University of Massachusetts Dartmouth

School for Marine Science and Technology

MISSPECIFYING THE MAGURO:

EVALUATING STOCK ASSESSMENT PERFORMANCE ON MIXED STOCKS OF ATLANTIC BLUEFIN TUNA

A Thesis in

Marine Science and Technology

by

Molly Randall Morse

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

August 2018

We approve the thesis of Molly Randall Morse

Date of Signature

Steven X. Cadrin
Professor and Chairperson, Department of Fisheries Oceanography
Thesis Advisor

Gavin Fay
Assistant Professor, Department of Fisheries Oceanography
Thesis Committee

Lisa A. Kerr
Research Scientist, Gulf of Maine Research Institute
Portland, ME
Thesis Committee

Miles A. Sundermeyer
Graduate Program Director, School for Marine Science and Technology

Steven Lohrenz
Dean, School for Marine Science and Technology

Tesfay Meressi
Associate Provost for Graduate Studies

Abstract

MISSPECIFYING THE MAGURO: EVALUATING STOCK ASSESSMENT PERFORMANCE ON MIXED STOCKS OF ATLANTIC BLUEFIN TUNA

by Molly Randall Morse

Accounting for patterns of movement and mixing in stock assessment of highly migratory species is important for managing these fisheries sustainably. Despite the growing wealth of information characterizing mixing between the western and eastern populations of Atlantic bluefin tuna (*Thunnus thynnus*), the stock assessment methods on which their management is based do not explicitly account for this critical aspect of stock structure. Simulation was used to test the performance of a virtual population analysis estimation model for estimating Atlantic bluefin tuna population abundance using pseudodata from an operating model incorporating movement and mixing. Model misspecification caused estimation models to frequently produce biased estimates of recruitment and spawning stock biomass. Western recruitment was significantly overestimated (~200% positive bias) but eastern recruitment was underestimated (~30% negative bias). Similarly, spawning stock biomass was underestimated for the eastern population (~70% negative bias) but overestimated for the western population (~100% positive bias). These biases appear to result from the model's inability to capture a net subsidy of the eastern population into western stock areas and fisheries. Estimation models applied to alternative operating model scenarios that modeled different potential recruitment trajectories and maturity assumptions behaved similarly. Models were better able to predict the size of mixed-population stocks than populations, suggesting that model predictions may be more effective for informing short-term trends in the resources available to fisheries than for implementing management decisions required for conservation of populations. The results suggest that stock mixing should be more explicitly considered in stock assessment of Atlantic bluefin tuna, and underscore the importance of testing stock assessment models and communicating their biases and limitations to managers.

Acknowledgments

Completion of this thesis would not have been possible without the guidance and support of my mentors, colleagues, classmates, family, and friends. First, I would like to thank my advisor, Steve Cadrin, for giving me the opportunity to work with him and for helping me fulfill my dream of attending an ICCAT meeting (not to mention presenting at it—twice!). He was a greater source of support and mentorship than I imagined from a graduate advisor, and I'm grateful he believed in me enough to entrust me with this project. I would like to thank my committee member Lisa Kerr, for helping me adapt the operating model to my research focus, and for being a role model of a successful woman in the marine science field. I am grateful for the guidance and expertise of my committee member Gavin Fay, who guided me through complicated nuances of simulation and stock assessment modeling. I also want to acknowledge my SMAST professors, whose teaching built a strong foundation for this research, and my SMAST peers, whose encouragement, advice, and knowledge were invaluable.

The many scientists of the ICCAT Bluefin Tuna Assessment group contributed to and provided feedback on this research, especially Clay Porch, Matt Lauretta, John Walter, Tristan Rouyer, Ai Kimoto, J.-J. Maguire, Ben Galuardi, and Dave Secor. I thank the NOAA Bluefin Tuna Research Program[1] for funding this project, and the ICCAT

[1] Grant awards NA11NMF4720108, NA13NMF4720059, NA13NMF4720061, NA14NMF4720085, NA15NMF4720108, and NA16NMF4720101. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author and do not necessarily reflect those of NOAA or the Department of Commerce.

Atlantic-Wide Research Program for Bluefin Tuna (GBYP) for supporting my training in the VPA-2BOX model.

My husband Ben Morse was an unwavering source of love and support to me, and selflessly endured the emotional swings that come with supporting a spouse in graduate school and the research updates replete with fisheries and modeling jargon. My parents Rich and Beth Pugh provided constant support and encouragement, and my brothers Brian and Jack gave programming tips and provided comic relief. I want to thank the family, friends, and others (especially Edward) who encouraged, challenged, and prayed for me, and indulged me by asking questions about tuna over the past two years. Finally, I thank God for giving me the energy, purpose, and grace to keep at it and see this through to fruition.

Table of Contents

List of Tables

List of Figures

Abbreviations

ASAP	age-structured assessment program
CV	coefficient of variation
F	fishing mortality rate
FRAER1	French aerial survey index 1
FRAER2	French aerial survey index 2
ICCAT	International Commission for the Conservation of Atlantic Tunas
IQR	interquartile range
MSE	management strategy evaluation
OSSRS	Optimized Step-Size Random Search
SS	Stock Synthesis
SSB	spawning stock biomass
VPA	virtual population analysis

Chapter 1: Introduction

Complexity in the spatial distribution and movement patterns of exploited fish populations poses challenges to stock assessment modeling and resource management (Cadrin and Secor 2009). These challenges become particularly salient for high-value migratory species targeted by several competing nations. The stakes are raised even higher when there is controversy over the robustness of stock assessment results and management decisions (Collette 2017). Understanding the patterns of movement and mixing for such highly migratory species and accounting for them in stock assessment and management is an important step in ensuring the sustainability of these fisheries.

1.1 Atlantic bluefin tuna life history

Atlantic bluefin tuna (*Thunnus thynnus*) is a highly migratory species of the family Scombridae that is native to the North Atlantic Ocean and Mediterranean Sea (Fromentin and Powers 2005). One of the world's most sought-after marine fish, the bluefin tuna—including Atlantic, Pacific, and Southern Ocean species—is known for its impressive size and prized flesh (Collette et al. 2011). The Atlantic species consists of western and eastern populations, which are known to spawn in the Gulf of Mexico and Mediterranean Sea, respectively, and exhibit spawning-site fidelity by returning to their own birthing areas to spawn, resulting in genetic isolation of the two populations (NRC 1994, Nemerson et al. 2000, ICCAT 2002, Block et al. 2005, Boustany et al. 2008, Rooker et al. 2014). Recent research also suggests even more complex population structure (Galuardi et al. 2010), and that there may be an additional western spawning area in the Slope Sea (Richardson et al. 2016) and separate eastern and western Mediterranean spawning populations (Carlsson et al. 2004, Boustany et al. 2008).

The western and eastern populations of Atlantic bluefin tuna mix extensively in the North Atlantic Ocean after spawning, but this mixing varies across space, time, and demographic groups (NRC 1994, Mather et al. 1995, ICCAT 2002). Rooker et al. (2008a, 2008b, 2014) used stable isotope analysis of otoliths to demonstrate significant and size-dependent trans-Atlantic migrations and natal homing of bluefin tuna. Siskey et al. (2016) used similar techniques to show fluctuating levels of stock mixing in the northwestern Atlantic over a 40-year time span (1974-2014). Block et al. (2005) used electronic tags to demonstrate trans-Atlantic movements of fish and spawning-site fidelity, and Walli et al. (2009) used archival tags to demonstrate ontogenetic changes in trans-Atlantic movement patterns. A comprehensive review of stock composition information from otolith chemistry, tagging, and genetics has suggested that stock composition varies inter-annually and regionally (Morse et al. 2018).

1.2 Stock mixing and management

Despite the growing wealth of information on the nature of mixing between the two populations, the stock assessment methods on which Atlantic bluefin tuna management is based do not explicitly account for this important aspect of stock structure. The International Commission for the Conservation of Atlantic Tunas (ICCAT)—the regional fishery management organization founded in 1966 to manage tuna and tuna-like species in the Atlantic Ocean—is responsible for conducting stock assessments for Atlantic bluefin tuna. In the 1970s and early 1980s, ICCAT managed

Atlantic bluefin tuna as a single North Atlantic stock, despite early studies that suggested separate western and eastern stocks with “a small and variable interchange of fish between them” (ICCAT 1980). Further evidence for separate western and eastern populations convinced ICCAT to shift to bluefin management according to a two-stock approach, dividing the stocks at the 45°W meridian in the North Atlantic (**Figure 1**), beginning in the early 1980s and continuing today (ICCAT 1981).

The separate western and eastern Atlantic bluefin tuna stock assessments and fishery management measures assume closed populations (i.e., “unit stocks”; Secor 2014) and do not account for mixing between separate western and eastern populations. Since 1982, western and eastern Atlantic bluefin tuna stock assessments have been conducted using separate virtual population analysis (VPA) models. In 2017, Stock Synthesis (SS; Methot and Wetzel 2013) was used in conjunction with VPA for the western stock assessment. Neither the VPA nor the SS application considers stock mixing. Alternative exploratory models that incorporated information on mixing were presented at the 2017 ICCAT bluefin tuna stock assessment session, but the data required for these models was deemed insufficient for their use as the primary basis for management advice (ICCAT 2017, Morse et al. 2018).

When mixing between distinct populations occurs, the relative contribution of each population within a specific area should be accounted for in management advice and stock assessment (Reiss et al. 2009). Model misspecification—in which the population structure assumed by the model does not match the true population characteristics—may lead to biased estimates of stock size (Ying et al. 2011, Kerr and Goethel 2014, Kerr et al. 2015). In turn, model misspecification has the potential to lead to inaccurate stock assessment and ineffective management measures that lead to overexploitation, overfishing of more vulnerable or less productive stocks, or failure to rebuild overfished stocks (Smedbol and Stephenson 2001, Fu and Fanning 2004, Kerr et al. 2010, Lorenzen et al. 2010, Goethel et al. 2011, Ying et al. 2011, Kerr et al. 2016b, Li et al. 2018). However, in some cases, a stock assessment model without spatial structure may perform well in estimating abundance and fishing pressure in the presence of spatial population structure (Szuwalski and Punt 2015), and the degree of the appropriate spatial scale of management will depend on species or population characteristics, such as the level of exchange among sub-populations (Smedbol and Stephenson 2001).

Concerns about accurately characterizing stock structure and movement are especially relevant for Atlantic bluefin tuna, which has been beset by scientific and political controversy and outcry from conservation organizations that the health of the species has been jeopardized by poor management and overexploitation (Collette 2017). If movement of Atlantic bluefin tuna is as complex as the data suggest, there may justifiable concern that simpler models may not capture the essential dynamics to produce accurate population estimates (Porch et al. 2001). Conservation efforts made at an inappropriate spatial scale can be ineffective and lead to overfishing of subpopulations (Kerr et al. 2010).

1.3 Modeling stock mixing

Since the 1990s, many efforts have been made to incorporate stock structure and mixing into Atlantic bluefin tuna stock assessments (NRC 1994). Butterworth and Punt (1994) investigated the effects of mixing between the western and eastern populations on estimates of abundance trends from the separate western and eastern stock VPAs, and found that relatively low mixing rates could produce significantly different results. Porch et al. (1995) built on this study by modeling Atlantic bluefin tuna using a two-area VPA that incorporated movement and tagging data to estimate mixing rates between western and eastern populations, and found that predicted abundance was sensitive to the rate of mixing, and the model fit best when mixing rates were low. However, the authors identified limited data quality and quantity, rather than model misspecification, as the major limiting factors to modeling mixing (Porch 1995, Porch et al. 1995).

Porch et al. (1998) used an advection-diffusion equation to simulate data for mixed Atlantic bluefin tuna stocks to test the performance of the two-area VPA relative to the separate western and eastern VPAs. The diffusion model used Beverton and Holt's (1957) box-transfer model and assumed that fish from one area move to another area and take on the behavior of resident fish:

$$N_{k,t+1} = \sum_{j} N_{j,t} T_{j \rightarrow k,t} e^{-(F_{k,t}+M)}$$

where abundance N is modeled over time t for fish in each area k as a function of the transfer rate T to area k for both areas j, M is natural mortality, and F is fishing mortality in each area (Butterworth and Punt 1994, NRC 1994, Porch 1995). Porch et al. (1998) found that the diffusion mixing model and separate western and eastern VPA approaches performed equally well when mixing was moderate, but the diffusion model performed better when mixing was greater.

As more information was developed on the nature of Atlantic bluefin tuna movement patterns, it was determined that the advection-diffusion and box-transfer model misrepresented the behavior of Atlantic bluefin tuna, which migrate to feed but return to natal grounds to spawn (ICCAT 2013). This behavior of sympatry with philopatry is more accurately described by an overlap model:

$$N_{s,t+1} = N_{s,t} \sum_{k} \tilde{T}_{s,k,t} e^{-(F_{k,t}+M)}$$

where $\tilde{T}$ is the fraction of one stock s that resides in area k at time t (Porch 2003, Cadrin and Secor 2009, Goethel et al. 2011). The results of stock assessments assuming the overlap model of fish movement were less sensitive to mixing (Porch 2005, ICCAT 2013).

Because analyses demonstrated that diffusion-based models do not consistently perform better than those without movement (Porch et al. 1998), stock assessment results were not sensitive to overlap-based models of mixing (Porch 2005, ICCAT 2013), and data related to mixing have been limited, ICCAT has continued with status quo management based on separate assessments under the assumption that mixing between the western and eastern populations is negligible (Porch et al. 2001, Kerr et al. 2016b, ICCAT 2017). However, additional data on the extent of mixing (Carlsson et al. 2004, 2007; Block et al. 2005; Boustany et al. 2008; Walli et al. 2009; Secor et al. 2012, 2014; Rooker et al. 2014; Siskey et al. 2016) and advances in model development and simulation testing justify continued evaluation of the performance of current Atlantic bluefin tuna stock assessment models with respect to population mixing.

1.4 Simulation testing stock assessment models

Simulation testing can be used to evaluate the performance of stock assessment models, particularly if there are concerns about model misspecification (NRC 1998, Deroba et al. 2015). As Hilborn and Walters (1992) observed, "model misspecification is unavoidable, and the question is how do we assess the possible implications of this misspecification on the reliability of parameter estimation." Simulation is valuable in that it can be used to mimic complex natural systems and can be manipulated in ways that are not possible or too costly in nature (Peck 2004). Simulation is also important for the development of alternative plausible operating models to account for uncertainty in our understanding of the natural system (Kerr et al. 2013, Goethel et al. 2016).

Simulation testing typically involves an operating model, an estimation model, and a method for comparing their outcomes (Hilborn and Walters 1992). The operating model includes realistic biological dynamics, process error, and observation error and reflects how modelers believe the real-world system behaves. An estimation model may use maximum likelihood, Bayesian, or least-squares to estimate model parameters by fitting to data and is generally simpler than the operating model. The modeler generates a large number of data sets (called *pseudodata*) from the operating model and then uses the estimation model to estimate parameters for each data set. From these data sets, the modeler can evaluate estimation model performance based on bias, confidence, and parameter confounding.

Simulation is a useful tool for assessing and comparing the performance of stock assessment models on fisheries data sets. Deroba et al. (2015) used simulation to evaluate the performance of stock assessment models for several fish stocks and assess the variability within models versus between models to characterize uncertainty in population estimates. The authors found that models were consistent in estimating trends but not the scale of absolute biomass. They also concluded that divergence in “self-tests” (consistency checks within assessment models) and “cross-tests” (checks between different assessment models) was common. Lack of robustness in cross-tests was indicative of differences in operating and estimation models, and could be used to test the performance of estimation models on pseudodata generated by operating models with different structural assumptions.

Simulation modeling has also been used to evaluate the consequences of spatial structure and connectivity among fish stocks on stock assessment outcomes (e.g., Kerr et al. 2010, Fay et al. 2011, Szuwalski and Punt 2015). Goethel et al. (2015) showed that a spatially-explicit tag-integrated model and a closed population model performed equally well in assessing a simulated yellowtail flounder-like fishery that comprised three flounder stocks, except when a long time series of reliable tagging data was available. Guan et al. (2013) evaluated the consequences of misspecified spatial structure and migration in the assessment of Atlantic herring, and concluded that estimation bias can result when spatial structure in population and fishing effort are misspecified. Fu and Fanning (2004) demonstrated that misspecified spatial structure might have contributed to the collapse of Atlantic cod stocks off Nova Scotia, and recommended monitoring of migration and stock dynamics and making adjustments to management when multiple sub-stocks are managed as a unit. Their findings indicated that the smaller of two co-managed stocks may be more likely to experience overfishing, and that managing sub-stocks separately may help prevent collapse of the smaller and more vulnerable stock.

Kerr and Goethel (2014) and Goethel et al. (2016) developed a comprehensive approach for constructing spatially-explicit operating models to test and validate stock assessment models. First, a spatial assessment model is fit to observed data to obtain best fit parameter estimates. Second, these parameter estimates are used to develop a spatially-explicit operating model. Third, the simulation is run to emulate various population scenarios and the performance of stock assessment models and management procedures is evaluated. This approach was used to demonstrate how simulation modeling was useful for testing alternative stock structure hypotheses and for examining the ecological, stock assessment, and fisheries management consequences of stock structure (Kerr and Goethel 2014).

1.5 Research objectives

Atlantic bluefin tuna are known to have complex spatial structure and intermixing western and eastern populations, which is not considered by the current stock assessment methods. In light of this mismatch, this research sought to answer the question, *To what extent might model misspecification of spatial structure and mixing dynamics affect stock assessment results of Atlantic bluefin tuna?*

Simulation was used to test the performance of VPA stock assessment models on simulated mixed-population western and eastern Atlantic bluefin tuna stocks. An operating model accounting for spatial structure and stock mixing (Kerr et al. 2016a,

2018) was used to generate pseudodata reflective of the typical quantity and quality of data available for stock assessment of Atlantic bluefin tuna. An additional alternative operating model scenario that considered uncertainties in life history parameters—spawning potential (i.e., maturity) and the stock-recruitment relationship—was designed to allow for simulation testing on different population trajectories. The estimation model, VPA-2BOX, was applied separately to pseudodata for western and eastern stocks, and model performance was analyzed by comparing the relative bias between "true" operating model values and estimation model estimates of population abundance and mixed-population stock abundance. Analysis of these metrics revealed how stock mixing might influence the quality of stock assessment results for Atlantic bluefin tuna.

Chapter 2: Methods

The simulation testing framework employed in this study was informed by Hilborn and Walters (1992), Kerr and Goethel (2014), and previous simulation studies (e.g., Deroba et al. 2015, Goethel et al. 2015). It is summarized in **Figure 2**.

2.1 Base case operating model structure

The operating model in this study was a revised version of the model developed by Kerr et al. (2016a, 2018), parameterized with updated data from the 2017 ICCAT stock assessments of western and eastern Atlantic bluefin tuna. The operating model was coded in R (version 3.3.2; R Core Team 2016). The model simulates two spawning populations, with the western population originating in the Gulf of Mexico and the eastern population originating in the Mediterranean Sea. The operating model is age-structured (ages 1-29) and simulates movement of fish across seven geographic zones (**Figure 1**) and over four seasonal quarters. Bluefin tuna from one area move to another, but spawn only in their own natal area, according to the overlap model structure.

Fish movement was conditioned on fishery-independent telemetry-based movement rates developed by Galuardi et al. (2018). The Multistock Age-Structured Tag-integrated assessment model (MAST; Taylor et al. 2011) that was used to model movement in the preceding Kerr et al. (2016a) study was unstable for the younger maturity scenarios examined (Kerr et al. 2018). Galuardi et al. (2018) developed a simulation framework (SatTagSim) to analyze data from 499 tagged Atlantic bluefin tuna to derive movement probability matrices for the seven-zone model over four seasons.

Based on where fish were initially tagged, they were split into western and eastern area releases (as defined by the ICCAT stock boundary at 45°W), and western fish were split into size categories (<185 cm and ≥185cm, based on the commercial size limit in the United States).

The base case operating model was initialized using estimates of abundance-at-age for the first year (1974) and conditioned on age 1 recruitment for all years (1974 to 2015) from ICCAT (2017) VPA results for Atlantic bluefin tuna. Although the VPA estimates of abundance represent western and eastern stock sizes, these values were used for the initial western and eastern population sizes in the simulation. An important implication of this assumption is that the simulated western and eastern populations reflect current perceptions of stock development that are based on the available data from mixed stocks. Fishing mortality rates-at-age from the ICCAT (2017) VPA were also used to condition the model to reflect the best understanding of exploitation history. Fishing mortality rates, which are estimated stock-wide by the VPA, were partitioned into the seven geographic zones and four seasonal quarters using the stock-specific fishing mortality estimates and fleet-specific relative age composition data from ICCAT (2017). Previous work showed that ICCAT (2017) stock assessment estimates of recruitment and fishing mortality rates were relatively insensitive to stock mixing (Morse et al. 2018).

Of the two spawning fraction scenarios considered by ICCAT (2017) for the western stock, the base case operating model assumed the older (i.e., ~50% of fish spawn at age 10; **Table 1**, **Figure 3**). Similar to a conventional maturity-at-age vector, the spawning fraction is “the proportion of fish contributing to the spawning output of the population as a function of age,” which allows for the possibility of older mature fish producing more offspring than younger mature fish (Porch and Hanke 2018). Fishery selectivity and catchability of all fleets were assumed to be constant with time. These and other operating model specifications are summarized in **Table 1**.

The operating model was simulated over the 42-year time span (1974 to 2015) common to both western and eastern 2017 ICCAT stock assessments to produce quarterly values for abundance-at-age, spawning stock biomass (SSB, in metric tonnes), and fishing mortality-at-age. Because operating model outputs were disaggregated by geographic zone, season, and population-of-origin, they were re-aggregated into annual time steps and stock areas for deriving input pseudodata for estimation models (see section 2.3). Operating model outputs for the bluefin tuna resource size were presented both as *stock view*, referring to the separate geographically-distinct western and eastern mixed-population stocks separated by the 45°W meridian as defined by ICCAT (1981), and as *population view*, referring to the separate genetically-distinct western and eastern populations originating in their respective natal grounds (Kerr et al. 2018). Stock view attributes were derived by summing the abundance or biomass of fish over all geographic zones contained in a stock area (zones 1-3 for the western stock area, 4-7 for the eastern stock area), and population view attributes were derived by summing over all fish originating in the respective spawning areas (Gulf of Mexico for the western population, Mediterranean Sea for the eastern population).

2.2 Alternative operating model scenario

Defining stock-recruitment relationships for western and eastern Atlantic bluefin tuna stocks has challenged scientists for several years (Porch and Lauretta 2016). Before

2017, scientists used different scenarios of stock-recruitment for projections for determining stock status reference points: high-recruitment potential (Beverton-Holt model) and low-recruitment potential (hockey stick model) scenarios in the west, and high-, medium-, and low-recruitment potential scenarios in the east (ICCAT 2014). In the most recent stock assessment, greater uncertainty in stock-recruitment relationships led to basing projections and reference points on estimates of average recruitment rather than on stock-recruitment models (ICCAT 2017).

Determining the age at which western Atlantic bluefin tuna mature and produce viable offspring—represented mathematically by the spawning fraction—has also proven difficult. ICCAT has considered two possible scenarios (**Table 1**, **Figure 3**): an older spawning fraction that assumes spawning contribution is a function of maturity and age, and is supported by evidence of only older fish observed spawning in the Gulf of Mexico (assumed in the base case operating model scenario); and a younger spawning fraction that assumes spawning contribution is a function of maturity alone and considers spawning by younger fish outside the Gulf of Mexico.

An alternative operating model was developed to simulate another "tuna-like" population development scenario, and to compare the robustness of estimation model performance on different representations of reality. The alternative scenario retained all the same characteristics of the base case scenario, except for two key differences: recruitment was calculated from the western and eastern hockey stick stock-recruitment models (**Eq. 1**; Kerr et al. 2016a) and maturity of the western population was based on the younger spawning fraction (**Table 1**, **Figure 3**). The hockey stick models represent low recruitment potential for the western population and medium recruitment potential for the eastern population. The hockey stick model (**Eq. 1**) assumes a linear increase in expected recruitment from the origin to a "hinge" level of SSB, SSB^*, above which expected recruitment is constant and independent of SSB (R_{max}; ICCAT 2014), where N is abundance denoted by subscripts y for year, a for age, z for geographic zone, q for seasonal quarter, and p for population. Recruitment stochasticity was incorporated as an error term and modeled as a random lognormal variate scaled to the approximate recruitment variability observed for each stock (Kerr et al. 2016a). The error term for the western population ($p = 2$) was $\varepsilon_{y,p=2} \sim Lognormal(0, \sigma^2_{p=2})$ where $\sigma^2_{p=2} = 0.37$, and the error term for the eastern population ($p = 1$) was $\varepsilon_{y,p=1} \sim Lognormal(0, \sigma^2_{p=1})$ where $\sigma^2_{p=1} = 0.43$.

$$N_{y,a=1,z,q=1,p} = \begin{cases} R_{max_p} \varepsilon_{y,p} & if\ SSB_{y,z,q=1,p} \geq SSB^*_p \\ \dfrac{R_{max_p}}{SSB^*_p} SSB_{y,z,q=1,p} \varepsilon_{y,p} & if\ SSB_{y,z,q=1,p} < SSB^*_p \end{cases} \quad \textbf{Eq. 1}$$

2.3 Observation model

To generate pseudodata of the typical quantity and quality available for the western and eastern Atlantic bluefin tuna stock assessments, an observation model was applied to the operating model outputs. The observation model generated pseudodata for stock-specific catch-at-age, indices of relative abundance, and relative age composition of indices. The geographic range, magnitude, and time frame of fishery-dependent and fishery-independent indices of relative abundance emulated those included in the 2017 ICCAT stock assessments.

Pseudodata were generated with random observation error (stochastic) and without (deterministic). For stochastic simulations, stock-specific catch-at-age ($C_{y,a,s}$, **Eq. 2**), indices of relative abundance ($I_{y,g}$, **Eq. 3**), and relative age composition of indices ($P_{y,a,g,s}$, **Eq. 4**) were generated with observation error ε,

$$C_{y,a,s} = \left(\sum_{\substack{West\ z=1:3 \\ East\ z=4:7}} \sum_{q=1}^{4} \sum_{p=1}^{2} N_{y,a,z,q,p} * \frac{F_{y,a,z,q}}{F_{y,a,z,q} + M_{a,q,p}} * \left[1 - e^{-(F_{y,a,z,q} + M_{a,q,p})}\right] \right) * e^{\varepsilon_{y,a,s}} \qquad \textbf{Eq. 2}$$

$$I_{y,g} = \left(\sum_{\substack{West\ z=1:3 \\ East\ z=4:7}} \sum_{a=1}^{29} \sum_{p=1}^{2} S_{a,g} N_{y,a,z,q,p} W_{a,p} Q_g \right) * e^{\varepsilon_{y,g}} \qquad \textbf{Eq. 3}$$

$$P_{y,a,g,s} = \left(\sum_{\substack{West\ z=1:3 \\ East\ z=4:7}} \sum_{q=1}^{4} \sum_{p=1}^{2} N_{y,a,z,q,p} * \frac{E_{y,g} Q_g S_{a,g}}{E_{y,g} Q_g S_{a,g} + M_{a,q,p}} * \left[1 - e^{-(E_{y,g} Q_g S_{a,g} + M_{a,q,p})}\right] \right) * e^{\varepsilon_{y,a,s}} \qquad \textbf{Eq. 4}$$

where N is abundance, F is fishing mortality rate, M is natural mortality rate, S is selectivity, W is weight, Q is catchability, and E is effort, which are disaggregated by year y, age a, geographic zone z, fleet g, seasonal quarter q, population p, and stock s (for more detail on derivation of equations, see Kerr et al. 2018).

Indices of relative abundance were derived from simulated fishing fleets and surveys that emulated the geographic scope and magnitude of index data used in the ICCAT (2017) stock assessment. Abundance values used in the calculation of indices (**Eq. 3**) were assumed to be from the beginning of the third quarter (reflecting the fall season when the majority of fishing effort occurs in the Atlantic bluefin tuna fishery), except for indices that measured relative abundance in spawning areas (zones 1 and 7), in which case the abundance was assumed to be from the beginning of the first quarter to reflect the spawning season. Pseudodata for the relative age composition of indices were derived from the relative F-at-age within fleets and years, where the index-specific fishing mortality rate $F_{y,a,g}$ is derived from the index-specific effort $E_{y,g}$, catchability Q_g, and selectivity $S_{a,g}$, which were input parameters to the operating model based on ICCAT (2017) VPA estimates. These age composition pseudodata are used by the VPA-2BOX estimation model to inform the age-selectivities of the corresponding indices of relative abundance (Porch 2003). **Appendix A** provides further details on the index relative age composition pseudodata.

Assuming catch and index data were lognormally-distributed, observation error was generated stochastically according to the normal distribution, $\varepsilon \sim N(0, \sigma^2)$. The observation error $\varepsilon_{y,a,s}$ for the catch-at-age of each stock s had a standard deviation σ_s calculated as the root mean square error over all ages and years (**Eq. 5**),

$$\sigma_s = \sqrt{\frac{1}{Y}\frac{1}{A}\sum_{y}^{Y}\sum_{a}^{A}[\ln x_i - \ln \hat{x}_i]^2} \qquad \textbf{Eq. 5}$$

where y and Y are the first and last years of the time series, respectively (e.g., $y = 1$ and $Y = 42$); a and A are the first age and plus group, respectively (e.g., $a = 1$ and $A = 16$); x is the observed value; $\hat{x}$ is the predicted value; and i denotes the individual observation. The observed and predicted catch-at-age values were derived from an exploratory age-structured assessment program (ASAP) analysis of Atlantic bluefin tuna data (Maguire et al. 2018).

The observation error $\varepsilon_{y,g}$ for each index of relative abundance g had a standard deviation σ_g calculated as the root mean square error over all years of the index time series (**Eq. 6**). The observed and predicted index values were derived from the results of the 2017 ICCAT bluefin tuna western and eastern stock assessments.

$$\sigma_g = \sqrt{\frac{1}{Y}\sum_{y}^{Y}[\ln x_i - \ln \hat{x}_i]^2} \qquad \textbf{Eq. 6}$$

The operating model produced age-based data for ages 1 to 29, but the western and eastern Atlantic bluefin tuna stock assessment models have age 16+ and age 10+ groups, respectively. So, catch-at-age data were summed across all ages from the plus group age to age 29, producing catch-at-age series for ages 1 to 16+ for the western stock and 1 to 10+ for the eastern stock.

2.4 Estimation model

The VPA-2BOX program (version 4.01, Porch et al. 2001) was used to perform the 2017 western and eastern Atlantic bluefin tuna stock assessments. The model has the capability to simultaneously analyze two intermixing populations, but has been principally used by ICCAT for single unit stock assessment of Atlantic bluefin tuna. VPA-2BOX is a calibrated VPA based on the ADAPT framework (Parrack 1986, Gavaris 1988), which is an age-structured model that estimates population abundance and mortality using backwards recursions to fit to catch, effort, and abundance data (Porch 2003). VPA-2BOX employs the Optimized Step-Size Random Search (OSSRS) algorithm to converge on the general vicinity of the global minimum (Sheela 1979), and then the Nelder-Mead simplex algorithm AMOEBA to find the values of the parameters that minimize the objective function (Press et al. 1992). The estimated parameters include terminal F (the fishing mortality rate on each age group—excluding the plus group—in the last year), F-ratio (the ratio of the fishing mortality on the plus group to the next younger age), variance scaling (weight applied to each index of relative abundance), and catchability (catch efficiency of each index of relative abundance).

Initial model settings were specified as in the ICCAT (2017) western and eastern Atlantic bluefin tuna stock assessments (**Table 1**). However, problems with model convergence for some pseudodata sets using the default ICCAT (2017) settings warranted the testing of alternative estimation model configurations (**Table 2**). Model convergence criteria were based on guidance from Zarrad et al. (2018) and Porch (2003) and included minimizing the objective function, first derivative test, parameter boundary constraints, and parameter correlation. Problems with convergence were largely based on estimation of terminal F parameters and variance scaling parameters for abundance indices. The terminal F parameters specify the fishing mortality rate (F) on each age group in the last year, and variance scaling parameters effectively weight the relative importance of each index of relative abundance. The majority of the western model runs based on the original ICCAT (2017) model settings did not converge. A revised configuration of the estimation model in which the bounds on all terminal F parameters were changed to allow lower values (**Table 2**) achieved a higher convergence rate and allowed for comparison of model estimates with the true operating model values. Some runs of the eastern model with original ICCAT (2017) model settings overfit to particular indices of relative abundance (the observed and predicted index values were equivalent for the French aerial survey indices 1 and 2, abbreviated as “FRAER1” and “FRAER2”) and estimated parameters poorly. In these cases, model fit was improved by revising the estimation model configuration by setting the variance scaling parameter for one index (FRAER2) equal to the value estimated for the other (FRAER1; **Table 2**). In total, two eastern and two western estimation model configurations were tested.

A few additional assumptions regarding the eastern estimation model were made. Relative age composition is used by VPA-2BOX to calculate selectivity (i.e., relative vulnerability-at-age) information for each index. Although the pseudodata for relative age composition were derived from the operating model using **Eq. 4**, the relative vulnerability for the western Mediterranean larval survey index was based on the spawning fraction vector in the same way as in ICCAT (2017). In addition, the VPA-2BOX input coefficients of variation (CVs) for each index were derived from ICCAT (2017) VPA input files as the average of all CVs for a given index across all years.

Five-hundred stochastic pseudodata sets from the operating model were run through the western and eastern VPA estimation models. For each realization, pseudodata were formatted using R into ASCII text input files required by VPA-2BOX program (T. Rouyer, Institut Françias de Recherché pour L’exploitation de la Mer, IFREMER, Sète, France, personal communication, 2017; see Porch 2003 for file formatting requirements). Similar to Deroba et al.’s (2015) “cross-tests,” estimation model results for stochastic runs of pseudodata with observation error were aggregated across realizations and compared for accuracy with the true values from the operating model to understand the effect of model misspecification on estimates of population size (**Figure 2**). Estimates from stochastic model runs (pseudodata with observation error) were compared to estimates from deterministic model runs (pseudodata without observation error) to understand estimation model sensitivity to observation error.

Relative error and absolute error of resource size estimates (θ_{est}, where θ may be SSB or recruitment) from cross-tests were calculated relative to the true population values (θ_{true}) of the operating model to reflect the closed-population assumptions of the VPA. Percent relative bias and absolute bias (**Eq. 7** and **8**) were calculated as the average relative and absolute error, respectively, of all 42 years (y) of all 500 simulated realizations (n). SSB estimates were also compared to the mixed-population stock view from the operating model. For these comparisons, operating model estimates of population and stock size from the third seasonal quarter were used to reflect the season (fall) when the largest proportion of fishing effort generally occurs in the actual Atlantic bluefin tuna fishery.

$$relative\ bias\ \% = \left(\frac{1}{500 * 42}\sum_{n=1}^{500}\sum_{y=1}^{42}\frac{\theta_{est_{n,y}} - \theta_{true_y}}{\theta_{true_y}}\right) * 100\% \qquad \textbf{Eq. 7}$$

$$absolute\ bias = \frac{1}{500 * 42}\sum_{n=1}^{500}\sum_{y=1}^{42}\theta_{est_{n,y}} - \theta_{true_y} \qquad \textbf{Eq. 8}$$

2.5 Self-test

To better understand the source of estimation bias in SSB, a self-test of the eastern estimation model was performed to complement the cross-tests of a spatially complex operating model and a spatially simple estimation model. The self-test was conducted to reveal how much of the estimation bias could be attributed to sources of bias in the VPA-2BOX model other than misspecification of spatial structure and fish movement. In a self-test, the operating model and estimation model have the same structural assumptions and model settings (Deroba et al. 2015), so the base case operating model in this study was simplified to eliminate spatial structure, seasonal structure, and fish movement. **Appendix B** provides further details on the methodology of the self-test. The results of the self-test were used to attribute estimation bias in eastern SSB to specific causes.

Chapter 3: Results

3.1 Operating model behavior

The base case operating model produced considerably different perceptions of western bluefin between the population and stock views, but perceptions of eastern bluefin were similar (**Figure 4**). Operating model time series of eastern population and stock SSB were similar in magnitude, but the western stock SSB was at least double the population SSB. For both the western and eastern populations and stocks, the operating model time series of SSB were larger in magnitude than ICCAT (2017) because of discrepancies in the realized fishing mortality rate produced by seasonal and spatial patterns in the operating model, but both population view trajectories reproduced the general trends in SSB from ICCAT (2017). SSB of the eastern population was consistently greater than the western population. Spawners from each population were more abundant within their respective stock areas, but crossed the stock management boundary in appreciable numbers (**Appendix C**). The base case operating model was conditioned on the ICCAT (2017) time series of recruitment estimates, so these recruitment time series were equivalent. Eastern recruitment (measured in the first quarter) was greater than western recruitment by approximately one order of magnitude throughout the time series (**Figure 4**).

Fishery yield was greatest in the Mediterranean Sea and in coastal areas of the eastern U.S. and western Europe and Africa, and many more tonnes of the eastern population were caught than the western population (**Appendix C**). Fish from both populations were caught by fisheries in both stock areas, and relatively more eastern population fish were caught in the western stock area than western population fish caught in the eastern stock area.

Although the operating model was conditioned on fishing mortality rates estimated by the ICCAT (2017) VPA, the realized fishing mortality rates in the operating model after disaggregation into the seven zones and re-aggregation by stock area were lower in magnitude. This was a result of deriving zone-based fishing mortality rates by partitioning them based on relative age composition data of indices from the ICCAT (2017) VPA model inputs. The interaction between the fishing mortality rates partitioned into the seven geographic zones and fish movement among these zones resulted in lower realized fishing mortality rates. Fishing mortality rates were generally greater but abundance and biomass were generally lower in the western stock area, especially the western Atlantic (geographic zone 3), relative to the eastern stock area. Operating model dynamics are depicted graphically and described in more detail in **Figure 4** and **Appendix C**.

The alternative operating model produced a different perception of western and eastern bluefin tuna population and stock sizes (**Figure 4**). The recruitment time series were generally of a lower magnitude and decreased over the entire time series. Values for SSB were generally lower than the base case and exhibited downward trends over the time series. These dynamics suggest that despite the younger spawning fraction for the western population, which should produce higher reproductive rates, the stock-recruitment relationships for both populations resulted in lower productivity than in the base case scenario. As in the base case operating model, western stock SSB in the alternative scenario was much higher in magnitude than population SSB.

3.2 Western estimation model performance

3.2.1 Model convergence

There was a wide range in convergence rates among the different western estimation models (**Table 3**). When the model settings were the same as ICCAT (2017), only 6% of runs converged with pseudodata from the base case operating model and 64% converged with pseudodata from the alternative operating model. The unconverged runs had parameter estimates that hit bounds, did not complete parameter estimation, or did not produce parameter correlation and covariance matrices. The model had the most difficultly in estimating terminal F and index variance scaling parameters. When the lower bound on terminal F parameters was decreased to allow for lower parameter estimates in the revised version of the estimation model, 86% of runs converged (**Table 3**).

None of the deterministic runs of the western estimation models converged, which was driven by model misspecification of the data structure. Index data that perfectly reflected the fleets and surveys in the spatially-complex operating model imposed some conflicting signals in the simpler estimation model, and convergence improved when observation error was added as in the stochastic runs.

3.2.2 Resource size estimates – Base case operating model

Results from the revised estimation model configuration instead of the default ICCAT (2017) configuration were used to assess the relative bias of estimates from the western estimation model because the revised model is similar to the estimation models applied by ICCAT (2017), only 6% of runs of the ICCAT (2017) configured model converged, and 86% of runs in the revised model converged. The revised estimation model produced estimates of recruitment and SSB that were within 22% of those from the ICCAT (2017) configured model. The revised estimation model replicated relative trends in recruitment from the operating model (i.e., increasing and decreasing trends at the same points in the time series), but overestimated the magnitude across the entire time series (**Figure 5**), producing estimates of recruitment positively biased by 192% on average and adding an average of about 280,000 age 1 fish each year over the entire time series (**Table 4**). Recruitment was poorly estimated in the last four years (2012 to 2015) for all estimation model configurations for both stocks, which is common in age-based stock assessments, so these estimates were excluded from the evaluation.

The revised estimation model produced estimates of SSB that were mostly between the true values for the western population and mixed-population stock SSB: SSB was overestimated by 107% (adding an average of approximately 25,000 tonnes each year over the entire time series) relative to the operating model population SSB but underestimated by -27% (missing an average of approximately 23,000 tonnes each year over the entire time series) relative to the stock SSB (**Table 4**, **Figure 6**). Average SSB was estimated to be greater than both population and stock SSB for the first four years of the time series. There was a large range (41% to 187%) in the percent relative bias in SSB estimates (**Figure 6**).

3.2.3 Resource size estimates – Alternative operating model

The percent relative bias in estimates of recruitment and population SSB obtained from fits to pseudodata from the alternative operating model was greater than from the base case operating model (**Table 4**). Recruitment was overestimated by 284% on average over the entire time series (**Table 4**, **Figure 5**). The percent relative bias in recruitment varied across the time series between 46% and 608% (**Figure 5**).

Population SSB was overestimated by 146% on average over the entire time series (**Table 4**), but estimates were between the true SSB values for the western population and mixed-population stock, with the exception of the first nine years (**Figure 6**). The range in the percent relative bias (129% to 170%) was less than that of the estimation model applied to the base case pseudodata (**Figure 6**).

3.3 Eastern estimation model performance

3.3.1 Model convergence

At least 95% of eastern estimation model runs converged for all operating model scenarios and estimation model configurations (**Table 3**). When model settings were the same as ICCAT (2017), 95% of all runs converged (**Table 3**), but the deterministic run failed to converge on realistic parameter estimates and overfit the French aerial survey index 2 (FRAER2). The predicted index values for the FRAER2 index were exactly the same as the observed values in the deterministic run. In addition, the variance scaling parameter for the FRAER2 index was estimated on the order of 10^{-15}, much lower than that of all the other indices from this estimation model and that of the FRAER2 variance scaling parameter from ICCAT (2017), which were on the order of 10^{-1} or 10^{-2}. By fixing the FRAER2 variance scaling parameter to the value estimated for the French aerial survey index 1 (FRAER1), model convergence improved marginally to 96% (**Table 3**).

Like the western estimation models, the deterministic runs of the eastern estimation model tended to behave the worst, overfitting the data more frequently and converging less often than the stochastic runs. The revised estimation model was the only configuration in which the deterministic run converged. The addition of observation error seemed to improve model convergence. These convergence problems were similar to those experienced by ICCAT for modeling the available eastern Atlantic bluefin tuna data (e.g., Zarrad et al. 2018).

3.3.2 Resource size estimates – Base case operating model

The converged runs of the estimation model with default ICCAT (2017) settings captured operating model relative trends but underestimated recruitment with an average bias of -33%, missing an average of approximately 800,000 age 1 fish each year over the entire time series (**Table 4**, **Figure 5**). The percent relative bias varied very little across the recruitment time series (**Figure 5**). Bias in recruitment was slightly worse in the revised estimation model (-36%; **Table 4**).

Both configurations of the eastern estimation model consistently underestimated SSB (**Table 4**). The percent relative bias of the ICCAT (2017) configured estimation model was -71% (missing an average of about 470,000 tonnes each year) from the population SSB and -69% (missing an average of about 420,000 tonnes each year) from the stock SSB over the entire time series (**Table 4**, **Figure 6**). The percent relative bias in population SSB was worse in the revised estimation model (-74%). The range in the annual percent relative bias was small across the time series (**Figure 6**).

3.3.3 Resource size estimates – Alternative operating model

The estimation model fitted to pseudodata from the alternative operating model performed similarly to the estimation model fitted to pseudodata from the base case operating model, and recruitment was estimated with an average relative bias of -29% over the entire time series (**Table 4**, **Figure 5**). Similar to the estimation model applied to pseudodata from the base operating model, the percent relative bias in recruitment varied little across the time series (**Figure 5**). Population SSB was underestimated by 72% (**Table 4**, **Figure 6**). The percent relative bias was relatively consistent like the base case (**Figure 6**).

3.3.4 Deterministic and stochastic simulations

There was a divergence in the estimates of eastern bluefin tuna stock size between the deterministic and stochastic runs. Of the converged runs of the revised estimation model (the only estimation model with a converged deterministic run), the average relative error in SSB of the deterministic run relative to the operating model population view was -39% over the entire time series, whereas the average relative bias of the stochastic runs was -74% (**Figure 7**). The difference between deterministic and stochastic runs was most evident in time series estimates of SSB but much smaller in estimates of recruitment. In abundance-at-age time series, the relative difference progressively increased from younger to older ages (**Figures 8** and **9**). Sensitivity runs showed that decreasing the standard deviation used in generating catch-at-age observation error from 0.86 to 0.3 could eliminate the divergence. Therefore, the effect is greatest at the oldest ages and appears to result from observation error in the catch-at-age.

Unlike the eastern estimation model, there was little to no difference between deterministic and stochastic estimates of SSB in the western estimation model (**Figure 7**). The only slight divergence between the deterministic and stochastic simulation runs was during the mid-1970s and between 2000 and 2015.

3.3.5 Self-test

The high degree of estimation bias in eastern SSB was unexpected because the population is considerably larger than the western population, and because the eastern estimation model has previously been shown to be less sensitive to structural error than the western model (Morse et al. 2018). Using the results of the self-test, estimation bias in eastern SSB (**Figure 6**) was attributed to specific causes. Although the relative bias in population SSB was -71% for stochastic runs of the cross-test with the estimation model configured as in ICCAT (2017; **Table 4**), it was only -39% for deterministic runs (**Figure 7**). As described in section 3.3.4 above, the 32% difference is attributable to the addition of observation error to the catch-at-age pseudodata. In comparison, the relative bias in population SSB for the self-test was only -30%, which suggests that a large portion of the total relative bias in SSB in the stochastic runs stems from other problems with the VPA-2BOX estimation model unrelated to model misspecification of spatial structure. The problems evidenced in this self-test include poor estimation of recruitment and terminal F parameters for younger age groups. In addition, VPA-2BOX has historically demonstrated instability when fit to the available eastern Atlantic bluefin tuna data in recent stock assessments (ICCAT 2014, 2017; Zarrad et al. 2018). For these reasons, bias in the eastern estimation model is likely due to these persistent problems in addition to stock mixing. **Appendix B** provides further details on the results of the self-test.

Chapter 4: Discussion

This study demonstrated how a simulation testing framework was successfully used to test the performance of stock assessment models when there is mixing among western and eastern populations of Atlantic bluefin tuna. The study sought to answer similar questions as those explored by ICCAT scientists in the 1990s and 2000s, but recently improved understanding of population biology, life history, and mixing rates allowed for advances in operating model parameterization. Developments in stock assessment simulation testing for fish stocks with complex spatial structure (Kerr and Goethel 2014) also provided a foundation upon which a spatially-simple stock assessment model could be tested using spatially-complex fishery pseudodata.

The results of simulation testing confirm previous observations that VPA models accurately reflect general abundance trends, but often produce biased magnitudes (Parrack 1986). It has been assumed that this bias results from indirect assignment of length to age for catch data and from model sensitivity to starting assumptions about the fishing mortality rate (Parrack 1986), but the results of this research suggest it can also be attributed to misspecification of spatial structure and stock mixing. The closed-population estimation models generally reflected the mixed-population stock and population trends in the operating model, but absolute estimates of population size were considerably biased.

Scientific advice based solely on these VPA estimates of Atlantic bluefin tuna stock size may provide misleading guidance to fishery managers and lead to ineffective management actions, particularly when population conservation is an objective. Because bias in estimates of the mixed-population stock size was relatively low, the current ICCAT approach to stock assessment and management of Atlantic bluefin tuna may inform trends in the stock resources that are available to fisheries. However, the current assessments may be misleading for avoiding overfishing or rebuilding populations, particularly for the smaller and more vulnerable western population, as evidenced by the failure of the VPA to accurately estimate recruitment and population SSB. Managers should be made aware of these and other model biases and limitations discovered through simulation testing (Deroba et al. 2015).

4.1 Operating models

Time series of SSB in both operating model scenarios showed that the perception of resource size depends on whether it is the population or stock unit being examined. The eastern stock view was only slightly lower in magnitude than the population view, but the western stock view was greater in magnitude than the population view by as much as 300% in some years. These results confirm previous observations that eastern-origin fish migrating to the western stock area and augmenting western catches provide an overly optimistic view of the potential productivity of western fish (Kerr et al. 2016a, 2018), and suggest that the western stock assessment model results might produce an inaccurate view of the western population leading to overfishing. Fish from the western population also migrated into the eastern stock area and were subject to eastern fisheries. These dynamics underscore the need for more accurate understanding and parameterization of Atlantic bluefin tuna movement and mixing dynamics in stock assessment, particularly for conservation of the smaller and more vulnerable western population.

The base case operating model was conditioned on the ICCAT (2017) stock assessment, so it was expected that the SSB time series would be similar. However, the differences in SSB trends and magnitudes were likely a direct result of the spatial structure present in the operating model and absent in the estimation model and of some simplifying assumptions in the operating model. For both the west and east, the base case operating model SSB was larger in magnitude than ICCAT (2017) because the effective fishing mortality rate (F) resulting from spatial allocation of F and spatial distribution of the population in the operating model was less than the aggregate F estimates from ICCAT (2017; **Appendix C**). Current efforts by ICCAT scientists to condition operating models for management strategy evaluation (MSE) of Atlantic bluefin tuna have had similar difficulties reflecting the ICCAT perception of stock development when incorporating stock mixing (Carruthers and Kell 2017; Carruthers and Butterworth 2018a,b).

Additional assumptions made in parameterization of the operating model may have contributed to differences between the operating model and the ICCAT (2017) stock assessment. The model inputs to ICCAT (2017) included time-varying weight-at-age for calculating SSB, but the operating model assumed a single vector for weight-at-age for the entire time series, which fell within the range of ICCAT (2017) weight-at-age values. The operating model was limited to the period 1974 to 2015—the only years for which data were available for the ICCAT (2017) western Atlantic bluefin tuna stock assessment—whereas the ICCAT (2017) eastern stock assessment included data beginning six years earlier in 1968. Despite the discrepancies in the times series of SSB between ICCAT (2017) and the base case operating model, the trends and magnitude were sufficiently similar to produce valuable comparisons for the goals of simulation testing.

The need to make certain operating model parameterization assumptions resulted in the use of parameter values that were not necessarily consistent with all available information. The operating model, which applies different biological parameter values to separate western and eastern populations (explicitly modeling population mixing), was conditioned on parameters estimated by the ICCAT (2017) stock assessments, which assume closed populations (implicitly assuming no population mixing). For example, parameterization of the stock-recruitment relationships used in the operating model relied on estimates of SSB and recruitment derived from mixed western and eastern populations. As a result, parameters generated using mixed-population data, which do not reflect true population-specific characteristics and dynamics, were applied to the respective unmixed populations in the operating model.

On the other hand, improving the internal consistency of certain aspects of the operating model of Atlantic bluefin tuna has proven challenging. Initial efforts to adapt the Multistock Age-Structured Tag-integrated assessment model (MAST; Taylor et al. 2011) to model the implications of stock mixing and life history uncertainty of Atlantic bluefin tuna (Kerr et al. 2016a) demonstrated that the model was unstable when particular scenarios were assumed, such as early maturity-at-age (Kerr et al. 2018). To avoid this instability, the operating model used in this study was conditioned on ICCAT estimates of recruitment and fishing mortality as well as fishery-independent estimates of movement (Galuardi et al. 2018). Fortunately, the recruitment and fishing mortality rates on which the current operating model was conditioned have been shown to be relatively robust to assumptions about stock mixing (Morse et al. 2018). As a result, the current operating model was stable and conditioned on the most up-to-date data, with the flexibility to model alternative perceptions of population dynamics.

This study used time series of recruitment and SSB as metrics of model performance, but other metrics were also considered. Because VPA-2BOX partitions fishing mortality estimates by age group, summary statistics such as apical F (maximum fishing mortality rate-at-age for each year) and harvest ratios between the estimation and operating model were considered. These statistics showed that operating model versus estimation model values were incomparable due to their different scales (seven geographic zones in the operating model, two stock areas in the estimation model) and different ages of maximum fishing mortality (age 29 in the operating model, plus groups 10+ and 16+ in the estimation model; **Figure 10**). Because of this, model performance was based on population size estimates. However, ICCAT bases stock status of Atlantic bluefin tuna on both stock size and fishing mortality rate reference points, so being able to test model performance on these parameters is an important next step (ICCAT SCRS 2017). Developing a better method for testing a model's ability to estimate fishing mortality rates where the resolution and age groupings of fishing mortality rates is different between the operating model and the estimation model should be a focus of future simulation studies.

Incorporation of stock-recruitment relationships and the younger spawning fraction for the western stock in the alternative operating model substantially altered population dynamics relative to the base case scenario, and facilitated the testing of estimation models on different population trajectories. Despite the younger spawning fraction for the western stock allowing fish to reproduce at a younger age, the stock-recruitment relationship resulted in overall lower simulated stock productivity than conditioning the model directly on ICCAT (2017) VPA estimates of recruitment. The stock-recruitment parameters in the operating model were those used in the 2014 benchmark stock assessment, so an alternative parameterization of the stock-recruitment model based on the addition of more recent recruitment and SSB data might produce a different perspective on population development. This ability to incorporate alternative parameterizations demonstrates the flexibility and generalizability of this operating model.

4.2 Estimation model performance and misspecification

Correctly specifying models is a challenge to all stock assessment scientists, especially for species with complex spatial structure. As Hilborn and Walters (1992) wrote, "Only the most naïve stock assessment biologist would actually believe he ever correctly specifies a model." For species like Atlantic bluefin tuna, in which two populations have strong connectivity, stock assessment models fit to each population are inherently misspecified (Jardim et al. 2018). Based on what is known about Atlantic bluefin tuna biology and migratory behavior (NRC 1994, Nemerson et al. 2000, ICCAT 2002, Block et al. 2005, Boustany et al. 2008, Rooker et al. 2014), the modeling approach of separate VPAs assuming closed populations is misspecified in the spatial dimension.

Because the operating model in this study was more complex spatially than the estimation model, model misspecification led to structural error in the estimation model results (Deroba et al. 2015). In both operating model scenarios and in all estimation model configurations, estimates of recruitment and SSB were biased. On the basis of *absolute bias*, the eastern model performed better than the western model because more recruits and biomass were added or lost than were added or lost to the western stock. However, the true (operating model) western population was smaller than the eastern population by an order of magnitude. On the basis of *percent relative bias*, the western estimation model performed worse than the eastern model, by some metrics doubling or tripling the biomass or number of recruits in the true population. This result suggests that bias in the estimation of western recruitment and SSB could pose a high risk of imposing unsustainable quotas on the western stock that would jeopardize the western population.

Failure of the western estimation model to accurately estimate recruitment and SSB in both operating model scenarios was driven by substantial subsidies of eastern population fish into the western stock area and the capture of some western population fish in eastern fisheries (**Appendix C**). Eastern population fish made up as much as 81% of the resource in the western stock area and accounted for as much as 77% of the total western fishery yield (**Appendix C**). As much as 11%—or more than 100 tonnes—of the yield of western population fish was taken by the eastern fishery (**Appendix C**). This dynamic has been observed in other Atlantic bluefin tuna modeling studies (e.g., Porch et al. 1998, 2001; Morse et al. 2018). The eastern estimation model was less sensitive to misspecification, converging at a higher rate on the mixed-stock pseudodata and producing more accurate estimates of population SSB and recruitment than the western model. The self-test demonstrated that the eastern estimates of SSB were less sensitive to structural error than indicated by the cross-test, which is explained by the small portion of the relatively large eastern stock mixing with the much smaller western stock. By contrast, the western estimation model is much more sensitive to misspecification and the effects of mixing, and western stock size estimates are more likely to be misleading of the true population dynamics.

Although eastern SSB was underestimated relative to both the population and stock views, western SSB was estimated at a magnitude between the two operating model views of SSB. This implies that VPA-2BOX can be expected to underestimate eastern SSB in terms of the population or stock size, and management actions based on eastern stock assessment results may pose a lower risk of overfishing. On the other hand, estimates of western SSB may be equally as informative (or uninformative) for either the stock or population size. The VPA-2BOX estimates of western SSB may be considered to roughly average out the estimation bias relative to the population view (adding an average of about 25,000 tonnes per year) and stock view (missing an average of about 23,000 tonnes per year), but may pose a relatively higher risk of overfishing than for the east.

Counterintuitively, the addition of observation error improved estimation model convergence. None of the deterministic western runs and only one of the deterministic eastern model runs converged when fitted to pseudodata generated without error, whereas the majority of stochastic runs converged. Because the model was misspecified, the structure of the pseudodata without observation error was generally incompatible with the estimation model configuration, but adding observation error to make the pseudodata “noisier” obscured signals in the pseudodata that conflicted with the estimation model and led to higher convergence rates.

Observation error in the catch-at-age pseudodata caused the deterministic and mean of the stochastic estimates of eastern SSB to diverge. Larger values for the standard deviation of the observation error on the catch-at-age as calculated from the residuals of an ASAP analysis of Atlantic bluefin tuna data ($\sigma = 0.86$) resulted in a greater divergence; decreasing this standard deviation to 0.3 produced no divergence. The divergence between deterministic and stochastic runs increased with older ages and was most marked in the age 10+ group and SSB, but was not observed in estimates of recruitment and abundance at younger ages. The divergence did not seem to change significantly when observation error was reduced in the indices of relative abundance pseudodata. Therefore, the divergence in eastern SSB and abundance at older ages appears to have been caused by sensitivity to signals in the catch-at-age pseudodata, which change based on the scale of observation error. In addition, the greater magnitude of catch numbers in the east than the west may have caused the observation error to have a proportionally larger effect on the eastern than the western model results. This model behavior, in combination with difficulties with convergence for the ICCAT assessments (Zarrad et al. 2018), is evidence for lack of identifiability in the VPA, in that the pseudodata are not informative enough for predicting all unknown parameters and estimating SSB accurately. The model’s unidentifiability, combined with model misspecification, resulted in negatively biased estimates of eastern SSB.

The western VPA as specified by ICCAT (2017) was particularly poorly-suited for fitting to spatially complex bluefin tuna-like pseudodata. Simulation testing a model for which ICCAT scientists have finely adjusted model settings to fit to the available data for Atlantic bluefin tuna proved challenging, and some model re-configurations (e.g., lowering parameter bounds) were needed to achieve convergence. In addition, the self-test of the eastern estimation model suggested that the VPA-2BOX model may perform poorly even without misspecification of spatial structure. This demonstrates the need for extensive adjustment of VPA-2BOX model settings to specific bluefin tuna data sets, and reinforces the understanding of the tendency of VPA-2BOX to not converge unless finely configured to data with a specific structure (Zarrad et al. 2018).

4.3 Management recommendations

Both the western and eastern VPAs were better at estimating the abundance of mixed-population stocks than of populations, the implications of which depend on the management goal. Some aspects of fishery management, such as short-term catch projections and quotas, may be accurately informed by estimation models that produce abundance estimates similar to the true values of mixed-population stocks. However, other medium- to long-term fishery management objectives, such as avoiding overfishing and rebuilding depleted populations, require accurate estimates of population abundance and production. Estimates of the eastern stock size were less biased than the western stock size relative to the true population view on a percent relative bias basis, so stock assessment results may be more effective for informing long-term management and rebuilding targets in the east than the west. In addition, the stock assessment demonstrated a consistent tendency to underestimate the size of the eastern bluefin resource, which might further protect the eastern population from overfishing.

In situations of high complexity and limited data like that of Atlantic bluefin tuna (ICCAT 2014), and when complex population dynamics are not fully understood, status quo management—although suboptimal—often becomes the default option (Kerr et al. 2016b). Evaluating the behavior and performance of stock assessment models used to inform management—especially models that are known to be or are suspected of being misspecified—should be a routine aspect of the stock assessment process so that information on model limitations and tendencies can be used to qualify the reporting of model results to managers (NRC 1998). This study demonstrated the unreliability of the status quo closed-population VPAs for assessing Atlantic bluefin tuna—particularly the western population—due to misspecified spatial structure. Several models that account for stock mixing have been considered by the ICCAT Atlantic bluefin tuna working group, but they have not been operationalized due to data constraints and model convergence issues (ICCAT 2017, Morse et al. 2018). An appropriate response to the results of this study would be continuing to develop stock assessment methods for Atlantic bluefin tuna that more accurately account for stock mixing and movement dynamics (Porch 1995, Taylor et al. 2011, Carruthers et al. 2016, Morse et al. 2018), and to consider the results of these methods for determining stock status and fishery management advice (ICCAT 2008, 2017).

4.4 Next steps

This study evaluated the performance of stock assessment methods for Atlantic bluefin tuna based on a simulation framework for testing the effect of mismatches between the assumed stock structure (represented by an estimation model) and the true state of nature (represented by an operating model). The reproducibility of the simulation method offers the opportunity to modify operating model conditioning and data structure and the use of different estimation models, such as a multi-stock VPA (e.g., VPA-2BOX with overlap; Porch 2003) or other statistical age-structured models (e.g., age-structured assessment program, ASAP; stock synthesis, SS). Efforts are underway to use this simulation framework to test the performance of a revised VPA that considers bluefin stock mixing by partitioning catch data into western and eastern biological populations, instead of stock areas, using stock composition data derived from otolith samples (Morse et al. 2018).

Next steps for this research include the incorporation of this simulation testing framework into MSE for Atlantic bluefin tuna. MSE will require modifying the operating model to project population dynamics beyond 2015, deriving reference points, and specifying alternative management strategies. Potential management strategies would include harvest control rules that meet ICCAT's objectives (Fromentin 2006). ICCAT is currently engaged in a MSE process for Atlantic bluefin tuna, and this research will continue with an aim to contribute to that effort (Carruthers and Butterworth 2018a).

References

Ailloud, L.E., Lauretta, M.V., Hanke, A.R., Golet, W.J., Allman, R.J., Siskey, M.R., Secor, D.H., and Hoenig, J.M. 2017. Improving growth estimates for Western Atlantic bluefin tuna using an integrated modeling approach. Fish. Res. **191**: 17-24. Available from http://dx.doi.org/10.1016/j.fishres.2017.02.013. Accessed June 2018.

Beverton, R.J.H., and Holt, S.J. 1957. On the dynamics of exploited fish populations. Fish. Invest. **2**: 19.

Block, B.A., Teo, S.L.H., Walli, A., Boustany, A., Stokesbury, M.J.W., Farwell, C.J., Weng, K.C., Dewar, H., and Williams, T.D. 2005. Electronic tagging and population structure of Atlantic bluefin tuna. Nature **434**: 1121-1127. doi: 10.1038/nature03463.

Boustany, A.M., Reeb, C.A., and Block, B.A. 2008. Mitochondrial DNA and electronic tagging reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). Mar. Biol. **156**: 13-24.

Butterworth, D.S., and Punt, A.E. 1994. The robustness of estimates of stock status for the western North Atlantic bluefin tuna population to violations of the assumptions underlying the associated assessment models. Collect. Vol. Sci. Pap. ICCAT **42**(1): 192-210.

Cadrin, S.X., and Secor, D.H. 2009. Accounting for spatial population structure in stock assessment: Past, present, and future. *In* The Future of Fisheries Science in North America. *Edited by* R.J. Beamish and B.J. Rothschild. Fish & Fisheries Series, Springer Science + Business Media B.V. pp. 405-426. doi: 10.1007/978-1-4020-9210-7.

Carlsson, J., McDowell, J.R., Carlsson, J.E.L., and Graves, J.E. 2007. Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. J. Hered. 98(1): 23-28. Available from https://doi.org/10.1093/jhered/esl046. Accessed May 2018.

Carlsson, J., McDowell, J.R., Díaz-Jaimes, P., Carlsson, J.E.L., Boles, S.B., Gold, J.R., and Graves, J.E. 2004. Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. Mol. Ecol. **13**: 3345-3356. doi: 10.1111/j.1365-294X.2004.02336.x.

Carruthers, T., and Butterworth, D. 2018a. ABT-MSE: An R package for Atlantic bluefin tuna management strategy evaluation. Collect. Vol. Sci. Pap. ICCAT **74**(6): 3553-3559.

Carruthers, T., and Butterworth, D. 2018b. Summary of a reference set of conditioned operating models for Atlantic bluefin tuna. Collect. Vol. Sci. Pap. ICCAT **73**(6): 3523-3541.

Carruthers, T., and Kell, L. 2017. Issues arising from the preliminary conditioning of operating models for Atlantic bluefin tuna. Collect. Vol. Sci. Pap. ICCAT **73**(6): 2171-2182.

Carruthers, T., Kimoto, A., Powers, J.E., Kell, L., Butterworth, D.S., Lauretta, M.V., and Kitakado, T. 2016. Structure and estimation framework for Atlantic bluefin tuna operating models. Collect. Vol. Sci. Pap. ICCAT **72**(7): 1782-1795.

Collette, B.B. 2017. Bluefin tuna science vague. Science **358**(6365): 879-880. doi: 10.1126/science.aar3928.

Collette, B.B., Carpenter, K.E., Polidoro, B.A. Juan-Jordá, M.J., Boustany, A., Die, D.J., Elfes, C. Fox, W., Graves, J., Harrison, L.R., McManus, R., Minte-Vera, C.V., Nelson, R., Restrepo, V., Schratwieser, J., Sun, C.-L., Amorim, A., M. Brick Peres, Canales, C., Cardenas, G., Chang, S.-K., Chiang, W.-C., de Oliveira Leite, N. Jr., Harwell, H., Lessa, R., Fredou, F.L., Oxenford, H.A., Serra, R., Shao, K.-T., Sumaila, R. Wang, S.-P., Watson, R., and Yáñez, E. 2011. High value and long life—Double jeopardy for tunas and billfishes. Science **333**(6040): 291-292. doi: 10.1126/science.1208730.

Cort, J.L. 1991. Age and growth of the bluefin tuna, *Thunnus thynnus* (L.) of the northeast Atlantic. Collect. Vol. Sci. Pap. ICCAT **35**(2): 213-230.

Deroba, J.J., Butterworth, D.S., Methot, Jr, R.D., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., and Hulson, P.-J.F. 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. **72**(1): 19-30. Available from https://doi.org/10.1093/icesjms/fst237. Accessed June 2017.

Fay, G., Punt, A.E., and Smith, A.D.M. 2011. Impacts of spatial uncertainty on performance of age structure-based harvest strategies for blue eye trevalla (*Hyperoglyphe antarctica*). Fish. Res. **110**(3): 391-407. doi: 10.1016/j.fishres.2011.04.015.

Fromentin, J.-M. 2006. ICCAT Manual Chapter 2.1.5: Atlantic bluefin tuna., ICCAT, Madrid, Spain. Available from https://www.iccat.int/Documents/SCRS/Manual/CH2/2_1_5_BFT_ENG.pdf. Accessed March 2018.

Fromentin, J.-M., and Powers, J.E. 2005. Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. Fish. Fish. **6**(4): 281-306. Available from https://doi.org/10.1111/j.1467-2979.2005.00197.x. Accessed March 2018.

Fu, C., and Fanning, L.P. 2004. Spatial considerations in the management of Atlantic cod off Nova Scotia, Canada. N. Am. J. Fish. Manag. **24**(3): 775-784. doi: 10.1577/M03-134.1.

Galuardi, B., Cadrin, S.X., Arregi, I., Arrizabalaga, H., DiNatale, A., Brown, C., Lauretta, M., and Lutcavage, M. 2018. Atlantic bluefin tuna area transition matrices estimated from electronic tagging and SATTAGSIM. Collect. Vol. Sci. Pap. ICCAT **74**(6): 2903-2921.

Galuardi, B., Royer, F., Golet, W., Logan, J., Neilson, J., and Lutcavage, M. 2010. Complex migration routes of Atlantic bluefin tuna (*Thunnus thynnus*) question current population structure paradigm. Can. J. Fish. Aquat. Sci. **67**(6): 966-976. doi: 10.1139/F10-033.

Gavaris, S. 1988. An adaptive framework for the estimation of population size. Canadian Atlantic Fisheries Scientific Advisory Committee Research Document 88/29, Saint Andrews, New Brunswick.

Goethel, D.R., Kerr, L.A., and Cadrin, S.X. 2016. Incorporating spatial population structure into the assessment-management interface of marine resources. *In* Management Science in Fisheries: An introduction to simulation-based methods. *Edited by* C.T.T. Edwards and D.J. Dankel. Routledge, New York, NY. pp 319-347.

Goethel, D.R., Legault, C.M., and Cadrin, S.X. 2015. Testing the performance of a spatially explicit tag-integrated stock assessment model of yellowtail flounder (*Limanda ferruginea*) through simulation analysis. Can. J. Fish. Aquat. Sci. **72**(4): 582-601. Available from https://doi.org/10.1139/cjfas-2014-0244. Accessed September 2017.

Goethel, D.R., Quinn II, T.J., and Cadrin, S.X. 2011. Incorporating spatial structure in stock assessment: Movement modeling in marine fish population dynamics. Rev. Fish. Sci. **19**(2): 119-136. doi: 10.1080/10641262.2011.557451.

Guan, W., Cao, J., Chen, Y., and Cieri, M. 2013. Impacts of population and fishery spatial structures on fishery stock assessment. Can. J. Fish. Aquat. Sci. **70**(8): 1178-1189. Available from https://doi.org/10.1139/cjfas-2012-0364. Accessed April 2018.

Hilborn, R., and Walters, C.J. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Springer, Medford, MA.

ICCAT. 1980. Report for biennial period, 1978-79 Part II (1979) English Version. ICCAT, Madrid, E.S. Available from https://www.iccat.int/Documents/BienRep/REP_EN_78-79_II.pdf. Accessed November 2017.

ICCAT. 1981. Recommendation on bluefin management measures 81-01. ICCAT. Available from https://www.iccat.int/Documents/Recs/compendiopdf-e/1981-01-e.pdf. Accessed November 2017.

ICCAT. 2002. ICCAT Workshop on Bluefin Mixing. Collect. Vol. Sci. Pap. ICCAT **54**(2): 261-352.

ICCAT. 2008. Report of the 2008 Atlantic bluefin tuna stock assessment session. ICCAT, Madrid, E.S., 23 June-4 July 2008.

ICCAT. 2013. First meeting of the working group of fisheries managers and scientists in support of the western Atlantic bluefin tuna stock assessment. ICCAT, Montreal, Q.C., 26-28 June 2013.

ICCAT. 2014. Report of the 2014 Atlantic bluefin tuna stock assessment session. ICCAT, Madrid, E.S., 22-27 September 2014.

ICCAT. 2017. Report of the 2017 ICCAT bluefin stock assessment meeting. ICCAT, Madrid, E.S., 20-28 July 2017.

ICCAT Standing Committee on Research and Statistics (SCRS). 2017. Report of the Standing Committee on Research and Statistics (SCRS). Madrid, E.S., 2-6 October 2017.

Jardim, E., Eero, M., Silva, A., Ulrich, C., Pawlowski, L., Holmes, S.J., Ibaibarriaga, L., De Oliveira, J.A.A., Riveiro, I., Alzorriz, N., Citores, L., Scott, F., Uriarte, A., Carrera, P., Duhamel, E., and Mosqueira, I. 2018. Testing spatial heterogeneity with stock assessment models. PLoS ONE **13**(1): e0190791. Available from https://doi.org/10.1371/journal.pone.0190791. Accessed May 2018.

Kerr, L.A., Cadrin, S.X., and Secor, D.H. 2010. Simulation modelling as a tool for examining the consequences of spatial structure and connectivity on local and regional population dynamics. ICES J. Mar. Sci. **67**(8): 1631-1639. Available from https://doi.org/10.1093/icesjms/fsq053. Accessed May 2017.

Kerr, L.A., Cadrin, S.X., Secor, D.H., and Taylor, N. 2013. A simulation tool to evaluate effects of mixing between Atlantic bluefin tuna stocks. Collect. Vol. Sci. Pap. ICCAT **69**(2): 742-759.

Kerr, L.A., Cadrin, S.X., Secor, D.H., and Taylor, N. 2015. Evaluating the effect of Atlantic bluefin tuna movement on the perception of stock units. Collect. Vol. Sci. Pap. ICCAT **71**(4): 1660-1682.

Kerr, L.A., Cadrin, S.X., Secor, D.H., and Taylor, N. 2016a. Modeling the implications of stock mixing and life history uncertainty of Atlantic bluefin tuna. Can. J. Fish. Aquat. Sci. **74**(11): 1990-2004. Available from https://doi.org/10.1139/cjfas-2016-0067. Accessed May 2017.

Kerr, L.A., and Goethel, D.R. 2014. Simulation modeling as a tool for synthesis of stock identification information. *In* Stock Identification Methods: Applications in Fishery Science, Second Edition. *Edited by* S.X. Cadrin, L.A. Kerr, and S. Mariani. Elsevier, Waltham, MA. pp. 502-533.

Kerr, L.A., Hintzen, N.T., Cadrin, S.X., Clausen, L.W., Dickey-Collas, M., Goethel, D.R., Hatfield, E.M.C., Kritzer, J.P., and Nash, R.D.M. 2016b. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. ICES J. Mar. Sci. **74**(6): 1708-1722. doi: 10.1093/icesjms/fsw188.

Kerr, L.A., Morse, M.R., Cadrin, S.X., and Galuardi, B. 2018. Application of an Atlantic bluefin tuna operating model to generate pseudodata for stock assessment testing. Collect. Vol. Sci. Pap. ICCAT **74**(6): 3405-3426.

Li, Y., Bence, J.R., and Brenden, T.O. 2018. Can spawning origin information of catch or a recruitment penalty improve assessment and fishery management performance for a spatially structured stock assessment model? Can. J. Fish. Aquat. Sci. Available from https://doi.org/10.1139/cjfas-2017-0523. Accessed April 2018.

Lorenzen, K., Steneck, R.S., Warner, R.R., Parma, A.M., Coleman, F.C., and Leber, K.M. 2010. The spatial dimension of fisheries: Putting it all in place. Bull. Mar. Sci. **86**(2): 169-177.

Maguire J.-J., Cadrin, S.X., Hanke, A., and Melvin, G. 2018. An exploration of bluefin tuna data in the North Atlantic (West and East + Mediterranean) with ASAP. Collect. Vol. Sci. Pap. ICCAT **74**(6): 3205-3223.

Mather, F.J., Mason, J.M., and Jones, A.C. 1995. Historical Document: Life History and Fisheries of Atlantic Bluefin Tuna. NOAA Technical Memorandum NMFS-SEFSC - 370, Miami, FL.

Methot, Jr., R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86-99. Available from https://doi.org/10.1016/j.fishres.2012.10.012. Accessed January 2018.

Morse, M., Cadrin, S., Kerr, L., Secor, D., Siskey, M., Arrizabalaga, H., Hanke, A., and Porch, C. 2018. An updated analysis of bluefin tuna stock mixing. Collect. Vol. Sci. Pap. ICCAT **74**(6): 3486-3509.

National Research Council (NRC). 1994. An assessment of Atlantic bluefin tuna. National Academy Press, Washington, DC.

National Research Council (NRC). 1998. Improving fish stock assessments. National Academy Press, Washington, DC.

Nemerson, D., Berkeley, S., and Safina, C. 2000. Spawning site fidelity in Atlantic bluefin tuna, *Thunnus thynnus*: the use of size-frequency analysis to test for the presence of migrant east Atlantic bluefin tuna on Gulf of Mexico spawning grounds. Fish. Bull. **98**(1): 118-126.

Parrack, M.L. 1986. A method of analyzing catches and abundance indices from a fishery. Collect. Vol. Sci. Pap. ICCAT **24**: 209-221.

Peck, S.L. 2004. Simulation as experiment: a philosophical reassessment for biological modeling. Trends Ecol. Evolut. **19**(10): 530-534. doi: 10.1016/j.tree.2004.07.019.

Porch, C.E. 1995. A two-area VPA with discrete mixing: Can we discriminate between mixing rates given the present condition of the data? Collect. Vol. Sci. Pap. ICCAT **44**(1): 198-208.

Porch, C.E. 2003. VPA-2BOX Version 3.01 User’s Guide. NOAA Sustainable Fisheries Division Contribution SFD/2003-2004, Miami, FL.

Porch, C.E. 2005. The sustainability of western Atlantic bluefin tuna: a warm-blooded fish in a hot-blooded fishery. Bull. Mar. Sci. **76**(2): 363–384.

Porch, C.E., and Hanke, A. 2018. Estimating the fraction of western Atlantic bluefin tuna that spawn by age from size frequency data collected on the Gulf of Mexico spawning grounds. Collect. Vol. Sci. Pap. ICCAT **74**(6): 3224-3233.

Porch, C.E., Kleiber, P., Turner, S., Sibert, J., Bailey, R., and Cort, J.L. 1998. The efficacy of VPA models in the presence of complicated movement patterns. Collect. Vol. Sci. Pap. ICCAT **50**(2): 591-622.

Porch, C.E., and Lauretta, M.V. 2016. On making statistical inferences regarding the relationship between spawners and recruits and the irresolute case of western Atlantic bluefin tuna (*Thunnus thynnus*). PLoS ONE **11**(6): e0156767. doi:10.1371/journal.pone.0156767.

Porch, C.E., Restrepo, V.R., Turner, S.C., and Scott, G.P. 1995. Virtual population analysis of Atlantic bluefin tuna incorporating movement and tagging data. Collect. Vol. Sci. Pap. ICCAT **44**(1): 183-190.

Porch, C.E., Turner, S.C., and Powers, J.E. 2001. Virtual population analysis of Atlantic bluefin tuna with alternative models of transatlantic migration: 1970-1997. Collect. Vol. Sci. Pap. ICCAT **52**(3): 1022-1045.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. Chapter 10: Minimization or Maximization of Functions. *In* Numerical Recipes in C: The Art of Scientific Computing, Second Edition. Cambridge University Press, Cambridge, United Kingdom. pp 408-412.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/. Accessed January 2017.

Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W.J. 2009. Genetic population structure of marine fish: Mismatch between biological and fisheries management units. Fish. Fish. **10**(4): 361-395. doi:10.1111/j.1467-2979.2008.00324.x.

Richardson, D.E., Marancik, K.E., Guyon, J.R., Lutcavage, M.E., Galuardi, B., Lam, C.H., Walsh, H.J., Wildes, S., Yates, D.A., and Hare, J.A. 2016. Discovery of a spawning ground reveals diverse migration strategies in Atlantic bluefin tuna (Thunnus thynnus). PNAS **113**(12): 3299-3304. Available from https://doi.org/10.1073/pnas.1525636113. Accessed January 2017.

Rooker, J.R., Arrizabalaga, H., Fraile, I., Secor, D.H., Dettman, D.L., Abid, N., Addis, P., Deguara, S., Karakulak, F.S., Kimoto, A., Sakai, O., Macías, D., and Neves Santos, M. 2014. Crossing the line: Migratory and homing behaviors of Atlantic bluefin tuna. Mar. Ecol. Prog. Ser. **504**: 265-276. doi: 10.3354/meps10781.

Rooker, J.R., Secor, D.H., DeMetrio, G., Kaufman, A.J, Belmonte Ríos, A., and Tičina, V. 2008a. Evidence of trans-Atlantic movement and natal homing of bluefin tuna from stable isotopes in otoliths. Mar. Ecol. Prog. Ser. **368**: 231-239. Available from https://doi.org/10.3354/meps07602. Accessed December 2016.

Rooker, J.R., Secor, D.H., DeMetrio, G., Schloesser, R., Block, B.A., and Neilson, J.D. 2008b. Natal homing and connectivity in Atlantic bluefin tuna populations. Science **322**(5902): 742-744. doi: 10.1126/science.1161473.

Secor, D.H. 2014. The unit stock concept: bounded fish and fisheries. *In* Stock Identification Methods: Applications in Fishery Science, Second Edition. *Edited by* S.X. Cadrin, L.A. Kerr, and S. Mariani. Elsevier, Waltham, MA. pp. 7-28.

Secor, D.H., Gahagan, B., and Rooker, J.R. 2012. Atlantic bluefin tuna population assignment based on otolith stable isotopes: The 2003 year-class within the U.S. North Carolina recreational fishery. Collect. Vol. Sci. Pap. ICCAT **68**(1): 212-222.

Secor, D.H., Rooker, J.R., Gahagan, B.I., Siskey, M.R., and Wingate, R.W. 2014. Depressed resilience of bluefin tuna in the western Atlantic and age truncation. Conserv. Biol. **29**(2): 400-408. doi: 10.1111/cobi.12392.

Sheela, B.V. 1979. An optimized step-size random search (OSSRS). Computer Methods in Applied Mechanics and Engineering **19**: 99-106.

Siskey, M.R., Wilberg, M.J., Allman, R.J., Barnett, B.K., and Secor, D.H. 2016. Forty years of fishing: Changes in age structure and stock mixing in northwestern Atlantic bluefin tuna (*Thunnus thynnus*) associated with the size-selective and long-term exploitation. ICES J. Mar. Sci. **73**(10): 2518-2528. doi:10.1093/icesjms/fsw115.

Smedbol, R.K., and Stephenson, R. 2001. The importance of managing within-species diversity in cod and herring fisheries of the north-western Atlantic. J. Fish Biol. **59**(Supplement A): 109-128.

Szuwalski, C.S., and Punt, A.E. 2015. Can an aggregate assessment reflect the dynamics of a spatially structured stock? Snow crab in the eastern Bering Sea as a case study. Fish. Res. **164**: 135-142. doi: 10.1016/j.fishres.2014.10.020.

Taylor, N.G., McAllister, M.K., Lawson, G.L., Carruthers, T., and Block, B.A. 2011. Atlantic bluefin tuna: A novel multistock spatial model for assessing population biomass. PLoS ONE **6**(12): e27693. Available from https://doi.org/10.1371/journal.pone.0027693. Accessed September 2017.

Walli, A., Teo, S.L.H., Boustany, A., Farwell, C.J., Williams, T., Dewar, H., Prince, E., and Block, B.A. 2009. Seasonal movements, aggregations and diving behavior of Atlantic bluefin tuna (*Thunnus thynnus*) revealed with archival tags. PLoS ONE **4**(7): e6151. doi: 10.1371/journal.pone.0006151.

Ying, Y., Chen, Y., Lin, L., and Gao, T. 2011. Risks of ignoring fish population spatial structure in fisheries management. Can. J. Fish. Aquat. Sci. **68**(12): 2101-2120. doi: 10.1139/F2011-116.

Zarrad, R., Walter, J., and Lauretta, M. 2018. VPA-2BOX model diagnostics used in the 2014 assessment of eastern Atlantic bluefin tuna. Collect. Vol. Sci. Pap. ICCAT **74**(6): 2809-2831.

Tables

Table 1. Operating model specifications.

	West	East
Simulation period	1974-2015 (quarterly)	
Number of geographic zones	7	
Geographic zones in stock area (**Figure 1**)	1-3	4-7
Number of fleets and survey indices	17	10
Age classes	1-16+	1-10+
Natural mortality vector	0.38, 0.30, 0.24, 0.20, 0.18, 0.16, 0.14, 0.13, 0.12, 0.11, 0.11, 0.11, 0.10, 0.10, 0.10	0.38, 0.30, 0.24, 0.20, 0.18, 0.16, 0.14, 0.13, 0.12, 0.10
Spawning fraction vector (**Figure 3**)	Young: 0, 0, 0.25, 0.5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 Old: 0, 0, 0, 0, 0, 0.001, 0.007, 0.039, 0.186, 0.563, 0.879, 0.976, 0.996, 0.999, 1, 1	0, 0, 0.25, 0.5, 1, 1, 1, 1, 1, 1
Growth parameters (length-at-age)	Richards model (Ailloud et al. 2017): $L_1 = 33.0$, $L_2 = 270.6$, $p = -0.12$, $A_1 = 0$, $A_2 = 34$, $K = 0.22$	Von Bertalanffy model (Cort 1991): $K = 0.093$, $L_\infty = 319$, $t_0 = -0.97$
Growth equation (length-weight)	$W_a = 0.0000177054 * L_a^{3.001251847}$	$W_a = 0.0000350801 * L_a^{2.878451}$
Hockey stick stock-recruitment parameters (**Eq. 1**, alternative operating model scenario)	$R_{max} = 84{,}363$ $SSB^* = 12{,}236$	$R_{max} = 1{,}889{,}896$ $SSB^* = 215{,}548$

Table 2. All estimation models and settings.

Stock	Estimation Model	Settings
West	ICCAT (2017)	equal to ICCAT (2017) western VPA
	Revised	decreased lower bound on all terminal F parameters from 1.0E-4 to 1.0E-7
East	ICCAT (2017)	equal to ICCAT (2017) eastern VPA
	Revised	estimated FRAER2 index variance scaling parameter equal to FRAER1

Table 3. Convergence rates (of 500 simulations) for different estimation models applied to corresponding operating model scenarios.

Stock	Operating Model Scenario	Estimation Model Configuration	Convergence Rate
West	Base case	ICCAT (2017)	6%
		Revised	86%
	Alternative	ICCAT (2017)	64%
East	Base case	ICCAT (2017)	95%
		Revised	96%
	Alternative	ICCAT (2017)	95%

Table 4. Percent relative bias and absolute bias (average of percent relative error and absolute error, respectively, over all years and all simulated realizations; from **Eq. 7** and **8**) in VPA model estimates of recruitment and SSB (converged runs only) relative to the operating model population and stock views. Recruitment bias calculations exclude years 2012-2015.

Stock	Operating Model Scenario	Estimation Model Configuration	Percent Relative Bias (Absolute Bias)		
			Recruitment (numbers)	SSB (tonnes)	
				Population	Stock
West	Base case	ICCAT (2017)	191% (288,597)	99% (25,198)	-25% (-21,237)
		Revised	192% (280,121)	107% (24,754)	-27% (-22,581)
	Alternative	ICCAT (2017)	284% (87,758)	146% (24,925)	-5% (1,560)
East	Base case	ICCAT (2017)	-33% (-802,232)	-71% (-467,662)	-69% (-420,326)
		Revised	-36% (-891,822)	-74% (-487,041)	-72% (-439,705)
	Alternative	ICCAT (2017)	-29% (-196,328)	-72% (-255,991)	-70% (-232,625)

Figures

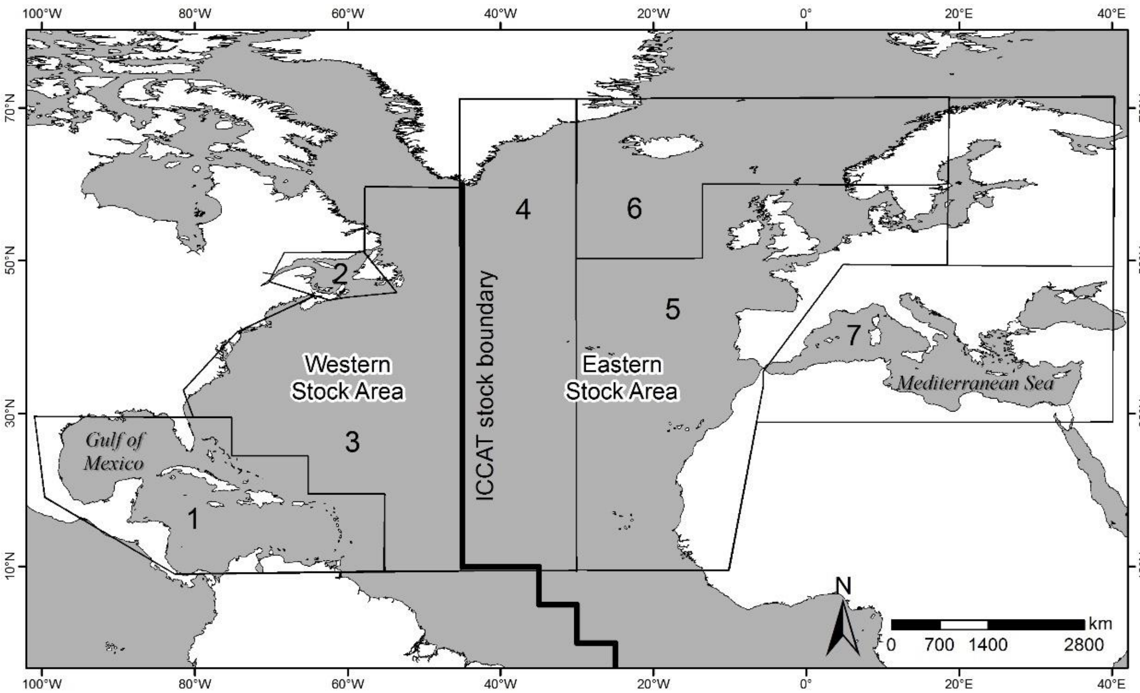


Figure 1. Western and eastern bluefin tuna stock management areas in the Atlantic Ocean (Fromentin 2006) and operating model geographic zones: Gulf of Mexico (zone 1), Gulf of St. Lawrence (zone 2), western Atlantic (zone 3), central Atlantic (zone 4), eastern Atlantic (zone 5), northeast Atlantic (zone 6), and Mediterranean Sea (zone 7; Kerr et al. 2016a).

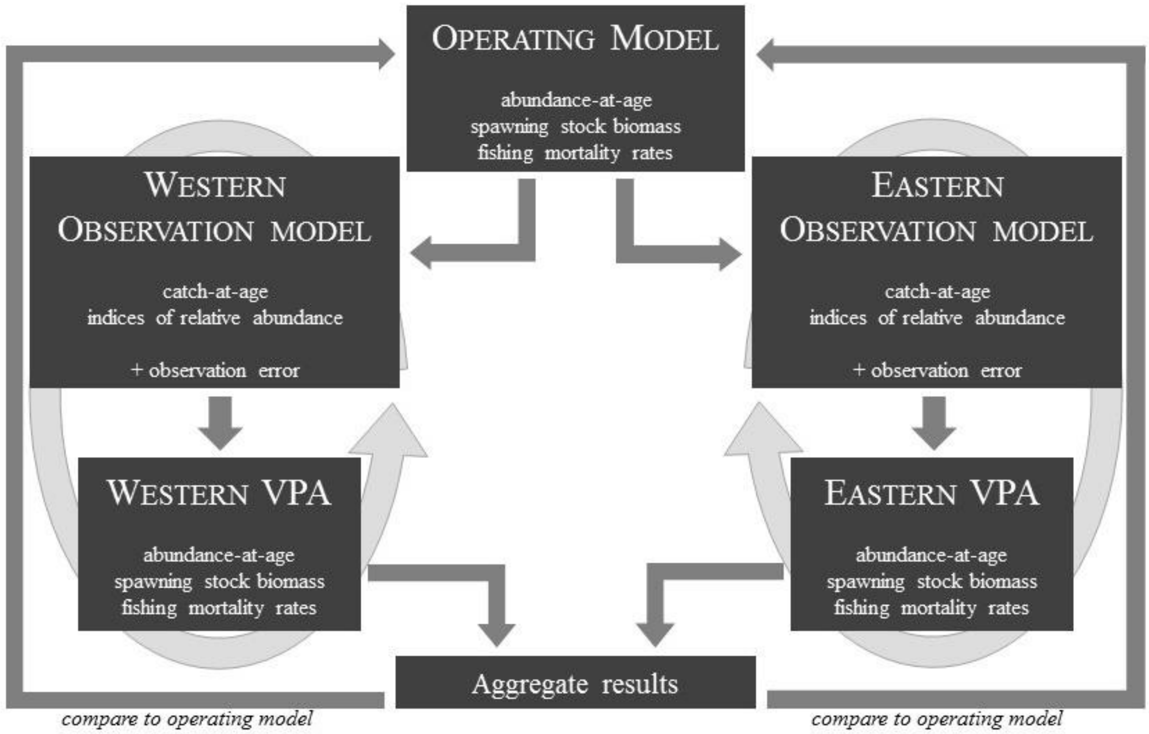


Figure 2. Simulation testing framework employed in this study. Light gray circular arrows represent replicated realizations (e.g., 500) with unique sets of random observation errors.

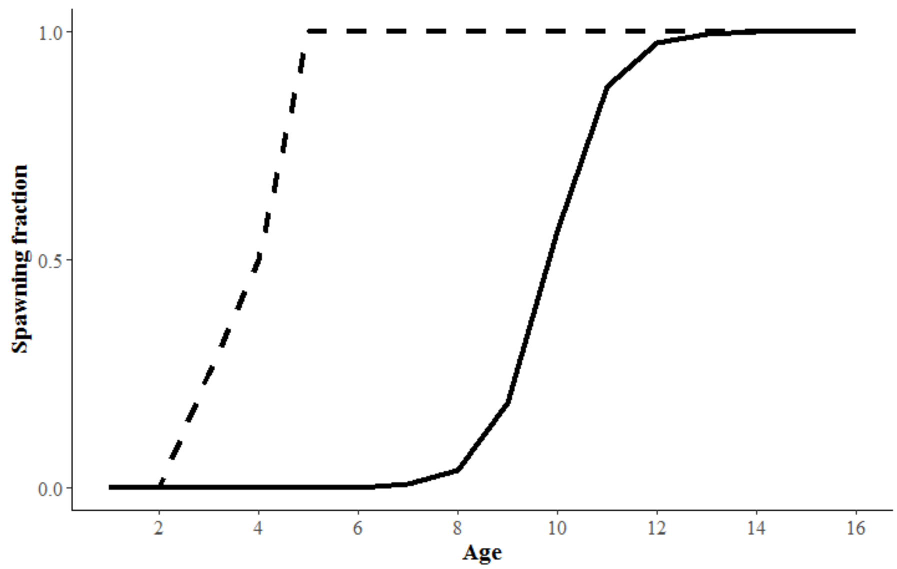


Figure 3. Spawning fraction scenarios. The solid line is the older western spawning fraction used in the base case operating model, and the dashed line is the eastern and western younger spawning fractions used in the alternative operating model scenario.

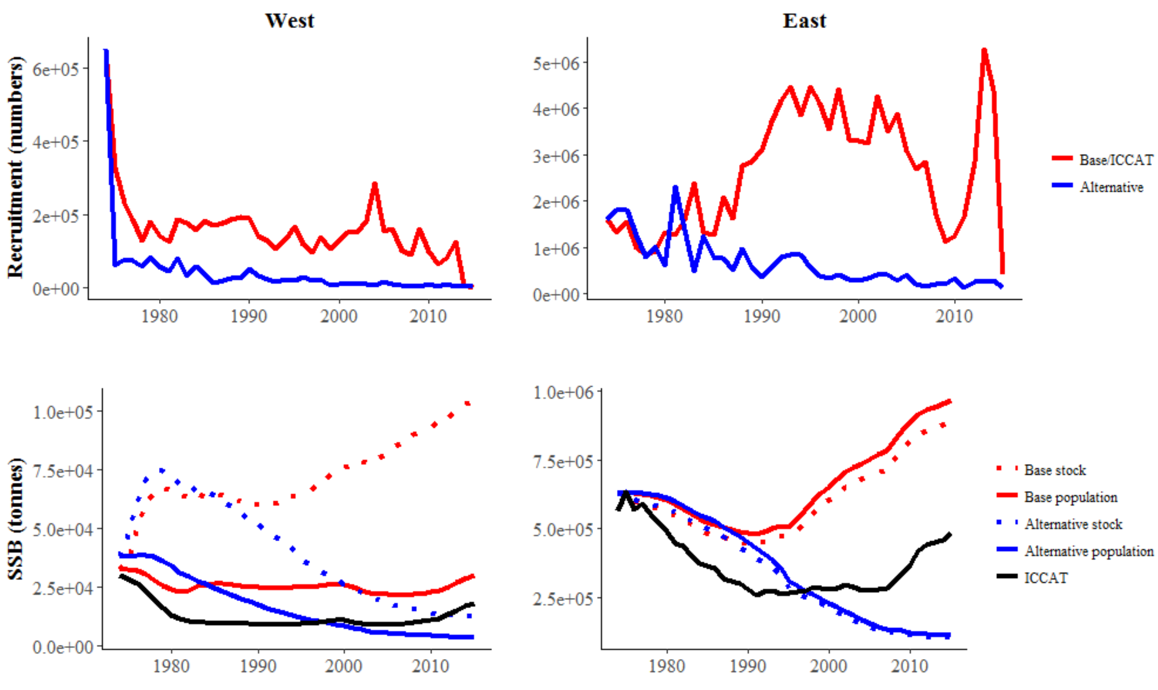


Figure 4. Base case and alternative operating model time series for recruitment and SSB for western and eastern population and stock units relative to the ICCAT (2017) stock assessment estimates.

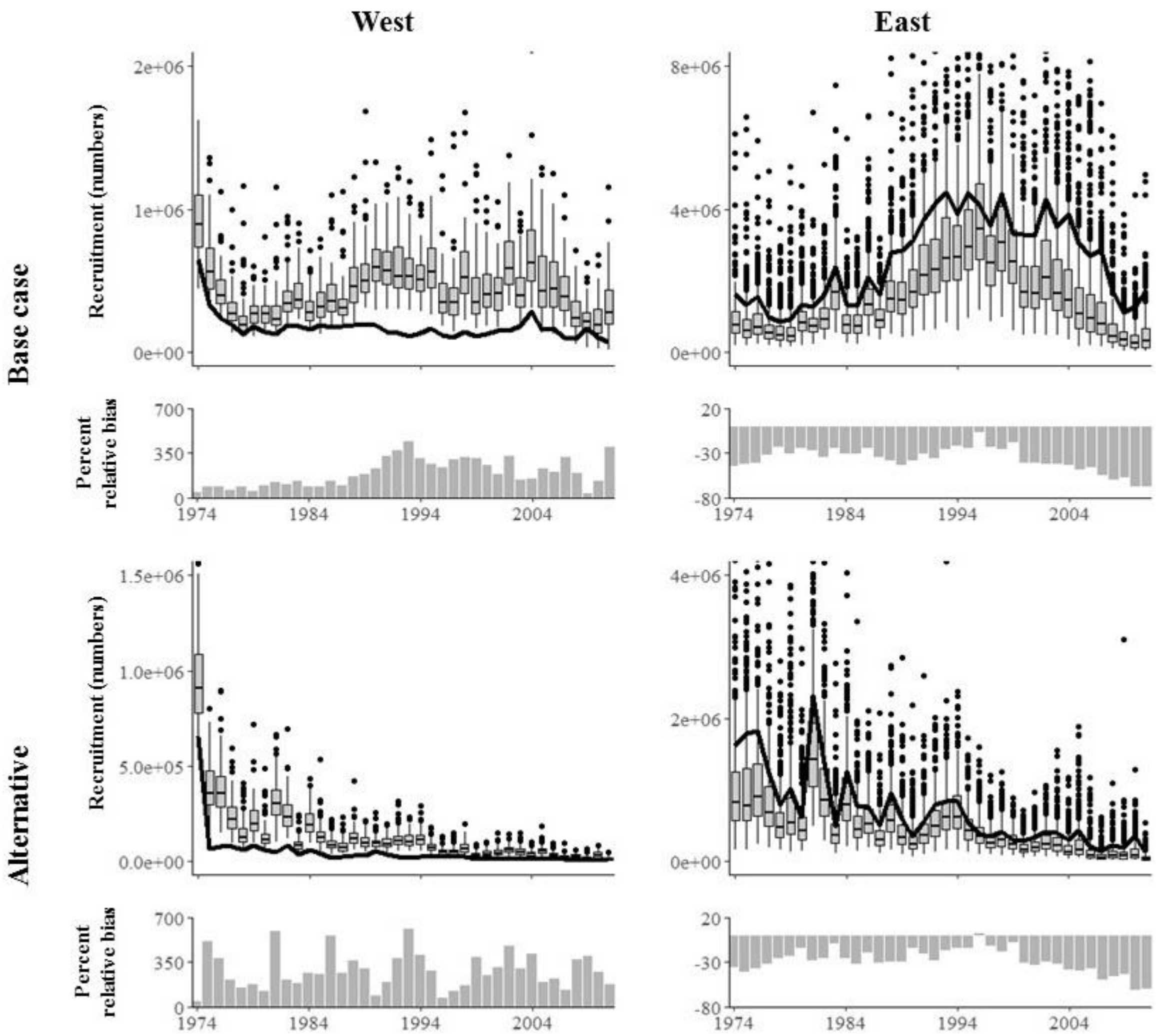


Figure 5. Boxplots show the distribution of VPA estimates of recruitment for western and eastern stocks using pseudodata from base case and alternative operating model scenarios. Boxplots display median, interquartile range (IQR, hinges), 1.5 * IQR (whiskers), and outliers (points). Solid black lines are the operating model time series. Bar charts show the mean annual percent relative bias across simulated realizations corresponding to each recruitment time series. Poorly estimated recruitment data in 2012-2015 are excluded and large outliers are not shown.

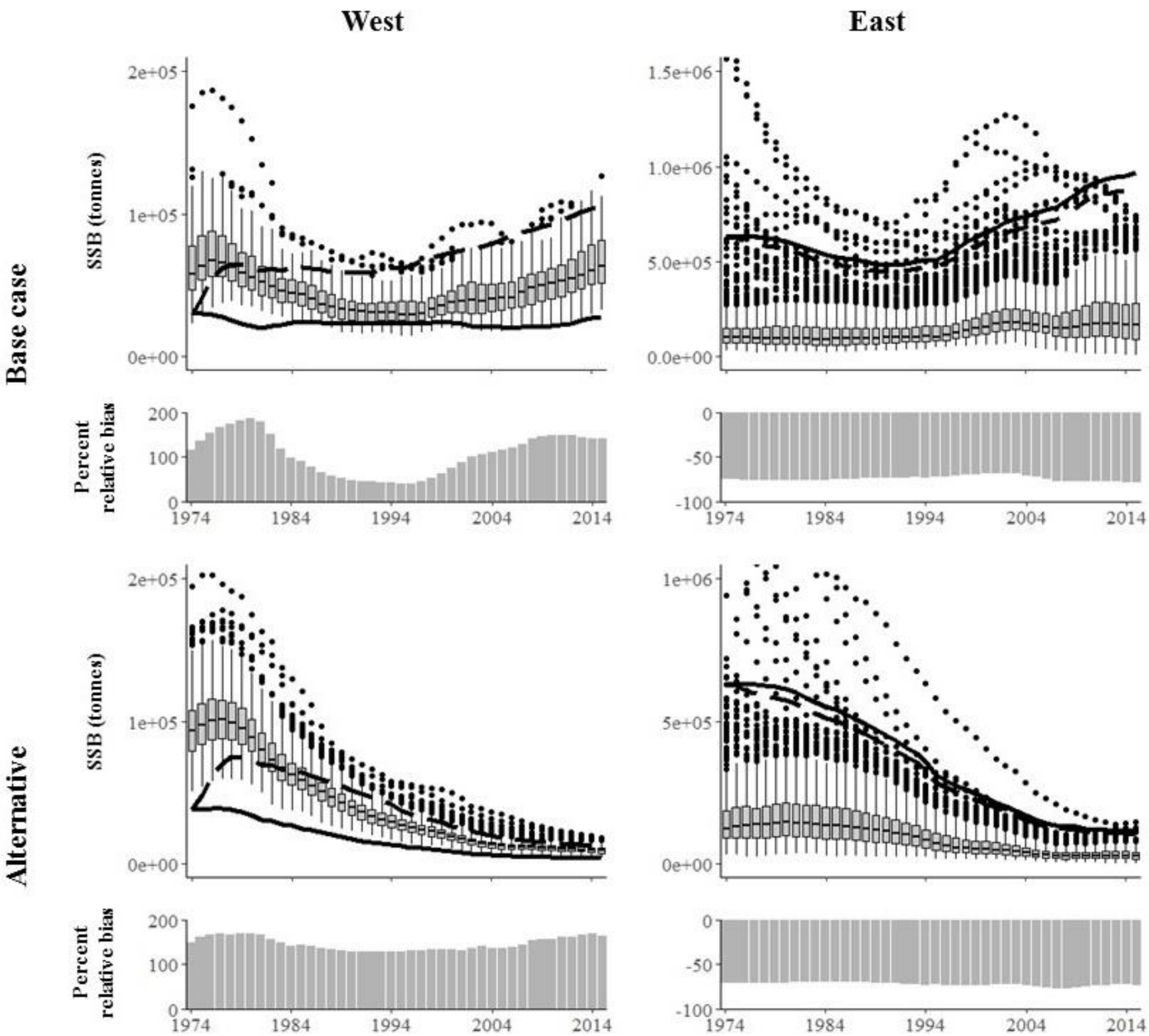


Figure 6. Boxplots show the distribution of VPA estimates of SSB for western and eastern stocks using pseudodata from base case and alternative operating model scenarios. Boxplots display median, interquartile range (IQR, hinges), 1.5 * IQR (whiskers), and outliers (points). Solid lines are the operating model population view and dashed lines are the stock view. Bar charts show the mean annual percent relative bias across simulated realizations corresponding to each SSB time series relative to the operating model population view. Large outliers are not shown.

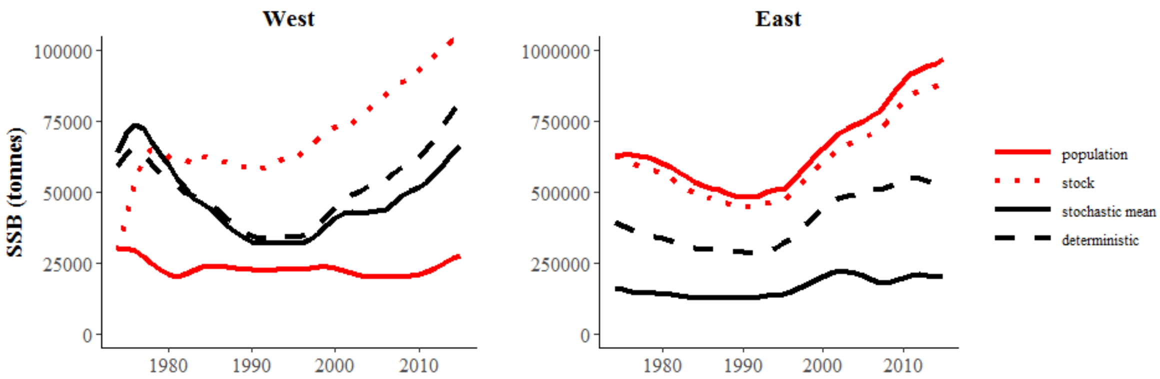


Figure 7. SSB time series of deterministic (black dashed) versus mean of stochastic runs (black solid) for converged runs only of the revised western and eastern estimation models. Shown with operating model population (red solid) and stock (red dotted) SSB.

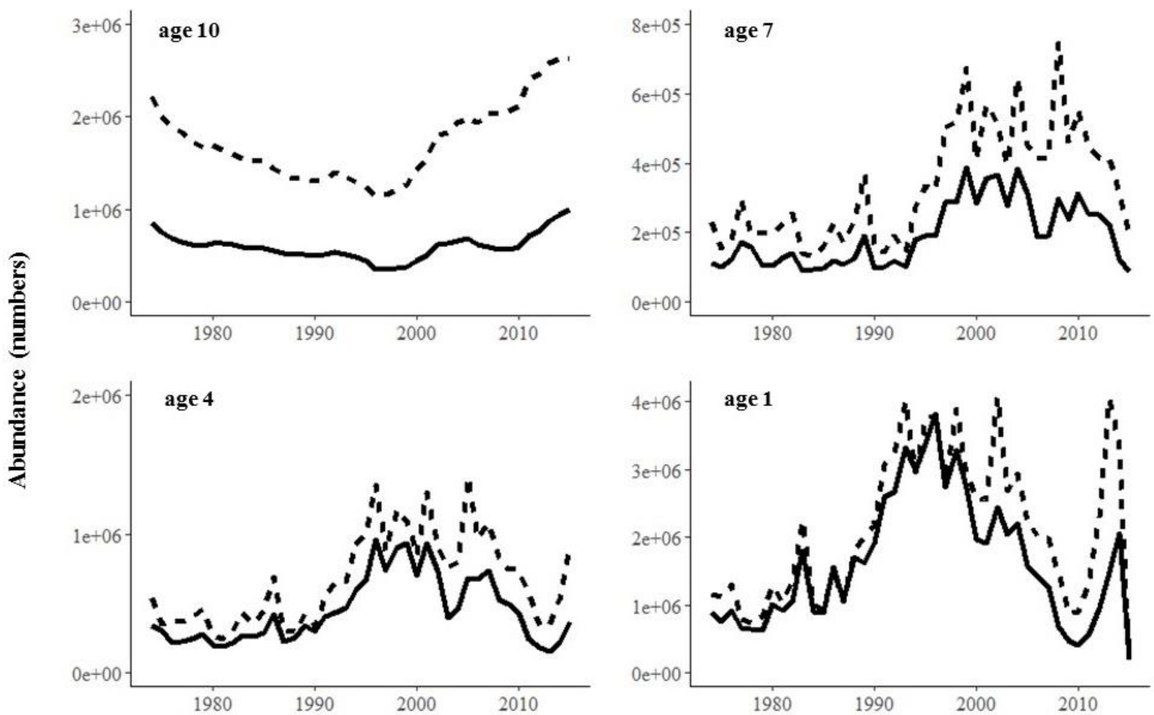


Figure 8. Comparison of abundance at ages 10, 7, 4, and 1 (i.e., recruitment) deterministic (dashed) and mean of stochastic runs (solid) for the revised eastern estimation model.

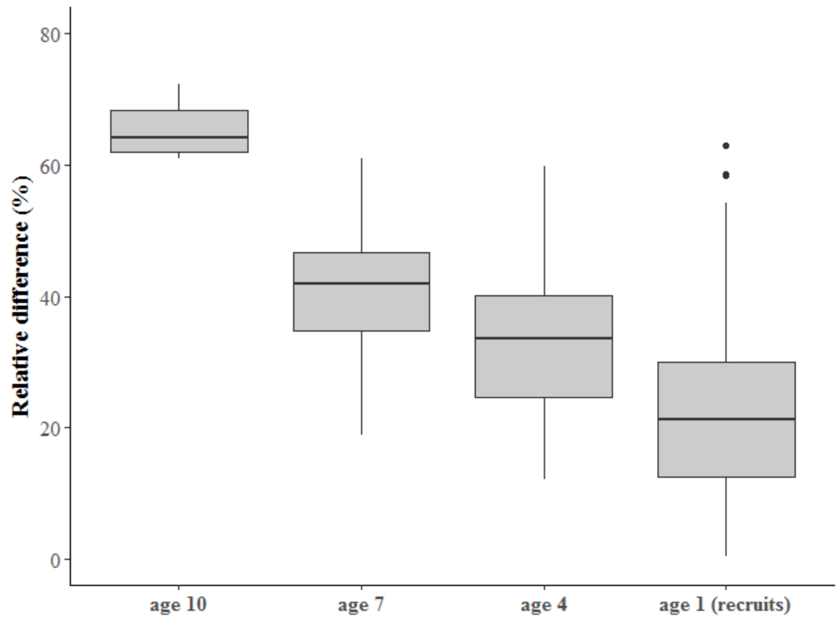


Figure 9. Mean annual relative difference between deterministic and mean stochastic abundance-at-age (**Figure 8**) for the revised eastern estimation model. Boxplots display median, interquartile range (IQR, hinges), 1.5 * IQR (whiskers), and outliers (points).

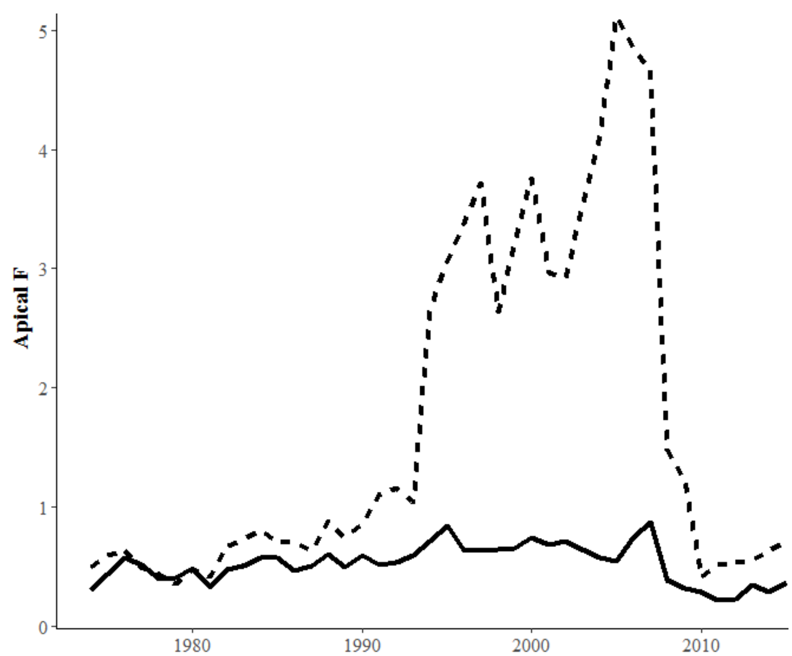


Figure 10. Apical F comparison between operating model (dashed line) and estimation model (solid line) for the eastern stock alternative operating model scenario.

Appendix A

Additional Information on Relative Age Composition Pseudodata

Some stock assessment models, such as statistical catch-at-age, partition catch-at-age data into fleet-specific partial catch-at-age, which are used to indicate the relative amount that each fleet contributes to the overall catch. However, the estimation model tested in this study, VPA-2BOX, is not a fleet-based model and thus does not use partial catch-at-age data in the same way a statistical catch-at-age model would. VPA-2BOX does provide the option to use indices of relative abundance, based on fleet or survey data, and can use additional data on the relative age composition of each index to inform the index selectivity-at-age (**Figures A1** and **A2**). These relative age composition data inform the model about which ages the index data should be applied. In Atlantic bluefin tuna stock assessments using VPA-2BOX, these relative age data have come from different sources, but primarily from catch-at-age data and maturity vectors (ICCAT 2017). For example, in the 2017 stock assessment of eastern Atlantic bluefin tuna, catch-at-age data were used to inform the selectivities of most indices and the eastern bluefin maturity-at-age vector was used to inform the western Mediterranean larval survey selectivity (ICCAT 2017).

Pseudodata used as the relative age composition for most western and eastern indices in this study were derived from the operating model using **Eq. 4**, and mimicked the method of using catch-at-age data to inform age-selectivity from the ICCAT (2017) stock assessments. Another possible method for deriving these pseudodata that would more closely resemble fleet-specific partial catch-at-age data, in which the partial catch-at-age of each fleet is a direct proportion of the overall stock-wide catch-at-age, would have required making several simplifying assumptions regarding the spatial and temporal structure of the operating model. This alternative approach would have required either aggregating the fishing mortality and natural mortality rates over geographic zones, populations, and seasonal quarters, or disaggregating the effort term ($E_{y,g}$) into the seven geographic zones and four seasonal quarters. The first option compromises the spatial and temporal structure of the operating model that is critical to this study, and the second option requires making several assumptions about the level of effort in each zone for each season.

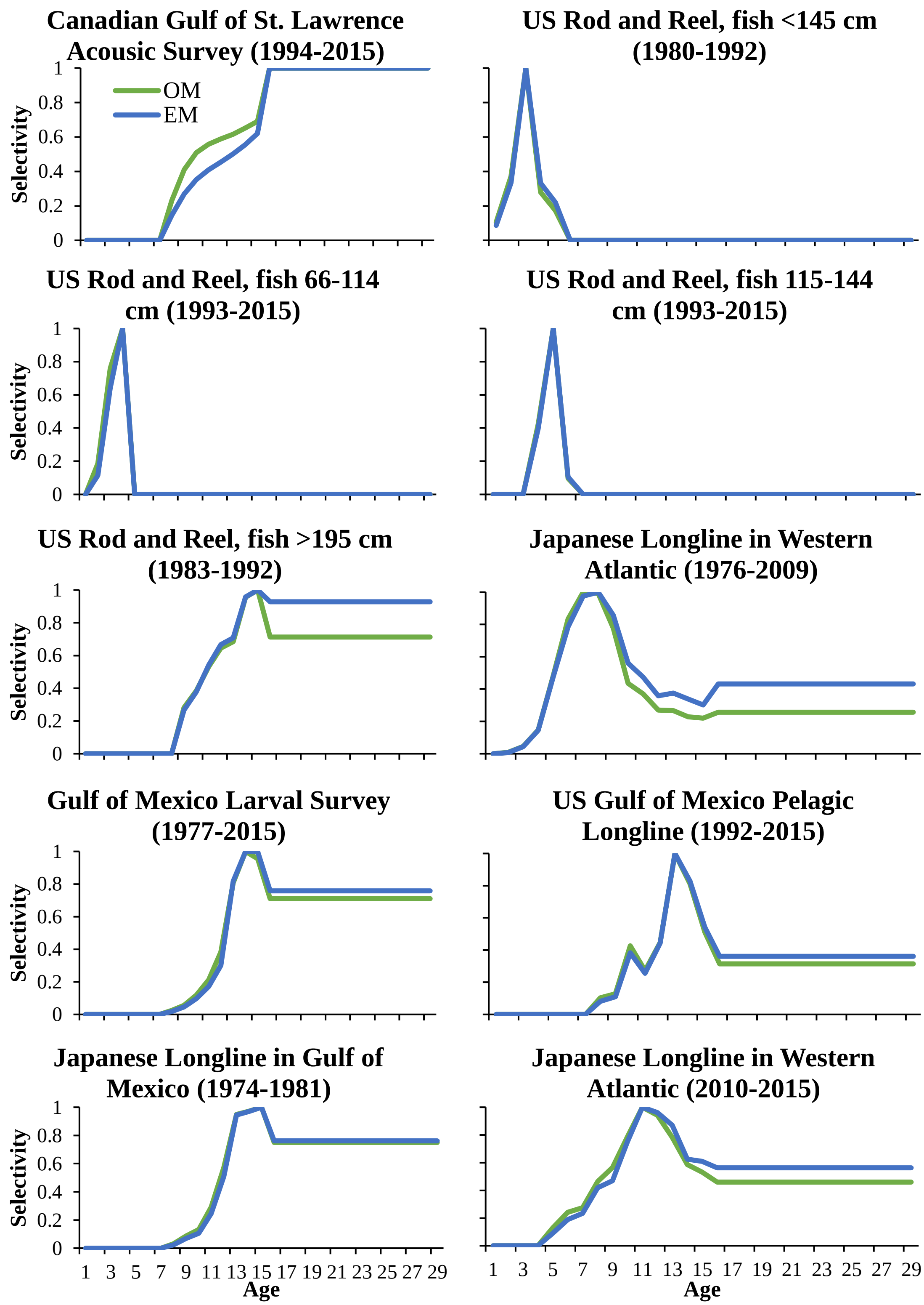


Figure A1. Selectivity-at-age series from the operating model (OM) and the deterministic run of the estimation model (EM) for all western fleets included as indices of relative abundance.

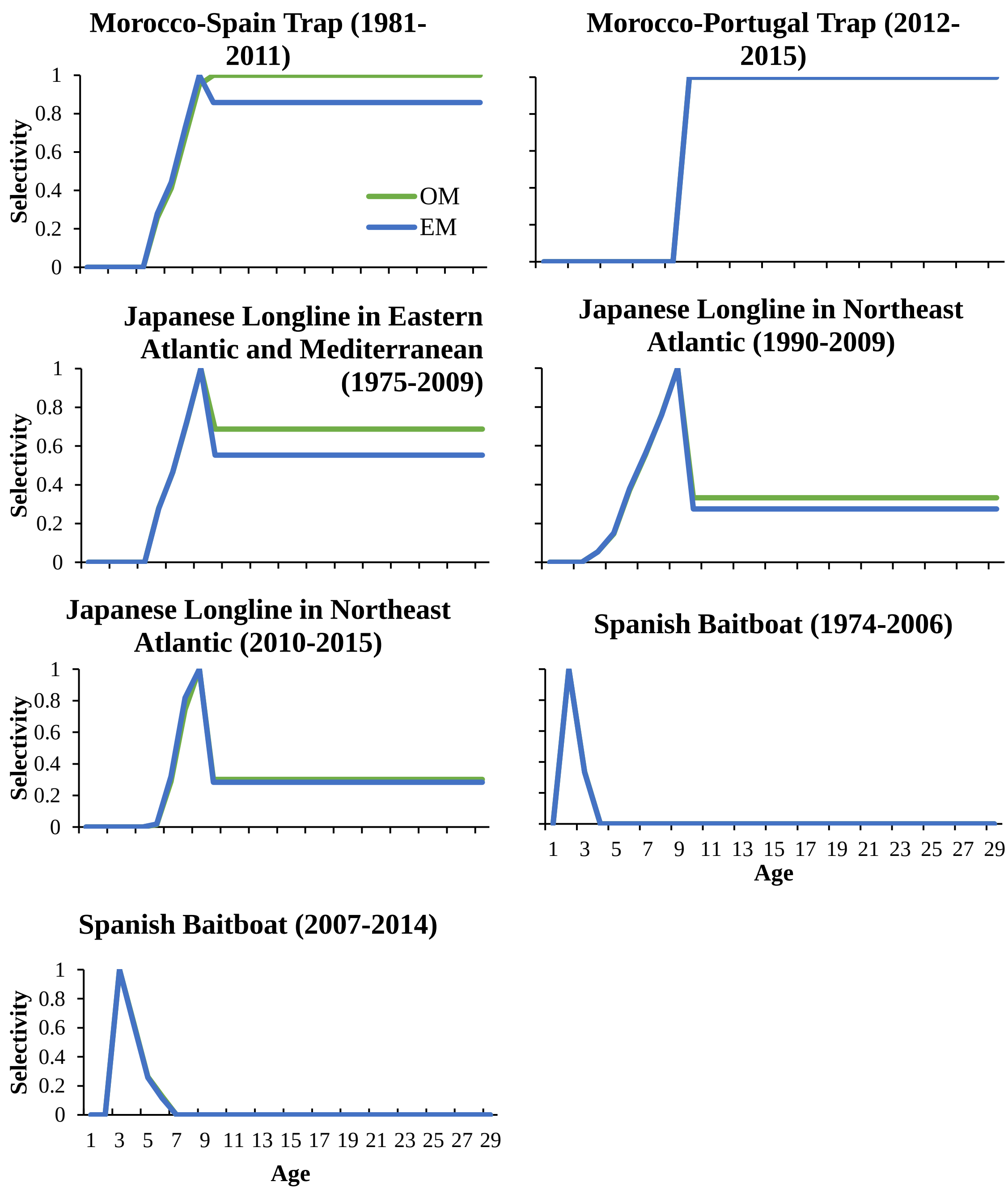


Figure A2. Selectivity-at-age series from the operating model (OM) and the deterministic run of the estimation model (EM) for all eastern fleets included as indices of relative abundance.

Appendix B

Self-test of Eastern Estimation Model

The eastern estimation model was observed to have greater bias in estimates of spawning stock biomass (SSB) than could be reasonably attributed to model misspecification of spatial structure. To determine the degree of estimation bias attributable to spatial structure and fish movement, a “self-test” was performed for the eastern estimation model, in which the operating model and estimation model had the same structural assumptions and model settings (Deroba et al. 2015). For the self-test, the operating model was modified to eliminate spatial structure and fish movement. All seven geographic zones were removed and zone-specific quantities were aggregated and generalized to the two larger stock areas. Seasonal structure was also eliminated to mimic the annual assumptions of the estimation model. The operating model was conditioned directly on ICCAT (2017) VPA estimates of abundance-at-age and fishing mortality-at-age for all ages and years 1974-2015. The estimation model used the revised configuration of the ICCAT (2017) eastern VPA-2BOX model (**Table 2**), in which the value of the variance scaling parameter for the French aerial survey index 2 was set equal to the estimated value for the French aerial survey index 1.

The self-test underestimated SSB by -30% (**Table B1**, **Figure B1**), in comparison to the cross-test which underestimated SSB by approximately -70% (**Table 4**). This suggests that although a large proportion of the population and stock SSB estimation bias can be explained by either model misspecification of spatial structure or other factors, e.g., observation error (see section 3.3.4), as evidenced by the cross-test, it can also be attributed to generally poor performance of VPA-2BOX for the eastern stock as shown by the self-test.

The bias associated with VPA-2BOX in the self-test may be attributed to several factors. The VPA-2BOX model for eastern Atlantic bluefin tuna has a recent history of poor performance and instability when fitted to the available data. In both the 2014 and 2017 stock assessments, VPA-2BOX demonstrated instability that was attributed to low quality data (ICCAT 2014, 2017), insufficient exploration of diagnostics, and improper parameter settings (Zarrad et al. 2018). As referenced in chapters 3 and 4 of this study, the high standard deviation of the observation error on the catch-at-age pseudodata ($\sigma_{east} = 0.86$)—which was retained in the self-test—also had an influence on the bias in estimates of SSB (**Figure B1**). The eastern estimation model in the self-test produced negatively biased estimates of SSB and positively biased estimates of age 1 recruitment (**Figure B1**). Recruitment in the most recent 5 years was especially poorly estimated (not shown), which was driven by the model's difficulty in estimating the terminal F parameter for age 1, and to a lesser extent, for age 2. The high CV (standard error of the estimator divided by the value of the estimate; Porch 2003) and the first derivative diagnostic test indicated poor estimation of these parameters.

Table B1. Percent relative bias (average of percent relative error over all years and all simulated realizations; from **Eq. 7**) in VPA model estimates of recruitment and SSB (converged runs only) for the self-test.

Stock	Operating Model Scenario	Estimation Model Configuration	Percent Relative Bias	
			Recruitment (numbers)	SSB (tonnes)
East	Self-test (no spatial or seasonal structure)	Revised	60%	-30%

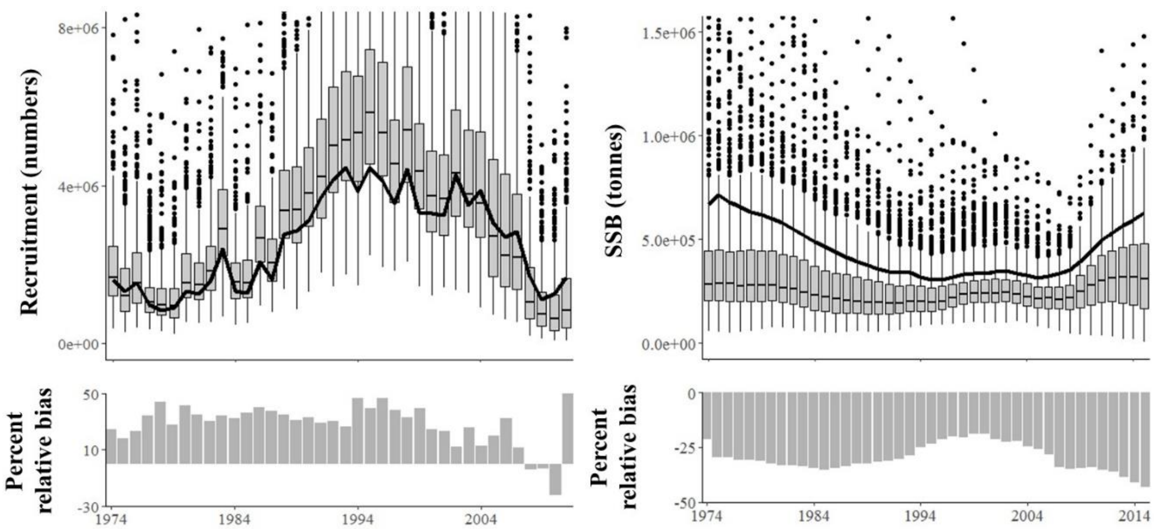


Figure B1. Boxplots show the distribution of VPA estimates of recruitment (age 1 abundance) and SSB from the eastern estimation model self-test. Boxplots display median, interquartile range (IQR, hinges), 1.5 * IQR (whiskers), and outliers (points). Solid black lines are the operating model time series. Bar chart shows the mean annual percent relative bias across simulated realizations corresponding to the recruitment or SSB time series relative to the operating model. Large outliers are not shown.

Appendix C

Operating Model Dynamics

To gain a more thorough understanding of operating model dynamics and patterns, results from the operating model were visualized using a series of additional plots of fishing mortality rates, abundance, biomass, and yield, similar to those presented in Kerr et al. (2016a).

Although the operating model was conditioned on fishing mortality rates estimated by the ICCAT (2017) VPA, the realized fishing mortality rates in the operating model after disaggregation into the seven zones and re-aggregation by stock area were lower in magnitude (where the fishing mortality rate is the weighted average fishing mortality rate over all age classes; **Figure C1**). This was a result of deriving zone-based fishing mortality rates by partitioning them based on the index relative age composition data from the ICCAT (2017) VPA model inputs. Operating model fishing mortality rates were generally greater in the western stock area and peaked in 1981, whereas eastern fishing mortality rates peaked in 1996 (**Figure C1**). Fishing mortality rates tracked known tuna migration patterns, with the greatest rates over the entire simulation area occurring in the third quarter in zone 3 (western Atlantic) where tuna are known to feed (**Figure C2**). The greatest fishing mortality in the first quarter (spawning season) was in the spawning zones 1 (Gulf of Mexico) in the 1970s and 7 (Mediterranean Sea) in the 1990s and 2000s (**Figure C2**).

Atlantic bluefin tuna abundance was generally greatest in zones 5 (eastern Atlantic) and 7 and least in zones 1 and 2 (Gulf of St. Lawrence) throughout each year (**Figure C3**). The greatest overall abundance was in the 1990s, 2000s, and, after decreasing in 2010, in 2014 (**Figure C3**). The average increase in overall abundance throughout the time series (**Figure C3**) coincided with an average decrease in fishing mortality rates (**Figures C1** and **C2**) over time (except for the high fishing mortality rates in the first quarter in zone 7 in the 1990s and 2000s).

SSB of the eastern population across the entire operating model area in the third quarter (when the majority of fishing mortality occurs, **Figure C2**) was consistently greater than the western population (**Figure C4**). SSB in the third quarter was generally greatest in zone 3 for the western population and in zone 5 for the eastern population (**Figure C4**). Generally, spawners from each population were more abundant within their respective stock areas, but did cross the stock management boundary in appreciable numbers and were distributed throughout all geographic zones during the third quarter (**Figures C4** and **C5**). In particular, SSB of western population fish was high in zones 4 (central Atlantic) and 5 and SSB of eastern population fish was high in zone 3 (**Figure C4**). Like SSB in the third quarter, the abundance of eastern recruits in the first quarter in zone 7 was greater than western recruits in zone 1 by approximately one order of magnitude (**Figure C6**).

Fishery yield was greatest in coastal areas of the eastern U.S. and western Europe and Africa and in the Mediterranean Sea (zones 5, 3, and 7), and many more tonnes of the eastern population were caught than the western population (**Figure C7**). Yield from the western population was greatest in zone 3 throughout the time series in the second, third, and fourth quarters, and moderate in zone 1 in the 1970s in the first and second quarters (**Figure C7**). Yield of the eastern population was greatest in zone 7 in the first quarter (spawning quarter) and greatest in zones 5, 3, 7, and 6 (northeast Atlantic) in the second, third, and fourth quarters (**Figure C7**). Fisheries in both stock areas caught fish from both populations, and relatively more eastern population fish were caught in the western stock area than western population fish caught in the eastern stock area (**Figure C7**).

In general, there were greater fishing mortality rates but lower abundance and biomass in the western stock area. Yields of both the western and eastern populations were substantial in the western Atlantic (zone 3). Eastern population fish migrated into the western stock area, particularly into the western Atlantic, and were subject to these high fishing mortality rates. Western population fish were also subject to the high fishing mortality rates in the western Atlantic, but were also subject to moderate fishing mortality rates in the central Atlantic in the eastern stock area.

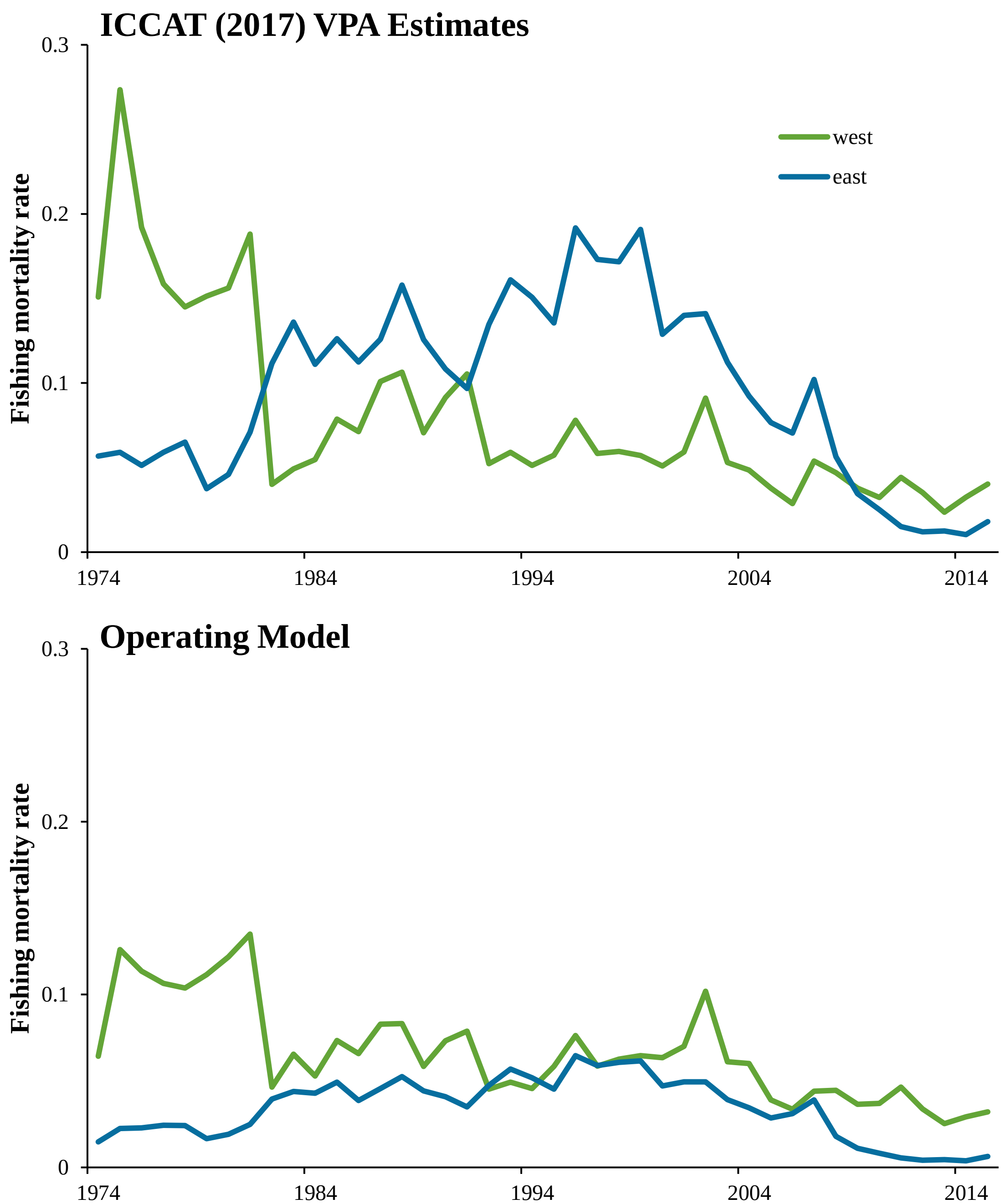


Figure C1. Comparison of fishing mortality rates between west and east stock areas from the ICCAT (2017) VPA results (top panel) and the operating model (bottom panel) over time. For both plots, annual values are the average fishing mortality rate over all age classes, and for the operating model plot, over the geographic zones contained in the stock area (zones 1-3 in the western area, zones 4-7 in the eastern area).

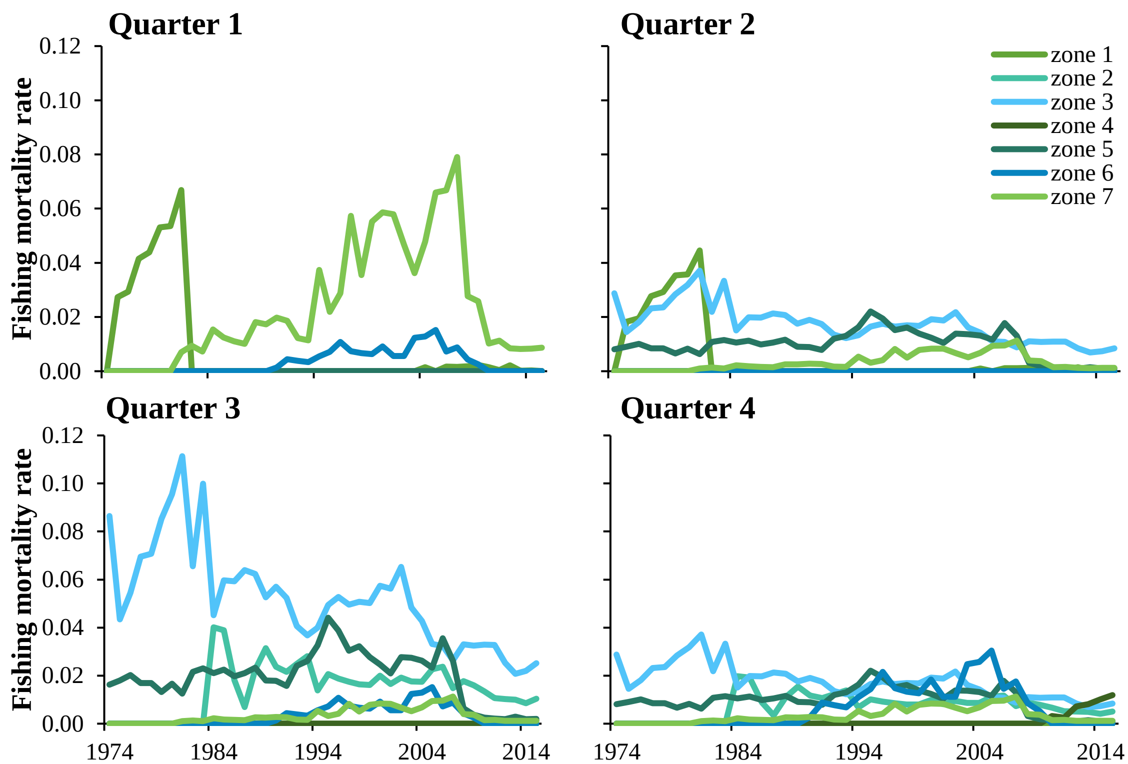


Figure C2. Comparison of fishing mortality rates among the four seasonal quarters and seven geographic zones over time. Annual values are the average fishing mortality rate over all age classes contained in the zone. (See **Figure 1** for all zone descriptions.)

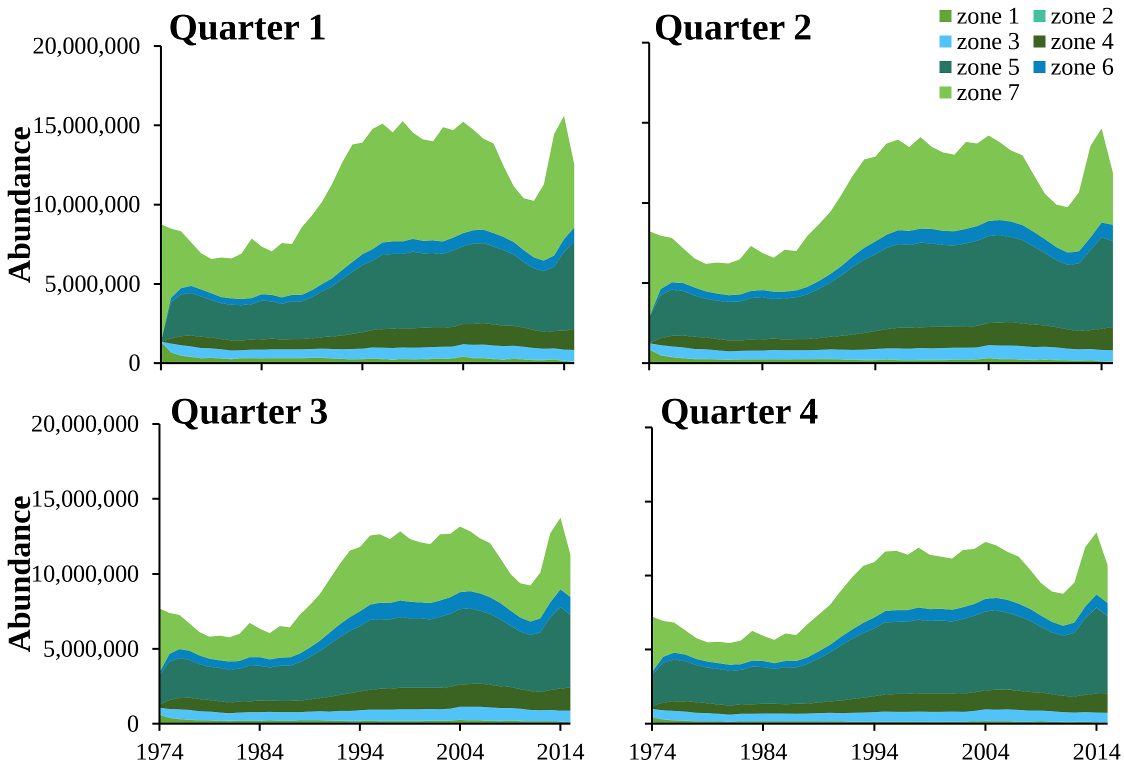


Figure C3. Comparison of Atlantic bluefin tuna abundance (western and eastern populations combined) among the four seasonal quarters and seven geographic zones in the operating model over time. (See **Figure 1** for all zone descriptions.)

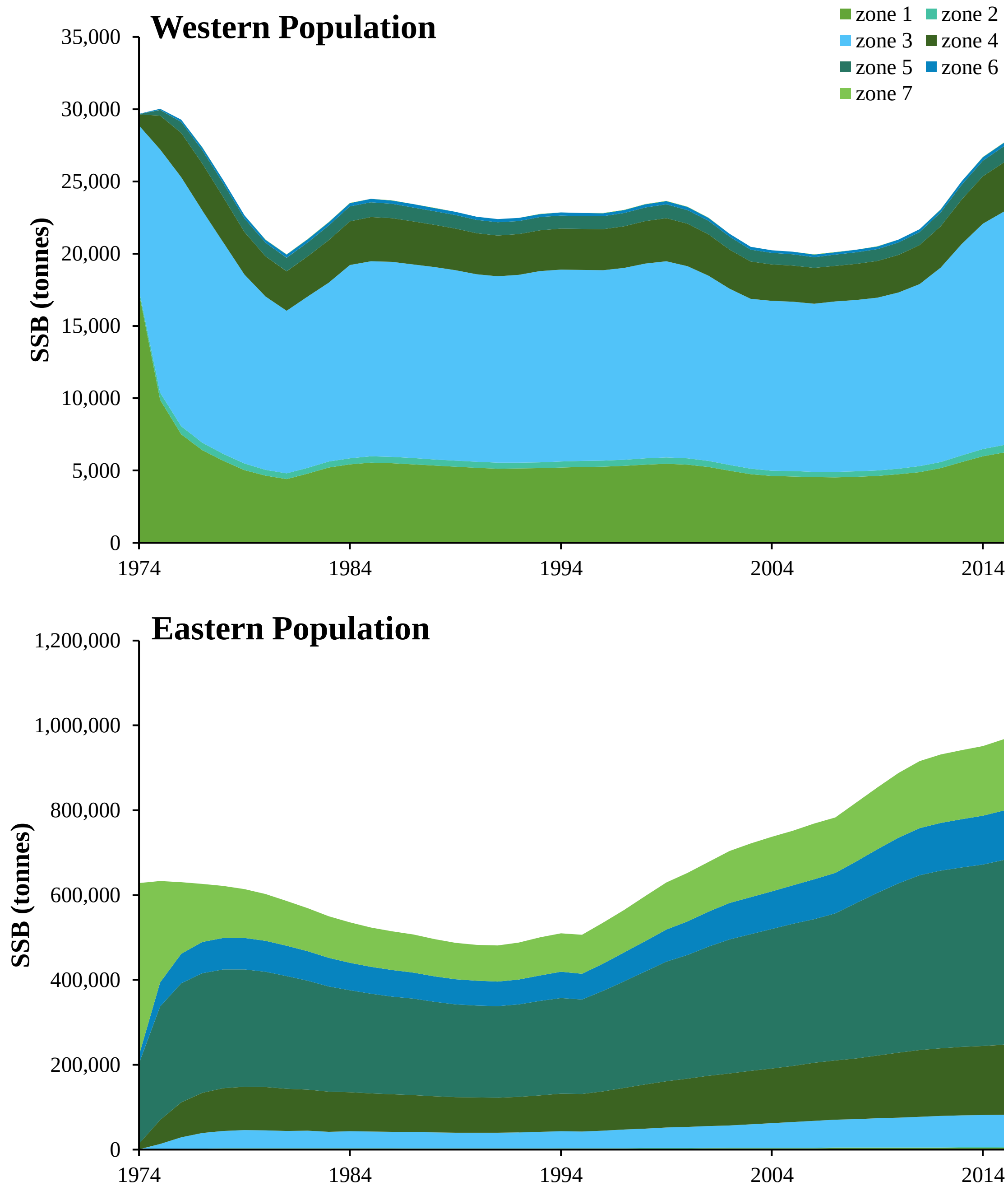
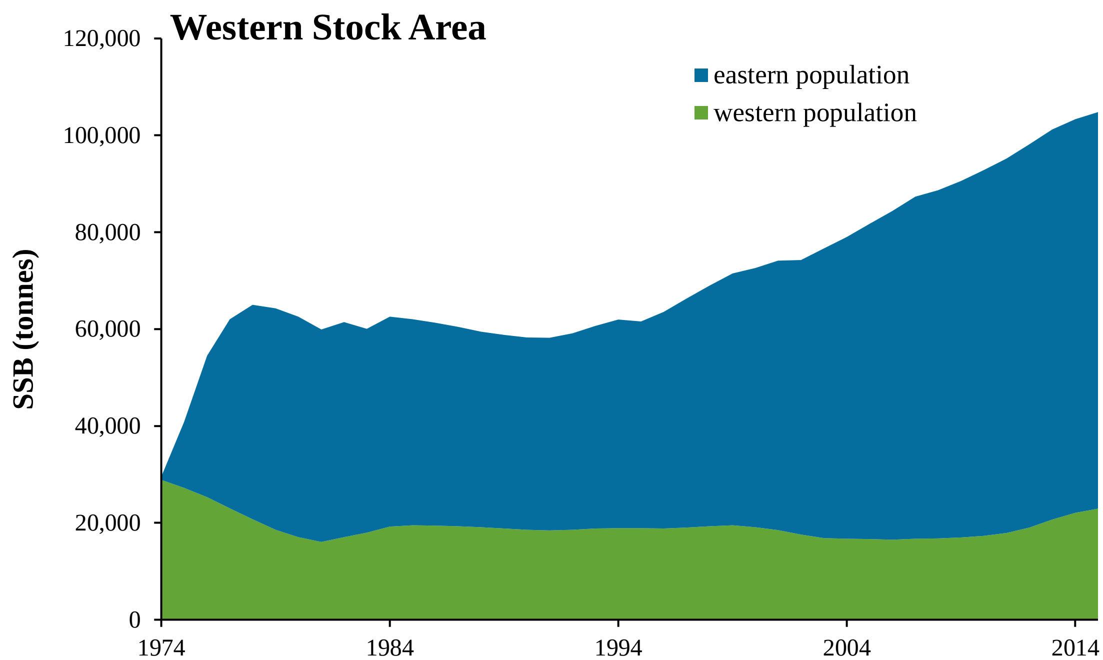


Figure C4. Third quarter SSB of the western and eastern populations among the seven geographic zones in the operating model over time. (See **Figure 1** for all zone descriptions.)



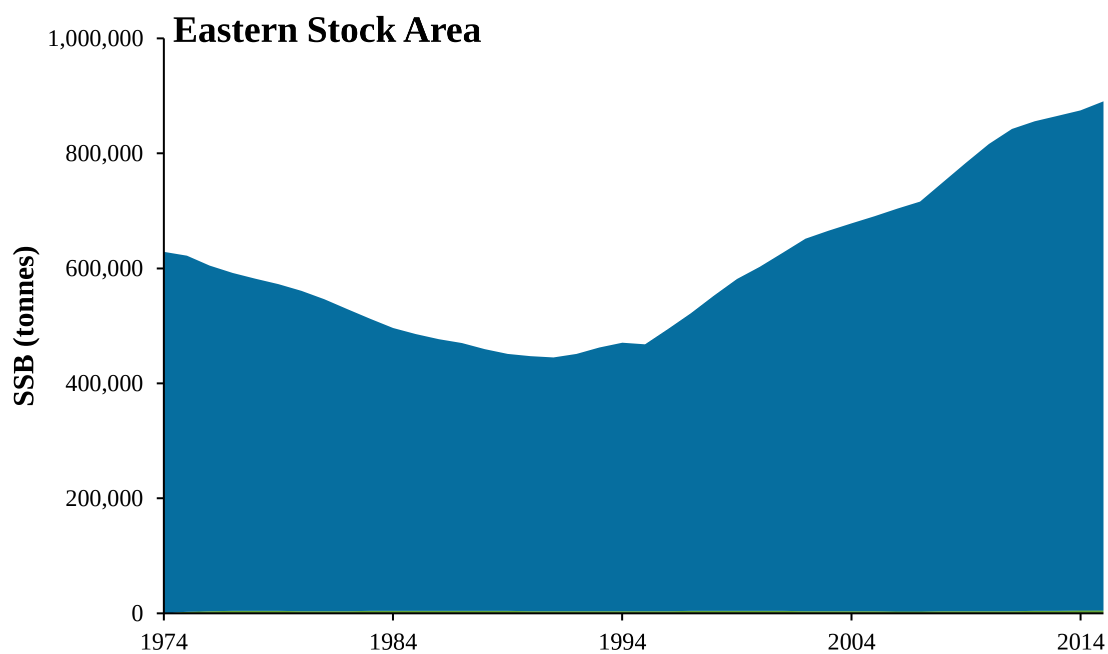


Figure C5. Third quarter SSB of western and eastern population fish in the western and eastern stock areas in the operating model over time.

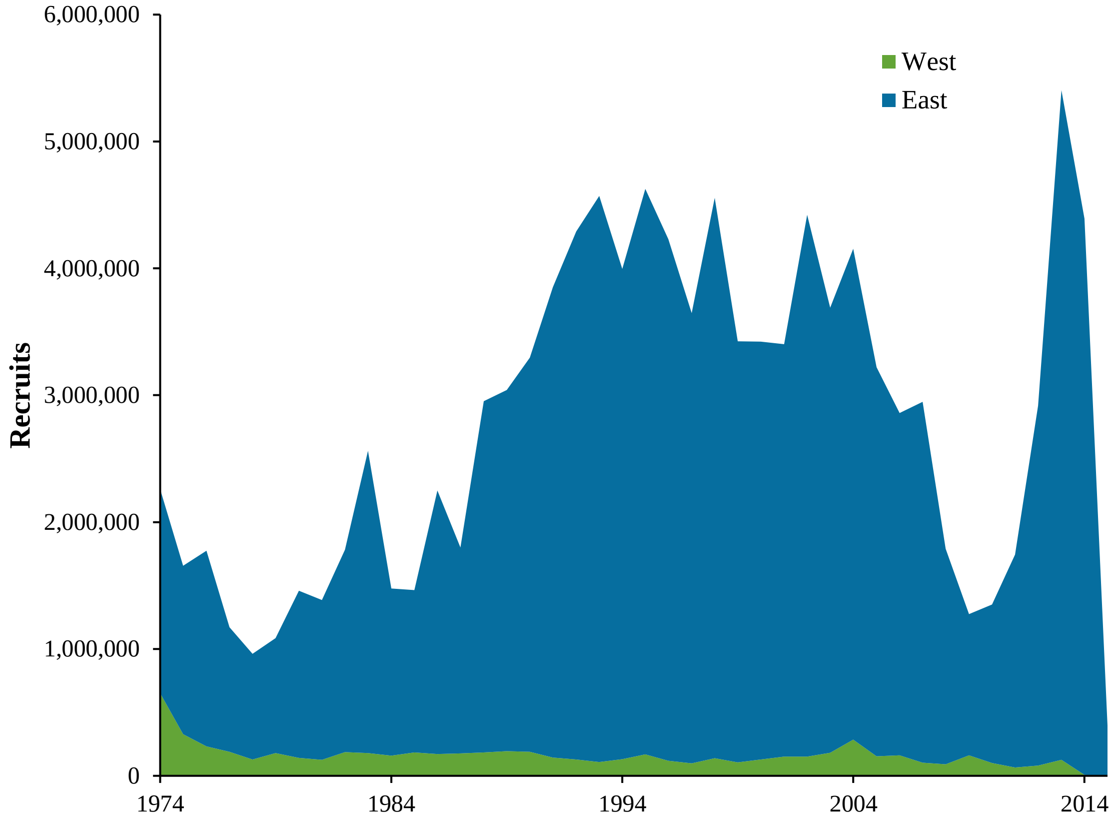


Figure C6. Number of recruits (age 1 fish) from the western and eastern populations from the first quarter in the respective spawning zones 1 and 7 (Gulf of Mexico and Mediterranean Sea) over time.

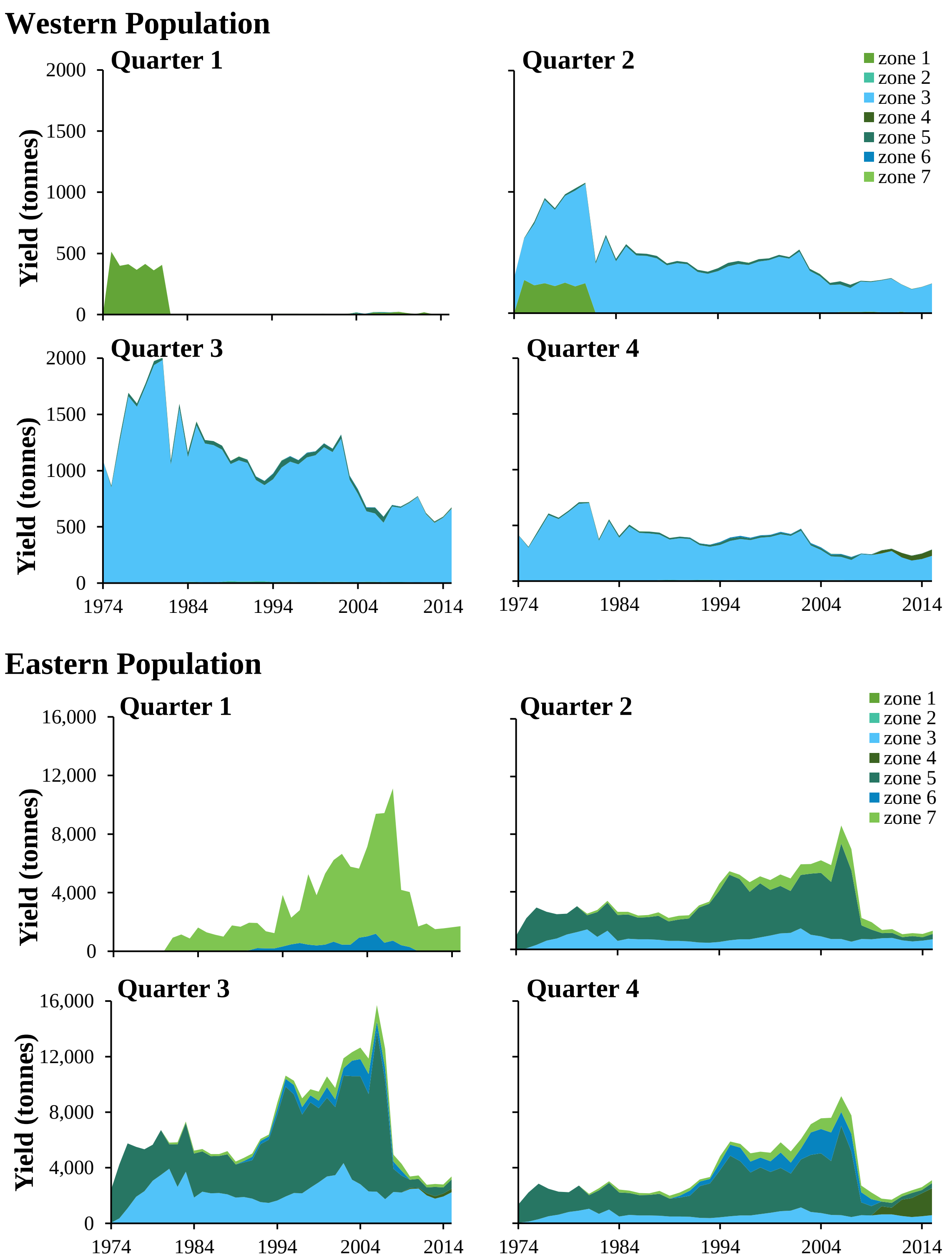


Figure C7. Comparison of fishery yield of the western and eastern populations in the operating model among the four seasonal quarters and seven geographic zones over time. (See **Figure 1** for all zone descriptions.)