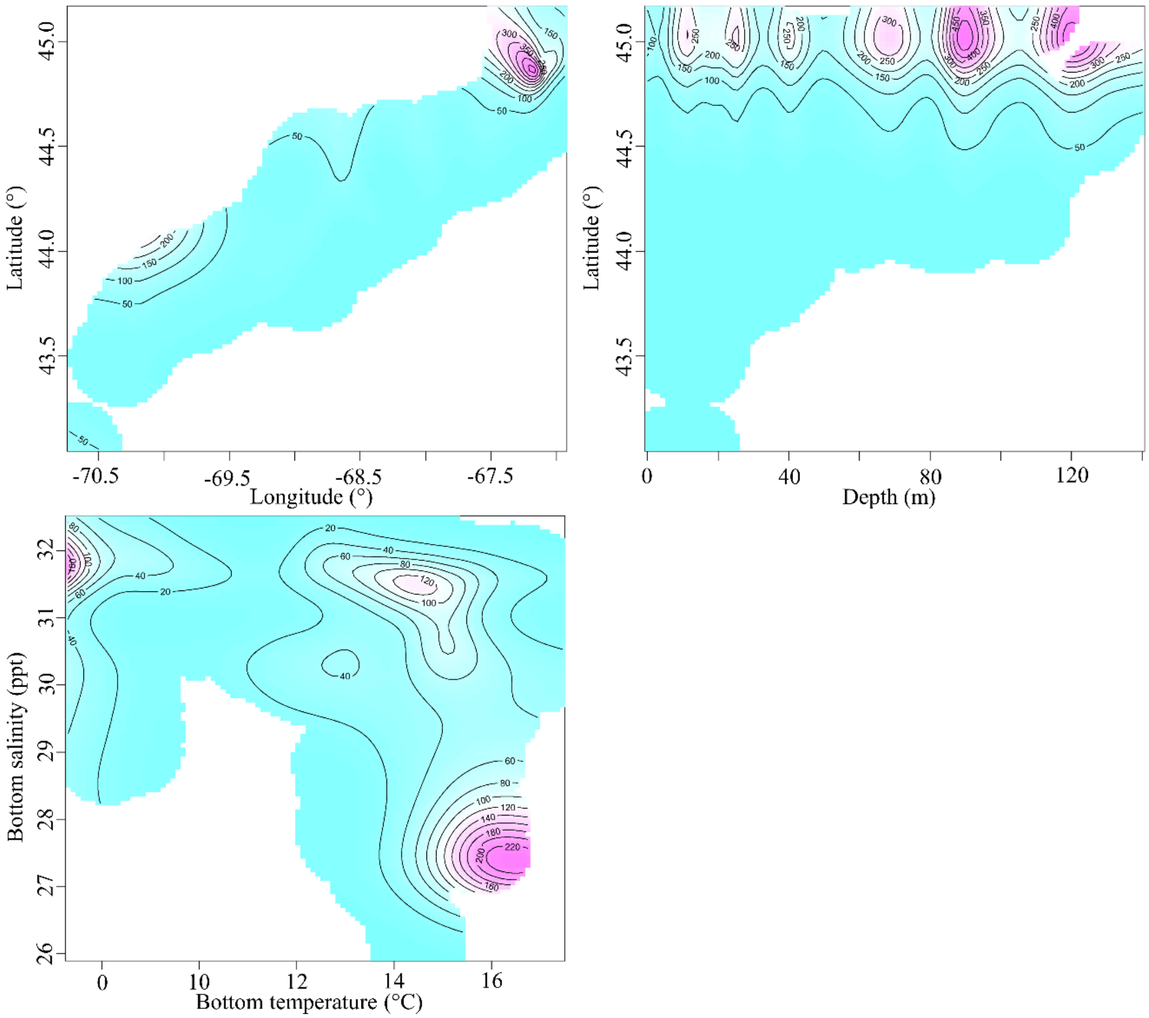


Figure 3.5: Partial generalized additive model plots depicting the significant interaction effects of bivariate environmental variables included in the modeling framework.

Model fitting and variable selection using BSP went through two iterations of term removal. Depth was removed during the first round and bottom temperature was removed during the second round. Deviance explained for the parsimonious BSP model was 48.60 %. Model fitting and variable selection using TPRS, similarly, penalized depth to a high degree (edf = 0.05, Table 3.3, Figure 3.4) to where it had a negligible effect on predictions. The effect of both bottom temperature (edf = 0.91) and longitude (edf = 0.94) on scallop density were reduced to almost linear relationships (Table 3.3, Figure 3.4). Deviance explained for the parsimonious TPRS model was 49.70% (Table 3.3).

Table 3.3: Generalized additive models for scallop (*Placopecten magellanicus*) density in the nearshore Gulf of Maine. Deviance explained by the model (Dev. Exp.), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Model terms are latitude (La), longitude (Lo), depth (De), bottom temperature (Bt), bottom salinity (BS), and current velocity (Cv). Edf = estimated degrees of freedom. Greyed out terms are variables that were removed from the model. * Denotes the highest performing model from each category (thin plate regression spline and backwards stepwise).

Model	edf	Dev. explained	AIC	BIC
Thin plate regression spline w/ penalty				
* s(La) + s(Lo) + s(De) + s(Bt) + s(Bs) + s(Cv) +s(La*Lo) + s(Bt*Bs) + s(De*La)	4.84, 0.94, 0.05, 0.91, 7.92, 5.03, 24.21, 21.11, 20.16	49.70%	42274.97	42888.10
Backwards stepwise				
s(La) + s(Lo) + s(De) + s(Bt) + s(Bs) + s(Cv) +s(La*Lo) + s(Bt*Bs) + s(De*La)	2.64, 1.00, 1.00, 1.00, 3.47, 3.85, 25.04, 21.12, 20.03	48.60%	42361.42	42919.23
s(La) + s(Lo) + s(De) + s(Bt) + s(Bs) + s(Cv) +s(La*Lo) + s(Bt*Bs) + s(De*La)	2.644, 1.00, 1.00, 3.47, 3.85, 25.04, 21.12, 21.03	48.60%	42358.74	42908.00
* s(La) + s(Lo) + s(De) + s(Bt) + s(Bs) + s(Cv) +s(La*Lo) + s(Bt*Bs) + s(De*La)	2.644, 1.00, 3.47, 3.85, 25.04, 22.16, 21.03	48.60%	42357.53	42902.99
s(La) + s(Lo) + s(De) + s(Bt) + s(Bs) + s(Cv) +s(La*Lo) + s(Bt*Bs) + s(De*La)	2.654, 1.00, 3.85, 25.07, 24.41, 21.09	48.40%	42370.16	42905.51

While overall model performance was similar between the parsimonious BSP and TPRS GAMs, TPRS model slightly outperformed BSP model across all evaluation criteria (Table 3.3, Figure 3.6). AIC and BIC were lower in the TPRS model (AIC = 42,274.97, BIC = 42,888.10), suggesting its superior performance over the BSP model (Table 3.3). Assessed by cross validation, the TPRS model showed higher predictive performance (α = -0.60, β = 1.02, R^2 = 0.42) compared to the BSP model (α = -2.278, β = 1.04, R^2 = 0.41; Figure 3.6). Therefore, the TPRS model was determined to be more appropriate than the BSP model in this study.

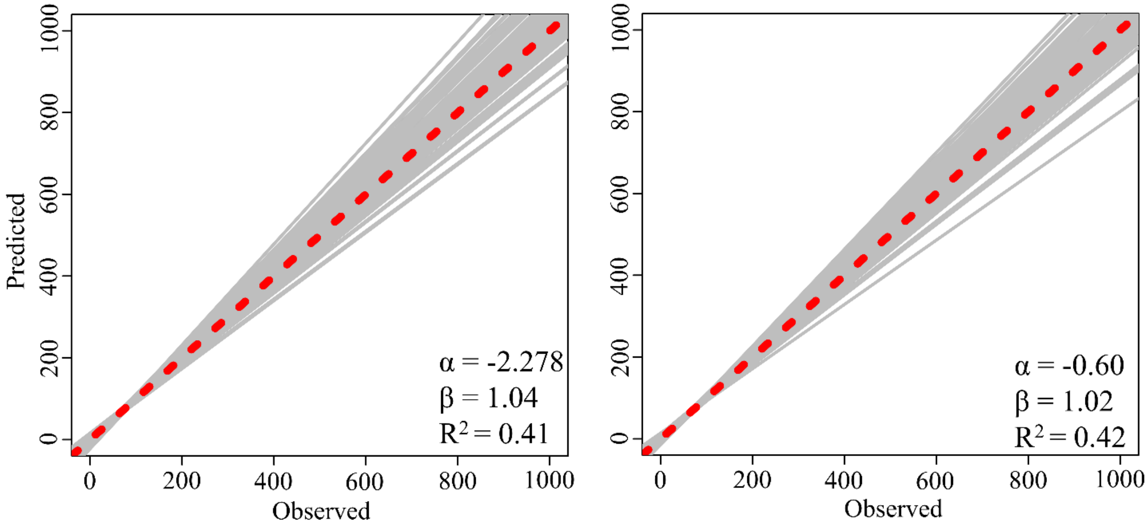


Figure 3.6: Diagnostic plots depicting the comparison of model performance between the parsimonious generalized additive model produced by backwards selection (left) and thin plate regression splines (right). Predictive performance was assessed by cross validation. A graphical summary of observed vs predicted scallop (*Placopecten magellanicus*) density based on100 runs of random data sampling are displayed.

3.4.2 Model predictions

The parsimonious TPRS GAM was used to map model predictions over the inshore GoM. Model results show that nonlinear relationships commonly exist between environmental variables and scallop density; however, both bottom temperature and longitude were reduced to decreasing, near-linear relationships by the TPRS model. The response curves for scallop density as a function of bottom salinity and current velocity were dome shaped, with salinity peaking between 27 – 32 ppt and current velocity peaking between 0.10 – 0.17 (m/s). Because depth was penalized out of the TPRS model, this response curve was flat.

Interactions included in this model (latitude-longitude, latitude-depth, bottom temperature-bottom salinity), identified using BRT analysis, were included to capture the common effect of single environmental variables operating within a more complex composite of factors directly controlling species distribution (Austin 2007; Araujo and Peterson 2012). For example, salinity in this model acts as a proxy for broad spatial patterns in scallop distribution which is shown by the strong interaction between salinity and temperature detected within the model, and likely corresponds to the origin of water mass existing in a given area (MacDonald and Thompson 1985a, 1985b; Macdonald et al. 1987).

Predicted scallop density was higher (> 20 scallops per m^2), in general, along inshore areas relative to offshore areas (< 5 scallops per m^2) (Figure 3.7). Select estuaries along the Maine coast show significantly higher (> 60 scallops per m^2) predicted scallop density relative to other inshore areas. In particular, the Cobscook Bay area shows high predicted density (> 100 scallops per m^2) relative to surrounding areas (Figure 7). Additionally, offshore waters in Western Maine show slightly higher predicted density (10 - 20 scallops per m^2) relative to Eastern Maine (< 10 scallops per m^2) (Figure 3.7). Over the 8-year study, predicted density shows a strong increasing trend along inshore areas and a decreasing trend in offshore areas (Figure 3.8). Cobscook Bay and Penobscot Bay show stronger increasing trends relative to other inshore areas and offshore waters in Western Maine show a stronger decreasing trend in predicted density relative to Eastern Maine.

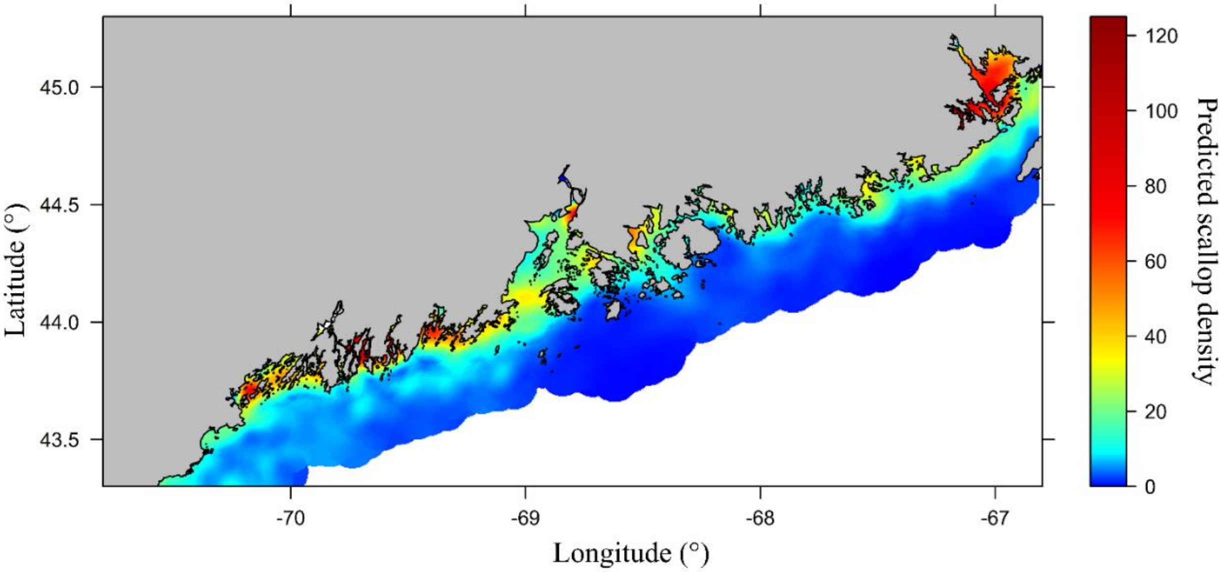


Figure 3.7: Map showing the spatial distribution of median predicted density of scallops (*Placopecten magellanicus*) in the nearshore Gulf of Maine from 2005-2013. The color ramp corresponds to predicted density (scallops per m2), where blue indicates low catches and red indicates high catches.

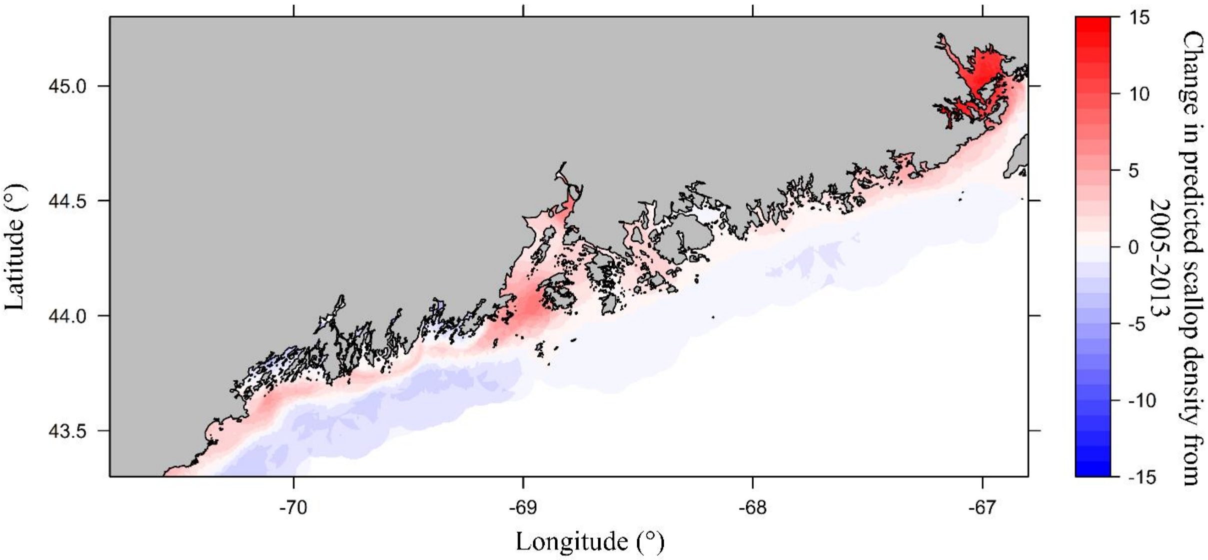


Figure 3.8: Map showing the temporal change in predicted density of scallops (*Placopecten magellanicus*) in the nearshore Gulf of Maine waters from 2005 – 2013. The color ramp corresponds to the degree of change in predicted density. Red areas have a positive change and blue areas have a negative change.

3.5 Discussion

Decision-making associated with conservation planning and fisheries management should use as much information and knowledge as possible to maximize the benefits of management actions (De Ornellas et al. 2011). The model developed in our study was designed to use an existing dataset describing both the abundance and distribution of Atlantic sea scallops to develop a climate-niche species distribution model (SDM). Our Tweedie GAM approach produced high quality predictions of abundance for scallops. Model outputs agree generally with consensus of the distribution of scallops along coastal Maine according to fishermen knowledge, the distribution of fishing effort in the area, as well as landings information (not able to be shown here due to confidentiality agreements). Thus, we consider this modeling effort a successful approach to predict the distribution of scallops across unsampled areas.

This study provides a regional projection in the distribution of scallops within the inshore GoM and hindcasts these projections back eight years. A dominant spatial trend made apparent by this climate-niche SDM is a decline in predicted density moving from inshore to offshore locations. This finding corresponds with habitat value for scallops being higher among inshore areas relative to offshore areas likely attributed to deteriorating food availability moving offshore (Torre et al. 2018). This is thought to represent a natural gradient of habitat quality (Sarro and Stokesbury 2009).

An increasing temporal trend in climate-driven scallop density was observed for inshore areas with a decreasing trend in offshore areas (Figure 3.8). Any change in density over time results from changes in dynamic environmental variables (bottom temperature, bottom salinity, and current velocity), suggesting that the composite of these 3 factors has changed favorably regarding scallop abundance from 2005-2013 in inshore areas and has changed unfavorably in offshore areas. These trends assume that scallop-environment relationships have remained consistent over the study period (Pearson and Dawson 2003; Crisp et al. 2009; Catullo et al. 2015). This general temporal trend is reflected in a recent study which depicts the spatiotemporal distribution of available *scallop* habitat in the GoM using a bioclimate envelope model (Torre et al. 2018).

The TPRS GAM, incorporating a Tweedie distribution for zero-inflated catch data was shown to be a useful prediction tool according to cross validation. The response curves in general agreed with known information about drivers of *scallop distribution* (Naidu and Anderson 1984;

Thouzeau et al. 1991; Wildish and Kristmason 1993; Stewart and Arnold 1994; Pilditch and Grant 1999; Hart and Chute 2004; Torre et al. 2018).

Depth in the case of this modeling framework was not significant, and so was penalized out of the TPRS model. In other studies, scallops have been shown to grow more slowly (MacDonald and Thompson 1985a; Thouzeau et al. 1991) and occur at reduced densities (Schick et al. 1988; Shumway and Parsons 2006) at deeper depths; however, the example given here is comprised of areas that were surveyed in less than 100 m, as opposed to scallops occupying deep areas in other studies (up to 170 m in the case of Schick et al. 1988). Since, within the current study, the dataset describing scallop distribution has low contrast in depth, it follows that depth is not a critical environmental component in this modeling framework.

The results show that, overall, a shrinkage approach where each variable in the full GAM was fitted with a thin plate regression spline including a thin plate spline penalty (TPRS) performed better across all model evaluation criteria than a classic, backwards selection procedure (BSP) for predicting scallop density. These results support findings from Marra and Wood (2011), which used simulated data to show that shrinkage approaches perform significantly better than competing approaches (including BSP) in terms of predictive ability, and are competitive in terms of variable selection performance. However, when using simulated data, the advantage of one modeling approach over another depends on the underlying structure of the simulated data, which makes it difficult to select a “best procedure” for all situations (Binder and Tutz 2008). For example, Marra and Wood (2011) suggest that BSP would be preferred over shrinkage approaches in situations where the data have particularly high information content. Hence, using real-world data to evaluate the utility of model fitting and variable selection methods, as they apply to species distribution modeling, adds an important dimension to comparisons made with simulated data.

In addition to direct model performance benefits offered by TPRS, compared to BSP, shown here and in Marra and Wood (2011), there are other considerations that need to be taken into account when choosing an appropriate method for model fitting and variable selection. Stepwise procedures, such as BSP, have the potential to be inconsistent due to high sensitivity to small variability in the response data, which can sometimes lead to very different subsets of chosen variables. Additional variation in application of stepwise procedures results from a dependence on the initial starting path chosen through the variable space (Marra and Wood 2011). Another downside of these procedures is that during variable selection and hypothesis testing using the selected model, p-values associated with model terms do not take into account variable selection uncertainty, and can therefore be misleading (Marra and Wood 2011). Conversely, shrinkage approaches have been shown to be a valid alternative to stepwise procedures in terms of consistency among iterations of application, and increased robustness to variability in the data (Marra and Wood 2011). Moreover, since variable selection in shrinkage approaches is carried out within a single step these methods are less computationally demanding, especially when using larger datasets (Leathwick et al. 2006; Hesterberg et al. 2008)

An important limitation of this modeling framework is that the development of species distribution models relies upon environmental data, and as with any environmental data there are several possible sources of error that could cause misrepresentation of model predictions. The current modeling framework relies particularly heavily upon FVCOM predictions. To evaluate performance of FVCOM within the study area, Tanaka and Chen (2016) and Li *et al.* (2017) performed comparisons between environmental monitors on observed temperature data to and

FVCOM predictions. They found that in general, FVCOM adequately captured broad spatial and temporal trends in bottom temperature and salinity which adds validity to the quality and accuracy of FVCOM predictions.

In this study, environment-density relationships for scallops were defined upon only four environmental variables when many physical, biological, and chemical conditions likely factor into the life history and distribution of this species. However, variables used to train the generalized additive model may have functioned as surrogates for factors directly controlling scallop distribution through physiological mechanisms (Austin 2007; Araujo and Peterson 2012). From the associations between variables we can infer the relationship between spatiotemporal variability of environmental factors, habitat quality, and resulting scallop distribution. For example, salinity in this model may act as a proxy for broad scale spatial patterns in scallop distribution due to the inherent relationship between salinity and the origin of water mass existing in a given area. Environmental predictors in this study were selected based on availability and assumed correlation with scallop density. As more comprehensive environmental data becomes available in the future, studies to develop a further detailed species distribution modeling approach could include additional variables such as pH, dissolved oxygen, predator-prey, and other food-web interactions to capture a more comprehensive representation of scallop ecology (Araújo and Luoto 2007).

The climate-niche SDM developed in this study establishes the ability to quantify relationships between a commercially important fish stock and the surrounding environment, which provides a tangible tool to visualize species distribution over space and time as well as to evaluate potential impacts of a changing GoM ecosystem. Also, our real-world evaluation of two common GAM selection and fitting procedures provides insights into the effectiveness of each method and can be incorporated into further research using GAMs. This modeling approach is highly generalizable to a variety of commercially important species and can advise conservation efforts for the scallop fishery in the GoM to help ensure the implementation of adaptive management strategies under uncertain climate conditions.

Chapter 4 - THE 2016 MAINE DEPARTMENT OF MARINE RESOURCES-UNIVERSITY OF MAINE NORTHERN GULF OF MAINE SEA SCALLOP DREDGE SURVEY: SURVEY DESIGN, ANALYSIS, AND BIOMASS ESTIMATES.

4.1 Project background and survey rationale

Fishery-independent surveys are used to gather information regarding the status of fish stocks in order to satisfy a central goal of fishery resource management: to estimate the abundance of populations over space and time. Addressing this goal constitutes a critical component of stock assessment and management (Smith and Lundy 2006; Liu et al. 2009) and is vital to establishing sustainable fisheries. The usefulness of survey information to provide reliable abundance estimates is greatly impacted by survey design and the allocation of sampling effort (Liu et al. 2009; Wang et al. 2009). Thus, great care must be taken to ensure an appropriate sampling design.

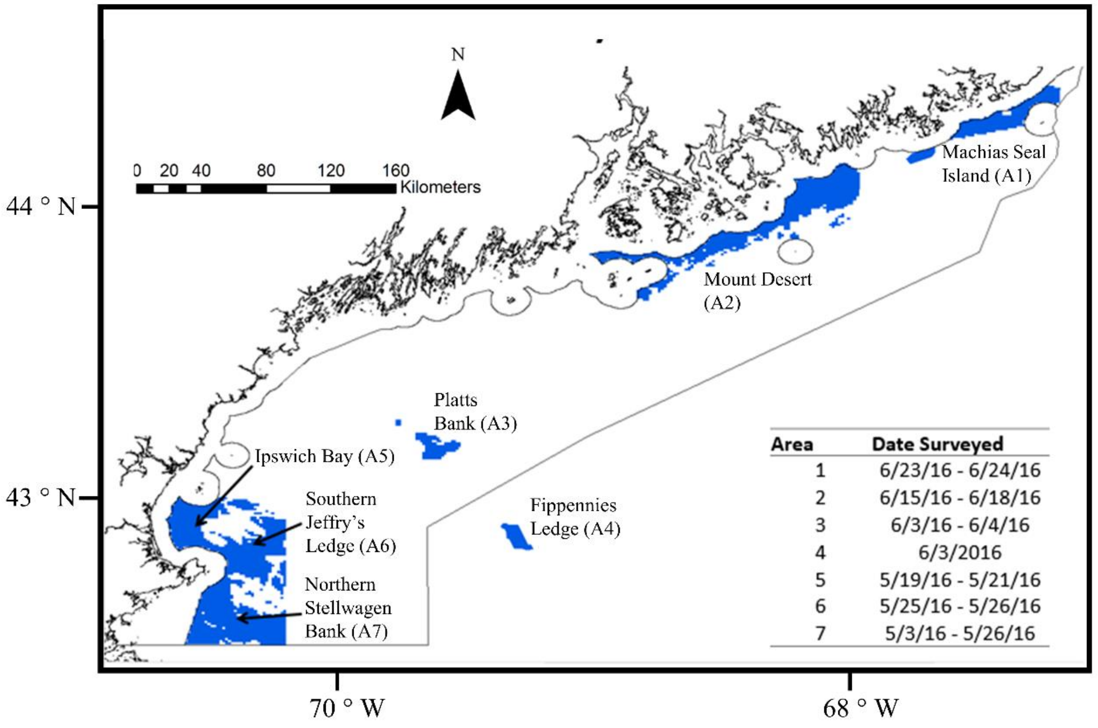
The Northern Gulf of Maine (NGOM) has limited fishery-independent data available with the only recent broad-scale surveying efforts being the two previous DMR-UMaine dredge surveys, which took place in 2009 and 2012. There were three surveys conducted between 1974 and 1984. However, these efforts were confined spatially, and no large-scale surveys were conducted between 1980 and 2008 (Spencer 1974; Serchuk 1984; Serchuk and Wigley 1984). A total allowable catch (TAC)-based management approach was established for the NGOM scallop fishery prior to 2010, which targeted a small number of General Category permit holders that were actively fishing this area during that time. Due to a lack of fishery-independent data, the initial TAC was determined using federal landings data, and subsequent TACs were informed by the 2009 and 2012 DMR-UMaine dredge surveys along with occasional sampling efforts by UMass Dartmouth (Stokesbury et al. 2010; Truesdell 2014).

This section goes over the details regarding the design of the 2016 Maine Department of Marine Resources-University of Maine Northern Gulf of Maine Sea Scallop Dredge Survey and analysis of resulting data. The following sections include a condensed overview of survey area selection and delineation, methodology pertaining to the allocation of sampling effort, and data analysis. The initial design of this survey was based off the two past surveying efforts conducted by Sam Truesdell at the University of Maine (Truesdell 2014). This survey was funded by the 2015 sea scallop Research Set-Aside (RSA) allocation. While the main purpose of this survey was to develop biomass estimates for the NGOM, the data generated by this survey was also used for the development of a species distribution model-computer simulation framework to optimize the design of fisheries surveys discussed in a subsequent chapter of this dissertation (See chapter 5).

4.2 Methods: data collection and analyses

4.2.1 Survey design

Following the previous surveys in 2009 and 2012, a stratified random sampling survey design was used to allocate sampling effort within survey areas. Stratified random sampling (Cochran 1997) is a widely used, highly effective method to estimate the abundance and distribution of different fish species. This type of design tends to be cost-effective and produce estimates of abundance with relatively higher precision compared to other sampling designs under the same level of effort. The effectiveness of this method is dependent upon how well survey strata are selected in reflecting the variability of the organism’s spatial distribution. Based on the 2009-2012 surveys, the strata were able to be delineated to closely match the distribution of scallops in the region. The number of sampling stations in each area reflected stratum size, expected resource abundance, and spatial variability. More intensive sampling took place in areas considered as having commercial importance as well as showing high abundance and spatial variability in the 2009 and 2012 surveys. Based on these criteria, areas of high priority included the federal waters near Machias Seal Island, Mount Desert Island, Platts Bank, Ipswich Bay, the northern part of Stellwagen Bank and the southern part of Jeffrey's Ledge. Fippennies Ledge was also sampled; however, this area falls outside of the NGOM management area, thus is was excluded from overall biomass estimates (Figure 4.1).



Area	Date Surveyed
1	6/23/16 - 6/24/16
2	6/15/16 - 6/18/16
3	6/3/16 - 6/4/16
4	6/3/2016
5	5/19/16 - 5/21/16
6	5/25/16 - 5/26/16
7	5/3/16 - 5/26/16

Figure 4.1: Northern Gulf of Maine (NGOM) management area along with survey areas from the 2016 Maine Department of Marine Resources - University of Maine Scallop Dredge Survey.

The total area of each survey area was delineated based on depth, which is often considered a key factor in the selection of strata boundaries as it influences many important aspects of habitat quality (e.g. water temperature, salinity, and food availability), and thus, abundance and distribution of fish species (Smith and Gavaris 1993). Historical surveys of the NGOM from the 1980s (Serchuk 1984; Serchuk and Wigley 1984), along with NMFS sea scallop survey and trawl survey data on Georges Bank and the GOM have identified depth thresholds. These survey efforts show that catch rates of scallops are much higher in shallow areas and indicate that scallop biomass drops off at depths > 100 m. Thus, delineation of survey areas followed these criteria to maximize sampling coverage on areas of productive scallop habitat (Smith and Gavaris 1993).

Total sampling effort within each area was allocated based on the importance to the fishery and expected biomass following the 2009-2012 survey data. As per Smith and Gavaris (1993), an appropriate allocation of sampling effort to the various strata can play a large role in improving the precision of abundance estimates. To further refine sampling effort allocation in the three strata where highest scallop biomass was expected (Ipswich Bay, Southern Jeffry's Ledge, Northern Stellwagen Bank), each area was split into three strata: high, medium, and low density, based on expected scallop biomass following VTR records, VMS data, as well as 2009 and 2012 survey data. Stratified sampling has been shown to increase the precision over simple random sampling by reducing the variance of estimates due to increased homogeneity in catches within a stratum versus between strata (Nielsen and Johnson 1983); i.e. more sampling effort is allocated to areas with higher expected variance to increase precision.

After defining the strata for each survey area, we analyzed 2009 and 2012 survey data to calculate the relative variance among strata. To determine the number of stations within each stratum, we used the Neyman equation:

$$n_h = n\frac{W_h S_h}{\sum_{h=1}^{H} W_h S_h}$$

where n is the total number of sampling stations for the survey area, H is the total number of strata, W_h is the proportion of stratum h area over the survey area, and S_h is the estimated standard deviation of historical data in stratum h. The Neyman's equation ensures that sampling effort is allocated to areas of high variance to increase precision of abundance indices and refine the resulting biomass estimates.

4.2.2 Data Collection

The 2016 UMaine/DMR NGOM Scallop Survey ran from 5/3/2016 to 6/24/2016. This coastwide dredge-based survey covered six areas within the NGOM management zone considered to have high current and historical scallop abundance as well as having high recent fishing intensity (Figure 4.1). Survey area covered 3805 km^2 and 230 tows were conducted in total. Gear used for this survey was a 7 ft. wide New Bedford-style chain sweep with 2-inch rings, 1¾ inch head bale, 125 mm twine top and 10 inch pressure plate. The dredge was unlined and equipped with rock chains. An onboard marine navigation program was used to chart tow locations. Pitch and roll of the drag, as well as depth and water temperature, was measured using Star-Oddi™ sensors mounted to the neck of the survey drag. This survey type allowed for the collection of data describing the distribution and abundance as well as the collection of shell-height meat-weight information.

Scallops in each tow were counted and weighed. At least 100 scallops (or all scallops if $n \leq 100$) from each tow was measured for SH. If $n > 1{,}000$, a subsample of 10% was measured. Juniper Allegro™ units with ports for digital calipers was used for rapid entry of shell height measurements and other tow data. Meats were extracted and individually weighed from a representative sample of 8-10 scallops from every tow to determine shell height-meat weight height relationships for each area which are necessary to generate biomass estimates. Shell depth and shell width were also measured on these scallops. All other bycatch and bottom type data was collected as described by DMR scallop survey protocol (Schick and Feindel 2005; Kelly 2012).

4.2.3 Data analysis

Using the swept area method and incorporating dredge efficiency coefficients, the total exploitable stock biomass/abundance was calculated for each survey area, accounting for the efficiency of the dredge and the approximate selectivity. For each survey area, the overall average abundance of swept area was estimated as:

$$\bar{X} = \sum_{h=1}^{H} W_h \bar{X}_h$$

where $\bar{X}_h$ is the average abundance of swept area for stratum h, H is the total number of strata, and W_h is proportion of the area of stratum h with respect to the survey area. Since scallop distribution within this region is known to be patchy, common distribution-based estimates of error are not appropriate and, as in the analysis of the past survey data, we used use bootstrapping procedures to estimate the confidence limits around the abundance estimates.

Total exploitable stock biomass/abundance was then estimated for each survey area, accounting for the efficiency of the dredge and the approximate selectivity. Since meat size for a given shell height is known to vary regionally within the Gulf of Maine, shell height-meat weight relationships were modeled individually for each sampling area using a generalized linear mixed model (GLMM) with sampling station as a random effect and depth and an environmental covariate after (Hennen and Hart 2012). We refer to the estimated stock biomass/abundance as exploitable because we depend heavily on active fisheries data in determining survey area and allocating sampling stations. For each survey area, the 95% confidence intervals for the estimated exploitable stock biomass/abundance were calculated using bootstrap methods.

4.3 Results

4.3.1 Scallop distribution and biomass estimates

Within recent years, the most heavily fished area within the NGOM is the southwestern part, from survey areas Ipswich Bay (IB), Southern Jeffries Ledge (SJ), and Northern Stellwagen Bank (NSB). During the 2016 survey, within the IB area, most scallops were found on the western side near the state waters boundary (Figure 4.2). In the NSB area, the main scallop biomass was found towards the south eastern boundary, and in the SJ area, the main scallop biomass was found towards the south western boundary (Figure 4.2).

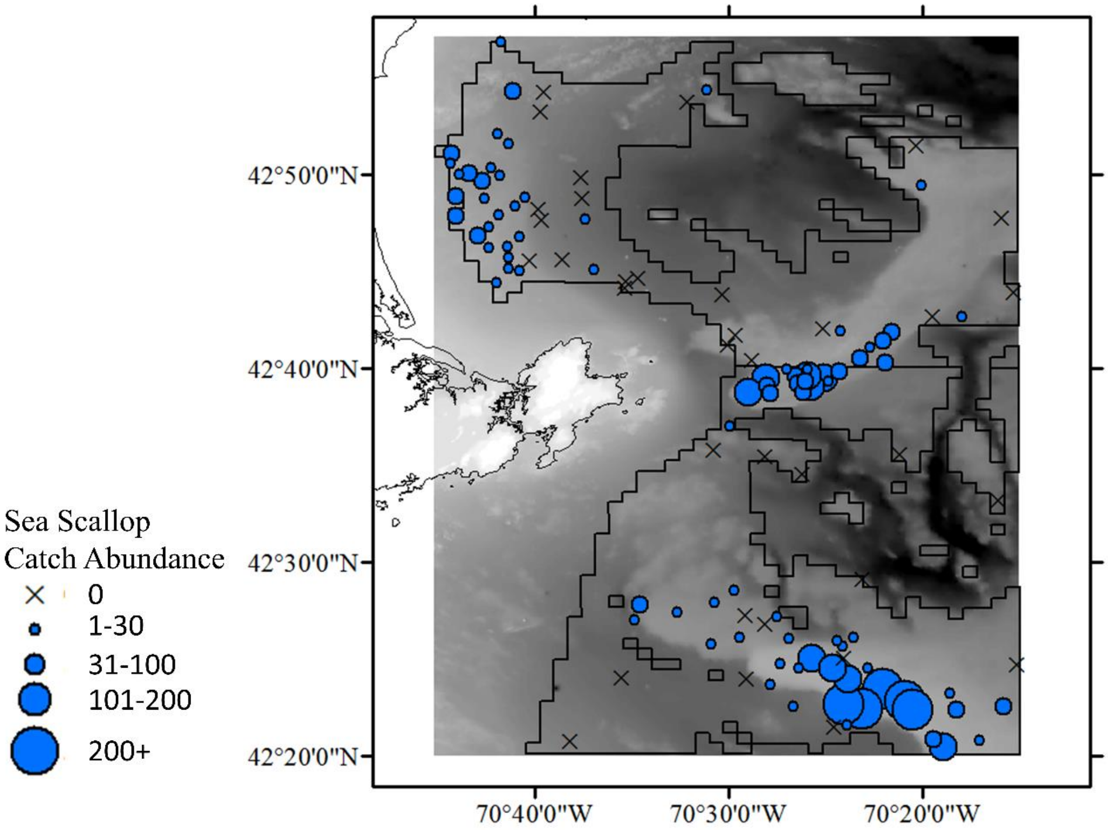


Figure 4.2: Distribution of sea scallop abundance on Ipswhich Bay (top right), Jeffries Ledge (middle) and Stellwagen Bank (bottom).

IB, SJ, and NSB had, multimodal shell height distributions in 2016 (Figure 4.3). The growth of the cohort observed in 2012 was evident in all areas (Figure 4.3). The mode shell heights on IB, SJ, and NSB were ~ 110 mm, ~95 mm, and ~90 mm respectively (Figure 4.4).

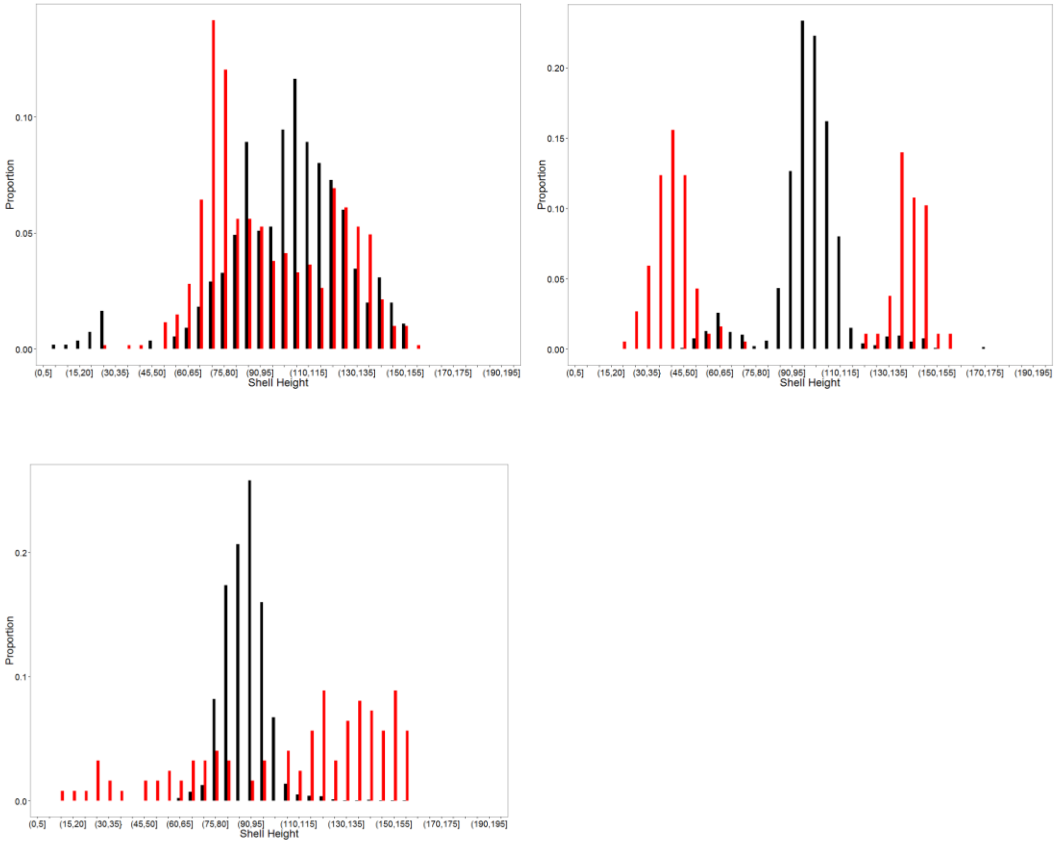


Figure 4.3: Shell height distribution in mm for Ipswich Bay (top left), Jeffries Ledge (top right), and Northern Stellwagen Bank (bottom) during 2012 (red) and 2016 (black).

The survey in Machias Seal Island encountered scallops spread fairly evenly across this area relative to the patchiness observed in the southern areas (Figure 4.4). The only persistent aggregation of scallops was near Machias Seal Island, within state waters. Two age classes of scallops were seen in this region and were between 75 and 110 mm and 130-150 mm respectively (Figure 4.5).

The survey in MDI encountered almost no scallops in 2016 as in past surveys (Figure 4.4). There were scallops to the south of this area near Mount Desert Rock in 2016, but this small region is within Maine state waters and not part of the NGOM.

The survey on Platts Bank found the highest biomass on the southwest portion (Figure 4.4). The growth of the cohort first observed in 2012 was evident (Figure 4.5); the mode shell height grew from ~75 mm in 2012 to ~110 mm in 2016. In both years there was a small proportion of scallops that were between 125 and 150 mm.

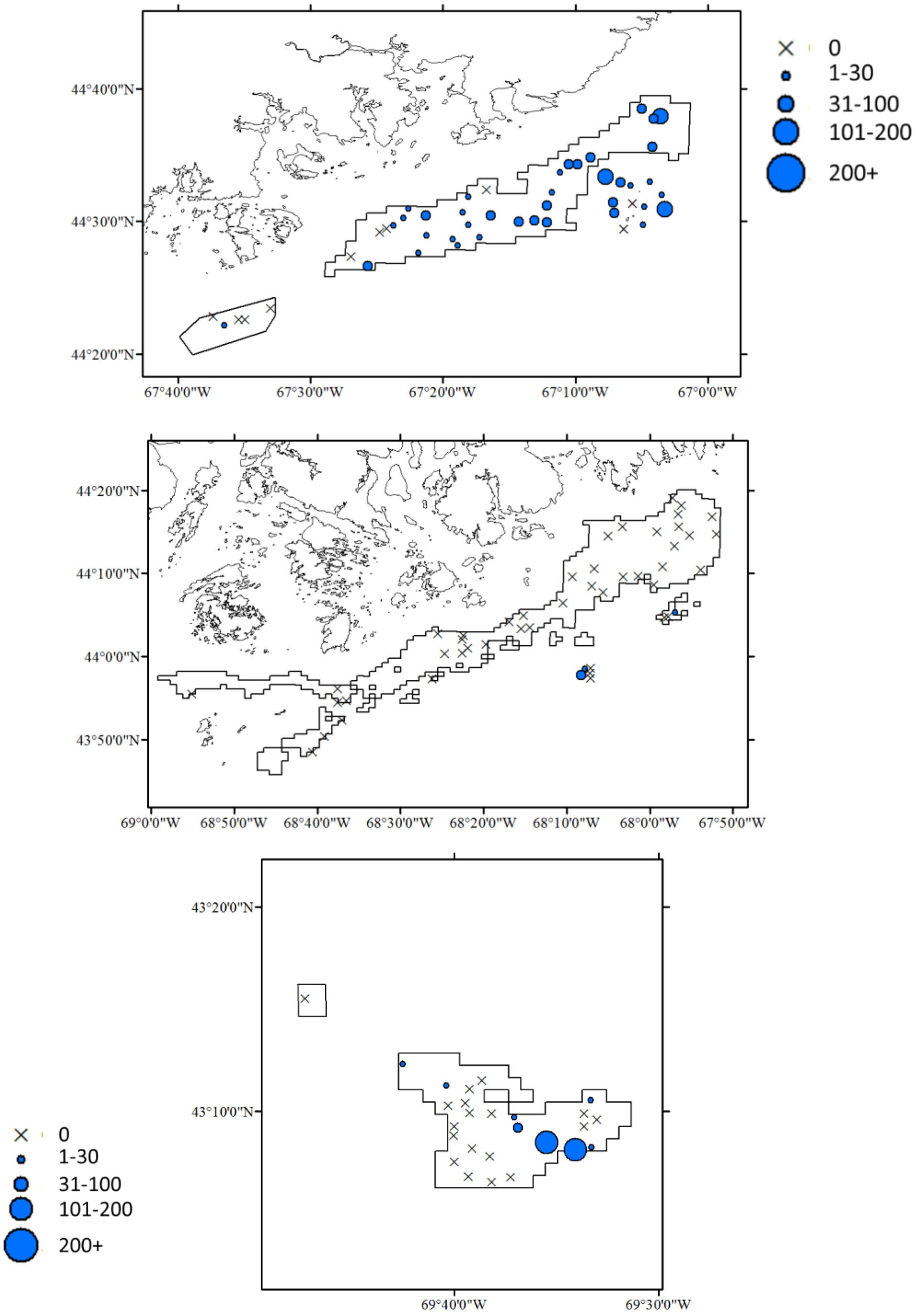


Figure 4.4: Distribution of scallop abundance on Machias Seal Island (top), Mount Desert Island (middle), and Platts Bank (bottom).

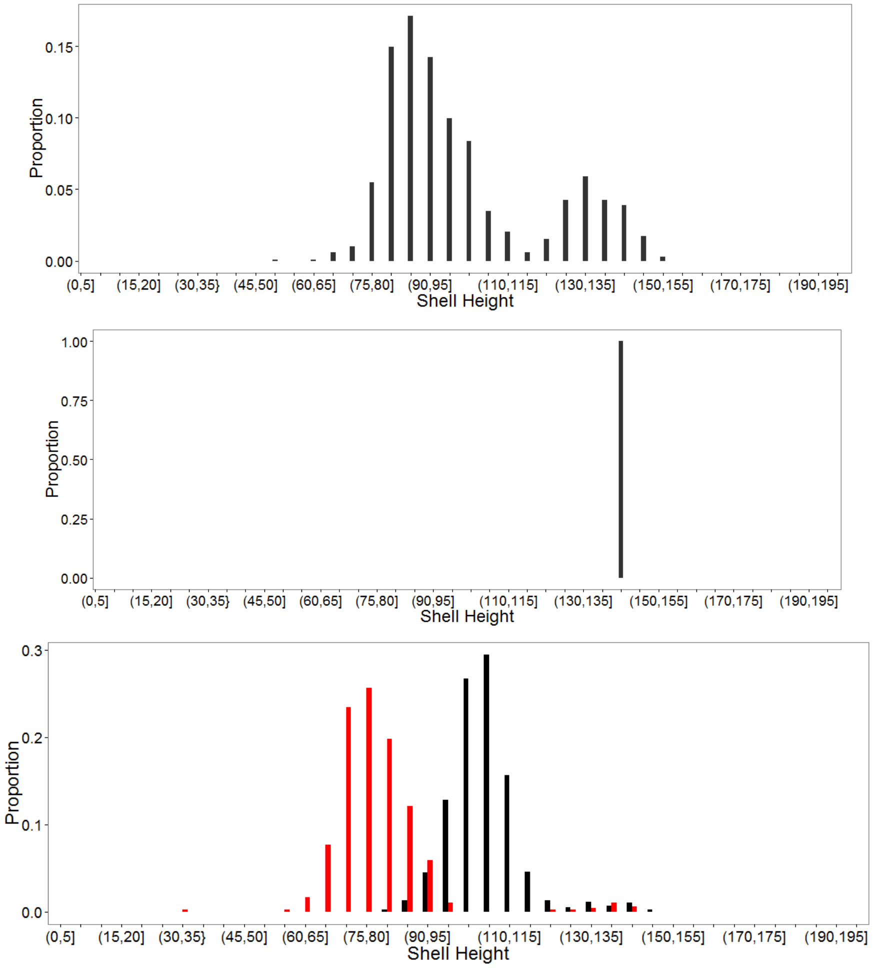


Figure 4.5: Shell height distribution in mm for Machias Seal Island (top), Mount Desert Island (middle), and Platts Bank (bottom) during 2012 (red) and 2016 (black).

The relationship between shell height and meat weight varied by area, as in past surveys (Figure 4.6). The best condition meats were in NSB and SJ, while the meats in MSI were clearly smaller for their size. Few samples of larger scallops were taken on PB, but those greater than 100 mm were of similarly poor condition to the scallops sampled in MSI.

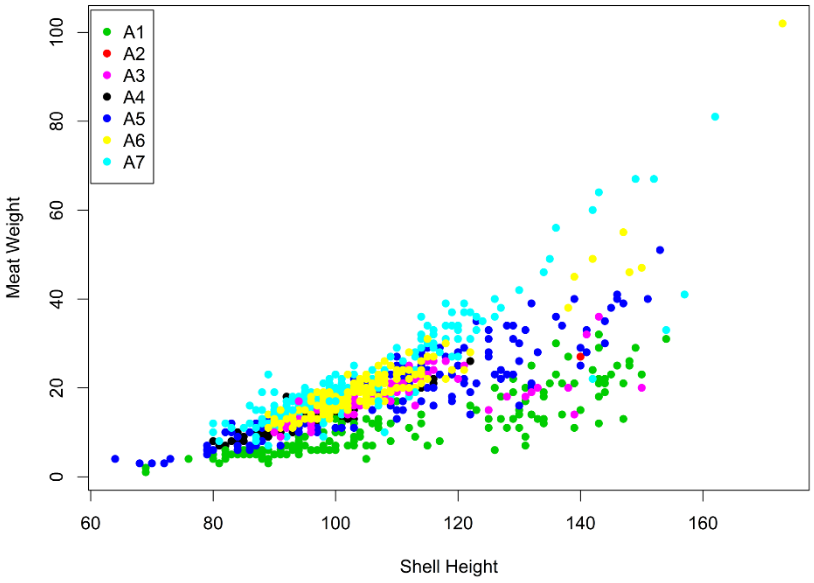
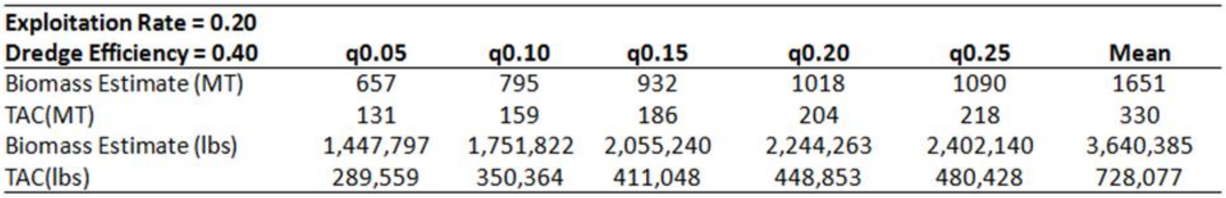


Figure 4.6: Relationship between shell height and meat weight in 2016 for the survey areas.

The highest scallop biomass from this survey was observed on Stellwagen Bank (Figure 4.7). Total biomass in the NGOM, assuming a dredge efficiency of 0.4 and selecting a conservative value (q 0.10 on the bootstrapped distribution), was estimated at ~ 795 MT (Table 1). Using an exploitation rate of 0.2, the removable biomass was calculated to be ~ 159 MT (Table 1).

Table 4.1: Biomass estimated from 2016 NGOM survey (F = 0.2, Dredge Efficiency = 0.4)



Exploitation Rate = 0.20 **Dredge Efficiency = 0.40**	**q0.05**	**q0.10**	**q0.15**	**q0.20**	**q0.25**	**Mean**
Biomass Estimate (MT)	657	795	932	1018	1090	1651
TAC(MT)	131	159	186	204	218	330
Biomass Estimate (lbs)	1,447,797	1,751,822	2,055,240	2,244,263	2,402,140	3,640,385
TAC(lbs)	289,559	350,364	411,048	448,853	480,428	728,077

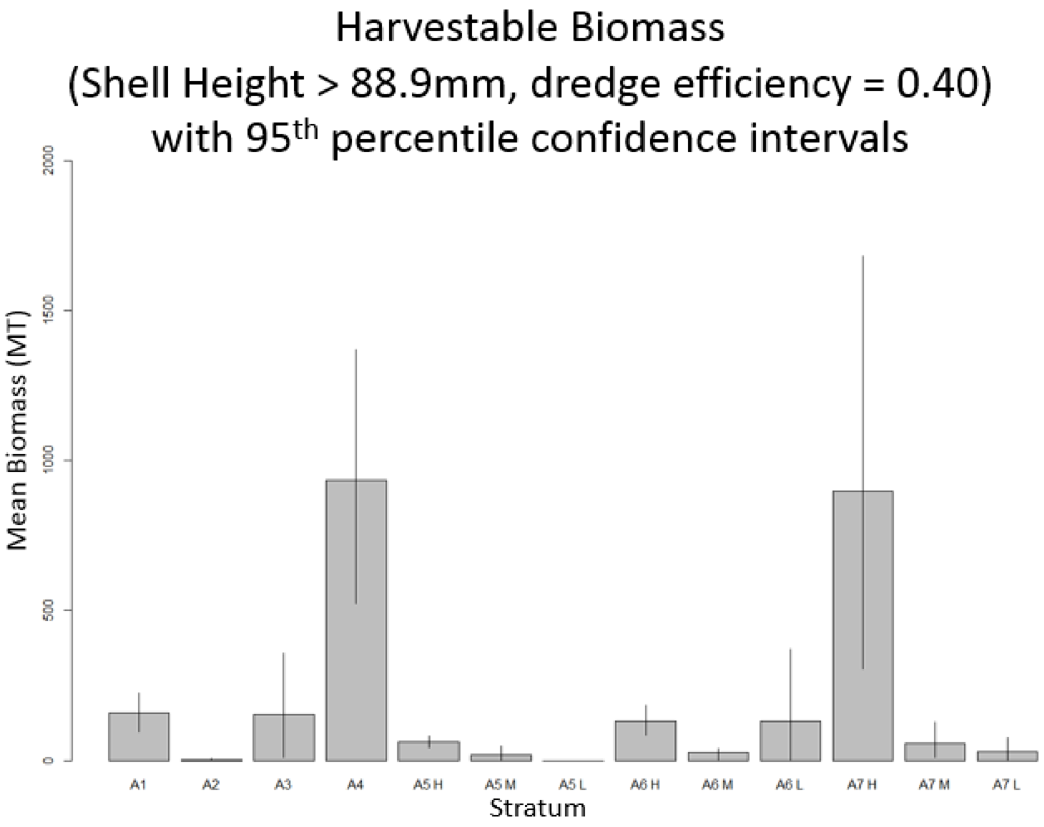


Figure 4.7: Estimated harvestable biomass from each survey area

4.4 Conclusion

The biomass estimates obtained through this survey design and analysis were used to recommend an appropriate level of removable biomass to the New England Fishery Management Council which informed the final TAC for the 2017 fishing season. Despite the limited amount of fishery-independent data available for scallops in this region, it is evident from the length frequency distributions that NGOM scallop populations are characterized by sporadic recruitment events. A large recruitment event was observed within the western NGOM during the 2009 and 2012 DMR-UMaine scallop surveys (Truesdell 2014). This recruitment is what drove high harvestable biomass in the Ipshwich Bay, Southern Jeffrey's Ledge, and Northern Stellwagen Bank areas.

The goal of this survey was to provide reliable biomass estimates of harvestable sea scallops within the NGOM in order to inform management actions for the 2017 fishing season. Since fishing activity remains high in this area, this survey is an ongoing process, so it remains important to use each iteration of survey information to optimize the design with regards to accuracy and precision of sea scallop biomass estimates. To this end, chapter 5 presents a coupled two-stage species distribution model (SDM) and computer simulation framework to evaluate and optimize future iterations of this dredge survey (Torre and Chen in review).

Chapter 5 - EVALUATION AND OPTIMIZATION OF FISHERIES SURVEY DESIGN USING COMPUTER SIMULATION

5.1 Abstract

We present a coupled two-stage species distribution model (SDM) and computer simulation framework to evaluate and optimize an ongoing dredge survey for Atlantic sea scallops (*Placopecten magellanicus*). Model-predicted distribution of *P. magellanicus* biomass is used as the underlying population upon which we simulate multiple sampling effort allocation schemes to evaluate the bias and precision of biomass estimates from the currently implemented survey design and test alternative approaches to select the top performers with respect to these criteria. The SDM developed in our study performed well predicting both the occurrence and biomass distribution of *P. magellanicus*. There was considerable difference, in terms of accuracy and precision, in the reliability of biomass estimates across survey designs. A marked improvement was observed in the relative root mean squared error (RMSE%) of biomass estimates from the best performing survey design (11.4) relative to the one currently implemented in the dredge survey (26.0), justifying changes be made for future surveying efforts. The framework developed in this study offers a dynamic and adaptive approach for evaluating and improving fishery survey design.

5.2 Introduction

A central goal of fishery resource assessment is to estimate the abundance of populations over space and time, which is critical in developing sustainable fisheries. Fishery-independent surveys are used to gather this information about fish stocks on distinct spatial and temporal scales, and are thus critical components of stock assessment and management (Smith and Lundy 2006; Liu et al. 2009). However, the usefulness of survey information to provide reliable stock assessments or hypothesis testing depends greatly on both the bias and precision of abundance estimates (Mier and Picquelle 2008), which are impacted by survey design and the allocation of sampling effort (Liu et al. 2009; Wang et al. 2009).

Within this context, the need to optimize the effectiveness of survey designs with respect to providing abundance estimates that are both unbiased and precise, and thus biologically realistic, is of principle importance to fisheries management (Wang et al. 2009; Hyun et al. 2018). However, this can be difficult when contending with large survey areas and limited resources due to the high costs associated with such efforts (McAllister and Pikitch 1997; Mier and Picquelle 2008; Wang et al. 2009). Thus, many surveying efforts default to implementing traditional designs with insufficient background information, and as a result, abundance estimates may be unreliable or misinterpreted (Wang et al. 2009; Hyun et al. 2018).

Advances in species distribution modeling (SDM) have brought these techniques to the forefront of conservation planning (Austin 2007; Buisson et al. 2010; Merow et al. 2014). Through quantification of species-environment relationships, an environmental profile is established that can be used to predict the distribution of a species across unsampled areas (Austin 2007; Reiss et al. 2015). Application of SDMs can be a cost-effective way to use existing data to expand upon the current understanding of a species distribution (Guisan and Thuiller 2005) and evaluate and refine survey methodology to provide more reliable abundance estimates. As such, the overall goal of this study was to provide a general framework to improve fishery survey design through the development and application of an SDM.

This study develops a two-stage modeling approach to produce an underlying distribution of species biomass upon which multiple survey designs are simulated and evaluated. The first stage of this approach uses an ensemble SDM to predict probability of occupancy which is then regressed in the second stage, along with other environmental covariates, using random forest regression to predict a spatial distribution of species biomass. This modeling approach makes the implicit assumption that species occurrence and biomass distribution are regulated by related environmental factors, but display different relationships with these factors (He and Gaston 2000). Therefore, a two-stage approach better characterizes both occurrence-habitat and biomass-habitat relationships and takes advantage of the greater amount of presence-absence data available relative to abundance or biomass data for many species (Hill et al. 2017). Similar two-stage approaches have been applied to terrestrial species and were successful in predicting species abundance (Hill et al. 2017; Mi et al. 2017).

We develop a coupled two-stage SDM and survey simulation framework designed for evaluating and optimizing survey designs. We applied this framework to a dataset describing both the presence-absence and biomass distribution of Atlantic sea scallops (*Placopecten magellanicus*) to evaluate and optimize an ongoing dredge survey. The model-predicted distribution of *P. magellanicus* biomass is used as the underlying population upon which we simulate and evaluate multiple survey designs. We sought to evaluate both the bias and precision of abundance estimates from the currently implemented survey design and test alternative sampling approaches to identify the best performing design with respect to these criteria. This study provides a straightforward, cost-effective framework for evaluating and improving fishery survey design.

5.3 Materials and Methods

5.3.1 Survey information

This study seeks to develop and test a framework to evaluate and improve an ongoing dredge-based fishery-independent survey of *P. magellanicus* that has been conducted during the spring of 2009, 2012, and 2016 through a joint effort between Maine Department of Marine Resources (DMR) and the University of Maine (Figure 5.1). The gear used for all surveys was an unlined, 7 ft New Bedford style drag with 2 in rings, 1.75 in head bale, 3.5 in twine top, 10 in pressure plate and rock chains. Tows were conducted at 3.5-4 knots and lasted 5 minutes. All combined surveys yielded a total of 361 tows and captured 12,745 individuals (Figure 5.1). Since *P. magellanicus* < 65 mm in shell height were not efficiently sampled with 2 in rings (Kelly 2012), these were excluded from SDMs. A subsample of the *P. magellanicus* catch (n=24) was retained, and meat samples taken, from approximately one third of all tows to estimate shell height-meat weight relationships and convert catch density into grams per meter.

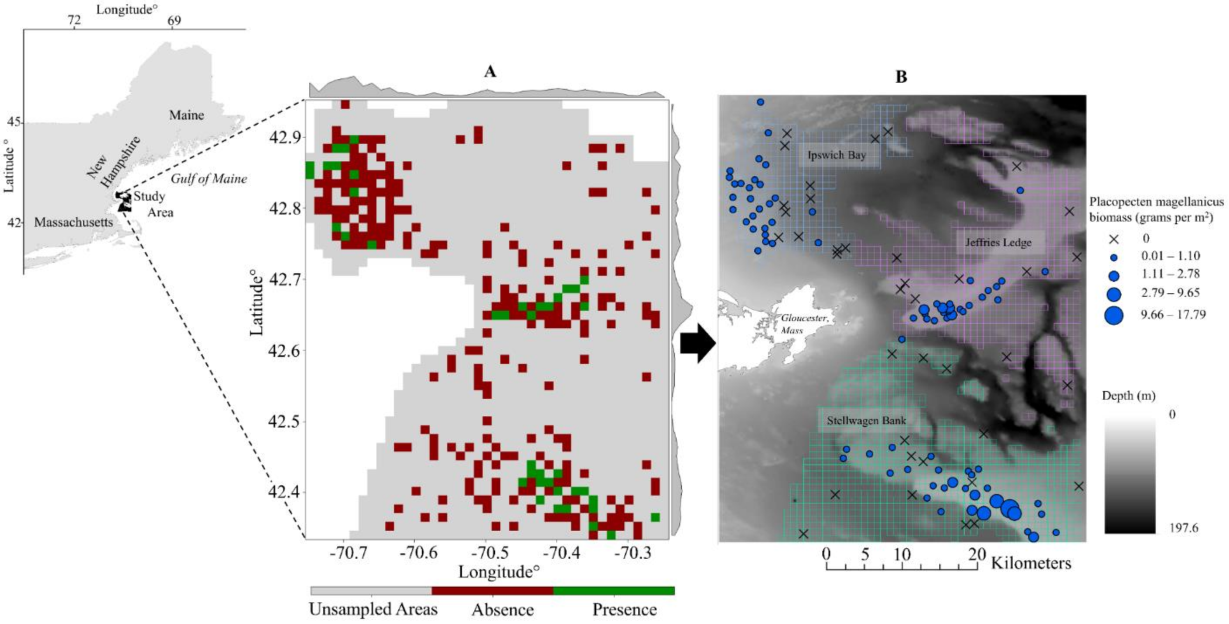


Figure 5.1: Spatial extent of the Maine Department of Marine Resources Scallop Dredge Survey from 2009, 2012, and 2016 which covers a ~ 2,000 km2 portion of the Northern Gulf of Maine (NGOM) Scallop Management area, off Gloucester Massachusetts, USA. A) Presence and absence of *Placopecten magellanicus* (catch density < 15 individuals per tow marks absence) from 2009, 2012, and 2016 surveys. B) Biomass distribution (grams

per m2) of Placopecten magellanicus from the 2016 survey. The survey area is comprised of three distinct areas, Ipswich Bay, Stellwagen Bank, and Jeffries Ledge.

The survey covers a ~ 2,000 km^2 portion of the Northern Gulf of Maine (NGOM) Scallop Management area, off Gloucester Massachusetts, USA (Figure 5.1). The total survey coverage comprises three distinct areas, Ipswich Bay, Jeffries Ledge, and Stellwagen Bank (Figure 5.1). Each area is divided into “High”, “Medium”, and “Low” density (*P. magellanicus*) strata. Strata delineation was set to follow perceived biomass distribution of *P. magellanicus* within the survey areas based on past survey data, direct input from fisherman, and vessel monitoring system (VMS) data. Each survey is allotted 120 tows divided evenly between the three areas. Within each area the allocation of sampling effort (number of dredge tows) follows a stratified random design with effort being divided amongst between high, medium, and low-density strata. Layout of the strata cannot be shown here due to confidentiality agreements, however it remained consistent for all survey designs tested in this study. Tow allocation among strata is carried out using the following equation, Neyman’s Allocation (Neyman 1934):

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^{H} W_h S_h}$$

where n is the total number of sampling stations for the survey area (40 per each area), H is the total number of strata, W_h is the proportion of stratum h area over the survey area, and S_h is the estimated standard deviation of historical data (from the 2016 UMaine-DMR survey) in stratum h. The incorporation of high, medium, and low-density strata coupled with the Neyman equation help to allocate more sampling effort, or tows, to areas with a high variance in *P. magellanicus* biomass (Neyman 1934).

5.3.2 Environmental data

The unstructured-grid Finite-Volume Community Ocean Model (FVCOM) configured in the Northwest Atlantic Shelf region was used to simulate monthly estimates of bottom temperature, bottom salinity, and current velocity. Temperature, salinity, and current velocity were input into the model as the average yearly values from 2009 to 2016 throughout the study area (Chen et al. 2011; Torre et al. 2018) (Figure 5.2). Li *et al.* (2017) used a collection of observed bottom temperatures collected by environmental monitors on lobster traps (eMOLT) to compare with FVCOM predictions in order to evaluate performance of FVCOM within the study area. They found some variability in FVCOM outputs relative to eMOLT observations, but that FVCOM adequately captured general spatial and temporal trends in bottom temperature and salinity. These findings add validity to the quality and accuracy of FVCOM predictions over broader spatial scales. Bathymetry data was obtained from the U.S. Coastal Relief Model (CRM) (NGDC, 1999; Poppe et al., 2005) (Figure 5.2). All environmental data was interpolated using ordinary kriging with a semivariogram function to produce continuous outputs at 1 km^2 resolution (Bailey and Gatrell 1995; R Core Team Development 2016).

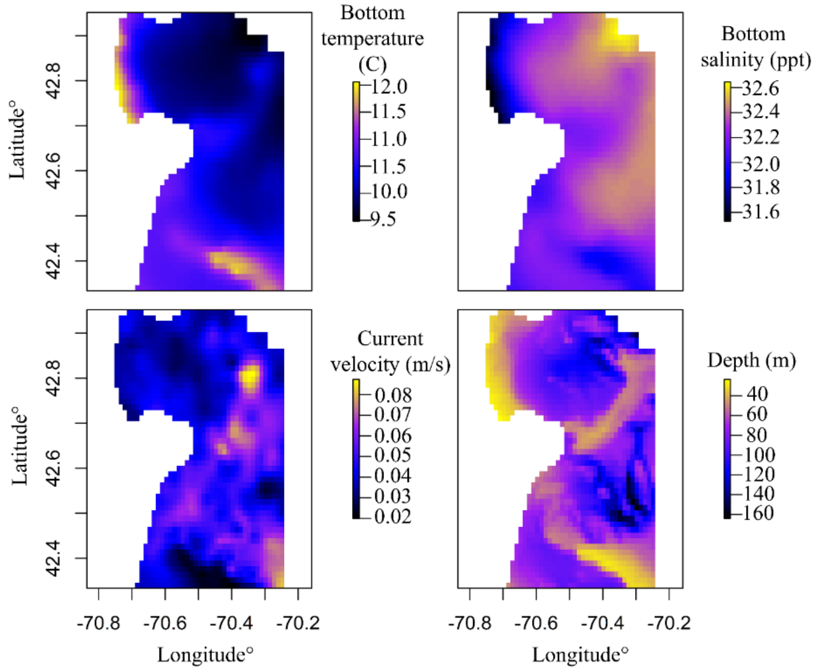


Figure 5.2: Average yearly values from 2009 to 2013 of bottom temperature, bottom salinity, and current velocity obtained from the unstructured-grid Finite-Volume Community Ocean Model (FVCOM) along with bathymetry data obtained from the U.S. Coastal Relief Model (CRM). All environmental data was interpolated using ordinary kriging with a semivariogram function to produce continuous outputs at 1 km^2 resolution.

5.3.3 Two-stage approach to modeling species biomass

Data from the University of Maine-Maine DMR Dredge Survey was used to model scallop distribution and abundance within the study area. This modeling framework was carried out in two stages, where the first stage develops an ensemble species distribution model (SDM) and the second stage couples these results with random forest regression to predict the distribution of *P. magellanicus* biomass. An overview of this modeling framework is shown in Figure 5.3.

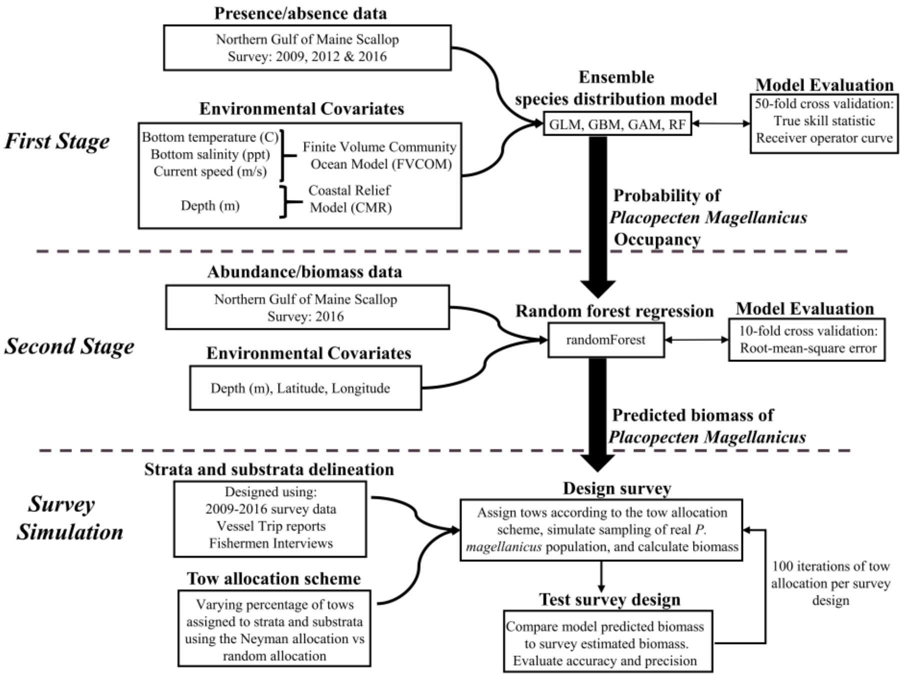


Figure 5.3: Schematic diagram of the modeling and computer simulation framework implemented in this study. All data exploration and modeling procedures were conducted within the R programming environment.

The first stage of the modeling framework uses the R package biomod2 (R Core Team Development 2016; Thuiller et al. 2016) to train ensemble SDMs that map probability of occupancy for *P. magellanicus* across the study area. This stage incorporates presence-absence data for *P. magellanicus* with environmental data (Table 5.1) to run several algorithms commonly used to map the distribution of species (generalized linear model, boosted regression tree, generalized additive model, classification tree analysis, artificial neural network, surface range envelope, flexible discriminant analysis, multiple adaptive regression splines, random forest, and maximum entropy; Miller 2010; Thuiller et al. 2016; Mi et al. 2017). Abundance data from the dredge survey were transformed into presence-absence using a cutoff of 15 individuals/swept area of tow, i.e. any tow that captured less than 15 individuals was considered an absence. This presence-absence criterion was implemented to better reflect the presence, or absence, of *P. magellanicus* beds rather than scattered individuals and to better reflect areas of suitable habitat.

All ten algorithms were run 20 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 (Thuiller et al. 2016). Algorithms were run 20 times each in order to limit computational demands while still achieving stable results. Using cross-validation, two model evaluation criteria, receiver operator curve (ROC) and the true skill statistic (TSS), were used to assess the performance of each algorithm, with higher values for each metric being an indication of better model fit (Hill et al. 2017; Mi et al. 2017). An ensemble distribution model was built using the best performing algorithms based on both the TSS and ROC. Additionally, the response curves for all models produced from each algorithm were visually assessed for validity. If model results produced by an algorithm differed greatly from consensus it was excluded from the ensemble model. Top performing algorithms (excluding those with implausible results) were ranked by TSS score and the top 50 were combined, using a weighted average of TSS scores, to produce the final ensemble model, which predicts probability of occupancy for *P. magellanicus* at 1 km^2 resolution in the study area. This model selection process was developed to produce a robust SDM that is biologically realistic and has high predictive power.

The second stage of this modeling framework uses a machine learning technique, random forest, from the R package randomForest (Liaw and Wiener 2002; R Core Team Development

2016) to regress probability of *P. magellanicus* occupancy with biomass data alongside additional environmental covariates to map the distribution of *P. magellanicus* biomass over the study area (Figure 5.3). Random forest regression was used because of its ability to handle a high proportion of zero-catch tows and skewness in CPUE data, along with its robustness to overfitting (Liaw and Wiener 2002; Prasad et al. 2006). Additionally, it has been shown to perform well using similar types of data inputs (Hill et al. 2017; Mi et al. 2017). Depth was included as an environmental covariate due to known decreases of growth and somatic production of *P. magellanicus* in deeper areas (MacDonald and Thompson 1985b; Cote et al. 1993; Stokesbury and Himmelman 1995). Random forest withholds data (OOB, out-of-bag) to validate each tree for the classification of OOB error, which is used for the calculation of root mean squared error (RMSE) (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).

5.3.4 Survey Simulation

The results from random forest regression were used to test the effectiveness of different survey designs in estimating the total biomass of *P. magellanicus* within the study area (Figure 5.3). The distribution of *P. magellanicus* biomass, as predicted by random forest regression, was treated as the “real” underlying population upon which the simulated surveys were conducted. For each simulated survey, 40 tows were conducted in each of the three areas (Ipswich Bay, Jeffries Ledge, Stellwagen Bank; Figure 5.1). Each of the tested survey designs was unique with respect to the proportion of sampling effort among high, medium, and low-density strata, that was allocated using the Neyman equation vs random allocation; i.e. a varying percent of sampling effort (90%, 80%, 70%, …0% of tows) was allocated using the Neyman equation and the remaining tows were allocated randomly among each strata. Table 5.2 shows all survey designs that were tested in this study. Survey-estimated biomass for each area was calculated using the data picked up from simulated tows by means of the following equation:

$$\bar{X} = \sum_{h=1}^{H} W_h \bar{X}_h$$

where $\bar{X}$ is the total biomass of the sampling area, $\bar{X}_h$ is the bootstrapped mean biomass of all sampled cells for strata h, H is the total number of strata, and W_h is the area of stratum h. Because scallops in this region follow a patchy distribution, typical distribution-based estimates of error were not appropriate, and a bootstrapping procedure was used to calculate mean biomass of all sampled cells within each strata (Efron and Tibshirani 1986). Total survey estimated biomass was calculated by taking the sum of biomass in all three areas.

For each survey design, 100 simulations were carried out and mean estimated biomass was calculated. The effectiveness of each survey design was ranked based on the relative root-mean-square-error (RMSE%), which is a measure of the uncertainty in survey-estimated biomass that takes account of both the bias and the precision (Harbitz et al. 2009), of biomass estimates with respect to the model-estimated biomass:

$$RMSE\% = \frac{\sqrt{bias^2(B_s) + var(B_s)}}{M} 100\%$$

where M is the model-estimated biomass, B_s is the survey-estimated biomass, and $bias^2(B_s)$ and $var(B_s)$ are the bias and variance (precision) of B_s relative to M. This equation effectively quantifies the effectiveness of each survey design using both the bias and precision of survey-estimated biomass with respect to model-estimated biomass. Survey designs were ranked according to the $RMSE\%$ score where lower scores corresponded to better survey performance.

5.4 Results

Nine out of the ten algorithms run to develop the ensemble SDM performed acceptably with regards to prediction capability (TSS score > 0.65; Figure 5.4; Hill et al. 2017, Mi et al. 2017). SRE was rejected straight away from the final distribution model due to a considerably lower ROC & TSS score than other algorithms (Figure 5.4).

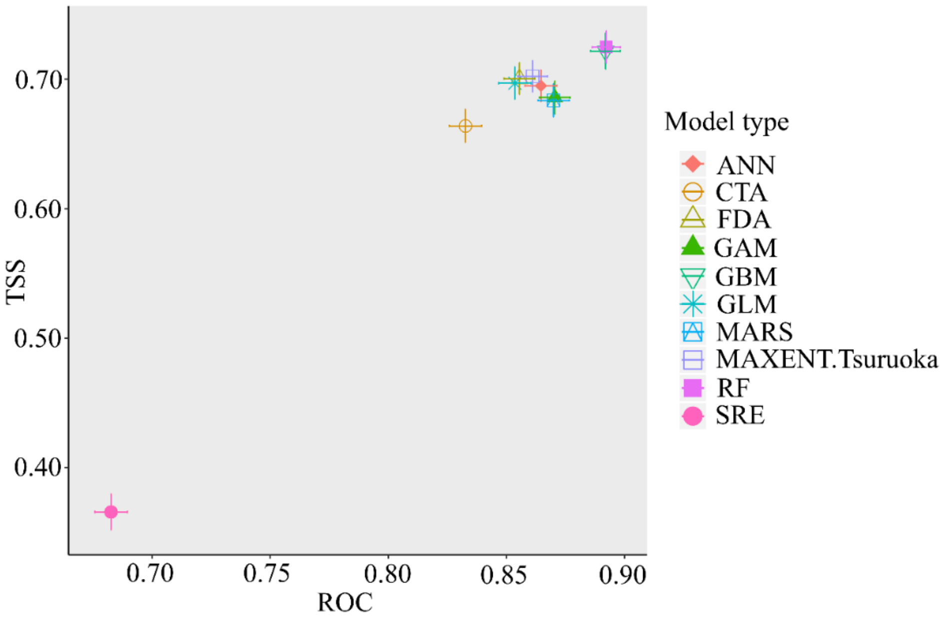


Figure 5.4: Diagnostic plot depicting the comparison of model performance among all algorithms used in the development of an ensemble species distribution model. Algorithms were assessed using both the true skill statistic (TSS) and the receiver operator curve (ROC).

Based on producing the highest combined TSS & ROC scores, along with having the most biologically plausible model results (Figure 5.5, 5.6), the four algorithms included in the final ensemble model were GLM, GBM, GAM, and RF (Figure 5.4). Each of these four algorithms had a ROC score between 0.85 and 0.90 and a TSS score between 0.68 and 0.72 (Figure 5.4).

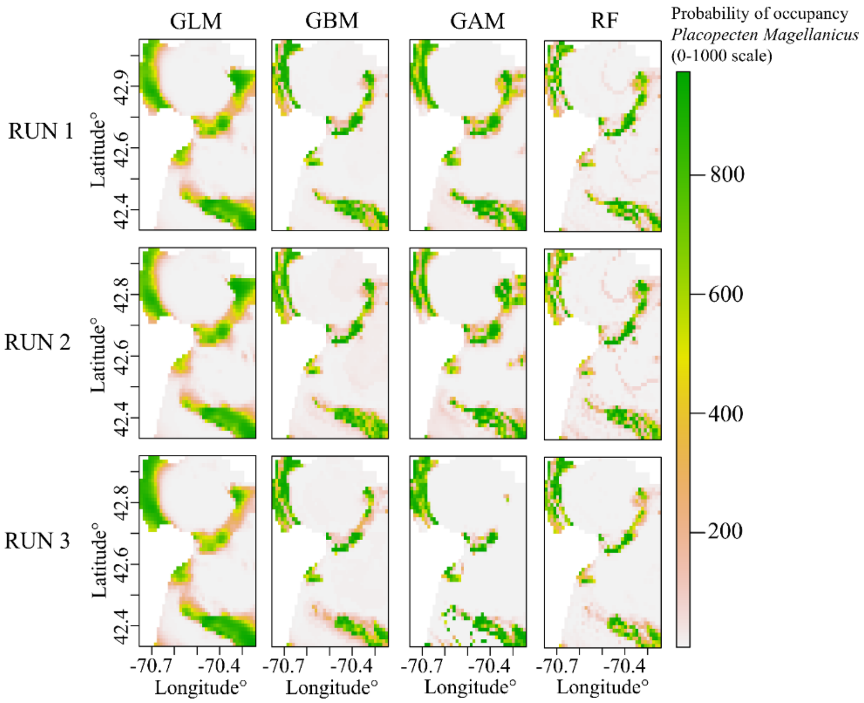


Figure 5.5: Predicted probability of occupancy of *Placopecten magellanicus* across the study area from top performing algorithms at 1 km^2 resolution: generalized linear model (GLM), boosted regression tree (GBM), generalized additive model (GAM), and random forest (RF). The first three runs are shown for each algorithm out of 20 total runs.

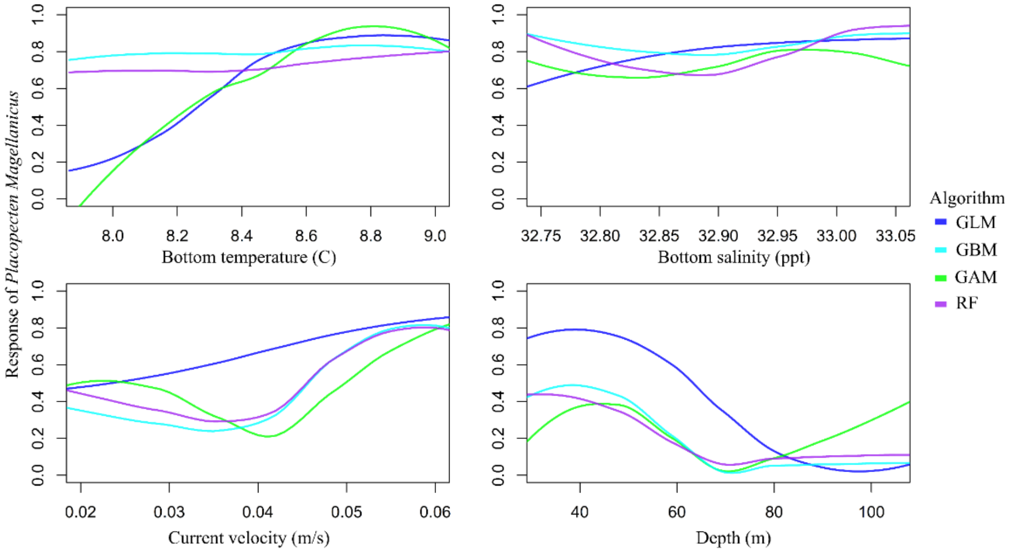


Figure 5.6: Plots describing the effect of environmental covariates on the probability of occupancy of *Placopecten magellanicus* among the top performing algorithms used in ensemble model development. The response curves describe the relationship between a given environmental term and catch densities of *Placopecten magellanicus.*

Since the final model was built using 100% of available data, evaluation information is not given for the ensemble model, because this would effectively test the model on the same data upon which is was built, thus resulting in misleading evaluation statistics (Hill et al. 2017). However, prediction accuracy of the four selected algorithms was good and they successfully predicted a substantial proportion of presence-absence points for *P. magellanicus* (Figures 5.4, 5.5).

All four algorithms included in the final ensemble model (GLM, GBM, GAM, and RF) show similar relationships overall between the response of *P. magellanicus* and each environmental variable (bottom temperature, bottom salinity, current velocity, and depth; Figure 5.6). Some deviation is shown among algorithms for Bottom temperature, where GBM and RF show a more favorable condition for *P. magellanicus* at lower temperatures (<8.4 °C), relative to GLM and GAM, and all four algorithms show a similar relationship at temperatures > 8.4 °C.

Random forest regression successfully predicted 30.2% of the variation of the distribution of *P. magellanicus* biomass within the study area and root-mean-square error (RMSE) was 0.35. Overall, random forest regression performed well for predicting *P. magellanicus* biomass. The general distribution of biomass within the study area agrees closely with both general knowledge and data from the fishery (Figure 5.7). Random forest regression predicted a total *P. magellanicus* biomass of 520.8 MT, a biologically plausible value that corresponds well with data from the fishery.

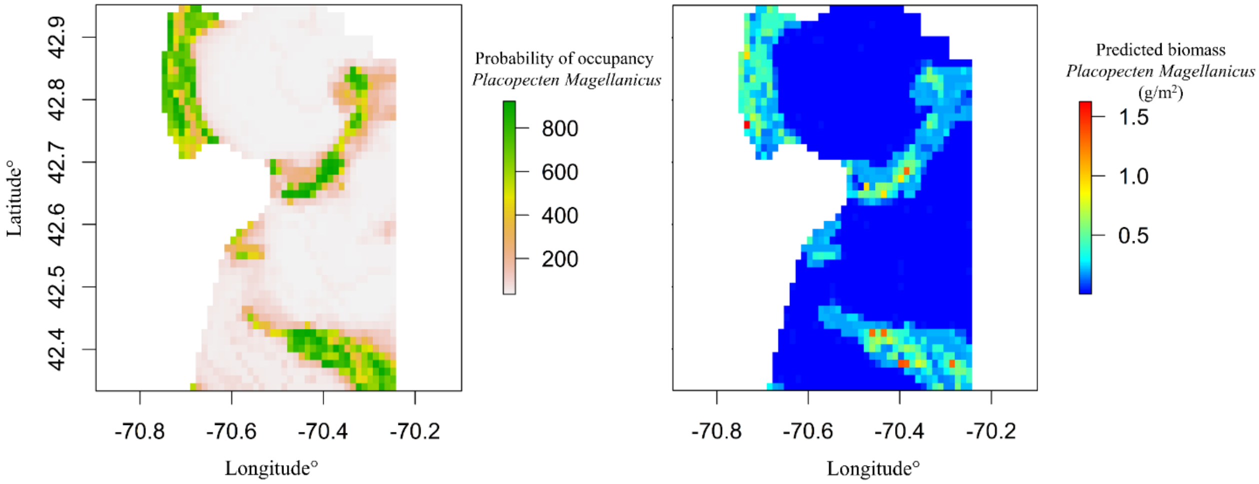


Figure 5.7: Left: Map showing the probability of occupancy for *Placopecten magellanicus* across the study area as predicted by the final ensemble model. The color ramp corresponds to predicted probability of occupancy value, where white indicates poor low probability of occupancy and green indicated high probability. Right: Map showing the spatial distribution of biomass for *Placopecten magellanicus* across the study area as predicted by random forest regression (right). Both maps are at a resolution of 1 km^2.

A total of 11 survey designs were tested in this study (Table 5.2). The best performing survey design allocated 50% of tows via the Neyman equation and 50% of tows randomly and the worst performing design allocated 100% of tows via the Neyman equation and 0% of tows randomly. In general, the best performing survey designs tended to split allocation more evenly between the Neyman allocation and random allocation than did the worst performing designs (Table 5.2). Figure 5.8 shows sample tow placement from the best two and worst two survey designs. Overall, the top designs appear to disperse tows to a greater degree than poor designs while still adequately covering high biomass areas; whereas survey designs that performed relatively poorly clustered sampling effort too tightly and left areas of biomass unsampled (Figure 5.8).

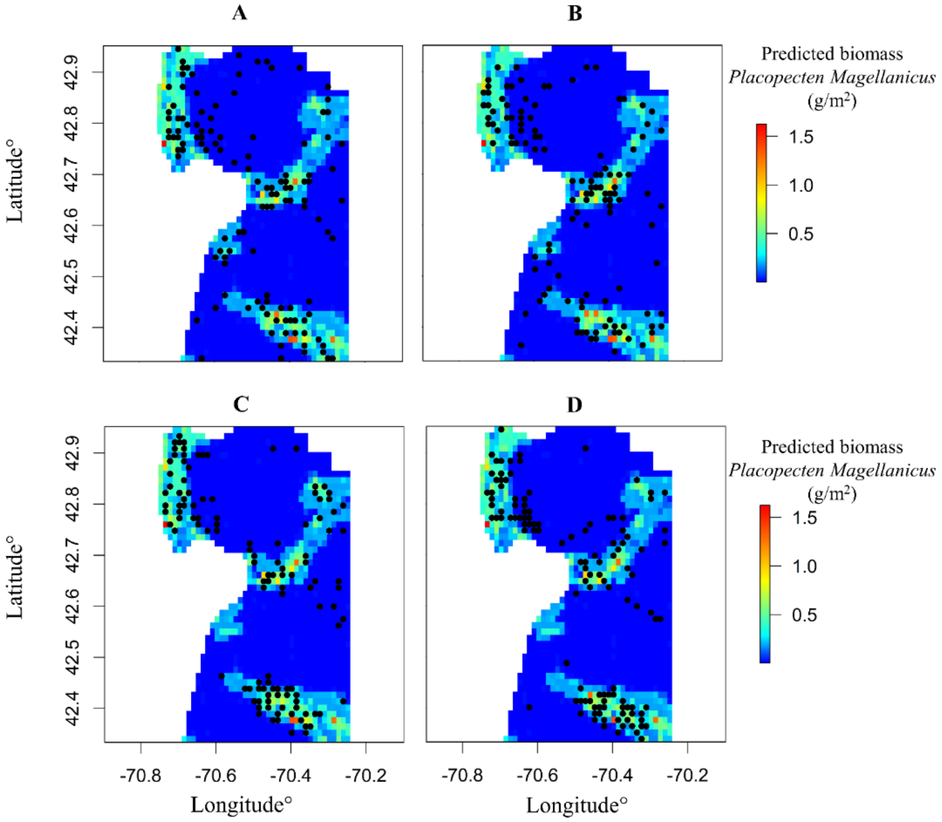


Figure 5.8: Sample tow allocation by best (A,B) and worst (C,D) performing survey designs. Simulated surveys are run upon the biomass distribution as predicted by the two-stage modeling framework (shown in figure 5.7). Tows are denoted by black circles. Each survey design was run 100 times to obtain mean accuracy and precision of survey-estimated biomass relative to model-estiamted biomass. Sampling allocation of each depicted survey is as follows: A) 50% Neyman equation, 50% random; B) 40% Neyman, 60% random; C) 90% Neyman, 10% random; D) 100% Neyman, 0% random.

There was considerable difference, in terms of accuracy and precision, in the reliability of biomass estimates across survey designs. The top performing design had a mean biomass estimate close (0.13% difference) to model-estimated biomass while the worst performing design had a mean biomass estimate that was 1.6% higher than model-estimated biomass. In terms of precision, the top performing model had a CV of 0.1 while the worst performing design had a CV of 0.3. Additionally, there was a marked improvement in the relative root mean squared error (RMSE%) of biomass estimates between the top performing survey designs (RMSE% of 11.4 for the top performing design) relative to the poorest performing designs (RMSE% of 26.0 for the worst performing design; Table 5.2).

5.5 Discussion

Conservation planning and decision-making should use as much available information and knowledge as possible to maximize the benefits of management actions (De Ornellas et al. 2011). The occurrence and biomass distribution models developed in our study were designed to use a limited dataset describing both the presence-absence and biomass distribution of *Placopecten magellanicus* to evaluate and refine an ongoing dredge survey. Our two-stage modeling approach produced high quality predictions of both occurrence, and biomass distribution for *P. magellanicus* in the study area. Model outputs closely agree with general consensus of the distribution of *P. magellanicus* biomass in the NGOM according to fishermen knowledge, the distribution of fishing effort in the area, as well as landings information (not able to be shown here due to confidentiality agreements). Thus, we consider the modeling framework a successful approach to utilizing existing survey data to predict biomass distribution of *P.*

magellanicus across unsampled areas and to produce a reliable model-predicted population upon which varying survey designs were evaluated.

An ensemble modeling approach was used during the first stage of this modeling framework to predict probability of occurrence for *P. magellanicus* in the study area. Ensemble models are being increasingly applied to species distribution modeling because they address a major source of uncertainty in SDM projections, that choice of model algorithm has a large impact on projections (Araújo and New 2007; Buisson et al. 2010; Forester et al. 2013). A growing body of literature suggests that the combination of individual algorithms yield lower mean error than any of the constituent parts (Araújo and New 2007; Buisson et al. 2010). Thus, ensemble models offer the distinct advantages of producing more reliable predictions as well as providing a more straightforward approach to model selection. The inclusion of a machine learning technique, random forest regression, as the second stage of this modeling framework offers additional benefits. Random forest does not make any assumptions about the shape of species-habitat relationships and is robust to overfitting, making it a powerful technique for applications such as this, where a high proportion of zero-catch and skewness in the CPUE data is present (Liaw and Wiener 2002; Prasad et al. 2006).

This two-stage modeling framework has an advantage over conventional single-stage models in that it establishes relationships between occurrence-habitat and biomass-habitat separately. This approach better characterizes each of these relationships and satisfies the concept that occurrence, and abundance or biomass are not influenced identically by environmental covariates (He and Gaston 2000; Nielsen et al. 2005), thus resulting in a more robust and biologically realistic model results. Similar two-stage models have been used on other species, in terrestrial environments, with promising results. Hill et al. (2017), produced ensemble

SDMs and modeled the abundance distributions for 25 British tree species. They found that a two-stage modeling approach predicted the abundance well in 20 of the 25 species with TSS scores between 0.31 and 0.83 (for ensemble occurrence model) and RMSEs generally less than 5 (for random forest regression models) (Hill et al. 2017). Another example of a similar two-stage modeling framework was developed for the Great Bustard (*Otis tarda dybowskii*) in Bohai Bay, China by Mi et al. (2017). Their model produced reliable predictions of occurrence and abundance for this species, with a ROC of 0.77 and RMSE of 26.54, and they were able to use these predictions to develop a guide for placement of protected areas. The models developed in our study had TSS scores between 0.68 and 0.72, ROC scores between 0.85 and 0.90, and a RMSE of 0.35 thus, establishing that this framework performed well for predicting both the occurrence and biomass distribution of *P. magellanicus.*

A critical factor to consider in this coupled modeling-simulation approach is that the simulations upon which different survey designs are evaluated are based entirely upon the model-derived distribution of scallop biomass. This makes interpretation of simulation results more difficult due to the fact that the random forest regression model only predicted ~30% of the variation in scallop biomass. Thus, while the results of the simulation component of this study should be of some limited interest in that they provide an example of how this type of modeling approach can be a valuable tool as applied to fisheries management, more emphasis should be placed on the framework itself than on the results of this study. However, while fine-scale interpretation of the simulation results should be avoided, the results from this study adequately establish a mis-match between the distribution of scallop biomass and the delineation of substrata in the study area. For this reason, future survey efforts in the area should incorporate an element of randomness in the design, which will help to alleviate this mis-match.

An important limitation to consider regarding this modeling approach is that the development of SDMs relies upon environmental data, which are subject to numerous sources of error and could cause misrepresentation of model predictions. The current modeling framework relies upon FVCOM predictions. To evaluate performance of FVCOM within the study area, Tanaka and Chen (2016) and Li *et al.* (2017) performed comparisons between observed temperature data from environmental monitors and FVCOM predictions. In general, they found that FVCOM adequately captured broad spatial and temporal trends in bottom temperature and salinity which adds validity to the quality and accuracy of FVCOM predictions used in this study.

The best performing survey designs, with regards to the accuracy and precision of survey-estimated biomass relative to model-predicted biomass, split the allocation of sampling effort between the Neyman allocation and random allocation. This resulted in tows that were spread more evenly among high, medium, and low-density strata while still having concentrated coverage of high biomass areas. Low performing survey designs either allocated the majority of tows via the Neyman equation or randomly, which resulted in sampling coverage either too concentrated, and likely missing areas of biomass, or too dispersed and not providing adequate coverage to areas of high biomass. The UMaine-DMR dredge survey is currently set up to allocate 100% of tows among strata via the Neyman allocation, which in our study was the lowest performing allocation scheme. This design tended to overestimate *P. magellanicus* biomass with a mean survey-estimated biomass of 529.3 MT compared with a model-predicted 520.8 MT. Additionally this allocation scheme had the lowest precision of any of the survey designs tested, with a standard error of 13.5 and a CV of 0.3. Conversely, the highest performing design, which allocated 50% of tows via the Neyman equation and 50% of tows randomly, had a mean estimated biomass closer to the model-predicted biomass (521.5 MT), and was more precise with a standard error of 5.9 MT and a CV of 0.1, constituting a substantial improvement over the setup for the current survey.

It is possible that low performance exhibited by those survey designs which allocated a high proportion of tows via the Neyman allocation was due to a mismatch between high, medium, and low strata delineation and model-predicted biomass distribution. If so, allocating more tows using the Neyman allocation could potentially cluster tows in low biomass areas while leaving high biomass areas with too few samples, resulting in biomass estimates with low accuracy and precision. Thus, these results justify further investigation into the delineation of strata within each area.

While the current study only tested a single feature of survey design, the framework developed here, which couples a two-stage modeling approach with a survey simulation, can be used to evaluate any aspect of survey design. Future studies could assess criteria such as the optimal number of total tows a survey needs to provide adequate coverage while maximizing efficiency, or overall sampling design, i.e. stratified random sampling vs systematic random sampling. Additionally, as mentioned above, the delineation of strata can be further refined to correspond to areas of high biomass as predicted by the modeling framework.

In conclusion, this framework for identifying species occurrence and the distribution of biomass, coupled with computer simulation provides a straightforward, cost-effective framework for evaluating and improving fishery survey design. In this study we evaluate a dredge survey for *P. magellanicus*, however, this framework is flexible to other species and survey types. Additionally, this modeling approach and survey simulation will become increasingly effective as future surveying efforts provide additional data that can be incorporated into the models. This allows for the further refinement of occurrence-environment and biomass-environment relationships that are crucial for predicting the underlying “real” population upon which survey design is evaluated. An ideal application of this framework would use an iterative process where early surveying efforts could be informed using limited available data for a species followed by constant refinement of each subsequent survey design using newly available data from past surveys. Thus, this framework offers a dynamic and adaptive approach to the design of fisheries surveys.

Chapter 6 – FITTING SPECIES DISTRIBUTION MODELING INTO BROADER ECOLOGICAL THEORY

Understanding species biogeography constitutes a critical component of ecological research, conservation planning, and fisheries management (Franklin 2010). Species distributions are known to be influenced by many interacting biotic and abiotic processes which can manifest as highly complex occurrence-environment relationships (Anderson et al. 2002; Chang et al. 2010; Boulangeat et al. 2012; Merow et al. 2014) and a key step in understanding the biogeography of species is identifying primary environmental factors that control the distribution of a species (Merow et al. 2014).

The biogeography of many marine species is currently shifting in response to climate related changes in habitat suitability (Nye et al. 2009; Howell and Auster 2012; Hollowed et al. 2013). Altered biogeography of a species poses several management challenges because changes in species distributions can move stocks into and out of fixed management boundaries (Gaines et al. 2018). If management does not adapt to shifting species distributions, underinformed practices can lead to overharvesting, even in fisheries that are currently managed well (Gaines et al. 2018). Therefore, understanding likely responses of important fish stocks to changes in climactic conditions is critical to the implementation of adaptive management measures (Hollowed et al. 2013). However, shifts in biogeography for many species remain largely unknown (Tompkins and Adger 2004).

The modeling framework developed in my dissertation research has elucidated aspects of the distribution of habitat for sea scallops as well as the biogeography for this species in the Gulf of Maine. Additionally, this research has further established the ability to quantify sea scallop-environment relationships, while providing several tangible tools to visualize the distribution of this species over space and time as well as to evaluate potential impacts of a changing Gulf of Maine ecosystem.

6.1 How this modeling framework fits into ecological niche theory

The distributional area of a species is an intricate expression of its ecology, determined by myriad factors operating at different intensities among different scales (Pearson and Dawson 2003; Soberon and Peterson 2005). Critical to conceptualizing species distribution modeling is the ecological niche concept (Sinclair et al. 2010). From this theory, we can break down the factors which determine the areas in which a species is found into four distinct classes: 1) Abiotic conditions - including aspects of climate, along with the physical environment which impose physiological limits on a species ability to persist in an area; 2) Biotic factors – a set of interactions with other species that modify the species' ability to maintain a population; 3) hard dispersal limitations from existing populations to new areas based on landscape configuration; and 4) the evolutionary capacity of populations to adapt to new conditions (Soberon and Peterson 2005). From this, we can assume that a species will be present in a particular area when three conditions are met (summarized in figure 6.1): 1) favorable abiotic conditions (A); 2) a suitable suite of species is present (B); and 3) the area is reachable from established populations (M) (Soberon and Peterson 2005). Region A represents the functional niche (FN), region B represents the area where appropriate interspecific interactions necessary for the species presence are met, and region M represents the area that is able to be dispersed to from established populations (Soberon and Peterson 2005). $A \cap B$ is the area where both abiotic and biotic conditions are suitable for the persistence of the species, and $A \cap B \cap M = P$, where P is equal to the area that has the right set of biotic and abiotic factors that is also physically accessible to the species (Soberon and Peterson 2005).

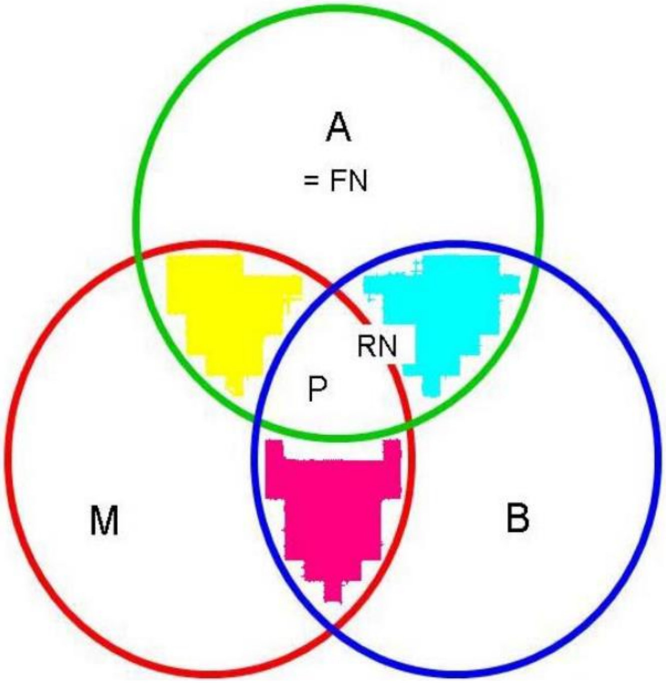


Figure 6.1:from (Soberon and Peterson 2005) – Shows a graphic representation of the ecological niche theory where region A represents the area that has favorable abiotic conditions, or the functional niche (FN), region B represents the area where appropriate interspecific interactions necessary for the species presence are met, and region M represents the area that is able to be dispersed to from established populations.

The chapters of my dissertation constitute an effort to visualize different aspects of this ecological niche concept. The development of a bioclimate envelope model in chapter 2 predicts the fundamental niche (*A*) for sea scallops in the Gulf of Maine, or the network of areas that display a favorable composite of abiotic conditions. This shows areas that scallops are likely able to persist in, relating to conditions where density-dependent fitness would be positive; however, the bioclimate envelope model does not predict *P*, or the areas where scallops are actually distributed. Conversely, the development of the generalized additive model in chapter 3 predicts the realized niche (*P*) or the actual distribution of sea scallops. The two stage species distribution modeling approach developed in Chapter 5 also predicts the realized niche for sea scallops over a small area in the Northern Gulf of Maine in the form of both expected probability of occupancy and the distribution scallop biomass. While biotic conditions are not specified directly in the modeling frameworks presented in either chapter 3 or 5 due to a lack of available data, proxies for these variables are included in the form of latitude and longitude. An important factor to consider in these modeling approaches is that model predictions are derived entirely from catch data made up of the adult life history stage. Thus, successful larval dispersal, settlement, and survival into the late juvenile to early adult life history stages are implied.

6.2 Model assumptions and potential issues associated with spatiotemporal modeling

Given the pervasiveness of uncertainty in ecological systems, it is important to consider the impacts that uncertainty can have on model results (Hoshino et al. 2014). Multiple types of uncertainty exist, inherent to models, that affect the accuracy and precision of model outputs including process error, observation error, and model error (Hoshino et al. 2014). These type of error each arise from different aspects of the modeling process from spatial and temporal stochastic error to the ability of the modeling framework to accurately reflect system dynamics (Hoshino et al. 2014).

Based on the importance of uncertainty in modeling studies, it is worthwhile to consider the assumptions and limitations of my dissertation research. One important limitation inherent to all modeling approaches used in my dissertation is that model development relied upon environmental data which introduces several possible sources of error that could cause misinterpretation of results. Species-habitat relationships were developed using environmental data from the Finite Volume Community Ocean Model (FVCOM; Chen et al. 2011). Since all data obtained through FVCOM are outputs from model simulations as opposed to measure values, the prediction accuracy of FVCOM needs to be considered when evaluating the validity of model outputs. To evaluate performance of FVCOM within the study area, Tanaka and Chen (2016) and Li et al. (2017) used a collection of observed bottom temperatures collected by environmental monitors on lobster traps (eMOLT) to compare with FVCOM predictions. They found that while some variability occurred in FVCOM outputs relative to eMOLT observations, FVCOM adequately captured general spatial and temporal trends in bottom temperature and salinity. These findings add validity to the quality and accuracy of FVCOM predictions over broader scales.

Species-environment relationships for sea scallops in the modeling framework for my dissertation were defined upon a limited number of environmental variables when in reality, many physical, biological, and chemical conditions likely factor into the life history and distribution of this species. However, as is often the case with species-environment modeling, certain variables used to train these models may be functioning as surrogates, or proxies, for factors directly controlling species distribution through physiological mechanisms (Austin 2007; Araujo and Peterson 2012). Thus, from the associations between variables we can infer the relationship between spatiotemporal variability of environmental factors, habitat quality, and resulting species distribution. For instance, salinity in these models may act as a proxy for broad scale spatial patterns in sea scallop distribution due to the inherent relationship between salinity and the origin of water mass existing in a given area. Another example of this concept is that latitude and longitude can be thought of as proxy variables to capture a wide range of covarying bioclimate factors (Guernier et al. 2004). Environmental predictors for each modeling approach were selected based on availability and assumed importance to the species. As more comprehensive environmental data becomes available in the future, studies to develop more detailed modeling approaches could include additional variables such as pH, dissolved oxygen, predator, and other food-web interactions to capture a more comprehensive representation of the ecology of sea scallops (Araújo and Luoto 2007).

Another important factor to consider in these modeling approaches is that the developed species-environment relationships are directly related to the observed abundance or presence of sea scallops, while not accounting for several additional factors that collectively act on the distribution of this species. For instance, not accounted for in these modeling approaches, is spatiotemporally variable fishing pressure, which acts directly on the density of this species. As a result, certain areas with quality habitat may still have lower scallop densities, and thus a lower probability of detected presence, which could impact modeled nonlinear presence-environment or abundance-environment relationships. An additional factor to consider is that the density of benthic species is highly influenced by larval supply, or dispersal during the pelagic larval stage, which is especially true for a sedentary species like sea scallops (Shumway and Parsons 2006). An important next step in this type of research would be to couple these species distribution models with a dispersion model to capture the effects of larval supply on large scale spatiotemporal patterns of species distribution. However, while these are important points to consider, as a result of the large spatiotemporal extent of observations used to train these models, it is likely that collectively, modeled species-environment relationships reflect preferred ranges of environmental variability for sea scallops.

Due to the limitations discussed above as well as the limited resolution of environmental data used in these models, it is unlikely that the approaches used throughout my dissertation would be able to resolve fine-scale species-environment patterns or have the ability to produce predictions that warrant high confidence over small spatial scales. Instead, the modeling approaches developed in my dissertation are more useful for exploring broader-scale spatio-temporal trends in habitat suitability and species distribution. Additionally, in general, while the model results and the inference upon these results should be of some interest, greater emphasis should be placed upon the modeling frameworks themselves that were developed in each chapter. These frameworks can be applied to other species and ecosystems as well as more precisely over a local scale, to help visualize fine-scale species-environment relationships. Being applied to areas where a higher density of both species data and environmental data is present will allow for high resolution model results and predictions that warrant a higher degree of confidence over fine spatial scales.

BIBLIOGRAPHY

Akaike, H. 1974. A new look at the statistical model identification. Trans. Autom. Contr. **19**: 716–723.

Anderson, R.P., Gómez-Laverde, M., and Peterson, A.T. 2002. Geographical distributions of spiny pocket mice in South America: Insights from predictive models. Glob. Ecol. Biogeogr. **11**(2): 131–141. doi:10.1046/j.1466-822X.2002.00275.x.

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.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Modell. **200**(1–2): 1–19. doi:10.1016/j.ecolmodel.2006.07.005.

Bailey, T.C., and Gatrell, A.C. 1995. Interactive spatial data analysis. *In* Ethical Perspectives. Longman Scientific & Technical. doi:10.2143/EP.1.2.630092.

Bayer, M., and Porter, W.F. 1988. Evaluation of a guild approach to habitat assessment for forest-dwelling birds. Environ. Manage. **12**(6): 797–801. doi:10.1007/BF01867606.

Beyer, H.L., Haydon, D.T., Morales, J.M., Frair, J.L., Hebblewhite, M., Mitchell, M., and Matthiopoulos, J. 2010. The interpretation of habitat preference metrics under use-availability designs. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **365**: 2245–2254. doi:10.1098/rstb.2010.0083.

Binder, H., and Tutz, G. 2008. A comparison of methods for the fitting of generalized additive models. Stat. Comput. **18**(1): 87–99. doi:10.1007/s11222-007-9040-0.

Bivand, R. 2013. classInt: choose univariate class intervals R package version 0. 1-22. Available from http://cran.r-project.org/package=classInt.

Boulangeat, I., Gravel, D., and Thuiller, W. 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecol. Lett. **15**(6): 584–593. doi:10.1111/j.1461-0248.2012.01772.x.

Bradbury, R.H., Green, D.G., and Reicheltl, R.E. 1986. Qualitative patterns and processes in marine ecology: a conceptual programme. Mar. Ecol. Prog. Ser. **29**: 299–304. doi:10.3354/meps029299.

Brown, S.K., Buja, K.R., Jury, S.H., Mark E. Monaco, and Banner, A. 2000. Habitat Suitability Index Models for Eight Fish and Invertebrate Species in Casco and Sheepscot Bays, Maine.

North Am. J. Fish. Manag. **20**(April 2016): 408–435. doi:10.1577/1548-8675(2000)020.

Buisson, L., Thuiller, W., Casajus, N., Lek, S., and Grenouillet, G. 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Chang. Biol. **16**(4): 1145–1157. doi:10.1111/j.1365-2486.2009.02000.x.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference. Springer Science + Buisness Media, New York.

Caddy, J. 1979. Long-term trends and evidence for production cycles in the Bay of Fundy scallop fishery. Rapp. el Proces-Verbaux des Reun. Conscil Int. pour 1'Exploration la Mer **15**: 97–108.

Caputi, N., Lestang, S. de, Flusher, S., and Wahle, R.A. 2013. The Impact of Climate Change on Exploited Lobster Stocks. *In* B. F. Phillips (Ed.), Lobsters: Biology, Management, Aquaculture & Fisheries, 2nd Edition. John Wiley & Sons, Ltd. pp. 84–112.

Carsen, A.E., Hatcher, B.G., Scheibling, R.E., Hennigar, a. W., and Taylor, L.H. 1995. Effects of site and season on movement frequencies and displacement patterns of juvenile sea scallops Placopecten magellanicus under natural hydrodynamic conditions in Nova Scotia, Canada. Mar. Ecol. Prog. Ser. **128**: 225–238. doi:10.3354/meps128225.

Catullo, R.A., Ferrier, S., and Hoffmann, A.A. 2015. Extending spatial modelling of climate change responses beyond the realized niche: Estimating, and accommodating, physiological limits and adaptive evolution. Glob. Ecol. Biogeogr. **24**: 1192–1202. doi:10.1111/geb.12344.

Chai, T., and Draxler, R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geosci. Model Dev. **7**: 1247–1250. doi:10.5194/gmd-7-1247-2014.

Chang, J.-H., Chen, Y., Holland, D., and Grabowski, J. 2010. Estimating spatial distribution of American lobster Homarus americanus using habitat variables. Mar. Ecol. Prog. Ser. **420**: 145–156. doi:10.3354/meps08849.

Chang, Y.-J., Sun, C.-L., Chen, Y., and Yeh, S.-Z. 2012. Habitat suitability analysis and identification of potential fishing grounds for swordfish, Xiphias gladius, in the South Atlantic Ocean. Remote Sens. **33**(23): 7523–7541.

Chen, C., Beardsley, R., Cowles, G., Qi, J., Lai, Z., Gao, G., Stuebe, D., Xu, Q., Xue, P., Ge, J., Ji, R., Hu, S., Tian, R., Huang, H., Wu, L., and Lin, H. 2011. An unstructured-grid, finite-volume community ocean model fvcom user manual. Massachusetts Institute of Technology Cambridge,.

Chen, C., Beardsley, R.C., and Cowles, G. 2006. An unstructured-grid, finite-volume coastal ocean model (FVCOM) system. Oceanography **19**(1): 78–89.

Cheung, W.W.L., Lam, V.W.Y., Sarmiento, J.L., Kearney, K., Watson, R., and Pauly, D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish Fish. **10**(3): 235–251. doi:10.1111/j.1467-2979.2008.00315.x.

Cochran, W.G. (n.d.). Sampling Techniques. 3nd Edition. John Wiley & Sons, Ltd, New York.

Cote, J., Himmelman, J.H., Claereboudt, M., and Bonardelli, J.C. 1993. Influence of density and depth on the growth of juvenile sea scallops (Placopecten magellanicus) in suspended culture. Can. J. Fish. Aquat. Sci. **50**: 1857–1869.

Crisp, M.D., Arroyo, M.T.K., Cook, L.G., Gandolfo, M.A., Jordan, G.J., McGlone, M.S., Weston, P.H., Westoby, M., Wilf, P., and Linder, H.P. 2009. Phylogenetic biome conservatism on a global scale. Nature **458**(7239): 754–756. Nature Publishing Group. doi:10.1038/nature07764.

Crossin, T.C., Al-Ayoub, S.A., Jury, S.H., Howell, W.H., and Watson III, W.H. 1998. Behavioral thermoregulation in the American lobster Homarus americanus. J. Exp. Biol. **201**: 365–74. Available from http://www.ncbi.nlm.nih.gov/pubmed/9427670.

Development, R.C.T. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Dickie, L. 1955. Fluctuations in abundance of the giant scallop, Placopecten magellanicus (Gmelin), in the Digby area of the Bay of Fundy. J. Fish. Res. Board Canada **12**: 797–857.

Drinkwater, K.F., and Gilbert, D. 2004. Hydrographic variability in the waters of the Gulf of St. Lawrence, the Scotian Shelf and the eastern Gulf of Maine (NAFO Subarea 4) during 1991-2000. J. Northwest Atl. Fish. Sci. **34**(September): 85–101. doi:10.2960/J.v34.m545.

Efron, B., and Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. **1**(1): 54–75.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. **77**(4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Forester, B.R., Dechaine, E.G., and Bunn, A.G. 2013. Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions. Divers. Distrib. **19**: 1480–1495. doi:10.1111/ddi.12098.

Frank, K.T., Perry, R.I., and Drinkwater, K.F. 1990. Predicted Response of Northwest Atlantic Invertebrate and Fish Stocks to CO 2 -Induced Climate Change. Trans. Am. Fish. Soc. **119**: 335–365. doi:10.1577/1548-8659(1990)119<0353.

Franklin, J. 2010. Mapping Species Distributions. Spatial Inference and Prediction. Cambridge University Press.

Gaines, S.D., Costello, C., Owashi, B., Mangin, T., Bone, J., Molinos, J.G., Burden, M., Dennis, H., Halpern, B.S., Kappel, C. V., Kleisner, K.M., and Ovando, D. 2018. Improved fisheries management could offset many negative effects of climate change. Sci. Adv. **4**(8): 1–9. doi:10.1126/sciadv.aao1378.

Giannoulaki, M., Pyrounaki, M.M., Liorzou, B., Leonori, I., Valavanis, V.D., Tsagarakis, K., Bigot, J.L., Roos, D., De Felice, A., Campanella, F., Somarakis, S., Arneri, E., and Machias, A. 2011. Habitat suitability modelling for sardine juveniles (Sardina pilchardus) in the

Mediterranean Sea. Fish. Oceanogr. **20**(5): 367–382. doi:10.1111/j.1365-2419.2011.00590.x.

Guisan, A., Edwards, T.C., and Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Modell. **157**: 89–100. doi:10.1016/S0304-3800(02)00204-1.

Guisan, A., and Thuiller, W. 2005. Predicting species distribution: Offering more than simple habitat models. Ecol. Lett. **8**(9): 993–1009. doi:10.1111/j.1461-0248.2005.00792.x.

Harbitz, A., Ona, E., and Pennington, M. 2009. The use of an adaptive acoustic-survey design to estimate the abundance of highly skewed fish populations. ICES J. Mar. Sci. **66**: 1349–1354. doi:10.1093/icesjms/fsp088.

Hart, D.R., and Chute, a S. 2004. Essential Fish Habiat Source Document: Sea Scallop, Placopecten magellanicus, Life History and Habitat Characteristics. *In* U.S. Dep. Commer. NOAA Technical Memorandum.

Hastie, T. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer.

Hattab, T., Albouy, C., Lasram, F.B.R., Somot, S., Le Loc'h, F., and Leprieur, F. 2014. Towards a better understanding of potential impacts of climate change on marine species distribution: A multiscale modelling approach. Glob. Ecol. Biogeogr. **23**: 1417–1429. doi:10.1111/geb.12217.

He, F., and Gaston, K.J. 2000. Estimating species abundance from occurrence. Am. Nat. **156**(5): 553–559. doi:10.1086/303403.

Hennen, D.R., and Hart, D.R. 2012. Shell height-to-weight relationships for Atlantic sea scallops (Placopecten magellanicus) in offshore U.S. waters. J. Shellfish Res. **31**(4): 1133–1144. doi:10.2983/035.031.0424.

Hesterberg, T., Choi, N.H., Meier, L., and Fraley, C. 2008. Least angle and penalized regression: A review. Stat. Surv. **2**: 61–93. doi:10.1214/08-SS035.

Hijmans, R., Philips, S., Leathwick, J., and J, E. 2017. dismo: Species distribution modeling. Available from https://cran.r-project.org/web/packages/dismo/dismo.pdf.

Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., and Brown, N. 2017. Abundance distributions for tree species in Great Britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. Ecol. Evol. **7**(4): 1043–1056. doi:10.1002/ece3.2661.

Hollowed, A.B., Planque, B., and Loeng, H. 2013. Potential movement of fish and shellfish stocks from the sub-Arctic to the Arctic Ocean. Fish. Oceanogr. **22**(5): 355–370. doi:10.1111/fog.12027.

Hoshino, E., Milner-Gulland, E.J., and Hillary, R.M. 2014. Why model assumptions matter for natural resource management: Interactions between model structure and life histories in fishery models. J. Appl. Ecol. **51**(3): 632–641. doi:10.1111/1365-2664.12225.

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.

Hyun, S., Maunder, M., and Rothschild B J. 2018. Importance of modelling heteroscedasticity of survey index data in fishery stock assessments. ICES J. Mar. Sci. **72**: 130–136. doi:10.1093/icesjms/fsu223.

Kelly, K.H. 2012. Results from the 2011 Maine sea scallop survey.

Kurihara, H. 2008. Effects of CO2-driven ocean acidification on the early developmental stages of invertebrates. Mar. Ecol. Prog. Ser. **373**: 275–284. doi:10.3354/meps07802.

Langton, R.W., Auster, P.J., and Schneider, D.C. 1995. A spatial and temporal perspective on research and management of groundfish in the northwest Atlantic. Rev. Fish. Sci. **3**(3): 201–229.

Leathwick, J.R., Elith, J., and Hastie, T. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol. Modell. **199**: 188–196. doi:10.1016/j.ecolmodel.2006.05.022.

Li, B., Tanaka, K.R., Chen, Y., Brady, D.C., and Thomas, A.C. 2017. Assessing the quality of bottom water temperatures from the Finite-Volume Community Ocean Model (FVCOM) in the Northwest Atlantic Shelf region. J. Mar. Syst. **173**(December 2013): 21–30. Elsevier B.V. doi:10.1016/j.jmarsys.2017.04.001.

Li, Y., Jiao, Y., and He, Q. 2011. Decreasing uncertainty in catch rate analyses using Delta-AdaBoost: An alternative approach in catch and bycatch analyses with high percentage of zeros. Fish. Res. **107**: 261–271. doi:10.1016/j.fishres.2010.11.008.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. R News **2**(3): 18–22.

Liu, Y., Chen, Y., and Cheng, J. 2009. A comparative study of optimization methods and conventional methods for sampling design in fishery-independent surveys. ICES J. Mar. Sci. **66**: 1873–1882. doi:10.1093/icesjms/fsp157.

Luoto, M., Pöyry, J., Heikkinen, R.K., and Saarinen, K. 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. Glob. Ecol. Biogeogr. **14**: 575–584. doi:10.1111/j.1466-822X.2005.00186.x.

Macdonald, A.B.A., Thompson, R.J., and Bayne, B.L. 1987. Influence of temperature and food availability on the ecological energetics of the giant scallop Placopecten magellanicus IV. Reproductive effort, value and cost. Oecologia **72**(4): 550–556.

MacDonald, B., and Thompson, R. 1985a. Influence of temperature and food availability on the ecological energetics of the giant scallop Placopecten magellanicus. II. Reproductive output and total production . Mar. Ecol. Prog. Ser. **25**: 295–303. doi:10.3354/meps025295.

MacDonald, B., and Thompson, R. 1985b. Influence of temperature and food availability on the ecological energetics of the giant scallop Placopecten magellanicus. I. Growth rates of shell and somatic tissue. Mar. Ecol. Prog. Ser. **25**: 279–294. doi:10.3354/meps025279.

Marra, G., and Wood, S.N. 2011. Practical variable selection for generalized additive models. Comput. Stat. Data Anal. **55**(7): 2372–2387. Elsevier B.V. doi:10.1016/j.csda.2011.02.004.

McAllister, M., and Pikitch, E. 1997. A Bayesian approach to choosing a design for surveying fishery resoirces: aplication to the eastern Bering Sea trawl survey. Can. J. Fish. … **54**: 301–311. Available from http://www.nrcresearchpress.com/doi/pdf/10.1139/f96-286.

Mendo, T., Lyle, J.M., Moltschaniwskyj, N. a, Tracey, S.R., and Semmens, J.M. 2014. Habitat characteristics predicting distribution and abundance patterns of scallops in D'Entrecasteaux Channel, Tasmania. PLoS One **9**(1): 1–9. doi:10.1371/journal.pone.0085895.

Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., Mcmahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., and Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography (Cop.). **37**(12): 1267–1281. doi:10.1111/ecog.00845.

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.

Mier, K.L., and Picquelle, S.J. 2008. Estimating abundance of spatially aggregated populations: comparing adaptive sampling with other survey designs. Can. J. Fish. Aquat. Sci. **65**(2): 176–197. doi:10.1139/f07-138.

Miller, J. 2010. Species Distribution Modeling. Geogr. Compass **6**: 490–509.

Mills, K.E., Pershing, A.J., Brown, C.J., Chen, Y., Chiang, F.-S., Holland, D.S., Lehuta, S., Nye, J.A., Sun, J.C., Thomas, A.C., and Wahle, R.A. 2013. Fisheries management in a changing climate: Lessons from the 2012 ocean heat wave in the northwest Atlantic. Oceanography **26**: 191–195. doi:http://dx.doi.org/10.5670/oceanog.2013.27.

Naidu, K., and Anderson, J. 1984. Aspects of scallop recruitment on St. Pierre Bank in relation to oceanography and implications for resource management. Can. Atl. Fish. Sci. Adv. Comm. Res. Doc. **84**(29): 9.

National Geophysical Data Center. 1999. U.S. Coastal Relief Model - Northeast Atlantic. doi:10.7289/V5MS3QNZ.

National Marine Fisheries Service, (NMFS). 2010. Marine fisheries habitat assessment improvement plan. Report of the National Marine Fisheries Service Habitat Assessment Improvement Plan Team. NMFS-F/SPO-108.

Neyman, J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. J. R. Stat. Soc. **97**(4): 558–625.

Nielsen, S.E., Johnson, C.J., Heard, D.C., and Boyce, M.S. 2005. Can models of presence-absence be used to scale abundance ? Two in life history case studies extremes considering. Ecography (Cop.). **28**: 197–208. doi:10.1111/j.0906-7590.2005.04002.x.

NMFS. 2016. Fisheries of the United States, 2016 Report. Silver Spring, MD.

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 continental shelf. Mar. Ecol. Prog. Ser. **393**: 111–129. doi:10.3354/meps08220.

De Ornellas, P., Milner-Gulland, E.J., and Nicholson, E. 2011. The impact of data realities on conservation planning. Biol. Conserv. **144**: 1980–1988. Elsevier Ltd. doi:10.1016/j.biocon.2011.04.018.

Overholtz, W.J., Hare, J.A., and Keith, C.M. 2011. Impacts of Interannual Environmental Forcing and Climate Change on the Distribution of Atlantic Mackerel on the U.S. Northeast Continental Shelf. Mar. Coast. Fish. **3**(1): 219–232. doi:10.1080/19425120.2011.578485.

Pearson, R.G., and Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species : are bioclimate envelope models useful ? Glob. Ecol. Biogeogr. **12**(5): 361–371. doi:10.1046/j.1466-822X.2003.00042.x/pdf.

Pershing, A.J., Alexander, M.A., Hernandez, C.M., Kerr, L.A., Le Bris, A., Mills, K.E., Nye, J.A., Record, N.R., Scannell, H.A., Scott, J.D., Sherwood, G.D., and Thomas, A.C. 2015. Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. Science (80-.). **350**(6262): 809–812. doi:10.1126/science.aac9819.

Pilditch, C.A., and Grant, J. 1999. Effects of variation in flow velocity and phytoplankton concentration on sea scallop (Placopecten magellanicus) grazing rates. J. Exp. Mar. Bio. Ecol. **240**: 111–136.

Poppe, L.J., Williams, S.J., and Paskevich, V.F. 2005. USGS east-coast sediment analysis: Procedures, database, and GIS data (ver. 2.0): U.S. Geological Survey Open-File Report 2005–1001. Available from http://woodshole.er.usgs.gov/openfile/of2005-1001/.

Posgay, J.A. 1981. Movements of tagged sea scallops on Georges Bank. Mar. Fish. Rev. **43**(4): 19–25.

Prasad, A.M., Iverson, L.R., and Liaw, A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems **9**: 181–199. doi:10.1007/s10021-005-0054-1.

R Core Team Development. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.r-project.org/.

Reiss, H., Birchenough, S., Borja, A., Buhl-mortensen, L., Craeymeersch, J., Dannheim, J., Darr, A., Galparsoro, I., Gogina, M., Neumann, H., Populus, J., Rengstorf, A.M., Valle, M., Hoey, G. Van, Zettler, M.L., and Degraer, S. 2015. Benthos distribution modelling and its relevance for marine ecosystem management. ICES J. Mar. Sci. **72**(2): 297–315.

Ridgeway, G. 2015. gbm: Generalized Boosted Regression Models. R package version 2.1.1. Available from https://cran.r-project.org/package=gbm.

Saba, V.S., Griffies, S.M., Anderson, W.G., Winton, M., Alexander, M.A., Delworth, T.L., Hare, J.A., Harrison, M.J., Rosati, A., Vecchi, G.A., and Zhang, R. 2015. Enhanced warming of the Northwest Atlantic Ocean under climate change. J. Geophys. Res. Ocean. **121**: 118–132. doi:10.1002/2015JC011346.Received.

Sagarese, S.R., Frisk, M.G., Cerrato, R.M., Sosebee, K.A., Musick, J.A., and Rago, P.J. 2014. Application of generalized additive models to examine ontogenetic and seasonal distributions of spiny dogfish (Squalus acanthias) in the Northeast (US) shelf large marine ecosystem. Can. J. Fish. Aquat. Sci. **71**(6): 847–877. doi:10.1139/cjfas-2013-0342.

Sarro, C.L., and Stokesbury, K.D.E. 2009. Spatial and temporal variation in the shell height/meat weight relationship of the sea scallop Placopecten magellanicus in the Georges Bank fishery. J. Shellfish Res. **28**(3): 497–503. doi:10.2983/035.028.0311.

Schick, D., Shumway, S., and Hunter, M. 1988. A comparison of growth rate between shallow water and deep water populations of scallops, Placopecten magellanicus(Gmelin, 1791), in the Gulf of Maine. Am. Malacol. Bull. **6**(1): 1–8. Available from http://www.csa.com/partners/viewrecord.php?requester=gs&collection=ENV&recid=1901053.

Schick, D.F., and Feindel, S.C. 2005. Maine scallop fishery: Monitoring and enhancement. Final report to the Northeast Consortium.

Serchuk, F.M. 1984. Results of the 1983 USA Sea Scallop research survey: distribution and abundance of Sea Scallops in the Georges Bank, Mid-Atlantic and Gulf of Maine Regions and biological characteristics of Iceland Scallops off the coast of Massachusetts.

Serchuk, F.M., and Wigley, S.E. 1984. Results of the 1984 USA sea scallop research vessel survey: status of sea scallop resources in the Georges Bank, Mid-Atlantic and Gulf of Maine regions and abundance and distribution of Iceland scallops off the southeastern coast of Cape Cod. Woods Hole, MA.

Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fish. Res. **93**(1–2): 154–162. doi:10.1016/j.fishres.2008.03.006.

Shumway, S.E., and Parsons, G.J. 2006. Scallops: Biology, Ecology and Aquaculture. Elsevier. doi:10.1007/s13398-014-0173-7.2.

Sinclair, S.J., White, M.D., and Newell, G.R. 2010. How useful are species distribution models for managing biodiversity under future climates? Ecol. Soc. **15**(1): 8.

Slacum, H.W., Burton, W.H., Methratta, E.T., Weber, E.D., Llansó, R.J., and Dew-Baxter, J. 2010. Assemblage Structure in Shoal and Flat-Bottom Habitats on the Inner Continental Shelf of the Middle Atlantic Bight, USA. Mar. Coast. Fish. **2**(1): 277–298. doi:10.1577/C09-012.1.

Smith, P.A. 1994. Autocorrelation in Logistic Regression Modelling of Species' Distributions.

Glob. Ecol. Biogeogr. Lett. **4**(2): 47–61.

Smith, S.J., and Gavaris, S. 1993. Improving the precision of abundance estimates of eastern Scotian Shelf Atlantic cod from bottom trawl surveys. North Am. J. Fish. Manag. **13**(1): 35–47.

Smith, S.J., and Lundy, M.J. 2006. Improving the precision of design-based scallop drag surveys using adaptive allocation methods. Can. J. Fish. Aquat. Sci. **63**(7): 1639–1646. doi:10.1139/f06-063.

Soberon, J., and Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers. Informatics **2**: 1–10. doi:10.1093/wber/lhm022.

Spencer, F. 1974. Final report: offshore scallop survey – Cape Ann, Massachusetts to Maine-Canada border. Augusta, ME.

Stewart, P., and Arnold, S. 1994. Environmental requirements of the sea scallop (Placopecten magellanicus) in eastern Canada and its response to human impacts. Can. Tech. Rep. Fish. Aquat. Sci. **2005**: 1–36.

Stokesbury, K.D.E., Carey, J.D., Harris, B.P., and O'Keefe, C.E. 2010. High densities of juvenile sea scallop (Placopecten magellanicus) on banks ad ledges in the central Gulf of Maine. J. Shellfish Res. **29**(2): 369–372. doi:10.2983/035.029.0212.

Stokesbury, K.D.E., and Himmelman, J.H. 1995. Biological and physical variables associated with aggregations of the giant scallop Placopecten magellanicus. Can. J. Fish. Aquat. Sci. **52**(4): 743–753. doi:10.1139/f95-074.

Store, R., and Kangas, J. 2001. Integrating spatial mulit-criteria evaluation and expert knowledge of GIS-based habitat suitability modeling. Landsc. Urban Plan. **55**: 79–93.

Sumaila, U.R., Cheung, W.W.L., Lam, V.W.Y., Pauly, D., and Herrick, S. 2011. Climate change impacts on the biophysics and economics of world fisheries. Nat. Clim. Chang. **1**(9): 449–456. Nature Publishing Group. doi:10.1038/nclimate1301.

Sutcliffe, W.H., Loucks, R.H., and Drinkwater, K.F. 1976. Coastal circulation and physical oceanography of the Scotian Shelf and the Gulf of Maine. J. Fish. Res. Board Can. **33**: 98–115. doi:10.1139/f76-012.

Tanaka, K., and Chen, Y. 2015. Spatiotemporal variability of suitable habitat for American lobster (Homarus Americanus) in Long Island Sound. J. Shellfish Res. **34**(2): 531–543. doi:10.2983/035.034.0238.

Tanaka, K., and Chen, Y. 2016. Modeling spatiotemporal variability of the bioclimate envelope of Homarus americanus in the coastal waters of Maine and New Hampshire. Fish. Res. **177**: 137–152. Elsevier B.V. doi:10.1016/j.fishres.2016.01.010.

Tanaka, K.R., Chang, J.-H., Xue, Y., Li, Z., Jacobson, L., and Chen, Y. 2018. Mesoscale climatic impacts on the distribution of Homarus americanus in the US inshore Gulf of Maine. Can. J. Fish. Aquat. Sci.

Thouzeau, G., Robert, G., and Smith, S.J. 1991. Spatial variability in distribution and growth of juvenile and adult sea scallops Placopecten magellanicus (Gmelin) on eastern Georges Bank (northwest Atlantic). Mar. Ecol. Prog. Ser. **74**: 205–218. doi:10.3354/meps074205.

Thuiller, A.W., Georges, D., Engler, R., and Breiner, F. 2016. Biomod2: Ensemble platform for species distribution modeling.

Tompkins, E.L., and Adger, W.N. 2004. Does adaptive management of natural resources enhance resilience to Ccimate change? Ecol. Soc. **9**(2): 10.

Torre, M.P., and Chen, Y. (n.d.). Evaluation and optimization of fisheries survey design using computer simulation. Can. J. Fish. Aquat. Sci.

Torre, M.P., Tanaka, K.R., and Chen, Y. 2018. A spatiotemporal evaluation of Atlantic sea scallop Placopecten magellanicus habitat in the Gulf of Maine using a Bioclimate Envelope Model. Mar. Coast. Fish. **10**(2): 224–235. doi:10.1002/mcf2.10022.

Townsend, D.W., Thomas, A.C., Mayer, L.M., Thomas, M.A., and Quinlan, J.A. 2004. Oceanography of the Northwest Atlantic Shelf (1, W). *In* The Sea: The Global Coastal Ocean: Interdisciplinary Regional Studies and Syntheses. Harvard University Press. pp. 1–57.

Tremblay, M.J. 1997. Snow crab (Chionoecetes opilio) distribution limits and abundance trends on the Scotia Shelf. J. Northw. Atl. Fish. Sci. **21**: 7–21. doi:10.2960/J.v21.a1.

Tremblay, M.J., and Sinclair, M. 1991. Inshore-offshore difference in the distribution of sea scallop larvae: implications for recruitment. ICES mar. sci. symp. **192**: 39.

Truesdell, S.B. 2014. Distribution, population dynmics and stock assessment for the Atlantic sea scallop (Placopecten magellanicus) in the northeast US. University of Maine.

Wang, Y.G., Ye, Y., and Milton, D.A. 2009. Efficient designs for sampling and subsampling in fisheries research based on ranked sets. ICES J. Mar. Sci. **66**: 928–934. doi:10.1093/icesjms/fsp112.

Wildish, D., and Kristmason, D. 1993. /. *In* Bivalve Filter Feeders in Estuarine and Coastal Ecosystem Processes. NATO ASI Series.

Wildish, D.J., and Saulnier, A.M. 1992. The effect of velocity and flow direction on the growth of juvenile and adult giant scallops. J. Exp. Mar. Bio. Ecol. **133**(1): 133–143. doi:10.1016/0022-0981(92)90032-6.

Willis, K.J., and Whittaker, R.J. 2002. Species Diversity -- Scale Matters. Science (80-.). **295**(February): 1245–1247. Available from www.sciencemag.org/cgi/content/full/295/5558/1245.

Wood, S.N. 2003. Thin plate regression splines. J. R. Stat. Soc. B **65**(1): 95–114. doi:10.1111/1467-9868.00374.

Wood, S.N. 2006. Generalized additive models: an introduction with R. Chapman & Hall, London.

Wood, S.N. 2011. Fast stable REML and ML estimation of semiparametric GLMs. J. R. Stat. Soc. Ser. B (Statistical Methodol. **73**(1): 3–36. doi:10.1111/j.1467-9868.2010.00749.x.

Xue, Y., Guan, L., Tanaka, K., Li, Z., Chen, Y., and Ren, Y. 2017. Evaluating effects of rescaling and weighting data on habitat suitability modeling. Fish. Res. **188**: 84–94. Elsevier B.V. doi:10.1016/j.fishres.2016.12.001.

Young, M., and Carr, M.H. 2015. Application of species distribution models to explain and predict the distribution, abundance and assemblage structure of nearshore temperate reef fishes. Divers. Distrib. **21**: 1428–1440. doi:10.1111/ddi.12378.

Zuur, A.F., Ieno, E.N., and Smith, G.M. 2007. Analysing Ecological Data. Springer Science & Business Media.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. 2009. Mixed effects models and extensions in ecology with R. Springer, New York.

BIOGRAPHY OF THE AUTHOR

Michael P. Torre was born in Camden, New Jersey on June 19, 1988. He was raised in Medford, New Jersey and graduated from Shawnee High School in 2006. He attended Elizabethtown College and graduated in 2010 with a bachelor's degree in biology. Michael worked for New Jersey Fish and Wildlife during 2010 -2011. He then went to Lewes, Delaware and entered a graduate program at the University of Delaware. He earned his master's degree in marine science in 2014. He is a candidate for the Doctor of Philosophy degree in Ecology and Environmental Sciences from the University of Maine in May 2019.