2022 BLUEFISH RESEARCH TRACK ASSESSMENT

PARTICIPANTS

Working Group

NAME	AFFILIATION
Michael Celestino	NJFW
Karson Cisneros	MAFMC
Katie Drew	ASMFC
Sam Truesdell	MADMF
Abigail Tyrell	NEFSC/Ocean Associates Inc.
Jessica Valenti	NOAA/Rutgers
Samantha Werner	NEFSC
Tony Wood	NEFSC

Chair-invited analytical participants

Sarah Gaichas	NEFSC
Jim Gartland	VIMS
Tim Miller	NEFSC
Joe Myers	ACCSP

Working Group meeting attendees

Alan Bianchi	NC-DEQ
Greg DiDomenico	Lund's Fisheries
Cynthia Ferrio	GARFO
James Fletcher	Unk
Jesse Hornstein	NYDEC
Nathan Jackson	Unk
Cynthia Jones	ODU
Mike Waine	American Sportfishing Association

ACKNOWLEDGEMENTS

The Working Group thanks: Sarah Gaichas and her collaborators on development of the forage fish index including Jim Gartland, Brian Smith, Elizabeth Ng, James Thorson. We thank Tim Miller for assistance with WHAM exploration, development, and support. We also thank Joe Myers, ASMFC's Bluefish Technical Committee and Fishery Management Plan Coordinator, Dustin Colson Leaning, for their data collection and analysis support as well as participating in the ASMFC State Data Review Workshop for bluefish. Michele Traver and Alexander Dunn provided helpful administrative support. Matt Seeley participated in the early WG meetings before moving on to the Western Pacific Fishery Management Council; José Moñtanez provided helpful interim support after M Seeley's departure. Eric Robillard contributed important information on bluefish ageing and oversaw or conducted the majority of the bluefish ageing for this assessment. Lou Carr-Harris assisted with development of a research recommendation.

EXECUTIVE SUMMARY

Term of Reference (TOR) #1: Identify relevant ecosystem and climate influences on the stock. Characterize the uncertainty in the relevant sources of data and their link to stock dynamics. Consider findings, as appropriate, in addressing other TORs. Report how the findings were considered under impacted TORs.

Temperature and photoperiod are the principal factors directing activity, migrations, and distribution of adult bluefish. Based on this mechanistic connection, quantitative indicators of optimal temperature were developed to better understand temperature trends during the bluefish spawning season. Sources of uncertainty are discussed. Analyses suggested that the spawning season may now extend later in the year compared to historical periods, though it is unclear how these changes in potential spawning season may affect bluefish recruitment. On the other hand, the amount of habitat in the optimal temperature range during the peak spawning month of July has not changed over time, indicating stability in spawning conditions and therefore possibly also in recruitment. A Vector Autoregressive Spatiotemporal (VAST) model was developed from the fall NEFSC bottom trawl survey to determine the fall centers of gravity of three bluefish size groups over time; analyses suggested systematic trends in large and medium bluefish, but not small bluefish. Temperature was tested as a covariate in the VAST model, but resulting poor model diagnostics were beyond the scope of the present working group to address.

Using a VAST framework, we also developed a forage fish index to evaluate changes in bluefish prey over time and space that could be used to inform survey and/or fishery availability in the bluefish stock assessment to inform annual deviations in catchability. Small pelagic forage species are difficult to survey directly, so we developed a novel method of assessing small pelagic fish aggregate abundance using predator diet data. The forage fish indices based on fall, spring, and annual datasets all show fluctuations in forage fish biomass, alternating between multiple years or decades with higher and lower levels.

Variability in bluefish life history processes was modeled by splitting life history data by semesters of the year, by decade, by geographic region, and by sex; results and sources of uncertainty are discussed. Natural mortality was updated for this assessment from one based on a "rule of thumb" estimate of 0.2 for all ages to Lorenzen weight-based age-varying estimates. Our findings were considered and/or incorporated into several subsequent TORs, including: spatial domain of the stock (TOR2), estimates of seasonal and regional catch weights (TOR2), development of survey indices of abundance with environmental covariates (TOR3), incorporation of the forage fish index into a companion assessment model (TOR4), updating natural mortality for use in the assessment model (TOR4), and informed several research recommendations (TOR7).

Term of Reference #2: Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.

The majority of commercial landings over the time series (1950-present) have been taken in the Mid-Atlantic region (New York, New Jersey, and North Carolina). The majority of recreational activity occurred from May to October, with specific seasonal patterns varying by state.

Recreational offshore (3-miles, or 4.8-km, or more from shore) areas account for only about 7% of total catch.

Total bluefish removals (total dead catch) have declined since the beginning of the time series. There was a slow increase from 1996 to 2010, but the declining trend has continued to the lowest values in the time-series in recent years. On average, commercial landings account for 14% of the total removals with commercial discards averaging only 0.2%. Dead commercial discards have not contributed to total removals in previous assessments, but since they have been identified as a source of uncertainty, they were included in this assessment. Total removals are dominated by the recreational fishery with recreational landings accounting for 71% of total removals, and recreational dead releases averaging 15% of total removals. The recreational dead release mortality rate was updated for this assessment through reexamination of the methods used in the previous assessment, and an updated literature review; the value changed from 15% to 9.4%. The recreational dead discard component of the catch was calculated using the season/region length frequency distributions developed from all of the recreational biological sampling data for released fish; this is a change from previous assessments to account for regional differences in fish size.

> *Term of Reference #3: Present the survey data used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, application of catchability and calibration studies, etc.) and provide a rationale for which data are used. Describe the spatial and temporal distribution of the data. Characterize the uncertainty in these sources of data.*

The WG participated in an ASMFC Bluefish Technical Committee workshop to review available state datasets. The WG explored standardizing fishery independent indices of abundance using environmental covariates in a GLM framework. However, the standardization process did not notably affect index trends or reduce interannual variability or index coefficients of variation, so the WG did not use the standardized indices in the base run and instead used the stratified arithmetic mean for surveys with a stratified random design and the geometric mean for surveys with a fixed station design. Bayesian hierarchical modeling was used to combine YOY indices into a single composite index, using the method developed by Conn (2010) that represents the coast wide recruitment dynamics of bluefish. Surveys included in the composite index were from NH Juvenile Finfish Seine Survey, RI Narragansett Bay Juvenile Finfish Beach Seine Survey, NY Western Long Island Seine Survey, NJ Delaware River Seine Survey, MD Juvenile Striped Bass Seine Survey, and VIMS Juvenile Striped Bass Seine Survey. In addition, the bluefish working group decided on 8 additional representative indices of bluefish abundance for the assessment:

1. NEFSC Fall inshore strata: 1985-2008 (age-0 – age-6+)
2. NEFSC Fall outer inshore strata (FSV Bigelow): 2009-2021 (age-0 – age-6+)
3. NEAMAP Fall Inshore trawl survey: 2007-2021 (age-0 – age-6+)
4. ChesMMAP trawl survey: 2002-2018 (age-0-3)
5. Pamlico Sound Independent Gillnet Survey; 2001-2021 (age-0 – 6+)
6. Marine Recreational Information Program CPUE: 1985-2021 (age-0 – age-6+)
7. SEAMAP Spring Inshore trawl survey: 1989-2021 (age-1)
8. SEAMAP Fall Inshore trawl survey: 1989-2021 (age-0)

Calculation of the MRIP CPUE was updated for this assessment. Bluefish trips were defined using a guild approach where a trip was considered a bluefish trip if it caught either bluefish or a species that was significantly positively associated with bluefish. This was a change from the previous benchmark assessment where effort was described using “directed trips,” which describe trips where bluefish were considered a target species.

Multinomial age length keys were also explored as part of this assessment. Seasonal multinomial age length keys (ALKs) reduced retrospective trends and improved convergence diagnostics in statistical catch at age models relative to alternative ALKs; additionally, the WG did not believe data were sufficient for higher resolution (e.g., regional) ALKs, and so seasonal multinomial ALKs were selected for use in the assessment.

> *Term of Reference #4: Use appropriate assessment approach to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Compare the time series of these estimates with those from the previously accepted assessment(s). Evaluate a suite of model fit diagnostics (e.g., residual patterns, sensitivity analyses, retrospective patterns), and (a) comment on likely causes of problematic issues, and (b), if possible and appropriate, account for those issues when providing scientific advice and evaluate the consequences of any correction(s) applied.*

The Woods Hole Assessment Model (WHAM), a state-space, age structured stock assessment model, was used as the base model to estimate annual fishing mortality, recruitment, stock biomass, and associated estimates of uncertainty, with data updated through 2021. A suite of model fit diagnostic plots were examined for each model of interest and model fits were examined using conventional residual diagnostics, as well as one-step ahead residual diagnostics. Retrospective patterns in model results were evaluated using Mohn’s rho values.

The final model configuration included a number of notable model and data changes since the previous peer reviewed model, including: a state-space model, updated natural mortality estimate, addition of new indices, including a newly estimated MRIP CPUE index, and addition of several selectivity blocks. Spawning stock biomass from the final base model starts in 1985 high and declines through the late 1990s, remains stable for several years before rising to a localized peak in 2008, declining through 2018, and rising in the years since. This pattern broadly reflects trends from the previously accepted model, albeit with differences in scale. Fishing mortality from the base model starts low in 1985 and rises quickly, then declines and varies without trend over much of the timeseries; fishing mortality reached a high in 2017, and has declined to timeseries lows since. The trend from the previously accepted model is broadly similar, albeit again, with some differences in scale, primarily in estimates of recruitment.

WHAM allows for incorporation of environmental covariates on the catchability of survey indices, and we explored a companion model that leveraged this capability. The companion model that used the forage fish index as a covariate on catchability of the MRIP index showed promise for continued development. The covariate led to an overall decreasing trend in catchability over time.

Term of Reference #5: Update or redefine status determination criteria (SDC; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY reference points) and provide estimates of those criteria and their uncertainty, along with a description of the sources of uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for reference points. Compare estimates of current stock size and fishing mortality to existing, and any redefined, SDCs.

Existing status determination criteria from the 2021 management track assessment (data through 2019) were F_{MSY} proxy = $F_{35\%}$ = 0.181 and SSB_{MSY} = 201,729 MT (1/2 SSB_{MSY} = $SSB_{THRESHOLD}$ = 100,865 MT). Updated reference points from the ASAP continuity run are F_{MSY} proxy = $F_{35\%}$ = 0.176 and SSB_{MSY} = 190,771 MT (1/2 SSB_{MSY} = $SSB_{THRESHOLD}$ = 93,386 MT).

Both $F_{35\%}$ and $SSB_{35\%}$ were calculated in WHAM using average recruitment over the time series (1985-2021), and 5-year averages for fishery selectivity, maturity and weights-at-age for SSB per recruit calculations. Reference points from the final model (BF28W_m7) were F_{MSY} proxy = $F_{35\%}$ = 0.248 (95% CI: 0.209 – 0.299) and SSB_{MSY} proxy = $SSB_{35\%}$ = 91,897 MT (95% CI: 66,219–127,534 MT); $SSB_{THRESHOLD}$ =1/2 SSB_{MSY} proxy = 45,949 MT (95% CI: 33,110–66,768 MT). The retrospectively adjusted values of terminal year F and SSB were within the 90% confidence bounds of the unadjusted values, indicating a retrospective adjustment was not necessary to determine stock status. The terminal year SSB was 55,344 MT (95% CI: 35,185 – 87,052 MT) which was above the $SSB_{THRESHOLD}$ and 60% of SSB_{MSY}. Full fishing mortality was 0.166 (95% CI: 0.103 – 0.268) in 2021, which was 67% of the $F_{35\%}$ reference point. Stock status determination based on the final model indicates that there is an 87% chance that the bluefish stock is currently not overfished and over-fishing is not occurring.

Status determination criteria	2021 Management track assessment	2022 research track assessment (continuity run)	2022 research track assessment (WHAM)
F_{MSY} proxy = $F_{35\%}$	0.181	0.176	0.248
SSB_{MSY}	201,729 MT	190,771 MT	91,897 MT
½ SSB_{MSY}	100,865 MT	93,386 MT	45,949 MT

Term of Reference #6: Define appropriate methods for producing projections; provide justification for assumptions of fishery selectivity, weights at age, maturity, and recruitment; and comment on the reliability of resulting projections considering the effects of uncertainty and sensitivity to projection assumptions.

Short-term projections were conducted in WHAM, and incorporated model uncertainty and auto-regressive processes in recruitment and numbers-at-age. The projections used 5-year averages for natural mortality, maturity, fishery selectivity and weights-at-age. Removals in 2022 were assumed to be equal to the 2022 ABC (11,460 MT), and projections were carried forward for years 2023-2025 with different fishing mortality and harvest assumptions: F = 0, $F_{status\ quo}$ = 0.166, $F_{35\%}$ = 0.248, and that harvest in each year is equal to the acceptable biological catch (ABC) in each year. The probability of SSB in 2025 being above the SSB threshold is > 80% for all scenarios explored. Catch advice will be updated as part of the 2023 Management Track assessment, but catch advice from WHAM under the most likely scenario explored for this research track assessment (MAFMC risk policy assuming CV = 100%) is expected to be stable, but lower, relative to 2022.

Term of Reference #7: Review, evaluate, and report on the status of research recommendations from the last assessment peer review, including recommendations provided by the prior assessment working group, peer review panel, and SSC. Identify new recommendations for future research, data collection, and assessment methodology. If any ecosystem influences from TOR 1 could not be considered quantitatively under that or other TORs, describe next steps for development, testing, and review of quantitative relationships and how they could best inform assessments. Prioritize research recommendations.

The SAW 60 WG reviewed the status of previous research recommendations and proposed new ones to address issues raised during WG meetings. Notable accomplishments relative to past research recommendations include: development of an MRIP index using a species-association method to identify bluefish trips, updating the estimate of natural mortality used in the assessment model, evaluating model results that aggregated all model input data at a seasonal and regional level of resolution, multiple fishery independent surveys were combined using VAST as part of this assessment, examination of differences in the calibrated and uncalibrated MRIP estimates of bluefish catch, spatial stratification of recreational release length frequencies when calculating the weight of dead recreational releases, and the migration to the WHAM framework will allow for continued exploration and testing of covariates influencing time-varying catchability and selectivity.

The WG proposed several new research recommendations to better understand bluefish dynamics and assessing the population through the current or future models. These include the following: expand collection of recreational release length frequency data, continue development and refinement of the forage fish / availability index as well as incorporation of this index in to a base model for bluefish management advice, initiate additional fisheries-independent surveys or fishery-dependent sampling programs to provide information on larger, older bluefish, continue coastwide collection of length and age samples from fishery-independent and -dependent sources, refinement and development of indices of abundance, and develop a recreational demand model.

Term of Reference #8: Develop a backup assessment approach to providing scientific advice to managers if the proposed assessment approach does not pass peer review or the approved approach is rejected in a future management track assessment.

A backup assessment approach is required to be in place as a hedge against a scenario where the primary catch-at-age model is not suitable for providing management advice. The bluefish Working Group chose the index-based method Ismooth (previously known as PlanBSmooth) as the backup model due to its performance in the analyses performed by the Index Based Model Working Group (NEFSC 2020) and because it has a history of application at the NEFSC as an approach that has been used to develop ABCs (e.g., Georges Bank cod, Gulf of Maine / Northern

Georges Bank and Southern Georges Bank / Mid-Atlantic monkfish). Briefly, this approach applies recent trends in an index or indices to recent dead catch to generate ABC advice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 ECOSYSTEM AND CLIMATE INFLUENCES

Term of Reference (TOR) #1: Identify relevant ecosystem and climate influences on the stock. Characterize the uncertainty in the relevant sources of data and their link to stock dynamics. Consider findings, as appropriate, in addressing other TORs. Report how the findings were considered under impacted TORs.

1.1 Ecosystem and Socioeconomic Profile (ESP)

An Ecosystem and Socioeconomic Profile (ESP) was used as a framework to address TOR1 in this research track assessment. ESP is a standardized framework to facilitate the inclusion of ecosystem and socioeconomic information in the stock advice process; it leverages existing information to understand the ecological and socioeconomic drivers of stock dynamics and to incorporate this diverse information into the stock advice process through the creation of ecosystem and socioeconomic indicators. This standard framework also facilitates the interpretation of data and allows future working groups to update the existing indicators in addition to creating and assessing new indicators.

The ESP process begins with (1) a systematic review of existing ecosystem and socioeconomic literature and identification of problem statements for the stock, followed by (2) development of conceptual models to outline the major drivers on the stock, (3) creation of indicators relevant to stock performance, (4) analysis of select indicators, and, lastly, (5) reporting out scientific advice. The scientific advice provided by an ESP can inform the stock assessment in multiple ways, ranging from providing additional context and research recommendations, to suggesting new covariates to include that can inform dynamic processes within assessment modeling.

The bluefish ESP includes a comprehensive literature review of bluefish life history and related ecosystem considerations relating to bluefish habitat, distribution, diet, predators, competitors, growth, and survival at each life stage. It also served as a review of the history of the bluefish stock assessment and relevant biological information that is used to make decisions relating to the assessment modeling. A conceptual model identifying the major drivers for different life stages of bluefish was developed from this review (Figure 1). Diet data collected from multiple scientific surveys were analyzed to determine the major prey and predators of bluefish, supplementing the literature review on this topic with the most recent data. Distributional and environmental data from multiple state and federal surveys were analyzed to understand where, when, and under what conditions bluefish of different life stages and size classes were found. Ecosystem and socioeconomic indicators were developed to better understand the current status of bluefish in the context of each of these dimensions, as well as to begin to probe potential mechanistic linkages between the environment and the status of the bluefish stock. Relevant results are summarized below; see Working Paper 1 (Tyrell et al. 2022) for the detailed report.

1.1.1 Diet

In the Northeast Fisheries Science Center (NEFSC) bottom trawl data, anchovies, butterfish, and squid were important prey items in all years. Sandlance, herring, bluefish, scup, and drum were important prey in some years in the NEFSC bottom trawl. Bay anchovy, butterfish, and striped anchovy were important prey species in the Northeast Area Monitoring and Assessment Program (NEAMAP) bottom trawl. Bay anchovy, spot, and menhaden were important prey species in the

Chesapeake Bay Multispecies Monitoring & Assessment Program (ChesMMAP) bottom trawl. Overall, >80% of bluefish diet was composed of fish, both by weight and by abundance. There were few records of bluefish in the stomachs of other species captured and sampled in these surveys.

1.1.2 Environment, Spatial Distribution, and Cohorts

Adult and juvenile bluefish are found primarily in waters less than 20 meters deep along the Atlantic coast (Shepherd and Packer 2006). The 2022 bluefish research track assessment Working Group (referred to herein as WG) investigated whether Gulf of Mexico bluefish were part of the unit stock being assessed for the 2022 research track assessment. This investigation did not identify any known systematic studies (e.g., tagging, genetics) that demonstrated bluefish migrations into or out of the Gulf of Mexico. A review of the Florida Fish and Wildlife's acoustic receiver network and the American Littoral Society's volunteer angler tagging program indicated that no bluefish tagged on the Atlantic coast were ever recaptured in the Gulf of Mexico. Marine Recreational Information Program (MRIP) and commercial landings queries indicated that, on average, total bluefish harvest or removals (landings plus dead releases) in the Gulf of Mexico is 3-4% of combined Gulf and Atlantic coast bluefish removals. Finally, a query of recreational harvest length frequency suggested similarities between the two regions, which did not support a WG hypothesis that "missing" lengths in some observed periodic bimodal length frequency distributions on the Atlantic coast might reside in the Gulf of Mexico. Therefore, no data suggest that the Gulf of Mexico is an important habitat for Atlantic bluefish.

MRIP data and state and federal scientific surveys supported the seasonal migration pattern of bluefish, with fish observed in more southern locations in the winter and migrating northward in spring and summer. Spawning was also recorded in spring and summer, with eggs observed in some years in May through August. Length data from scientific surveys supported the presence of multiple sub-annual young-of-the-year cohorts in some years, although precise quantification of spring-spawned versus summer-spawned cohorts was generally not possible due to spatiotemporal variation in sampling effort and low sample sizes. Juveniles may be estuarine dependent (Munch 1997) although they also occur in nearshore ocean waters (Taylor et al. 2006); juvenile habitat use may vary by cohort (Taylor et al. 2007; Wuenschel et al. 2012). Adults use both estuarine and ocean environments and favor warmer water temperatures although they are found in a variety of hydrographic environments (Ross 1991; Shepherd and Packer 2006; Wuenschel et al. 2012). Small (≤30.3 cm) bluefish were generally found in the highest abundance along the Atlantic coast between Long Island and North Carolina. Medium (30.3-50.0 cm) bluefish were generally found in the highest abundance along the Atlantic coast as well as on Georges Bank. Large (≥50.0 cm) bluefish were generally found in the highest abundance in Southern New England and on Georges Bank.

In recent years, stakeholders have reported that larger bluefish are staying offshore and are less abundant inshore. The American Littoral Society Fish Tagging Program's bluefish data were analyzed to assess whether larger fish (> 18 inches or 46 cm) are being tagged and released or recaptured more frequently offshore. Analyses did not show that larger fish are being tagged and released or recaptured more frequently offshore in recent years (Working Paper 2, Valenti 2022a); however, very few large bluefish were tagged and released or recaptured in the last five years, so the sample size was notably small. The disparity between the stakeholder reports and the results of this analysis could be due to limitations inherent to volunteer fish tagging program data, including low sampling of bluefish and variability in angler effort and reporting.

A Vector Autoregressive Spatiotemporal (VAST; Thorson and Barnett (2017); Thorson (2019)) model was developed from the fall NEFSC bottom trawl survey to determine the fall centers of gravity of three bluefish size groups over time. Center of gravity analyses showed that medium bluefish are moving north and east at an average rate of 1.1 km/year, and large bluefish are moving north at an average rate of 0.2 km/year and east at an average rate of 0.5 km/year. The center of gravity of small bluefish did not have a trend. This distribution change may support anecdotes about large bluefish moving offshore in recent years. Further research is needed to fully understand bluefish distribution, as the 2020 and 2021 NEFSC fall bottom trawl surveys did not catch enough bluefish to be included in this study. Additionally, temperature was tested as a covariate in the VAST model, but resulting poor model diagnostics were beyond the scope of the present working group to address. See Working Paper 3 (Tyrell 2022) for more details.

1.1.3 Ecosystem Indicators

1.1.3.1 Temperature

Bluefish can tolerate temperatures ranging from approximately 11°-30.4°C, however they exhibit stress, such as an increase in swimming speed, at both extremes (Olla and Studholme 1971; Klein-MacPhee 2002). The literature indicated temperature and photoperiod are the principal factors directing activity, migrations, and distribution of adult bluefish (Olla and Studholme 1971, Taylor et al. 2007). Based on this mechanistic connection, quantitative indicators of optimal temperature were developed to better understand temperature trends during the bluefish spawning season.

The spawning season may now extend later in the year compared to historical periods. Bluefish spawning has been recorded at 18-25.6°C (Norcross et al. 1974). In the greater Mid Atlantic Bight and Southern New England regions, the first day when 75% or more of the sea surface reaches 18°C has remained stable over time, while the last day when 75% of the sea surface is above 18°C has occurred later in the year over time, currently persisting into mid and late October (in contrast to the beginning of October in the 1980s); the total number of days with 75% or more of the sea surface above 18°C has increased over time. It is unclear how these changes in potential spawning season may affect bluefish recruitment. There were no notable correlations between first, last, or number of days above 18°C and the composite young-of-the-year index used in the model (Section 3.1.2) or modeled recruitment (NEFSC 2019). However, the surveys used to characterize young-of-the-year bluefish may not fully document spawning that occurs in the fall due to a mismatch in survey timing and the recruitment of fall-spawned bluefish to estuaries; furthermore, most surveys do not capture the smallest bluefish (<15 cm length) that have been spawned most recently.

In contrast to trends in the potential spawning season, the amount of habitat in the optimal temperature range during the peak spawning month of July has not changed over time, indicating stability in spawning conditions and therefore possibly also in recruitment; however, the amount of habitat with colder-than-optimal temperatures (<18°C) has decreased, while the amount of habitat with warmer-than-optimal temperatures (>25.6°C) has increased. The amount of area in the Central Atlantic with optimal bluefish spawning temperatures in July was marginally positively correlated with bluefish recruitment (modeled recruitment in the 2015 assessment), while the amount of area with warmer-than-optimal temperatures was negatively correlated with bluefish recruitment. Although the amount of area with optimal temperatures in July has remained consistent over time, future ocean warming may eventually decrease the proportion of the Central Atlantic with optimal bluefish spawning temperatures as more areas warm above 25.6°C and fewer or no areas are left below 18°C.

1.1.3.2 Natural Mortality

Proxies for natural mortality were investigated to the extent possible. Relative condition of the small, medium, and large bluefish size groups was determined over time. Condition of large bluefish was found to be increasing over time, while the condition of medium bluefish had no change over time. Relative condition of small bluefish decreased slightly in the spring only, but remained above one, indicating good condition. Condition could be considered as a proxy for natural mortality, and generally indicates that mortality sources other than fishing have not increased compared to historical conditions, supporting the use of time-invariant natural mortality in the assessment model.

Bluefish predators are not well sampled, but existing data suggest that bluefish are not currently experiencing higher predation risk relative to historical conditions.

1.1.3.3 Condition and Recruitment

An increasing trend in the relative condition of large bluefish may be beneficial for bluefish recruitment, as larger and fatter bluefish may produce more eggs and/or more high-quality eggs; however, further research is needed to quantify the relationships between fecundity and length, weight, and age, which are currently not well documented.

1.1.4 Socioeconomic Indicators

Despite lower catches in recent years, bluefish remains one of the top recreational fisheries on the U.S east coast in terms of total catch, and therefore likely helps support a robust recreational fishing industry.

Although management was fairly stable in terms of catch limits and trip limits until the bag limit reductions implemented in 2020, the recreational fishery has shifted to catch-and-release rather than catch for harvest. Recreational landings in weight have decreased over time, with landings in 2021 being less than 10% of landings in 1981. Over the same time period, the total recreational catch (harvest plus all released bluefish) has decreased from 76 million fish in 1981 to 30 million fish in 2021.

Neither commercial nor recreational catch typically exceed catch limits, though in both 2020 and 2021 there were recreational catch limit overages of 32% in 2020 and 41% in 2021. However, the Acceptable Biological Catch (ABC) has generally decreased each year since it was implemented in 2011 due to stock condition. Therefore, recent decreases in catch and landings may be attributable to management actions rather than lack of interest in the bluefish fishery.

1.2 Forage Fish Index

The objective of this work was to create a forage fish index to evaluate changes in bluefish prey over time and space that could be used to inform survey and/or fishery availability in the bluefish stock assessment to inform annual deviations in catchability. Changing distribution and abundance of small pelagic prey may drive changes in predator distributions, affecting predator availability to fisheries and surveys. However, small pelagic forage species are difficult to survey directly, so we developed a novel method of assessing small pelagic fish aggregate abundance using predator diet data.

We used piscivore diet data collected from multiple bottom trawl surveys within a Vector Autoregressive Spatio-Temporal (VAST) model (Thorson and Barnett 2017, Thorson 2019) to assess trends of small pelagic forage species on the Northeast US shelf. This approach uses survey-sampled predator stomach contents as observations to develop a survey index for forage fish, following Ng et al. (2021), which used predator stomach data to create a biomass index for a single prey, Atlantic herring.

We adapted the approach of Ng et al. (2021) to generate an index for bluefish prey in aggregate rather than a single prey species. Further, we include inshore and offshore regions by combining two regional bottom trawl surveys, the NEFSC survey and the NEAMAP survey, as was done previously for summer flounder biomass (Perretti and Thorson 2019). Finally, since bluefish themselves are sparsely sampled by the surveys, we aggregate all predators that have a similar diet composition to bluefish to better quantify bluefish prey biomass.

Methods and results are summarized below; for more detail, see Working Paper 4 (Gaichas et al. 2022).

1.2.1 Forage Fish in Bluefish Diets

Using NEFSC bottom trawl survey diet data from 1973-2021, 20 small pelagic groups were identified as major bluefish prey, with 10 or more observations in bluefish stomachs over the entire 48-year period. In descending order of observations, bluefish prey are: longfin squids (*Doryteuthis* formerly *Loligo* sp.), anchovy family (Engraulidae), bay anchovy (*Anchoa mitchilli*), Atlantic butterfish (*Peprilus triacanthus*), Cephalopoda, striped anchovy (*Anchoa hepsetus*), red eye round herring (*Etrumeus teres*), sandlance (*Ammodytes* sp.), scup (*Stenotomus chrysops*), silver hake (*Merluccius bilinearis*), shortfin squids (*Illex* sp.), Atlantic herring (*Clupea harengus*), herring family (Clupeidae), bluefish (*Pomatomus saltatrix*), silver anchovy (*Engraulis eurystole*), longfin inshore squid (*Doryteuthis pealeii*), Atlantic mackerel (*Scomber scombrus*), flatfish (Pleuronectiformes), weakfish (*Cynoscion regalis*), and Atlantic menhaden (*Brevoortia tyrannus*).

Prey categories such as “fish unidentified”, “Osteichthyes”, and “unidentified animal remains” were not included in the prey list. Although unidentified fish and Osteichthyes can comprise a significant portion of bluefish stomach contents, unidentified fish in other predator stomachs may not represent the same types of unidentified fish in bluefish stomachs.

1.2.2 Predators Feeding Similarly to Bluefish

All size classes of 50 fish predators captured in the NEFSC bottom trawl survey were grouped by diet similarity to identify the size classes of piscivore species with the most similar diet to bluefish in the region. Diet similarity analysis was completed using the Schoener similarity index (Schoener 1970; B. Smith, pers. comm.), and is available via the NEFSC food habits shiny app. The WG evaluated several clustering methods to develop the predator list (see this link with detailed cluster results).

The nineteen predators with highest diet similarity to bluefish from the NEFSC diet database (1973-2020) included Atlantic cod, Atlantic halibut, buckler dory, cusk, fourspot flounder, goosefish, longfin squid, shortfin squid, pollock, red hake, sea raven, silver hake, spiny dogfish, spotted hake, striped bass, summer flounder, thorny skate, weakfish, and white hake.

The NEAMAP survey operates closer to shore than the current NEFSC survey. The NEAMAP dataset includes predators sampled by the NEFSC survey and adds two species, Spanish mackerel and spotted sea trout, not captured by the NEFSC survey offshore but included as bluefish-like predators based on WG expert judgement of diet similarity to bluefish. Predator size classes included are listed in Table 2 of Working Paper 4 (Gaichas et al. 2022).

1.2.3 Datasets

The mean weight of forage fish per predator stomach at each location was calculated by combining weight across the 20 forage fish (bluefish prey) groups found in stomachs from all 22 piscivores (including bluefish) at each surveyed location. Data for each station included station ID, year, season, date, latitude, longitude, vessel, mean bluefish prey weight (g), mean piscivore length (cm), number of piscivore species, and sea surface temperature (degrees C). Because approximately 10% of survey stations were missing in-situ sea water temperature measurements, National Oceanic and Atmospheric Administration Optimum Interpolation Sea Surface Temperature (NOAA OI SST) V2 High Resolution Dataset (Reynolds et al. 2007) data provided by the NOAA PSL, Boulder, Colorado, USA, from their website at https://psl.noaa.gov were used to fill gaps. For survey stations with in-situ temperature measurements, the in-situ measurement was retained. For survey stations with missing temperature data, OI SST was substituted for input into VAST models.

Models were developed combining all data for the year (“Annual”) and with separate data for “Spring” (collection months January - June) and “Fall” (collection months July-December) to align with seasonal stratification used in the bluefish stock assessment. Modeled years included 1985-2021 to align with other data inputs in the bluefish stock assessment.

1.2.4 VAST Modeling

VAST is structured to estimate fixed and random effects across two linear predictors, which are then multiplied to estimate an index of the quantity of interest. Following the methods of Ng et al. (2021), we applied a Poisson-link delta model to estimate expected prey mass per predator stomach. We used a higher resolution (500 knots, estimated by k-means clustering of the data) to define the spatial dimensions of each seasonal model. Two step model selection first compared whether the data supported estimation of spatial and spatio-temporal random effects, and then evaluated whether catchability covariates improved fits. Best fit models included spatial and spatio-temporal random effects, with predator mean length, number of predator species, and sea surface temperature as catchability covariates; that is, these covariates all influenced the observation process rather than the distribution or abundance of prey. Detailed results of model selection are available in Working Paper 4 (Gaichas et al. 2022).

Similar to findings of Ng et al. (2021), a vessel effect was not supported, but the inclusion of the predator length covariate may more directly account for vessel differences in predator catch that affect stomach contents than modeling a vessel catchability covariate directly. Similar to our results, Ng et al. (2021) found that predator length covariates were strongly supported as catchability covariates (larger predators being more likely to have more prey in stomachs). In our aggregate predator dataset, we also found strong support for including the number of predator species in a tow as a catchability covariate. The rationale for including number of predator species is that more species "sampling" the prey field at a particular station may result in a higher encounter rate (more stomachs with bluefish prey). Water temperature was also supported as a catchability covariate, perhaps because temperature affects predator feeding rate and fish distribution.

1.2.5 Spatial Forage Indices

Spring, fall, and annual prey indices were split into inshore and offshore areas to quantify changing prey availability over time in areas available to the recreational fishery and the bottom trawl survey. First, we define a partition that includes survey areas relevant to the bluefish assessment (Mid Atlantic and Georges Bank). Within this partition,

- To evaluate bluefish availability to the NEFSC bottom trawl survey, two inshore-offshore strata partitions were created to account for the NEFSC survey vessel change in 2008. Inshore and offshore strata partitions included:
    - Albatross inshore stations (historically included in the Albatross NMFS bottom trawl index developed for the bluefish assessment)
    - Bigelow inshore bluefish index stations (historically included in the Bigelow NMFS bottom trawl index developed for the bluefish assessment)
    - Offshore bluefish index stations (the same for both vessels, and considered for addition to the NMFS bottom trawl bluefish indices in 2022)
- To evaluate bluefish availability to the MRIP catch-per-unit-effort (CPUE) index, recreational fishery strata partitions included:
    - shoreline to 3 miles offshore (state waters)
    - offshore of 3 miles (federal waters)

NEFSC survey strata definitions are built into the VAST "northwest-Atlantic" extrapolation grid. We defined additional new strata to address the recreational inshore-offshore 3-mile boundary, and incorporated them into a custom extrapolation grid so that the forage indices could be calculated and bias corrected (Thorson and Kristensen 2016) for all strata within VAST.

Full VAST model results for Fall, Spring, and Annual models, along with diagnostics, are available in Working Paper 4 (Gaichas et al. 2022). Here we show the forage fish index for the Fall model. The index is calculated for several regions relevant to the bluefish assessment:

- Albatross New (AlbNew) includes all inshore and new offshore survey strata (largest area)
- Albatross Old (AlbOld) includes all inshore survey strata
- Bigelow New (BigNew) includes the subset of inshore survey strata that can be sampled by the R/V Henry Bigelow plus new offshore strata
- Bigelow Old (BigOld) includes the subset of inshore survey strata that can be sampled by the R/V Henry Bigelow
- StateWaters includes the coastline to 3 nautical miles offshore (smallest area)

1.2.6 WHAM model example covariates: forage index time series, fall

Comparison of inshore and offshore spatial forage indices shows higher abundance of forage fish in state waters than in the subset of inshore strata that can be sampled by the R/V Henry Bigelow (Figure 2). Highest forage abundance is in the largest area, which includes all inshore survey strata as well as new offshore strata proposed for use in the bluefish assessment. The forage fish indices based on fall, spring, and annual datasets all show fluctuations in forage fish alternating between multiple years or decades with higher and lower levels. In general, the fall forage indices were higher at the beginning of the time series (mid-1980s), dropping to lower levels in the mid- and late-1990s, then showing cyclical fluctuations up until 2021, but never returning to high levels observed in the mid-1980s.

1.3 Life History Parameters

A single dataset was created from life history data collected by fishery-independent and fishery-dependent sampling by NMFS and Atlantic coast states and agencies. These data included ages, lengths, weights, and maturity observations. Life history processes were modeled, including mean length- and weight-at-age, modeled age-length relationships (e.g., Figure 3), allometric growth (e.g., Figure 4) and maturity-at-age (Table 1).

1.3.1 Age-Length and Length-Weight Relationships

Parameter estimates from the different life history models and expected values for size-, weight- and maturity-at-age and weight-at-length were generally consistent with analyses performed during the 2015 benchmark as well as with other previous research. Variability in the life history processes was modeled by splitting the data by semesters of the year, by decade, by geographic region (north and south, defined as Maine-Virginia and North Carolina-Florida, respectively), and by sex. Seasonal differences in length-at-age, weight-at-length, and maturity-at-age were apparent from these data; consistent with first principles, bluefish tended to be larger, weigh more and were more likely to be mature for their age during the second semester of the year relative to the first semester. Inter-decadal changes and differences by sex were less evident from the data. See Working Paper 5 (Truesdell et al. 2022) for more information including figures and tables outlining these findings.

1.3.2 Maturity-at-Age

The bluefish maturity schedule, in combination with the estimated total weight by age class, is used to estimate spawning stock biomass. The WG examined a variety of approaches to calculate maturity-at-age; all were based on using logistic generalized linear models (GLMs). The WG surmised through fitting models using different iterations of the data that the 2015 benchmark assessment had most likely employed federal data only (i.e., NMFS survey and port sampling data) to determine this ogive. The 2022 WG computed ogives using only federal data through 2021 and using state and federal data combined. The ogives that used both state and federal data estimated a mid-year maturity schedule by including time of year in the GLM; see Working Paper 5 (Truesdell et al. 2022) for more details. Ultimately, the differences in the versions of maturity ogives were not dramatic (Table 1) and the WG decided to use the same schedule that was implemented during the 2015 benchmark for primary model runs as this had been previously reviewed.

1.3.3 Natural Mortality

In the absence of direct natural mortality estimates for bluefish, the WG evaluated life history and longevity-based estimators of natural mortality. These methods included: maximum-age based methods, life history methods (using von Bertalanffy parameters), and length- or weight-based methods. The length and weight data described in the life history working paper (Truesdell et al. 2022a) were used for these calculations. For a detailed comparison of methods and natural mortality (M) estimates, see Working Paper 6 (Tyrell and Truesdell 2022).

The WG decided not to rely on natural mortality methods based on von Bertalanffy parameters, following the reasoning of Then et al. (2015). Based on the updated analyses using the maximum-age based methods of Hewitt and Hoenig (2005) and Then et al. (2015), the WG agreed that the "rule of thumb" estimate of 0.2 for all ages used in the 2015 benchmark assessment was too low. Ultimately, the WG decided to proceed with the Lorenzen weight-based age-varying natural mortality method because these estimates were in line with both the Hewitt-Hoenig (2005) and Then et al. (2015) estimates, and furthermore retained biological realism. The Lorenzen (1996) estimates using empirical weight-at-age (M range = 0.27-0.85) were higher at all ages than the age-constant value of 0.2 used in the previous benchmark assessment (Figure 5).

1.4 Uncertainty

Some of the ecosystem indicators identified in the ESP were not developed due to uncertainty around the underlying mechanistic connection to bluefish and/or the data source, and are described in detail in the ESP (Working Paper 1 Tyrell et al. 2022). For example, the WG proposed a large predator index to inform natural mortality, but could not locate sufficient data to create this index. The WG also identified overwinter survival as a bottleneck on survival to age 1, but could not develop any indicators of overwinter survival due to uncertainty in the locations where juvenile bluefish overwinter. Furthermore, the main source of long-term, large-scale fishery independent data for bluefish are bottom trawls (e.g., the NEFSC bottom trawls, NEAMAP), which are not well suited to capturing a large, fast-moving pelagic species like bluefish. As a result, the available data on the spatio-temporal distribution and movement of larger, older bluefish and the associated environmental indicators is limited. All of these proposed indicators are documented in the ESP and can be revisited as further data become available during future stock assessments.

The ESP identified several mechanistic linkages between the environment and bluefish stock dynamics and developed several quantitative environmental indicators. The development of these indicators is a first step towards including ecosystem variability in the assessment model to reduce uncertainty. With the bluefish stock assessment model now in a Woods Hole Assessment Model (WHAM) modeling framework, these environmental linkages can be tested in the future by including environmental covariates in model sensitivity runs, informing processes such as catchability, selectivity, natural mortality, and recruitment.

While the fall forage fish indices are temporally aligned with bluefish assessment inputs and both temporally and spatially aligned with two trawl survey indices used in the assessment, improvements in spatial overlap with recreational fisheries and other survey indices could be considered in the future to reduce uncertainty in associating forage fish with the bluefish MRIP abundance index. The current forage index does not cover inland waters, aside from Narragansett

Bay and Buzzards Bay. Diet data are available for Chesapeake Bay from the ChesMMAP survey, which could be added to the VAST model in the future. Less diet information is available for the portion of the bluefish range south of Cape Hatteras, although some collections have taken place. Investigation of sources of diet information, or possibly direct forage fish surveys for inland and southern areas would be worthwhile to see whether data are adequate to cover the full range of bluefish.

A key recommendation for future treatment of the life history data is to account for variability in spatio-temporal observations, as numerous fishery-independent and fishery-dependent sampling programs contributed to the available life history information, each with different sampling intensity across the Atlantic coast and across seasons and years. The VAST model developed for the NEFSC fall bottom trawl (Working Paper 3 Tyrell 2022) is a first step towards addressing some of this uncertainty caused by the spatiotemporal variability, but more work is needed to resolve the issue.

1.5 Incorporating Findings into Impacted TORs

TOR2 (catch data): The WG elected to omit Gulf of Mexico bluefish catch data from the assessment, based on the review of movement and distribution data; this is consistent with previous assessments for bluefish (NEFSC 2015). To capture seasonal and regional variations in growth and availability/distribution of bluefish, the WG used a seasonal length-weight relationship and seasonal-regional length frequencies to describe the age structure of the commercial and recreational catch.

TOR3 (survey data): The WG explored standardizing survey indices using generalized linear models (GLMs) parameterized with several environmental covariates (depending on the data collected during the survey), such as temperature, salinity, and dissolved oxygen. Ultimately, the WG found that the trends and uncertainty in the standardized survey indices were similar to the trends and uncertainty in the nominal indices, and decided to use nominal indices because they are simpler to maintain and update in management track assessments (e.g., versus possible future update GLM convergence issues). See TOR 3 and Working Paper 7 (Celestino et al. 2022a) for more details.

TOR4 (fishing mortality, recruitment, spawning stock biomass), TOR5 (stock status), TOR6 (projections): The forage fish index was incorporated into a companion model run as a covariate for catchability associated with the MRIP index. This companion model had good diagnostics, but was not put forward as the primary model due to concerns that it did not capture forage fish trends in the South Atlantic Bight, among other issues (see Section 4). The companion model was used to generate population estimates for comparison with the primary model, and generally showed similar results to the primary model, and its continued exploration and development is a high priority research recommendation (see TOR 7). The age-structured primary model used time-varying size-at-age from the observed average weight-at-age by year and fleet to calculate total and spawning stock biomass, to reflect the observed interannual variability in growth and condition. The primary assessment model also used the age-specific natural mortality schedule that was developed under TOR 1. Additionally, the assessment model was shifted into the WHAM platform in part due to WHAM's ability to incorporate environmental covariates in future model updates.

TOR7 (research recommendations): The ecosystem information compiled for TOR1 was used to inform several research recommendations under TOR7, most notably suggesting further sampling to resolve spring-spawned and summer-spawned cohorts, associated environmental drivers of relative cohort strength, and possible effects on the bluefish population, as well as testing additional environmental covariates in the WHAM modeling framework.

2 CATCH

> *Term of Reference #2: Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort. Characterize the uncertainty in these sources of data.*

For more detailed information on commercial and recreational data collection and analysis, see Working Papers 8-10.

2.1 Commercial Removals

2.1.1 Commercial Landings Data Collection

Commercial landings (1950 to present) for all species on the Atlantic coast are maintained in the Atlantic Coastal Cooperative Statistics Program (ACCSP) Data Warehouse. The Data Warehouse is an online database of fisheries dependent data provided by the ACCSP state and federal partners. The Data Warehouse was queried on 31 May 2022 for all commercial bluefish landings (monthly summaries by state, gear and market category) from 1985-2021 for Florida (Atlantic coast), Georgia, South Carolina, North Carolina, Virginia, Maryland, Delaware, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine.

2.1.2 Commercial Landings

Over the last approximately 40 years, commercial landings from the bluefish fishery ranged from a high of 7,162 MT (15.8 million pounds) in 1988 and have steadily declined to a low of 1,090 MT (2.4 million pounds) in 2021 (Figure 6). During this time, commercial landings have been consistently lower than the recreational catch and accounted for on average approximately 14% of the total removals in weight (Table 2, Figure 6). Amendment 1 to the bluefish Fishery Management Plan (FMP) was implemented in the year 2000 and the commercial fishery has been regulated by quota since this time. Gill nets are the dominant commercial gear used to target bluefish and average approximately 50% of the bluefish commercial landings from 1982 to 2021; this gear is fished primarily in the Mid-Atlantic and Florida. Other commercial gears, including hook & line, pound nets, seines, and trawls, collectively account for approximately 50% of the commercial landings.

Regional variations in commercial fishing activity are linked to the seasonal migration of bluefish. The majority of commercial fishing activity in the North and Mid-Atlantic occurs from late spring to early fall when bluefish are most abundant in these areas. As water temperatures decrease in late fall and winter, bluefish migrate south. Peak landings in the South Atlantic occur in late fall and winter. The majority of commercial landings over the time series (1950-present) have been taken in the Mid-Atlantic region (New York, New Jersey, and North Carolina); approximately 65% of the coast-wide total landings have been taken by these three states since

1982. Florida accounted for a larger percent of commercial catch historically (early 1980s) but has accounted for a diminishing proportion of landings over time.

2.1.3 Commercial Biological Sampling

Commercial fisheries from Maine to Virginia were sampled as part of the NEFSC data collection program (1985-2021). In addition, Virginia, North Carolina, and Florida have collected age and length data from their commercial fisheries to characterize the catch from the late 1980s onward. Since 2012, states that account for more than 5% of total coastwide landings have been required to collect 100 paired age and length samples, although those samples may come from any combination of commercial, recreational, or fishery independent sources. Sampling details were modified in 2020; see ASMFC 2020 and 2021 for more detail on state sampling programs. Length frequency data for Maine – North Carolina were expanded according to total landings in weight by market category and quarter. Biological data collection for the bluefish fishery south of North Carolina was sparse. Florida landings were characterized by North Carolina length frequencies from 1985-1991 due to lack of sampling in Florida; from 1992-2021, Florida samples were expanded by half-year (hereafter referred to as "season"). Landings from South Carolina and Georgia were generally negligible across the time series; when they occurred, they were pooled with Florida landings and characterized using the length frequency data used for Florida.

2.1.4 Commercial Length Frequency Distribution

The length frequency distribution from the commercial fisheries is characterized by a bi-modal distribution for much of the time-series. In the more recent years (2012-2021), the larger mode is reduced, leading to a skewed distribution with a peak around 35 cm. This pattern in bluefish length frequency has been observed in some years of the recreational harvest length frequencies, and the recreational discard length frequencies. The bi-modal pattern is likely a result of low availability to the fisheries of age 3 to age 4 bluefish. Bluefish are known to school by size class and it is speculated that movement dynamics at this age/size range affects availability of these fish. Much of this size cohort could be staying in the south (SC-FL) or offshore in certain years, and since the dominant fisheries for bluefish are coastal and north of Cape Hatteras, North Carolina, this would account for a reduced availability of this size/age class.

2.1.5 Commercial Discards

Previous bluefish technical committees (TCs) and working groups have concluded that commercial discards for bluefish along the Atlantic coast were insignificant, and historically this portion of the commercial catch has been ignored. The 2022 research track WG concluded that although commercial discards are a small fraction of the total catch, they should still be estimated and included in the commercial catch totals. To estimate commercial discards for bluefish, the Standardized Bycatch Reporting Method (SBRM) approach (Wigley et al. 2007) was applied, using the combined (D2) estimator. Commercial discard rates from 1989-2021 were calculated by half-year, gear, mesh, and region. A commercial discard mortality estimate of 32% was estimated via a literature review and meta-analysis based on the relevant gear types for the bluefish fishery and applied to the annual discards. See Appendix I to Working Paper 8 (Wood 2022a) for more details. Commercial landed lengths were used to characterize the size structure of the dead discards in all years due to the absence of adequate discard length samples.

Commercial bluefish dead discards have ranged from a high of 166 MT in 1996, to a low of 7 MT in 2017 (Table 2, Figure 6). Trawl and gillnet fisheries account for almost all of the discards, with small contributions from handline, longline, and midwater fisheries. Observed trips where bluefish was a primary target averaged around 1,800 trips per year over the time series. Commercial bluefish discards average 1.5% of the commercial catch by weight, and 0.2% of the total catch. While this portion of the catch is insignificant, the inclusion of these data will allow future shifts in magnitude to be monitored and accounted for in the assessment and more closely represent commercial allocations in catch accounting.

2.1.6 Commercial CAA and WAA

Seasonal length-weight parameters (Figure 4; Working Paper 5 Truesdell et al. 2022) were used to calculate numbers at length for the commercial catch. Final commercial catch-at-age (CAA) and weight-at-age (WAA) matrices were calculated using the annual seasonal multinomial age length keys (Section 3.3.1; Working Paper 14 Celestino et al. 2022b). The commercial catch is predominately comprised of age-1 and age-2 bluefish.

2.2 Recreational Removals

2.2.1 Recreational Data Collection

Estimates of recreational harvest and live releases for bluefish come from the NOAA Fisheries Marine Recreational Information Program, which uses a combination of effort surveys and angler-intercept surveys to develop those estimates (Papacostas and Foster 2018). This program was historically known as the Marine Recreational Fishery Statistics Survey (MRFSS), but was renamed in 2013 as NOAA Fisheries began making improvements to the survey design and estimation methods to address concerns identified by a National Academies review of the program (NRC 2006).

In 2018, MRIP transitioned from the Coastal Household Telephone Survey (CHTS) of effort to a mail-based survey, the Fishing Effort Survey (FES), following three years of side-by-side benchmarking. The CHTS and the FES only estimate effort for the private angler mode; the for-hire mode is covered by a separate survey, the For-Hire Survey (FHS). The FES produced consistently higher estimates of effort than the CHTS, so MRIP calibrated the historical estimates of catch and effort from the CHTS to the new scale of the FES estimates to provide a consistent time series (Papacostas and Foster 2018). The calibration model included fixed annual and seasonal effects as well as random effects and included information on trends in state-specific population size for the full time series and the prevalence of wireless/cell phone only households by state from 2007-2014. The calibration process also included the 2013 changes to the angler-intercept survey and corrections for the historical inconsistencies in the MRFSS intercept survey design (Papacostas and Foster 2018).

This increase in effort translated into an increase in total catch for bluefish, in both harvest and live releases. The overall trends in harvest and live releases were generally the same between the calibrated and uncalibrated time series, but the calibrated estimates were consistently higher. For a more detailed review of MRIP changes over time and the impacts of the calibration, see Working Paper 9 (Drew 2022a).

2.2.2 Recreational Harvest

Recreational harvest estimates of bluefish have averaged around 20,000 MT (44.1 million pounds) annually since 1985. From the 1980s to the early 1990s, recreational harvest declined by about 60%. The 1985-1989 average harvest was 52,064 MT (114.8 million pounds), while the 1990-1994 harvest averaged 22,285 MT (49.1 million pounds). Recreational harvest estimates continued to decline at a somewhat slower rate until reaching a low of 10,695 MT (23.6 million pounds) in 1999, increasing to 21,269 MT (46.9 million pounds) in 2010, and steadily decreasing since then to a value of 5,471 MT (12.1 million pounds) in 2021 (Table 2, Figure 6). In 2021, recreational anglers along the Atlantic coast caught 6.2 million bluefish, a 34% decrease from 2020.

The majority of recreational activity occurred from May to October, with the peak activity in July and August and almost 70% of the bluefish harvest being taken between July and October. The seasonal pattern varies by state, however, with more northern states seeing a peak in the summer and more southern states seeing peaks at the beginning and end of the year, reflecting both differing effort patterns by state and differing availability to states as bluefish migrate.

MRIP assigns catch to three fishing areas based on where anglers report doing the majority of their fishing: inland (which includes bays and estuaries like Long Island Sound, Chesapeake Bay, and Albemarle Sound), near-shore ocean (state waters less than three 3 miles from the shore), and offshore ocean (federal waters three miles or more from shore). About 51% of the catch of bluefish on a coast-wide basis came from inland waters, followed by near-shore ocean (42%) (Figure 7). Offshore ocean is only about 7% of the total catch. The inland portion of the harvest has been decreasing in recent years, with a concurrent increase in near-shore ocean harvest (Figure 7). For a detailed analysis of the spatial distribution of bluefish based on MRIP catch information see Working Paper 10 (Drew 2022b).

The majority of recreational harvest comes from the private boat and shore-based fishing modes (Figure 8). Less than 10% of the catch came from for-hire boats over the time-series.

2.2.3 Recreational Discards/Dead Releases

MRIP estimates of bluefish released alive have ranged from a low of 5.2 million fish (1988) to a high of 42.5 million fish (2001) from 1985-2021. Recreational release estimates have generally increased in proportion to harvested fish over the time series, increasing from approximately 19% of the total coast-wide catch in 1985 to over 80% in 2021. These releases represent both regulatory discards as well as voluntary releases by anglers practicing catch-and-release fishing.

About 48% of recreational bluefish releases on a coast-wide basis came from inland waters, and 48% from nearshore waters (Figure 7). Offshore ocean is only about 4% of the total releases. For a detailed analysis of the spatial distribution of bluefish harvest and releases based on MRIP data see Working Paper 10 (Drew 2022b).

The majority of recreational live releases comes from the private boat and shore-based fishing modes (Figure 8). Less than 10% of the releases came from for-hire boats over the time-series.

2.2.3.1 Recreational Release Mortality Rate

Estimating recreational catch-and-release mortality of bluefish is an important component of the stock assessment process given the popularity of catch-and-release angling in this fishery and the direct influence of release mortality on the total allowable catch. The literature reviews and analyses completed for the 2015 benchmark assessment (NEFSC 2015) were updated to re-assess the appropriateness of the 15% bluefish recreational release mortality estimate. From the updated literature reviews, no additional bluefish-specific release mortality papers were discovered, and one additional release mortality review paper (which was used for a meta-analysis) was discovered. Eleven exclusion criteria were applied to each bluefish-specific study and the studies within the review paper to determine which studies were suitable for inclusion in the bluefish-specific analysis and the meta-analysis. Three bluefish-specific studies passed the exclusion criteria. The individual mortality estimates from these three studies were used to calculate the mean (± standard error) bluefish-specific release mortality estimate, which was 9.4% ± 0.6%. From the review paper literature tables, 19 studies passed the exclusion criteria. The 22 individual mortality estimates from these 19 studies were used to calculate the mean (± standard error) meta-analysis release mortality estimate, which was 9.7% ± 1.9%.

The bluefish-specific release mortality estimate of 9.4% was used for this assessment. See Working Paper 11 (Valenti 2022b) for the full review and analysis.

2.3 Recreational Biological Sampling

2.3.1 Recreational Harvest

Recreational landings are sampled for length as part of the MRIP program. The MRIP length samples were used to expand recreational harvest per half-year season. In some years of the time series bluefish harvest lengths exhibit a bi-modal distribution, with a peak of fish around 35 cm, and a smaller peak around 70 cm. This trend has diminished in recent years but is consistent with trends seen in the commercial length frequency distributions. The bi-modal pattern is a result of an apparent low availability to the fisheries of age-3 to age-4 bluefish. Bluefish are known to school by size class and it is likely that unobserved movement dynamics at this age/size range affects availability of the population. It is possible a larger portion of the population at these sizes are staying south or offshore each year. Since the dominant fisheries for bluefish are coastal and north of Cape Hatteras, North Carolina, this would account for reduced availability of this size/age class.

The size of bluefish harvested by the recreational fishery varied by state and mode, with more northern states harvesting a wider range of sizes with a higher proportion of large bluefish than more southern states, which rarely harvest bluefish larger than 50 cm in fork length (Figure 9). In addition, bluefish harvested by the shore mode in states from Massachusetts through New York had a distinct peak of smaller fish around 15-20 cm that were not harvested by the private and for-hire boat modes. Young-of-the-year “snapper” bluefish are typically found inshore and are often targeted by shore-based anglers in the northern states. From New Jersey southward, that peak of smaller fish disappeared and the shore and boat mode length frequencies almost completely overlapped each other (Figure 9).

2.3.2 Recreational Dead Releases

MRIP conducts limited at-sea observing on headboat trips to collect lengths of fish released alive. To characterize the length frequency of the dead releases, the MRIP observer data were supplemented with lengths from the American Littoral Society (ALS) volunteer angler tagging program (by definition released fish), volunteer angler logbook programs in RI, CT, and NJ, and a volunteer angler tagging program in SC. See Atlantic States Marine Fisheries Commission (ASMFC) (2021) for more details on the state volunteer angler programs.

The recreational dead discard component of the catch was calculated using the season/region length frequency distributions developed from all of the recreational biological sampling data for released fish. For each year, expanded lengths were calculated by season/region and summed to get a seasonal total length distribution. Seasonal length-weight parameters (see above) were then used to calculate total seasonal weight and summed for a total annual release weight. A discard mortality estimate of 9.4% (Section 2.2.3.1) was applied to calculate the weight of dead discards for the total catch.

When the samples were pooled across season and region without weighting by removals, as was done for SARC 60, the harvest and release length frequency distributions appeared fairly distinct, with harvested fish centered around a smaller mean size than released fish (Figure 10). However, when stratified by region and season, the length frequencies for the harvested and the released fish were generally similar, with the exception of the northern region in the second half of the year, which had a peak of smaller fish in the harvest and a peak of larger fish in the releases (Figure 11). The majority of the release lengths were from the northern region; the southern region was not well sampled, particularly in recent years (see Working Paper 8 Wood 2022a for more details on sample size), and the differences in length frequency by region made it important to stratify the releases by region as well as season. Of note, in season/region/year cells where n < 30 fish, the cumulative length frequency of released alive fish was used as a proxy, instead of borrowing from another region or season.

Recreational releases/discards in 2021 were estimated at 14,792 MT, and after adjusting for a 9.4% mortality rate, the resulting discard loss was 1,391 MT. Recreational discard loss in weight has ranged from a low of 905 MT in 1988, to a high of 7,271 MT in 2001 (Table 2, Figure 6).

2.3.3 Recreational CAA and WAA

Final recreational harvest-at-age, dead releases-at-age, and weight-at-age matrices were calculated using the annual seasonal multinomial age length keys (Section 3.3.1; Working Paper 14 Celestino et al. 2022) applied to the harvest and dead release length frequencies. The recreational harvest-at-age and dead-releases-at-age were summed to calculate the total recreational dead catch-at-age. The recreational catch is predominately comprised of age-0, age-1 and age-2 bluefish.

2.4 Total Removals

Total bluefish removals (total dead catch) by component are presented in Table 2 and Figure 6. Overall, total removals have declined since the beginning of the time series. There was a slow increase from 1996 to 2010, but the declining trend has continued to the lowest values in the time-series in recent years (Figure 6). On average, commercial landings account for 14% of the total removals with commercial discards averaging only 0.2%. Total removals are dominated by the recreational fishery with recreational landings accounting for 71% of total removals, and recreational dead releases averaging 15% of total removals.

3 SURVEY DATA

> *Term of Reference #3: Present the survey data used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, application of catchability and calibration studies, etc.) and provide a rationale for which data are used. Describe the spatial and temporal distribution of the data. Characterize the uncertainty in these sources of data.*

The ASMFC Bluefish TC held a workshop in November 2021 to review the available state datasets for bluefish with the goal of evaluating their utility for this assessment, including fishery independent surveys. Metrics used to evaluate the datasets included the length of the time series, the geographic coverage, the quality and consistency of the survey design, and the prevalence of bluefish in the dataset, as measured by the percent positive tows or hauls for bluefish. Detailed descriptions of the surveys considered and the TC evaluations are available in the State Data Review Workshop Report (ASMFC 2021). The WG participated in the workshop and reviewed the final recommendations of the TC as to which datasets to include, exclude, or explore further. The WG's final decisions on which indices to include are summarized in Table 3 and Table 4. The surveys covered the majority of the bluefish range on the Atlantic coast, ranging from the Gulf of Maine in the north to Cape Canaveral, Florida in the south (Figure 12).

The suite of indices used in the base model was similar to what was used in SARC 60 (NEFSC 2015). Two new indices were added: the SEAMAP age-1 index (Section 3.1.1) and the ChesMMAP age-0+ index (Section 3.2.3). Two indices were dropped: the New Jersey Ocean Trawl Survey (NJ OT) and the Connecticut Long Island Sound Trawl Survey (CT LISTS). The NJ OT survey was dropped on the recommendation of the state data providers, as it was dominated by age-0 fish and did not seem to be adequately tracking age-1+ abundance (ASMFC 2021). The CT LISTS survey was removed for similar reasons: it covered a smaller spatial area than other trawl surveys in the model and was dominated by age-0 fish with little information on age-1+ fish. In addition, inclusion of the index resulted in worse model diagnostics without significantly affecting population estimates (Section 4.3.1).

The WG explored standardizing the fishery independent indices of abundance using environmental covariates in a GLM framework. However, the standardization process did not significantly affect index trends or reduce interannual variability or index coefficients of variation (CVs), so the WG did not use the standardized indices in the base run and instead used the stratified arithmetic mean for surveys with a stratified random design and the geometric mean for surveys with a fixed station design; see Working Paper 7 (Celestino et al. 2022a) for a detailed write-up of the process and results. The exception to this decision was the SEAMAP survey; see Section 3.1.1 below.

3.1 Recruitment Indices

For detailed descriptions of survey methods, see the ASMFC State Data Review Workshop Report (ASMFC 2021).

3.1.1 SEAMAP Age-0 and Age-1 Indices

The Southeast Area Monitoring and Assessment Program (SEAMAP) Coastal Trawl Survey has sampled the coastal zone off the southeast U.S. between Cape Hatteras, North Carolina and Cape Canaveral, Florida with a standardized protocol since 1990. A stratified random sampling design is used, with strata based on latitude and water depth. The SEAMAP survey encounters both age-0 and age-1 bluefish, with the age frequency varying by season. The spring survey is dominated by age-1 bluefish, while the fall survey is dominated by age-0 bluefish (Figure 13). Therefore, separate indices were developed for age-0 (fall-caught) and age-1 (spring-caught) bluefish.

SEAMAP used a GLM to calculate the indices for bluefish. The GLM standardization was able to smooth an exceptionally large value in the nominal index for age-0 bluefish which improved the correlation between the SEAMAP age-0 and lagged age-1 indices and the correlation between the SEAMAP age-0 and the composite age-0 indices (Section 3.1.2). In addition, due to vessel, weather, and funding issues, sampling in the northern-most strata of the survey has dropped off in recent years. Those strata have the highest abundance of bluefish in the SEAMAP survey, and the use of latitude in the standardization accounts for the decline in sampling in those strata. Therefore, the WG used the standardized age-0 and age-1 indices developed by SEAMAP for both indices (Zimney and Smart 2022).

The age-0 and age-1 indices have generally varied without trend over the time series; strong and weak year classes can be tracked from the age-0 to age-1 index in several years (Figure 14).

3.1.2 Composite Young-of-Year (YOY) Index

States from New Hampshire to Virginia conduct seine and trawl surveys for juvenile finfish that capture YOY bluefish. These surveys are noisy and cover small geographical areas, compared to the range of bluefish. Bayesian hierarchical modeling was used to combine these indices into a single composite index, using the method developed by Conn (2010), which represents the coast wide recruitment dynamics of bluefish. A composite index developed from state trawl YOY surveys (Table 4) was also explored, but it was not well correlated with the age-0 catch or any of the other indices and was not used in the assessment model. See Working Paper 12 (Drew 2022c) for details of the analysis. The surveys included in the composite index are described below.

Overall, the composite index did not show a strong trend over the time series; the early years were higher than later years, but also had more uncertainty around them (Figure 14).

3.1.2.1 New Hampshire Juvenile Finfish Seine Survey

The New Hampshire Juvenile Finfish Seine Survey samples at 15 fixed stations during June through November. The stations are spread throughout the New Hampshire coast, including the Hampton/Seabrook Estuary, Little Harbor, the Piscataqua River and Little Bay/Great Bay. Historical catches have ranged from 2.3 – 22 cm total length, all classified as YOY using a 25 cm size cutoff. Samples from November and December were removed from the analysis due to no positive catches. The survey has run from 1997 through the present. The nominal index was calculated as a geometric mean catch per tow with bootstrapping (n = 1000) to estimate the annual CVs. The index varied without trend (Figure 15).

3.1.2.2 Rhode Island Narragansett Bay Juvenile Finfish Beach Seine Survey

The Rhode Island Narragansett Bay Juvenile Finfish Beach Seine Survey currently samples 18 fixed stations throughout the bay; the survey began with 15 stations and added one additional station in each of 1990, 1993 and 1995. The survey began in 1988 and runs from June through October. A 25 cm size cutoff was used as the threshold to identify YOY bluefish. The nominal index was calculated as a geometric mean catch per tow with bootstrapping (n = 1000) to estimate the annual CVs. The early part of the time series was characterized by considerable variability. Catches were generally stable from 2010-2016, dropped during 2017 and 2018, and have since increased (Figure 15).

3.1.2.3 New York Western Long Island Seine Survey

The New York Department of Environmental Conservation Western Long Island Beach Survey has employed a consistent methodology since 1987 to sample sites at fixed stations within western Long Island bays: Little Neck and Manhasset Bay on the north shore of Long Island, and Jamaica Bay on the south shore (1984-present). Other bays have been sampled on a shorter time frame but were not included in this index. The nominal index was calculated as a geometric mean catch per tow with bootstrapping (n = 1000) to estimate the annual CVs. The index has generally varied without trend over the time series (Figure 15).

3.1.2.4 New Jersey Delaware River Seine Survey

The New Jersey Fish and Wildlife Delaware River Seine Survey is a fixed station beach seine survey conducted in three regions of the Delaware River. It targets age-0 striped bass, but bluefish are also captured in the brackish to tidal freshwater regions of the river. A 25 cm length cutoff is used to identify age-0 bluefish. The bluefish YOY index was reported as the geometric mean number of YOY bluefish per seine haul of samples collected from mid-June through September in region 1, with bootstrapping (n = 1000) to estimate the annual CVs; samples taken in October through November were excluded as YOY bluefish are rarely captured in those months. The index included data from 2002-2021, although 2020 was missing due to the COVID-19 pandemic. The index generally varied without trend, but the three lowest values in the time series occurred in 2016, 2018, and 2021 (Figure 15).

3.1.2.5 Maryland Juvenile Striped Bass Seine Survey

The Maryland Department of Natural Resources Juvenile Striped Bass Seine Survey is a fixed station survey that samples in major striped bass spawning areas in Maryland’s portion of the Chesapeake Bay from July – September. A subset of 13 sample sites was selected for the development of a juvenile bluefish index from 1981 to present. The nominal index was calculated as a geometric mean catch per tow with bootstrapping (n = 1000) to estimate the annual CVs. The index is variable but has shown a declining trend over time, with low catch rates and a low proportion of positive hauls in recent years (Figure 15).

3.1.2.6 Virginia Institute of Marine Science Juvenile Striped Bass Seine Survey

The Virginia Institute of Marine Science Juvenile Striped Bass Seine Survey is a fixed station survey that samples from July – September in the James, York, and Rappahannock Rivers, as well as in the main tributaries of these systems. The nominal index was calculated as a geometric mean catch per tow with bootstrapping (n = 1000) to estimate the annual CVs. The index showed a period of higher recruitment from the late 1980s to the late 1990s, followed by a period of lower recruitment from the early 2000s forward, although 2019-2021 have been higher (Figure 15).

3.2 Age 0+ Indices

3.2.1 Northeast Fisheries Science Center (NEFSC) Fall Inshore Trawl Survey

Since 1963, the NEFSC has conducted a standardized bottom trawl survey during the fall and spring along the northeastern continental shelf of the United States in the area comprising the Western Scotian Shelf of the Gulf of Maine, south to Cape Lookout, North Carolina. The survey uses a stratified random design. There was a vessel change in 2009 from the F/RV Albatross to the F/RV Bigelow, which resulted in the loss of historical inshore strata from the survey area, all of which are now sampled by the Northeast Area Monitoring and Assessment Program (NEAMAP) via the Mid-Atlantic/Southern New England Nearshore Trawl Survey (Section 3.2.1), the Massachusetts Division of Marine Fisheries Bottom Trawl Survey (ASMFC 2021) and Maine-New Hampshire Inshore Trawl Survey (ASMFC 2021). For more information on the NEFSC bottom trawl survey design, see Avarovitz (1981) and NEFSC (2015).

Bluefish are predominately caught during the fall in the inshore strata south of the Gulf of Maine, so fall inshore strata from Cape Hatteras to Cape Cod were used to build two indices for bluefish, one for the Albatross years (1985-2008) and one for the Bigelow years (2009-2021). The indices were calculated as the stratified mean catch-per-tow. The Albatross index showed high variability at the beginning of the time series followed by a generally increasing trend from the mid-1990s to the mid-2000s (Figure 16). The Albatross index declined from 2005 to the end of that time series in 2009, and the Bigelow has shown a consistent decline over its entire time series from 2009-2021 (Figure 17).

The fall stratified mean length frequencies of the Albatross and Bigelow indices were apportioned to ages by applying the annual fall age-length key (Section 3.3.1). The age-structure of the Albatross and Bigelow indices was dominated by age-0 fish (Figure 16); the Bigelow had a higher proportion of age-1 fish than the Albatross did, but was still dominated by age-0 and age-1 fish (Figure 17).

3.2.2 NEAMAP Mid-Atlantic/Southern New England Nearshore Trawl Survey

The NEAMAP Mid-Atlantic/Southern New England Nearshore Trawl Survey uses a stratified random design to sample the coastal ocean from Martha’s Vineyard, MA to Cape Hatteras, NC since the fall of 2007. NEAMAP conducts two cruises per year, one in the spring and one in the fall, and samples inshore areas that were lost from the NEFSC Bottom Trawl Survey with the vessel change in 2009. The index was calculated as the stratified mean catch-per-tow for the fall cruise where the bluefish catch and proportion positive tows were higher. The index has been variable with a somewhat declining trend, reaching a time-series low in 2019 before increasing in 2020 and 2021 (Figure 18).

The fall stratified mean length frequency of the NEAMAP index was apportioned to ages by applying the annual fall age-length key (Section 3.3.1). The age-structure of the NEAMAP index was dominated by age-0 bluefish (Figure 18).

3.2.3 Chesapeake Bay Multispecies Monitoring & Assessment Program (ChesMMAP)

The Chesapeake Bay Multispecies Monitoring & Assessment Program (ChesMMAP) uses a stratified random design to sample the mainstem of Chesapeake Bay every other month from March through November. The survey underwent a vessel change in 2019, and the calibration work has not been completed. As a result, the ChesMMAP index for bluefish includes data from 2002-2018. The index was calculated as the stratified mean catch-per-tow for the May through November cruises, where the bluefish catch and proportion positive tows was highest. The index has generally varied without trend over the time series (Figure 19).

The length frequency of the ChesMMAP index was stratified by season (May-June and July-November). The seasonal length frequencies were apportioned to ages by applying the appropriate seasonal age-length key, and the final age composition of the index was calculated by summing the seasonal index age compositions. The age-structure of the ChesMMAP index was dominated by age-0 and age-1 bluefish and had no observations greater than age-3 (Figure 19).

3.2.4 North Carolina Pamlico Sound Independent Gill Net Survey (PSIGNS)

The North Carolina Division of Marine Fisheries Pamlico Sound Independent Gill Net Survey (PSIGNS) uses a stratified random design to sample the Pamlico Sound estuary from mid-February to mid-December. Bluefish is the second most commonly caught species in the survey. The index was calculated as the stratified mean catch-per-set for all months. The index increased from 2001 through 2007 and then declined to a time-series low in 2015; subsequent years have increased slightly, and 2019 was an extremely high value (Figure 20).

The length frequency of the PSIGNS index was stratified by season (February-June and July-December). The seasonal length frequencies were apportioned to ages by applying the appropriate seasonal age-length key, and the final age composition of the index was calculated by summing the seasonal index age compositions. The age-structure of the PSIGNS index was dominated by age-1 and age-2 fish, and had the highest proportion of older fish of all fishery-independent indices used in the assessment (Figure 20).

3.2.5 MRIP Recreational CPUE

The MRIP dockside intercept program dataset was used to develop recreational total catch-per-unit-effort (i.e., harvest plus live releases in numbers) as an index of abundance for bluefish. Bluefish trips were defined using a guild approach where a trip was considered a bluefish trip if it caught either bluefish or a species that was significantly positively associated with bluefish. This was a change from the previous benchmark assessment where effort was described using "directed trips," which describe trips where bluefish were considered a target species. The CPUE was standardized using a zero-altered negative binomial model with year, state, wave, state-wave interaction, mode (e.g., shore, private boat, charter), area fished, kind of day (i.e., weekday or weekend), and angler avidity as factors and angler-hours per trip as an effort offset. For more information on the MRIP CPUE development, see Working Paper 13 (Drew 2022c).

The MRIP CPUE peaked at the beginning of the time-series, declining through the mid-1990s after which it showed a stable to slightly increasing trend until 2016 (Figure 22). It has declined in recent years. The choice of trip definition (guild trips vs. directed trips) and standardization model (zero-altered negative binomial vs. negative binomial) resulted in significant changes in overall trend compared to the index developed during the last benchmark (directed trips standardized with a negative binomial model with no interaction terms). While the indices showed roughly similar trends – declining through the mid-1990s before stabilizing and increasing somewhat – the MRIP CPUE used in this assessment showed much more contrast than the continuity run index used in the SARC 60 assessment, starting out at a higher level, declining to lower levels, and not recovering as much after the decline (Figure 21).

The age-structure of the MRIP CPUE was developed from the recreational catch and release information, as the CPUE used both harvested and released alive fish in the calculation of the catch per unit effort. The recreational harvest numbers-at-age matrix was combined with the recreational live release numbers-at-age matrix. Unlike the recreational removals matrix, the live releases numbers-at-age were not scaled by the release mortality rate. The MRIP CPUE had a broader age-structure than the fishery independent indices (Figure 22).

3.3 Age and Length Data

3.3.1 Age-Length Keys

The WG evaluated multinomial age-length keys (ALKs) relative to traditional ALKs, and ALKs resolved at a seasonal as well as season-region level of resolution; for complete details, see Working Paper 14 (Celestino et al. 2022b). Briefly, multinomial ALKs were explored as an objective, repeatable, and efficient way to fill gaps in ALKs. The data to construct the ALKs for bluefish were sparse early in the timeseries, and throughout the time series when subset to a season (January-June and July-December) and region (Florida-North Carolina and Virginia-Maine) level of resolution. The multinomial approach to developing ALKs (Gerritsen et al. 2006) has been explored across a number of stocks assessed by ASMFC and NOAA Fisheries, and is available in modelling software [e.g., weakfish (ASMFC 2019), Stock Synthesis (Methot and Wetzel 2013)].

Age and length data collected by fishery-independent and fishery-dependent sampling by NMFS and Atlantic coast states were compiled and used to construct ALKs. When developing and comparing ALKs, the WG developed various borrowing and multinomial model configuration rules. Final multinomial ALKs were constructed using all data (all years, seasons, regions combined) input through a single model, with terms for year and season (or year, season, and region for exploration of the seasonal-regional ALKs). All spring age-0 fish were removed from the dataset prior to running multinomial models, which helped the performance of model predictions relative to biological expectations (e.g., minimized the probabilities of spring age-0 fish in ALKs). Traditional ALKs (i.e., non-model based ALKs such as those used in the 2015 assessment) were only constructed at the year-season level due to concerns related to the sparse nature of data at the year-season-region level of resolution that would require a large number of decisions related to data borrowing. Spring age-0 fish were also removed from that dataset.
The influence of multinomial and traditional ALKs, at the season and season-region level of resolution was evaluated in a statistical catch-at-age framework. Total catch and indices of abundance were apportioned into ages using seasonal traditional ALKs, seasonal multinomial

ALKs, and season-region multinomial ALKs. Statistical catch-at-age model performance was similar among model runs (Table 5). The scale of retrospective patterning was comparable between models that used multinomial ALKs, but higher for the model that used traditional ALKs. After extensive discussions, the WG did not believe data were sufficient to support an ALK model at a seasonal-regional level of resolution. All sample sizes are reduced as data are subset to finer spatial and temporal resolutions.

While Akaike's Information Criteria (AIC) among the ALK multinomial models supported seasonal-regional ALKs (vs seasonal multinomial ALKs), the WG did not have high confidence in partitioning data at this level of resolution, and so supported the use of the seasonal multinomial ALKs applied to season-region length frequencies for the catch and indices. The statistical catch at age model results suggested less retrospective patterning with the seasonal multinomial ALKs compared to the traditional seasonal keys, and so the WG supported use of the seasonal multinomial ALK for continued modelling (and ultimately the base model).

4 ASSESSMENT METHODS

> *Term of Reference #4: Use appropriate assessment approach to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Compare the time series of these estimates with those from the previously accepted assessment(s). Evaluate a suite of model fit diagnostics (e.g., residual patterns, sensitivity analyses, retrospective patterns), and (a) comment on likely causes of problematic issues, and (b), if possible and appropriate, account for those issues when providing scientific advice and evaluate the consequences of any correction(s) applied.*

4.1 History of the Bluefish Assessment

A statistical catch at age assessment model was first used to assess bluefish and provide management advice in 2005, at the Stock Assessment Workshop 41 review (NEFSC 2005). Prior to this review, several model types were explored including a modified Delury model, a surplus production model, a VPA, and catch-at-age models. At the time, the Bluefish TC concluded that age-based models such as a VPA or catch-at-age were the most appropriate for the bluefish assessment and age-based models have been used since.

At the last benchmark assessment in 2015, a number of changes were made to the data structure and assessment model. Major changes included fitting to the age composition of the surveys (as opposed to age-specific indices), separating total catch into two fleets (commercial and recreational), updated maturity-at-age information, splitting the Bigelow and Albatross survey time series into two indices, and changing MRIP index selectivity from independent estimates at-age to a logistic curve. The final model was reviewed during SAW/SARC60 (NEFSC 2015) and has been used to provide management advice since 2015.

The most recent operational assessment for bluefish took place in 2021, with data through 2019. Based on this assessment update of the 2015 benchmark model, the bluefish stock was overfished and overfishing was not occurring relative to the updated biological reference points. Spawning stock biomass was estimated to be 95,742 MT in 2019, about 47.5% of $SSB_{35\%}$

(201,729 MT) and 95% of the threshold (100,865 MT). Fishing mortality was estimated to be 0.172, which was 95% of $F_{35\%}$ (0.181). Average recruitment from 1985-2019 was 46 million age-0 fish. The terminal year estimates for fishing mortality and spawning stock biomass adjusted for retrospective error were within the 90% confidence bounds of the terminal year estimates, indicating no retrospective adjustment was needed for stock status determination.

4.2 Bluefish Research Track 2022 Model Introduction

The Research Track (RT) 2022 model building procedure for bluefish was accomplished over multiple steps. The majority of the model bridge was built using ASAP (Age Structured Assessment Program, Legault and Restrepo 1999), which was the previously approved assessment model. ASAP is an age-structured model that uses forward computations, assuming separability of fishing mortality into year and age components, to estimate population sizes given observed total catches, catch-at-age, and indices of abundance. Bluefish are modeled as age-0 through age-6+, with ages six and older pooled into a plus group. The separability assumption is partially relaxed by allowing for fleet-specific computations and by allowing the selectivity-at-age to change in blocks of years. Weights are specified for different components of the objective function, which allows for configurations ranging from relatively simple age-structured production models to fully parameterized statistical catch-at-age models. The objective function is the sum of the negative log-likelihood of the fit to various model components. Catch-at-age and survey age composition are modeled assuming a multinomial distribution, while most other model components are assumed to have lognormal error. Specifically, lognormal error is assumed for: total catch in weight by fleet, survey indices, stock recruit relationship, and annual deviations in fishing mortality. Recruitment deviations are also assumed to follow a lognormal distribution, with annual deviations estimated as a bounded vector to force them to sum to zero (this centers the predictions on the expected stock recruit relationship, or on mean recruitment when steepness is fixed at one). For more technical details, the reader is referred to the technical manual (Supporting documentation: ASAP manual, Legault 2012).

Early WG discussions led to the decision that the bluefish assessment model should be shifted into a new modeling framework, the Woods Hole Assessment Model (WHAM: Miller et al 2016, Miller and Hyun 2018, Miller et al. 2018). WHAM is a general state-space age-structured assessment model that is able to include environmental and other covariate effects on population processes. The shift from ASAP to WHAM allowed more flexibility, including the estimation of observation and process error, and the propagation of random effect parameters in stock projections. The final ASAP model was transitioned into its "ASAP-like" WHAM model counterpart, which was parameterized so that it was essentially identical to the ASAP model; after this initial WHAM model was fit, a suite of models that included random effects on the numbers-at-age were fit, and model selection via AIC was used to select a best model. Environmental indices based on a VAST analysis of forage fish availability along the east coast (Section 1.2) were also explored as covariates on the catchability of survey indices.

4.3 Bluefish Research Track 2022 Model Bridge

4.3.1 ASAP Modeling

The first step in modeling in ASAP was to conduct a continuity run, which updated the current assessment model with data through 2021. A base model was then constructed by adding new data (CAA and WAA) and indices to the continuity run, keeping the same model settings and weights. A model bridge was then built from the base model to a final ASAP model by changing model data formulation, specifications, and weighting inputs. In total, about 80 variations of ASAP models were explored during this bridge building procedure. The model steps with the most important changes that provide a linear path from the base model to the final ASAP model are presented below. See Working Paper 15 (Wood 2022b) for a detailed description of the bridge-building process, results, and diagnostics. Working Paper 16 (Wood 2022c) includes the complete diagnostic plots for the major milestone runs, including the final ASAP model; the diagnostic plots for the final ASAP model alone are linked below.

The continuity model run was carried out as update of the SAW/SARC 60 benchmark final model, which is the model currently used for management advice. Total catch, catch-at-age, weight-at-age, and indices-at-age were updated for 2020 and 2021 using previously established data protocols (Figure 23). Retrospective pattern for the continuity run was examined for F, SSB, and recruitment using 7-year peels. The analysis showed consistent and significant pattern in the estimates of F and SSB, with Mohn's rho values of -0.277 and 0.294, respectively. Recruitment estimates exhibited lower retrospective pattern that was inconsistent over the peels, with a Mohn's rho estimate of 0.170. The continuity run had poor convergence diagnostics; a jitter analysis indicated that when the parameter initial values were varied randomly, only 130 of 200 realizations of the model reached the same final objective function value as the base run and there were 18 non-converged models. Gradient values were also poor for a number of the runs that did converge, with the majority of maximum gradient values being greater than 0.0001, a value often used as threshold for an acceptable model (Carvalho et al. 2021). This diagnostic was not explored in the 2015 benchmark assessment.

The switch to multinomial age-length keys had a significant impact on estimates of SSB. The multinomial keys had the effect of spreading numbers-at-age in the older ages to younger ages, especially with the plus group. This had the result of lowering the SSB as the total biomass of mature fish was reduced. The multinomial keys substantially improved the convergence diagnostics for this model. All previous models that used the traditional keys had poor convergence diagnostics. A jitter analysis of starting parameter values showed that the previous model step with traditional keys failed to converge 52 times and only found the original model solution 129 times out of 200 jitter runs. Conversely, the model with multinomial keys was very robust to the original objective function solution and did not seem overly sensitive to the initial starting values; it failed to converge only 6 times, and found the original objective function 193 times, out of 200 jitter runs. All of the alternate objective function values were higher than the original objective function value.

The change from the directed trips method to the guild approach (Section 3.2.5) to develop the MRIP index was another significant change for the data going into the model. The MRIP index has historically been the most important index in the assessment model and effectively scales the model because of the assumed logistic selectivity. Without this flat-top selectivity, the model is able to create cryptic biomass in the older ages and can produce unrealistic results. Due to the importance of this index, small changes in trend have dramatic impacts on the scale of model results. The continuity MRIP index (i.e., using the directed trips effort and the previous standardization model) remained fairly flat throughout the time-series. Shifting to the guild approach for the index calculation resulted in a much different trend, with the guild approach index starting out at higher values and declining to lower levels compared to the continuity run directed trips index (Figure 21).

The overall effect of this new index on the model was a decrease in SSB and an increase in F. The switch to the MRIP guild index significantly reduced the retrospective pattern in SSB from 40% to 25%. The retrospective pattern in fishing mortality was also reduced from 36% to 21%. Convergence diagnostics in the ASAP framework for the model with the guild approach MRIP index were very good, with 191 of 200 jitter runs finding the original model solution and 8 non-convergences. All of the alternate objective function values were higher than the original objective function value.

The change from an age-constant natural mortality of 0.2 to the higher, age-varying estimates of M from the Lorenzen (1996) method increased SSB, decreased F, and greatly increased recruitment, as would be expected. The retrospective pattern was increased for all model results with the change in natural mortality (Table 6). Convergence diagnostics were very similar between the age-constant and age-varying M.

The previous assessment model specified a single selectivity block for each fleet for the entire time-series. To address the retrospective pattern and the patterning in the catch-at-age residuals, a selectivity block was added in each fleet beginning in the year 2000, which is the year Amendment 1 to the bluefish fishery management plan was implemented. An additional selectivity block was added in the recreational fleet from 2011-2021, to align with the increasing trend in the proportion of the recreational catch from the southern region, which tends to catch a smaller size range of bluefish (Working Paper 9 Drew 2022a). The addition of new selectivity blocks increased SSB estimates and reduced retrospective pattern.

Model fit diagnostics for the Connecticut Long Island Sound Trawl (CT LIST) survey index indicated a somewhat poor fit early on in the time-series, with two blocks of residuals from 1985-2000. This survey also caused issues with the estimation of retrospective peels, with some peels giving gradient estimates >0.001, indicating poor or no convergence. The removal of the CT LIST survey resulted in slight increases in both SSB and recruitment, and little change in fishing mortality. There was a small improvement in the retrospective pattern in SSB, and a small increase for the pattern in fishing mortality; this model did not have the retrospective peel convergence issues that occurred in previous models that included the CT LIST survey.

The WG chose model BF24 as the final bluefish ASAP model configuration, prior to migration into the state-space framework of the Woods Hole Assessment Model. A full suite of input, results, diagnostic, retrospective and MCMC plots are available for this run as part of Working Papers 15 and 16 (Wood 2022b and 2022c) and as a standalone file which can be downloaded or viewed from the following link: BF24 plots. When reviewing the ASAP plots, note that ASAP numbers the age classes starting with age-1, but the first age in the bluefish model is age-0. Therefore, all ages in the figures are increased by one relative to the biological age-class they represent (ASAP age-1 is really age-0, ASAP age-2 is really age-1, etc.). A brief summary of main model results is presented below.

The final ASAP model fleet selectivity-at-age estimates for the two fleets each show a decrease in selectivity at middle ages (ages 3-4), with selectivity increasing at older ages. The final selectivity block in the recreational fleet (2011-2021) has more of a domed shape, with older fish having much lower selectivity than previous blocks (Figure 24). Final ASAP model estimates for the index selectivities show a rapid decrease in selectivity after age-0. A few of the indices have higher selectivity towards larger/older fish, the most important being MRIP and PSIGNS, and to a lesser extent the Bigelow survey.

Abundance results from model BF24 showed a maximum of 424 million fish in 1985, declining to 166 million in 1995, and then increasing to a peak of 311 million in 2006. Total abundance declined from the peak in 2006 to a low of 147 million in 2016, a small peak to 208 million in 2018, and a terminal year estimate of 169 million fish. Spawning stock biomass started from a high of 208,791 MT in 1985 and declined over the time-series to a low of 44,931 MT in 2018, and increased since to a value of 63,320 MT in 2021. The majority of the spawning stock biomass is ages 5 and 6+ (30-60%) for the entire time-series. Fully selected fishing mortality in 2021 was 0.159, compared to an average full F from 1985 to 2021 of 0.354. Estimates of F have varied over the time-series from a peak in 1987 of 0.519 to the lowest value of 0.159 in 2021. Estimates of recruitment have remained steady over the time series, fluctuating around an average value of 127 million fish. Recruitment has been below average for the past 12 years, and was estimated at 95 million fish in 2021.

Retrospective pattern for the final model was examined for F, spawning stock biomass, and recruitment. There was a notable retrospective pattern in both SSB (Mohn's rho = 0.326) and fishing mortality (Mohn's rho = -0.277), with very little in recruitment (Mohn's rho = 0.017). Shifting this assessment model into the state-space framework of WHAM and estimating random effects helped to improve the retrospective diagnostics of this model.

The variation in the final ASAP model results for F and SSB was determined using a Markov chain Monte Carlo (MCMC) with 1000 iterations and a thinning factor of 2000 (2,000,000 iterations). Trace plots for both SSB and F show little to no patterning. There is no significant autocorrelation in the SSB or F chains. Terminal year 90% confidence intervals (CI) from the MCMC ranged from 49,856 to 71,780 MT, with a median estimate of 60,338 MT. The 2021 SSB point estimate from the final model (63,320 MT) is slightly higher than the median estimate from the MCMC distribution. The 90% CI around the terminal year F ranged from 0.112 and 0.231. The point estimate from the final model (0.159) is nearly identical to the median estimate (0.160) from the MCMC distribution.

Model BF24 had good convergence diagnostics with 192/200 jitter runs finding the original model solution, and 4 non-convergences (Figure 25).

4.3.2 Woods Hole Assessment Model (WHAM) Modeling

The Woods Hole Assessment Model (WHAM: https://github.com/timjmiller/wham) is a state-space age-structured stock assessment model developed at the Northeast Fisheries Science Center (NEFSC, Stock and Miller, 2020). WHAM is a flexible model framework that can be configured as a traditional statistical catch-at-age model, which allows for bridge building transitions from models like ASAP. In addition to the traditional catch-at-age approach, WHAM allows for the estimation of state-space effects, including annual transitions in the numbers-at-age, age and time varying random effects on natural mortality or selectivity, and the ability to incorporate environmental effects as covariates on population processes.

The final bluefish model from the ASAP model bridge (model BF24) was moved into WHAM for further model exploration. The WG made the decision to finish model exploration in WHAM because of its flexible framework, specifically allowing for the estimation of random effects on recruitment and numbers-at-age. A desirable feature of the state-space framework is that these models tend to have lower retrospective pattern in model results, and more realistic estimates of uncertainty (Stock and Miller, 2020). Model BF24 had a notable retrospective pattern in both SSB and F (Table 6) and this was a primary driver for moving the bluefish model into WHAM.

In addition to improving retrospective pattern, the final bluefish model was shifted into WHAM to explore environmental covariate links on the catchability of different surveys indices. Forage fish indices were developed using a VAST model (Section 1.2; Working Paper 4 Gaichas et al. 2022) and explored as environmental covariates on the catchability (q) of NEFSC survey indices and the MRIP catch-per-unit-effort index.

The focus of the model exploration in WHAM was to refine the final bluefish model from ASAP, and not continue building a model bridge. This refinement focused on models with random effects on recruitment and numbers-at-age. The models explored had different options for treating the yearly transitions in survival (numbers-at-age):

1. Deterministic survival: a traditional statistical catch-at-age (SCAA) model, recruitment in each year is estimated as independent fixed effect parameters.
2. Recruitment deviations (random about mean) are random effects
    a. Random effects are independent, uncorrelated: model subscript **_m2** going forward
    b. Autoregressive (AR1) by year (autocorrelated): model subscript **_m3** going forward
3. Full state-space model where survival of all ages are random effects
    a. Random effects are independent, uncorrelated: model subscript **_m4** going forward
    b. Autoregressive (AR1) deviations by year: model subscript **_m5** going forward
    c. Autoregressive (AR1) deviations by age: model subscript **_m6** going forward
    d. Autoregressive deviations by age and year (2D AR1): model subscript **_m7** going forward

To assess the fit and results of each model, a series of diagnostic criteria were applied. First, models were designated as converged if the maximum gradient was less than 1e-10 and the hessian matrix was invertible. Next, a model selection process using AIC was carried out to choose a best model among models with comparable likelihood structures. Convergence properties of the best models chosen by AIC were further explored using a jitter approach analogous to the approach used in ASAP. Parameter starting values were randomly generated using the model covariance matrix to develop random normal distributions around the MLE parameter estimates as well as a distribution scaling factor, which alters the spread of the distribution around the potential starting values by scaling the variance. Similar to the ASAP jitter approach, 200 iterations of the model were carried out to test model sensitivity to the initial parameter guesses and investigate convergence. The 200 realizations of the model objective function and gradient were examined to see how robust the model was to the starting values.

A suite of model fit diagnostic plots were also examined for each model of interest. Model fits were examined using conventional residual diagnostics, as well as one-step ahead residual diagnostics (OSA), which are more appropriate for state-space models with correlated parameters (Trijoulet et al. 2023). Finally, retrospective pattern in model results was evaluated using Mohn's rho values (Mohn 1999) calculated from 5-year model peels (Miller and Legault 2017, ICES 2020).

When reviewing the WHAM plots, note that WHAM (similar to ASAP) numbers the age classes starting with age-1, but the first age in the bluefish model is age-0. Therefore, all ages in the figures are increased by one relative to the biological age-class they represent (WHAM age-1 is really age-0, WHAM age-2 is really age-1, etc.).

4.3.2.1 Model BF24W: Run the final ASAP model as a traditional SCAA model in WHAM

The first step in WHAM modeling was to run the ASAP final model (BF24) as a traditional statistical catch-at-age model. A comparison of model results from the final ASAP model and BF24W show nearly identical results (Figure 26). The slight differences in model results can be attributed to different objective function and minimization algorithms between the two model frameworks.

One-step ahead residual diagnostics for the fleets indicate that the input CV of both fleets might be too broadly specified, with very tight blocking around 0 for the commercial fleet (fleet 1), and poor quantile distributions for both fleets (Figure 27).

4.3.2.2 Model BF26W to BF28W: Reduce CV around fleets

This series of models reduced the CV around fleet 1 by a factor of 0.5 (BF26W), the CV around fleet 2 by a factor of 0.5 (BF27W), and then both fleets' CVs by a factor of 0.5 (BF28W).

4.3.2.3 Model BF28W with different for NAA deviations specifications

Modal BF28W was used as a starting point to explore random effects models and the inclusion of environmental covariates on the catchability of selected survey indices.

The base statistical catch-at-age model (BF28W) and 6 state-space models (BF28W_m2 – BF28W_m7) with different options for treating the yearly transitions (survival) in recruitment and numbers-at-age were evaluated and compared (Table 7). Convergence diagnostics for each model run were examined and model selection via AIC was used to select a "best" model among the 6 models with comparable likelihood structures. Based on AIC selections, all of the top models were full state-space models, where survival of all ages were random effects with different correlation structures (Table 7). The model with the lowest AIC was BF28W_m7, which included correlation in the random effects by year and age (2D AR1). Model BF28W_m5 was very close in AIC but not within 2 AIC units of BF28W_m7 and was not considered equivalent based on model selection. Model BF28_m4 and BF28_m6 had similar model results but were noticeably higher in AIC.

Numbers-at-age deviations were correlated by age and year for the best model according to AIC, and were correlated by year for the next best model. The correlation by age was low and showed series of positive, negative and positive values from age-2 to age-4 in the middle of the time-series (Figure 28). The negative correlation between these ages is likely a result of the changing availability over time of this size class to the fisheries.

Results from the top 3 state-space models (BF28W_m7, BF28W_m5, and BF28W_m4) and the base statistical catch-at-age model (BF28W) showed good agreement among the model results (Table 7). The base model differed slightly in estimates of full F and SSB from 2008-2015 and in SSB again at the end of the time-series from 2016-2021, where SSB trended higher for this model (Figure 29). There were differences in the fleet selectivity block estimates, most notably with the base model in comparison to the state space models (Figure 30). In the final recreational selectivity block, the base model selectivity pattern was more domed, which likely resulted in the higher SSB estimates seen at the end of the time-series for this model. Index selectivity across the models showed differences mainly in those indices that catch older, larger bluefish. Those indices are the NEFSC Bigelow, PSIGNS, and ChesMMAP survey (Figure 31).

The final bluefish assessment model chosen by the working group was model BF28W_m7. A full presentation of parameter tables, input data, results, diagnostic, and retrospective plots are included in Working Paper 17 (Wood 2022d) and can also be downloaded or viewed separately from the following link: BF28W_m7_plots. A brief summary of results of the final model with selected plots are included below.

The final model fleet selectivity-at-age estimates for the two catch fleets showed a decrease in selectivity at middle ages (ages 3-5), with selectivity increasing at older ages. There was a decrease in the selectivity of these middle ages over time in the recreational selectivity blocks (Figure 32). Most of the index selectivities showed a domed selectivity after age-0. The MRIP CPUE index had a flat top logistic selectivity and was fully selected for the older ages. Both the NEFSC Bigelow index and the PSIGNS index had higher selectivity on the older, larger fish than the other fishery-independent indices (Figure 33).

Total abundance estimates from model BF28W_m7 peaked at a high of 599 million fish in 1985, declined to 162 million fish in 1995, and then increased to 269 million fish in 2005. Total abundance declined from 2005 to a low of 144 million in 2016, a small peak to 177 million in 2018, and a terminal year estimate of 162 million fish. Spawning stock biomass started from a high of 218,291 MT in 1985 and declined over the time-series to a low of 41,377 MT in 2018, and increased since then to 55,343 MT in 2021 (Figure 34). The majority of the spawning stock biomass is ages 5 and 6+ (30-60%) for the entire time-series. Fully selected fishing mortality in 2021 was 0.166, compared to an average full F from 1985 to 2021 of 0.309. Estimates of F have varied over the time-series from a peak in 2018 of 0.456 to the lowest value of 0.166 in 2021 (Figure 34). Estimates of recruitment remained stable over the time series, fluctuating around an average value of 128 million age-0 fish. Recruitment has been below average for the past 12 years, and was estimated at 87 million age-0 fish in 2021.

Retrospective pattern for the final model was examined for F, spawning stock biomass, and recruitment. Model BF28W_m7 exhibited a significantly improved retrospective pattern when compared to model BF24, the final ASAP model. The retrospective pattern was considered minor for SSB (Mohn's rho = 0.130), fishing mortality (Mohn's rho = -0.096), and recruitment (Mohn's rho = -0.063).

Model BF28W_m7 had excellent convergence diagnostics. Three sets of jitter analyses at increasing scale values of 1, 2, and 3 (the increase in scale broadens the distribution around the potential starting values by scaling the variance) were conducted. At a scale value of 1 (using variance estimates directly) 200/200 models converged at the original objective function. At a scale value of 2, all models converged, with 193/200 at the original objective function. Other objective function solutions were nearly identical to the original solutions (original objective function was 1468.54, other converged solutions were at 1468.69, 1468.72, and 1468.78). At a scale value of 3, all models converged with 155/200 jitter runs finding the original model solution and most of the other objective functions solutions very close to the original objective function (Figure 35). For comparison, the ASAP jitter analyses were only conducted at a scale of 1.

A historical retrospective analysis showing the model results from the 2015 benchmark assessment, 2021 operational assessment, BF01, the continuity run model, and BF28W_m7 (the final model) is presented in Figure 36.

4.3.2.4 Companion Model BF28WE: Environmental covariate on catchability of survey indices

One of the main reasons the bluefish assessment model was moved into WHAM was to explore the incorporation of environmental covariates on the catchability of different survey indices. Forage fish indices were developed using a VAST model (Section 1.2; Working Paper 4 Gaichas et al. 2022) and explored as environmental covariates on the catchability (q) of NEFSC survey indices and the MRIP CPUE index. These models are still under development and are being briefly presented as companion models for preliminary review. It is hoped that further exploration of these environmental models will lead to future improvements in the assessment.

The application of the forage fish indices as covariates on the catchability of the NEFSC science center surveys had mixed results. The forage fish index for the catchability of the NEFSC Albatross survey was explored as both a random walk and auto-regressive (AR1) process and each caused problems with the convergence of all models. Standard error around the covariate was explored using both the VAST estimated standard errors as an input standard error to the model, or allowing WHAM to estimate a single standard error of the covariate shared among time steps. All of the model runs either did not converge, or had issues with the hessian matrix calculations.

The forage fish index for the Bigelow survey did not have the same convergence issues as the Albatross index. The forage fish index was fit as a covariate on the Bigelow index catchability assuming a random walk over the time-series. All models with the forage fish covariate converged, but these models had worse fits than the base model according to AIC.

The application of the forage fish index to the MRIP CPUE index catchability was successful when implemented as an autoregressive (AR1) process over the time series with WHAM estimating a single shared standard error. The inclusion of the forage fish index improved the fit of all models (m2-m7), and model selection via AIC chose the time-varying catchability version of BF28W_m7 as the best model. This model will be referred to as model BF28W_m7ecov (where "ecov" refers to environmental covariate). Model BF28W_m7ecov had improved AIC of 2 units over BF28W_m5ecov, and by 5.6 units over BF28W_m7. The results from these top 3 models and the base model (BF28W) are presented in Figure 37.

A full presentation of parameter tables, input data, results, diagnostic, and retrospective plots are available for the best model BF28W_m7ecov are included in Working Paper 17 (Wood 2022d) and can also be downloaded or viewed separately from the following link: BF28W_m7ecov_plots.

The use of the forage fish index as a covariate on catchability led to an overall decreasing trend in catchability over time (Figure 38). The MRIP index is important in scaling the biomass results, and the lower availability at the end of the time-series led to higher recent biomass estimates from the environmental model. Spawning stock biomass started from a high of 181,804 MT in 1985 and declined over the time-series to a low of 52,697 MT in 2018, and increased since then to a value of 74,549 MT in 2021. Fully selected fishing mortality in 2021 was 0.126, compared to an average F from 1985 to 2021 of 0.271. Estimates of F have varied over the time-series from a peak in 1987 of 0.503 to the lowest value of 0.126 in 2021. Estimates of recruitment have remained stable over the time series, fluctuating around an average value of 143 million age-0 fish. Recruitment has been below average for the past 12 years, and was estimated at 106 million age-0 fish in 2021.

4.3.2.5 Model BF28W sensitivity analyses

A number of sensitivity runs of the final model (BF28W_m7) were explored. The model results and retrospective pattern results from each of these runs are presented in Table 8.

The sensitivity of the final model to the indices was explored in several ways. First, each index was removed individually, and the model was re-run to gauge the effect. Results from this series of models are in Table 8 and Figure 39. The final model was not overly sensitive to any single index, which was a shift from past bluefish assessment models. The bluefish assessment used to be heavily weighted towards the MRIP CPUE index. In many cases the model would not converge without this index included, or the model would scale the biomass to an unrealistic magnitude to find a model solution. This was no longer the case with model BF28W_m7, which converged without the MRIP CPUE index and found a solution that is in agreement with all the other index sensitivity runs (Figure 39). The model results appeared to be most sensitive to the removal of the PSIGNS index, which is an important index for tracking the abundance of older fish. Removal of this index significantly reduced SSB and increased F.

Two other index sensitivity runs were explored. Model based (GLM) versions of the indices (Working Paper 7 Celestino et al. 2022a) were substituted into the model for a sensitivity run. This change had very little impact on the model results and retrospective results.

Next, NEFSC indices that included some offshore strata were substituted into the model. The NEFSC survey encounters larger bluefish offshore in some years, and these "offshore" indices were explored to see impact of including bluefish observations from these offshore strata. The results from this sensitivity run were similar to the final model results, with both recruitment and SSB scaled upwards a small amount.

The next group of sensitivities focused on how recreational discard lengths were developed. First, a sensitivity was run that borrowed recreational discard (MRIP B2) lengths across regions, as opposed to using a cumulative length by season/region for years where the number of lengths sampled was less than 30. This sensitivity did not have good convergence properties and the hessian was not positive definite. This was due to changes in some of the fleet selectivity-at-age estimates, with some hitting the bound of 1.0. Further development could improve this model and results but were beyond the scope of a sensitivity analysis.

Next, recreational harvest lengths were borrowed for season/region years where the number of recreational discard lengths sampled was less than 30. The results of this model were nearly identical to the final model run. Finally, recreational length proportions from harvested fish (MRIP AB1) were used in place of the dead release lengths (instead of the i9, ALS, and VAS lengths). This model sensitivity also produced very similar results to the final model, with a slightly reduced recruitment and SSB, and slightly increased F (Figure 40).

Other sensitivity runs that were explored included:

1. Using the MRIP directed trip index instead of the Guild index
2. Setting the MRIP index to estimate selectivity-at-age instead of estimating a logistic curve
3. Assuming 15% recreational discard mortality instead of 9.4%
4. Assuming both the upper and lower confidence bounds for the Lorenzen M estimates

Results from each of these one-off sensitivities are presented in Table 8.

5 STATUS DETERMINATION CRITERIA

Term of Reference #5: Update or redefine status determination criteria (SDC; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY reference points) and provide estimates of those criteria and their uncertainty, along with a description of the sources of uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for reference points. Compare estimates of current stock size and fishing mortality to existing, and any redefined, SDCs.

In a meeting of the Mid-Atlantic Fishery Management Council Scientific and Statistical Committee (SSC) following the 2015 benchmark assessment for bluefish, the SSC stated, "…the F_{MSY} proxy of $F_{40\%}$ might be inappropriate for Bluefish, a highly productive stock…". Citing two studies as support, the SSC used $F_{35\%}$ to set the overfishing limits for 2016-2018. The two papers the SSC cited (Rothschild et al. 2012; Thorson et al. 2012) were read and evaluated for support to use $F_{35\%}$ for bluefish in the 2022 assessment update. The WG agreed that the literature supported the use of $F_{35\%}$ for bluefish and continued the use of $F_{35\%}$ as the F_{MSY} proxy.

Many species managed in the Greater Atlantic region that use per-recruit reference points use $F_{40\%}$ (e.g., many groundfish species with analytical assessments, Atlantic herring, scup and black sea bass). However, bluefish is not the only example of a species that currently uses $F_{35\%}$ as summer flounder also uses this reference point.

5.1 Stock Status from the Continuity Run, BF01

Stock status was first determined using the continuity run model, which is the current accepted model for providing management advice, and would be used in absence of the research track assessment. Reference points were calculated using the non-parametric yield and SSB per-recruit long-term projection approach assuming 5-year averages for fishery selectivity, maturity and weights-at-age for SSB per recruit calculations. The cumulative distribution function of the 1985-2021 recruitment estimates were resampled to provide future recruitment estimates for the projections and used to estimate the SSB_{MSY} reference point associated with $F_{35\%}$ from a 100-year projection.

Existing reference points from the 2021 management track assessment (data through 2019) were F_{MSY} proxy = $F_{35\%}$ = 0.181 and SSB_{MSY} = 201,729 MT (1/2 SSB_{MSY} = $SSB_{THRESHOLD}$ = 100,865 MT). Updated reference points from the continuity run are F_{MSY} proxy = $F_{35\%}$ = 0.176 and SSB_{MSY} = 190,771 MT (1/2 SSB_{MSY} = $SSB_{THRESHOLD}$ = 93,386 MT).

A retrospective adjustment of the terminal year results for F and SSB resulted in these values being outside of their 90% MCMC confidence bounds. The retrospective pattern in F and SSB was considered major (SSB_{rho} = 0.29, F_{rho} = -0.28, based on 7-year peel) and required a retrospective adjustment to determine stock status. The 2021 retrospective adjusted F was 0.222 and falls above F_{MSY}. The 2021 retrospective adjusted value for SSB was 70,900 MT, and is lower than $SSB_{THRESHOLD}$. The results from the continuity run model indicate that the bluefish stock is overfished, and over-fishing is occurring (Figure 41). The over-fishing status has changed since the 2021 management track assessment. This change is a result of increased retrospective for F in the updated continuity run model, resulting in a retrospective adjustment that increased the terminal F value.

5.2 Stock Status from the Final Research Track Model, BF28W_m7

Both $F_{35\%}$ and $SSB_{35\%}$ were calculated internally in WHAM using average recruitment over the time series (1985-2021), and 5-year averages for fishery selectivity, maturity and weights-at-age for SSB per recruit calculations. The 5-year average was selected for those parameters to capture the most recent conditions while still smoothing some interannual variability; the full time-series of recruitment was chosen to fully capture the range of possible recruitment, given that there did not appear to be a significant regime shift in recruitment levels for bluefish over the time series. $F_{35\%}$ explicitly accounts for uncertainty from selectivity; $SSB_{35\%}$ explicitly accounts for uncertainty from selectivity and average recruitment. Uncertainty in the reference points associated with the 2D-AR1 process is implicitly accounted for through its impacts on selectivity and average recruitment. Additional sources of uncertainty in reference points not explicitly accounted for include uncertainty associated with the remaining SPR calculation inputs (e.g., natural mortality, maturity, and average weights-at-age).

Reference points from the final model (BF28W_m7) were F_{MSY} proxy = $F_{35\%}$ = 0.248 (95% CI: 0.209 – 0.299) and SSB_{MSY} proxy = $SSB_{35\%}$ = 91,897 MT (95% CI: 66,219–127,534 MT); $SSB_{THRESHOLD}$ =1/2 SSB_{MSY} proxy = 45,949 MT (95% CI: 33,110–66,768 MT). The retrospectively adjusted values of terminal year F and SSB were within the 90% confidence bounds of the unadjusted values, indicating a retrospective adjustment was not necessary to determine stock status (Figure 42). The terminal year SSB was 55,344 MT (95% CI: 35,185 – 87,052 MT) which is above the $SSB_{THRESHOLD}$ and 60% of SSB_{MSY}. Full fishing mortality was 0.166 (95% CI: 0.103 – 0.268) in 2021, which is 67% of the $F_{35\%}$ reference point (Figure 43). Accounting for uncertainty in reference points and terminal year F and SSB estimates, stock status determination based on the final model indicates that there is an 87% chance that the bluefish stock is currently not overfished and over-fishing is not occurring (Figure 44).

A comparison of stock status results from 2015 benchmark model, 2021 operational assessment, the current assessment continuity run, and the final model from this assessment is presented in Table 9.

6 PROJECTION METHODS

> *Term of Reference #6: Define appropriate methods for producing projections; provide justification for assumptions of fishery selectivity, weights at age, maturity, and recruitment; and comment on the reliability of resulting projections considering the effects of uncertainty and sensitivity to projection assumptions*

Short-term projections were conducted in WHAM, and incorporate model uncertainty, auto-regressive processes and uncertainty in recruitment and numbers-at-age. Removals in 2022 were assumed to be equal to the 2022 ABC (11,460 MT), and projections were carried forward for years 2023-2025 with different fishing mortality and harvest assumptions: F = 0, $F_{status\ quo}$ = 0.166, $F_{35\%}$ = 0.248, and that harvest in each year is equal to the acceptable biological catch (ABC) in each year. The annual ABC values were derived using projected OFL catch and applying the Mid-Atlantic Fishery Management Council (MAFMC) risk policy with an assumed OFL CV (MAFMC, 2020). In recent years, the ABC for bluefish has been developed using an OFL CV = 100%. Projections were carried out assuming an OFL CV of 100% and 60%.

Fishing at $F_{35\%}$ caused a decrease in biomass over the projected years, from 65,805 MT in 2022 to 61,784 MT in 2025 (Table 10). The catches associated with fishing at the reference point (OFL catch) ranged from 13,909 MT to 13,584 MT (Table 11). The probability of the stock being over the biomass threshold in 2025 was 0.84 for the $F_{35\%}$ projection.

The most realistic projections are the F status quo projection, and the MAFMC risk policy projection at an assumed CV of 100%. The risk policy approach is how management specifications are currently developed for bluefish. The probability of the stock being over the biomass threshold in 2025 was 0.93 for the F status quo projection, and 0.88 for the risk policy approach.

The projections use 5-year averages for natural mortality, maturity, fishery selectivity and weights-at-age. The 5-year average was selected for those parameters to capture the most recent conditions while still smoothing some interannual variability; the full time-series of recruitment was chosen to fully capture the range of possible recruitment, given that there did not appear to be a significant regime shift in recruitment levels for bluefish over the time series. Projections were not retrospectively adjusted, as the adjusted terminal year estimates of F and SSB fell within the 90% confidence intervals of the unadjusted values (Figure 42). The sensitivity of these projection assumptions were tested using 3-year, and 10 year averages. The projections are not overly sensitive to these assumptions. Assuming a 3-year averages leads to ~7.0% decrease in biomass, and ~6.0% decrease in catch when compared to the 5-year average. Assuming a 10-year average for the projection input results in a <1.0% difference in all results when compared to the 5-year average.

A final projection was carried out at $F_{rebuild}$. The bluefish stock is currently under a rebuilding plan, with a target date of 2028. $F_{rebuild}$ for the stock is currently set a 0.166 and a projection through 2028 was done assuming this value in each year. The 2028 SSB resulting from this projection is 79,215 MT, which is 86% of the biomass target (91,897 MT).

7 RESEARCH RECOMMENDATIONS

> *Term of Reference #7: Review, evaluate, and report on the status of research recommendations from the last assessment peer review, including recommendations provided by the prior assessment working group, peer review panel, and SSC. Identify new recommendations for future research, data collection, and assessment methodology. If any ecosystem influences from TOR 1 could not be considered quantitatively under that or other TORs, describe next steps for development, testing, and review of quantitative relationships and how they could best inform assessments. Prioritize research recommendations.*

7.1 Status of Previous Research Recommendations

Some research recommendations were repeated in various documents (e.g., 2015 benchmark, SSC documents); for brevity, where the same, or substantially similar recommendations were made, we consolidated them under a single heading (e.g., 2015 benchmark), but noted the additional documents in which the recommendation was raised.

7.1.1 Research recommendations from SAW60 (NEFSC 2015)

7.1.1.1 High Priority

Recommendation: Determine whether NC scale data from 1985-1995 are available for age determination; if available, re-age based on protocols outlined in ASMFC (2011); if re-aging results in changes to age assignments, quantify the effects of scale data on the assessment

WG Response: The WG spoke with NC technical staff who endeavored to find the historical structures for ageing (in addition to 1985-1995 scale samples, otolith samples through 2000 were also included in the search). NC staff reached out to multiple additional agency staff at multiple offices throughout the state and the samples were not found.

Recommendation: Develop additional adult bluefish indices of abundance (e.g., broad spatial scale longline survey or gillnet survey); initiate fishery-dependent or fishery independent sampling of offshore bluefish populations to reduce reliance on MRIP sampling [also recommended in July 2015 and July 2021 SSC reviews]
WG Response: As part of the current research track assessment, ASMFC solicited data from state, federal, and academic partners as well as stakeholders via a press release and public data workshop. However, no new fishery-independent indices that capture very large fish, or offshore fish, were identified. The WG engaged with stakeholders to understand the extent of possible adult bluefish interactions with the offshore longline tilefish fishery, who indicated offshore interactions do occur with bluefish, but not consistently across states, and no clear trend was identified.

Recommendation: Expand age structure of SEAMAP index
WG Response: The WG added an age-1 index of abundance from the SEAMAP survey to the assessment; other ages classes were rarely encountered.

7.1.1.2 Moderate Priority

Recommendation: Investigate species associations with recreational angler trips targeting bluefish (on a regional and seasonal basis) to potentially modify the MRIP index used in the assessment model; Explore alternative definitions for targeting for calculating CPUE (e.g., directed trips or directed trips + incidental harvest) [also recommended in July 2015 and July 2021 SSC reviews]
WG Response: The WG developed an MRIP index using a species-association method to identify bluefish trips as well as a directed trips approach; see Section 3.2.5 and Working Paper 13 (Drew 2022d) for more details.

Recommendation: Explore age- and time-varying natural mortality from, for example, predator prey relationships; quantify effects of age- and time-varying natural mortality in the assessment model
WG Response: The WG evaluated a suite of life history and environmental data approaches to estimating age-constant and age-varying natural mortality and selected the Lorenzen (1996) age-varying approach for the base model; see Working Paper 6 (Tyrell and Truesdell 2022). The WG explored trends in large scale predator data (e.g., Shortfin mako, *Isurus oxyrinchus*) to potentially inform time-varying natural mortality, but concluded data were not sufficient to support a time-varying M at this time.

- *Next steps: If relevant predator abundance information becomes available in the future, a predator index could be used to inform time-varying natural mortality. The working group produced a condition index for bluefish of three size groups, which could be used to inform time-varying natural mortality in a WHAM model in the future; the working group prioritized using the forage fish index as a WHAM covariate in this research track assessment.*

Recommendation: Continue to evaluate the spatial, temporal, and sector-specific trends in bluefish growth and quantify their effects in the assessment model.

WG Response: The WG explored life history characteristics over various temporal and spatial scales (Working Paper 5 Truesdell et al. 2022), and constructed age-length keys and length-frequencies at the seasonal and regional level. While the age-length data were too sparse to support the season-region keys (Working Paper 14 Celestino 2022b), the catch length frequencies for both harvest and dead discards were stratified to the season and region level, an advance from the 2015 benchmark where only the harvest was stratified at that level.

Recommendation: Continue to examine alternative models that take advantage of length-based assessment frameworks. Evaluate the source of bimodal length frequency in the catch (e.g., migration, differential growth rates).
WG Response: The WG did not believe a length-based approach would improve the stock assessment for bluefish given the improvements in the age data collection across the coast and the longer time series of otolith-only data, and so did not pursue a length-based modelling approach. In addition, the WG did not have the type of data that would support a size transition matrix. The WG investigated whether the bimodal length frequency in the catch could be attributed to mid-size bluefish migrating to the Gulf of Mexico, but did not find support for this hypothesis. To some extent, with the development, expansion, and continuation of the coastwide biological collection program (Amendment I to the FMP), the bimodal pattern has become less frequent but has not disappeared; the WG suggests that this remains a research recommendation. Tagging programs (e.g., traditional, satellite) could provide additional insights.

Recommendation: Modify thermal niche model to incorporate water temperature data more appropriate for bluefish in a timelier manner [e.g., sea surface temperature data & temperature data that cover the full range of bluefish habitat (SAB and estuaries)].
WG Response: The 2015 analysis of the centers of biomass (COB) indicated that COB positions were correlated with variations in body size and abundance, but not temperature, and the annual proportion of thermal habitat suitability surveyed did not exhibit consistent, systematic trends. Therefore, the WG did not update the thermal niche model for this assessment, but included temperature as a covariate in the VAST forage fish index to serve a similar function as a covariate to inform catchability of the indices; see Working Paper 4 (Gaichas et al. 2022) for more details.

7.1.2 Research recommendations from SSC (July 2015)

Recommendation: Develop Bluefish-specific MSY reference points or proxies.
WG Response: Bluefish-specific MSY proxy reference points were developed for this assessment (see Section 5 above).

Recommendation: Low frequency environmental variability may have caused changes in the timing of the movement of juvenile Bluefish through the region that, in turn, may have affected availability. Changes in the selectivity of age-0 Bluefish in the survey relative to water column or surface temperature and date should be examined.
WG Response: The WG investigated the influence of temperature effects on bluefish as part of the ESP. See Section 1.1 and Working Papers 1 and 3 (Tyrell et al. 2022, Tyrell 2022) for more detail. However, more work, including additional bluefish data collection, needs to be done to incorporate this information into the assessment model framework in a quantitative way (see also

Section 7.2 below). The WHAM framework will allow for continued exploration and testing of covariates influencing time-varying catchability and selectivity.

- *Next steps: Additional survey data in the late fall would be needed to determine whether bluefish spawning is extending later in the year, which may be possible due to warmer temperatures extending later in the fall. Environmental covariates on recruitment could be incorporated into WHAM to test for improvements to model fit.*
- *Next steps: Further VAST models could be developed that incorporate additional scientific surveys, e.g., ChesMMAP and NEAMAP. The effect of environmental variability and timing of sampling could also be further investigated with VAST models, which can account for the day of sampling using a catchability covariate and can account for environmental variability using density covariates.*

Recommendation: Evaluate methods for integrating disparate indices produced at multiple spatial and temporal resolutions into a stock-wide assessment model, especially for a migratory species like Bluefish [also a July 2021 SSC review recommendation]
WG Response: The WG continued the use of the Conn (2010) approach to develop a single recruitment index from multiple state seine surveys as a means to addressing this research recommendation; see Section 3.1.2 and Working Paper 12 (Drew 2022c). The WG also explored using VAST to develop a forage fish index from multiple surveys (Section 1.2 and Working Paper 4 Gaichas et al. 2022) and to develop a standardized index with a single time series from the NEFSC Albatross and Bigelow vessels (Section 1.1.2 and Working Paper 3 Tyrell 2022); both approaches need more development before they can be incorporated into the base model of the assessment (see also Section 7.2 below).

- *Next steps: The bluefish Albatross-Bigelow VAST index could be further developed with environmental covariates (such as temperature). Additionally, multiple surveys could be combined in the VAST index.*

7.1.3 Research recommendations from SSC (July 2021)

Recommendation: A primary source of uncertainty is the recreational catch time series. The MRIP trend does not seem consistent with hypothesized reasons for differences between the mail and phone surveys. This historical correction to the MRIP estimates for bluefish should be explored further to evaluate the causes of differences from other species and to consider their plausibility.
WG Response: The WG examined differences in the calibrated and uncalibrated MRIP estimates of bluefish catch and found that while the magnitude of the calibration effect differed by mode and state, overall, we do generally see differences over time consistent with the hypothesized reasons for differences between the mail and phone surveys, and similar to trends in other mid-Atlantic species like summer flounder (*Paralichthys dentatus*), tautog (*Tautoga onitis*), and striped bass (*Morone saxatilis*). More detail is available in Working Paper 9 (Drew 2022a).

Recommendation: Investigate whether and how the selectivity pattern in discards has changed over time; the SSC questioned the methods for estimating the weight of recreational discards and the disparity between the use of volunteer angler data and the assumptions used by GARFO.
WG Response: For this assessment, the WG stratified released length frequency by region when calculating the weight of dead recreational releases to account for differences in the size

structure of removals and the release length samples between the regions. In addition, during the course of the present research track assessment, the WG communicated with the Greater Atlantic Regional Fisheries Office (GARFO) staff to ensure there is no longer a discrepancy between how the assessment estimates the weight of dead recreational releases and how that component is estimated for management; the agreed upon methods are consistent with other managed species (e.g., black sea bass, summer flounder).

Recommendation: Investigate patterns and trends in recent recruitments; the SSC noted low recruitment estimates in 2019 and asked whether it was possible to detect shifts between spring vs late summer recruiting cohorts.
WG Response: The WG's review of recruitment data largely suggested that data were not clear or sufficient to resolve whether there has been a shift between spring versus late summer recruiting cohorts; see Working Paper 1 (Tyrell et al. 2022) for a more detailed review of recruitment information available for bluefish.

- *Next steps: In order to quantitatively distinguish between spring-spawned and summer-spawned bluefish cohorts, regular seasonal sampling targeting small (<10cm) bluefish would need to be conducted over the broader Mid-Atlantic region and would have to extend later into the fall than current surveys.*

Recommendation: Long term environmental variability may have caused changes in the timing of the movement of juvenile Bluefish and the distribution of adults throughout the region that, in turn, may have affected availability.
WG Response: The WG explored development of VAST index of small pelagic fish aggregate abundance via predator diet data as a covariate for bluefish availability (Section 1.2 and Working Paper 4 Gaichas et al. 2022) and incorporating environmental covariates into index development via VAST (Section 1.1.2 and Working Paper 3 Tyrell 2022) and GLM-based standardization (Working Paper 7 Celestino et al. 2022a); both approaches need more development before they can be incorporated into the base model of the assessment (see also Section 7.2 below).

- *Next steps: More formal examination of time series changepoints and relationships of the forage indices with other ecosystem indicators will be explored during the NEFSC's 2023 State of the Ecosystem report development cycle, and can be included in future bluefish assessments.*
- *Next steps for Albatross-Bigelow VAST: there are additional VAST model changes that can be explored to better understand the influence of environmental covariates on bluefish distribution. The VAST model presented in this report could be further developed to successfully incorporate environmental covariates such as temperature.*

7.2 New Research Recommendations

7.2.1 High Priority

Expand collection of recreational release length frequency data
Recreational release mortality accounts for approximately 15% of total removals in weight in recent years, but information on the size structure of released fish is limited, particularly in the South Atlantic. The assessment now stratifies length frequency of released fish by region, but requires borrowing across years with low sample sizes (n<30), and this borrowing should be minimized or avoided where possible to better capture year class effects. Expansion and

promotion of volunteer angler survey programs would be one option to reduce this source of uncertainty in the assessment.

Continue development and refinement of the forage fish / availability index as well as incorporation of this index into a base model for bluefish management advice
Preliminary modelling that incorporated the forage fish index suggested an improved model fit relative to a model without the index. The forage fish index could provide information on availability of bluefish to different surveys and fisheries, and could potentially help the model resolve conflicts between indices that occur more offshore and indices that occur more inshore. Additional work could include:

- Investigate sources of piscivore diet data for "inland waters" (Chesapeake Bay, Delaware Bay, Long Island Sound) to integrate into the model, potentially providing more insight into availability to the MRIP index. (ChesMMAP has diet data; other surveys or studies should also be investigated)
- Investigate sources of piscivore diet data south of Cape Hatteras to expand to full bluefish range
- Investigate other potential environmental covariates (e.g., higher resolution SST)
- Continue modelling within the WHAM framework to resolve issues identified in TOR4
- Continue to explore environmental linkages to catchability, selectivity, recruitment, and natural mortality using WHAM

Initiate additional fisheries independent surveys and/or fishery-dependent sampling programs to provide information on larger, older bluefish
This remains a high priority given the limited information on older (e.g., age 2+) bluefish collected by existing fishery independent surveys. This item addresses the need to adequately characterize dynamics of older fish that are currently not well sampled by fishery independent trawl surveys, as well as to understand the extent of summer and fall spawning and the contribution of these fish to year classes. This item also would help address unresolved issues identified above (e.g., relative cohort strengths, offshore movements, environmental effects). Further engagement with stakeholders can help identify areas of incidental bluefish catch in offshore fisheries and inform potential development of voluntary or required reporting programs and data sources.

Continue coastwide collection of length and age samples from fishery-dependent and fishery-independent sources.
The availability of bluefish to different fisheries varies throughout the year along the coast; in order to accurately characterize the age-structure of the removals, adequate samples, stratified spatially and temporally, need to be collected. The increased sampling at the state level as a result of Amendment 1 to the Bluefish FMP improved the available data and reduced gaps in the ALK. Current sampling levels should be maintained at a minimum.

7.2.2 Medium Priority

Further index of abundance development and refinement

The large number of indices input into the model sometimes provide conflicting signals and add additional parameters that need to be estimated. Exploring environmental drivers of bluefish distribution and exploring index consolidation using VAST or other modeling approaches could provide more coherent indices and/or provide information on catchability covariates to resolve conflicting signals and improve model fits.

Develop a recreational economic demand model

Recreational demand models can inform managers of the likely economic and biological implications of alternative regulatory and stock conditions. Given the large role recreational effort plays in the bluefish fishery, efforts in developing recreational demand models should be prioritized to develop measures that will meet both biological and socioeconomic goals for the bluefish fishery (Appendix 1 of Working Paper 1 Tyrell et al. 2022).

7.2.3 Low Priority

Development of an updated recreational release mortality study

Given the importance of recreational releases in the bluefish fishery for both accurately estimating total catch and therefore population scale and in the correct allocation of dead catch to the commercial and recreational sectors, reducing uncertainty on the release mortality estimate is important, especially if it has changed over time with changing angler behavior. The WG discussed: (1) examination of release mortality based on study factors (hook type, fish length, etc.), and (2) a comparison of release mortality estimates generated from a variety of methods.

A coordinated tagging program to help understand migration patterns potentially contributing to patterns in length frequency distributions.

To the extent that spatiotemporal variation in availability is contributing to the bimodal length frequency distribution, this could help resolve a source of uncertainty in the assessment. The WG was not able to resolve the source of bimodal length frequency distributions and has hypothesized offshore migration or summer/fall residency in southern waters makes those size classes of bluefish unavailable to fisheries and surveys. A coordinated fishery-independent tagging program could also help to more definitively resolve migrations between the Gulf of Mexico and Atlantic coast. A coordinated program could also address the release mortality recommendation above and provide a different source of growth information to compare to age-based methods.

Commercial discard length frequency data

There are currently no length data to characterize the length frequency of commercial discards. This source of mortality is small relative to other sources of fishing mortality, but does represent a source of uncertainty.

8 BACKUP ASSESSMENT APPROACH

> *Term of Reference #8: Develop a backup assessment approach to providing scientific advice to managers if the proposed assessment approach does not pass peer review or the approved approach is rejected in a future management track assessment.*

A backup assessment approach is required to be in place as a hedge against a scenario where the primary catch-at-age model fails peer review. Such alternative models could include biomass dynamic-type models (e.g., as used for red crab), swept area approaches (e.g., witch flounder), catch curves, index-based methods (e.g., Georges Bank Atlantic cod) or other approaches. In one case, a statistical catch-at-age approach was put forward as the primary model and modifications that were still in the statistical catch-at-age framework were suggested as a backup approach (e.g., American Plaice Research Track).

The Working Group chose the index-based method Ismooth (previously known as PlanBSmooth; see Chris Legault's GitHub repository for more information) as the backup model due to its performance in the analyses performed by the Index Based Model Working Group (NEFSC 2020) and because it has a history of application at the NEFSC as an approach that has been used to develop ABCs (e.g., Georges Bank cod, Gulf of Maine / Northern Georges Bank and Southern Georges Bank / Mid-Atlantic monkfish).

In general, this approach applies recent trends in an index or indices to recent dead catch to generate ABC advice. There are two steps in the process. The model calculates an average of normalized indices that are selected for inclusion, applies a loess smooth to those values, fits a linear model to the final three years of log-transformed smoothed data, and extracts the slope of the fit. The results are then applied to recent dead catch levels. In this case, the previous three years of dead catch are averaged and the Ismooth exponentiated slope is multiplied by that average to generate the advice.

Ismooth was one of a number of data poor approaches examined by the NEFSC Index Based Methods Working Group (NEFSC 2020, Legault et al., n.d.). The primary focus of this Working Group was to quantify the performance of data poor approaches in circumstances that led to severe retrospective errors in statistical catch-at-age models. That group found that none of the data poor methods outperformed a retrospectively adjusted catch-at-age model over the long-term, but also concluded that the Ismooth approach performed reasonably well relative to other methods, especially with respect to maintaining an acceptable level of SSB and constraining F. Thus, the Ismooth approach represents a reasonable choice if the statistical catch-at-age model were to fail.

The WG simulated the performance of Ismooth relative to historical bluefish ABCs that were based on results of the ASAP model; see Working Paper 18 (Truesdell 2022) for additional information. In general, the retrospective advice calculated by the Ismooth model was correlated with the actual ASAP-derived ABCs that were recommended for management use by the SSC, especially when the MRIP index was included when developing the Ismooth advice (Figure 45). Accordingly, as a one-off ABC tool (i.e., when differences between approaches do not compound over time), Ismooth offers similar advice to the previously accepted statistical catch-at-age model (ASAP) given the historical indices that were used to compile the Ismooth estimate.

The WG explored other data-limited approaches for estimating sustainable yield including Depletion-Corrected Average Catch (DCAC; MacCall 2009) and Depletion-Based Stock Reduction Analysis (DBSRA; Dick and MacCall 2011) as was done in the previous benchmark

(NEFSC 2015). However, McCall (2009) noted that the DCAC method is not recommended for species where natural mortality is greater than approximately 0.20. Because of that, DCAC was dropped from further consideration as an alternative model for bluefish given the updated natural mortality rate. The DBSRA model produced significantly higher estimates of biomass and sustainable yield than the age-structured model and had a low rate of accepted runs when parameterized with updated bluefish life history information; because of this and concerns about the underlying surplus production model framework of the DBSRA, the WG did not recommend this approach for providing alternate catch advice.

Swept area approaches were also investigated but given the importance of the recreational sector a method that could incorporate the MRIP index was preferable. In addition, bluefish catchability and selectivity in trawl nets is not well understood which would decrease confidence in a trawl survey-only swept area approach. Catch curves were considered but were not recommended by the WG as a backup approach, as the WG did not know of other assessments that used catch curves to produce catch advice.

The WG does not anticipate the need for the backup approach to be applied. The scenarios in which the selected catch-at-age model would be abandoned are limited to new data that caused complete convergence failure, the discontinuation of critical data streams or logistical issues that precluded the model fitting process altogether. In the case of severe retrospective errors, the Index Based Model Working Group found that a retrospectively adjusted statistical catch-at-age model did not perform worse than index or data poor approaches, so this potential issue is not expected to cause a transition to the backup assessment. If new data causes major issues with model fitting, modifications within the catch-at-age framework (e.g., data weighting, random effect structure, etc.) would be exhausted before moving to the backup assessment approach. Such changes could be implemented through an Expedited or Enhanced management track peer review.

9 REFERENCES

ASMFC. 2011. Proceedings of the Atlantic States Marine Fisheries Commission Bluefish Ageing Workshop. ASMFC, Alexandria, VA. 26p. Available online at: http://www.asmfc.org/uploads/file/bluefishAgeingWorkshopReport_May2011.pdf

ASMFC. 2019. Atlantic States Marine Fisheries Commission, Weakfish Stock Assessment Update report. Arlington, VA. 86 pp. Available online at: http://www.asmfc.org/uploads/file/5de7fc7c2019WeakfishAssessmentUpdate.pdf

ASMFC. 2020. Memorandum from the bluefish Technical Committee to the bluefish Management Board, regarding recommendations on bluefish age and length frequency data sampling. 5 p. Available online at: http://www.asmfc.org/uploads/file/6022ea5dBluefishTC_Report_Sept2020.pdf

ASMFC. 2021. Report from the 2021 Bluefish State Data Review Workshop. Arlington, VA. 115p. Available online at: https://www.asmfc.org/uploads/file/63487cfe2021BluefishStateDataReviewReport.pdf

Avarovitz, T.R. 1981. A brief historical review of the Woods Hole Laboratory trawl survey time series. Pages 62–67 In: W.G. Doubleday and D. Rivard (editors), Bottom trawl surveys. Canadian Special Publication of Fisheries and Aquatic Sciences 58. 273 p.

Carvalho, F., H. Winker, D. Courtney, M. Kapur, L. Kell, M Cardinale, M. Schirripa, T. Kitakado, D. Yemane, K. R. Piner, and M. N. Maunder. 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research 240, p.105959.

Conn, PB. 2010. Hierarchical analysis of multiple noisy abundance indices. Canadian Journal of Fisheries and Aquatic Sciences 67:108-120.

Dick, E. J. and A. D. MacCall. 2011. Depletion-based stock reduction analysis: a catch-based method for determining sustainable yields for data-poor fish stocks. Fisheries Research 110(2): 331-341.

Gerritsen, H.D., McGrath, D. and Lordan, C. 2006. A simple method for comparing age-length keys reveals significant regional differences within a single stock of haddock (*Melanogrammus aeglefinus*). ICES Journal of Marine Science, 63: 1096-1100.

Hewitt, D. A., and J. M. Hoenig. 2005. Comparison of two approaches for estimating natural mortality based on longevity. Fishery Bulletin 103 (2): 433.

Klein-MacPhee, G. 2002. Bluefish: family Pomatomidae. In Bigelow and Schroeder's fishes of the Gulf of Maine (B. B. Collette, and G. Klein-MacPhee, eds.), p. 400–406. Smithsonian Institution Press, Washington, DC.

Legault, C. M. 2012. User manual for ASAP 3. 22 p.

Legault, C. M. and V. Restrepo. 1999. A flexible forward Age-Structured Assessment Program. ICCAT Coll. Vol. Sci. Pap. 49.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49 (4): 627–42.

MAFMC. 2020. OMNIBUS Acceptable Biological Catch and Risk Policy Framework Adjustment. Environmental Assessment. 116 pp. Document available at this link.

McCall, A. 2009. Depletion-corrected average catch: a simple formula for estimating yields in data poor situations. ICES Journal of Marine Science 66:2267-2271.

Methot, R.D. and Wetzel, C.R. (2013). Stock Synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86-99. https://doi.org/10.1016/j.fishres.2012.10.012

Miller, T. J., Hare, J. A., and Alade, L. A. 2016. A state-space approach to incorporating environmental effects on recruitment in an age-structured assessment model with an application to Southern New England yellowtail flounder. Canadian Journal of Fisheries and Aquatic Sciences 73(8): 1261-1270. Doi: 10.1139/cjfas-2015-0339

Miller, T. J. and Hyun, S-Y. 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Canadian Journal of Fisheries and Aquatic Sciences 75(5): 691-703. Doi: 10.1139/cjfas-2017-0035

Miller, T. J., O'Brien, L., and Fratantoni, P. S. 2018. Temporal and environmental variation in growth and maturity and effects on management reference points of Georges Bank Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences 75(12): 2159-2171. Doi: 10.1139/cjfas-2017-0124

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES Journal of Marine Science 56(4): 473-488.

Munch, S. 1997. Recruitment dynamics of bluefish, *Pomatomus saltatrix*, on the continental shelf from Cape Fear to Cape Cod, 1973-1995. M.S. thesis, State University of New York at Stony Brook, NY, 127 p.

National Research Council (NRC). 2006. Review of Recreational Fisheries Survey Methods. Washington, DC: The National Academies Press.

NEFSC. 2015. 60th Northeast Regional Stock Assessment Workshop (60th SAW) assessment report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 15-08; Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026 http://doi.org/10.7289/V5W37T9T

NEFSC. 2020. Draft report of the Index Based Methods Working Group, submitted for peer review November 20th 2020. 128 pp. Report available at this link.

Ng, E. L., J. J. Deroba, T. E. Essington, A. Grüss, B. E. Smith, and J. T. Thorson. 2021. Predator stomach contents can provide accurate indices of prey biomass. ICES Journal of Marine Science 78(3):1146–1159.

Norcross, J. J., S. L. Richardson, W. H. Massmann, and E. B. Joseph. 1974. Development of young bluefish (Pomatomus saltatrix) and distribution of eggs and young in Virginian coastal waters. Transactions of the American Fisheries Society 103(3): 477-497.

Olla, B.L. and A.L. Studholme. 1971. The effect of temperature on the activity of bluefish, *Pomatomus saltatrix* L. Biological Bulletin 141:337-349.

Perretti, C. T., and J. T. Thorson. 2019. Spatio-temporal dynamics of summer flounder (Paralichthys dentatus) on the Northeast US shelf. Fisheries Research 215:62–68.

Papacostas, K. J., and J. Foster. 2020. A replication approach to controlled selection for catch sampling intercept surveys. Fisheries Research 229:1-9.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax. 2007. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. Journal of Climate 20(22):5473–5496.

Ross, M.R. 1991. Recreational Fisheries of Coastal New England. University of Massachusetts Press, Amherst, MA.

Rothschild, B. J., Y. Jiao, and S. Y. Hyun. 2012. Simulation study of biological reference points for Summer Flounder. Transactions of the American Fisheries Society 141(2):426-436

Schoener, T. W. 1970. Nonsynchronous Spatial Overlap of Lizards in Patchy Habitats. Ecology 51(3):408–418.

Shepherd, G.R. and D.B. Packer. 2006. Essential Fish Habitat Source Document: Bluefish, *Pomatomus saltatrix*, Life History and Habitat Characteristics 2nd edition. NOAA Technical Memorandum, NMFS-NE-198:100.

Stock, B. C., and Miller, T. J. 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research, 240: 105967. Doi: 10.1016/j.fishres.2021.105967

Taylor, D.L., Rowe, P.M., and K.W. Able. 2006. Habitat use of the inner continental shelf off southern New Jersey by summer spawned bluefish (*Pomatomus saltatrix*). Fishery Bulletin 104:593-604.

Taylor, D.L., Nichols, R.S., and K.W. Able. 2007. Habitat selection and quality for multiple cohorts of young-of-the-year bluefish (*Pomatomus saltatrix*): Comparisons between estuarine and ocean beaches in southern New Jersey. Estuarine, Coastal and Shelf Science 73:667-679.

Then, A. Y., J. M. Hoenig, N. G. Hall, D. A. Hewitt, and Handling editor: E. Jardim. 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science 72 (1): 82–92.

Thorson, J. T. 2019. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research 210:143–161.

Thorson, J. T., and L. A. K. Barnett. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science 74(5):1311–1321.

Thorson, J. T., J. M. Cope, T. A. Branch, and O. P. Jensen. 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Canadian Journal of Fisheries and Aquatic Sciences 69(9):1556-1568.

Thorson, J. T., and K. Kristensen. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fisheries Research 175:66–74.

Trijoulet, V., C. M. Albertsen, K. Kristensen, C. M. Legault, T. J. Miller, and A. Nielsen. 2023. Model validation for compositional data in stock assessment models: Calculating residuals with correct properties. Fisheries Research 257, p.106487.

Wigley S. E., Rago P. J., Sosebee K. A., Palka D. L. 2007. The analytic component to the Standardized Bycatch Reporting Methodology Omnibus Amendment: sampling design end estimation of precision and accuracy (2nd edition). U.S. Dep. Commer., Northeast Fish. Sci. Cent. Ref. Doc . 07-09; 156 p. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026.

Wuenschel, M.J., Able, K.W., Buckel, J.A., Morley, J.W., Lankford, T., Branson, A.C., Conover, D.O., Drisco, D., Jordaan, A., Dunton, K., and D.H. Secor. 2012. Recruitment patterns and habitat use of young-of-the-year bluefish along the United States east coast: insights from coordinated coastwide sampling. Reviews in Fisheries Science 20:80-102.

10 TABLES

Table 1. Maturity-at-age through age-6 as calculated using various approaches. "Benchmark 2015" refers to the ogive used in the previous assessment, "NMFS 2022" refers to analyses performed during the 2022 research track assessment using data through 2021 but from federal sources only, and "Midyear model" refers to the GLM that was fit using federal and state data together.

Age	Benchmark 2015	NMFS 2022	Midyear model
0	0.00	0.000	0.000
1	0.40	0.417	0.456
2	0.97	0.965	0.926
3	1.00	0.999	0.995
4	1.00	1.000	1.000
5	1.00	1.000	1.000
6+	1.00	1.000	1.000

Table 2. Total removals of bluefish in metric tons by sector, 1985-2021.

Year	Commercial Landings	Commercial Discards	Recreational Landings	Recreational Dead Releases	Total Catch
1985	6,124		47,754	1,045	54,923
1986	6,657		75,470	1,611	83,738
1987	6,579		64,160	2,012	72,750
1988	7,162		36,475	905	44,542
1989	4,740	29	36,464	1,279	42,511
1990	6,250	32	31,553	1,976	39,811
1991	6,138	116	26,766	2,486	35,506
1992	5,208	38	22,533	1,769	29,548
1993	4,819	32	16,396	2,369	23,617
1994	4,306	162	14,176	3,140	21,783
1995	3,629	81	13,381	2,516	19,607
1996	4,213	166	10,760	2,756	17,895
1997	4,113	53	12,638	3,640	20,444
1998	3,741	74	15,414	2,995	22,224
1999	3,335	79	10,695	6,863	20,972
2000	3,660	83	11,141	6,289	21,174
2001	3,956	23	15,121	7,271	26,370
2002	3,116	37	13,904	4,581	21,638
2003	3,361	22	15,053	2,120	20,556
2004	3,673	62	17,570	4,744	26,050
2005	3,213	26	17,945	4,055	25,239
2006	3,354	34	16,912	5,708	26,009
2007	3,390	27	18,382	5,815	27,614
2008	2,731	22	17,410	5,428	25,591
2009	3,119	33	18,339	4,767	26,258
2010	3,304	87	21,269	6,384	31,044
2011	2,454	95	15,706	3,815	22,070
2012	2,212	14	15,291	2,833	20,350
2013	1,977	12	15,732	2,472	20,194
2014	2,251	18	12,324	2,880	17,473
2015	1,917	14	13,725	3,689	19,345
2016	1,946	14	10,634	1,837	14,431
2017	1,876	7	15,620	1,793	19,297
2018	1,105	8	5,857	1,579	8,548
2019	1,359	10	6,800	1,702	9,871
2020	1,112	9	5,923	1,253	8,296
2021	1,090	12	5,471	1,391	7,963

Table 3. Fishery-independent indices accepted by the Bluefish Working Group. "In Conn" indicates the index is part of the composite YOY index.

State	Index	Used in 2015?	Use in 2022?
NH	NH Seine Survey	Yes (in Conn)	Yes (in Conn)
RI	Beach seine (Narragansett Bay)	Yes (in Conn)	Yes (in Conn)
NY	WLIS Seine Survey	Yes (in Conn)	Yes (in Conn)
NJ	DE R. Seine Survey	Yes (in Conn)	Yes (in Conn)
MD	Striped Bass Seine Survey	Yes (in Conn)	Yes (in Conn)
VA	NEAMAP	Yes	Yes
VA	ChesMMAP	No	Yes
VA	Juv. Striped Bass Seine	Yes (in Conn)	Yes (in Conn)
NC	PSIGNS	Yes	Yes
SC	SEAMAP	Yes	Yes

Table 4. Fishery-independent surveys analyzed and excluded by the Bluefish Working Group.

State	Index	Used in 2015?	Use in 2022?	TC Comments	WG Comments
MA	MA Inshore Trawl Survey	No	No	Explore additional standardization; consider as a YOY index if trawl Conn dataset is expanded	Trawl Conn not used
RI	Trawl – Seasonal	No	No	Explore additional standardization; consider as a YOY index if Conn dataset is expanded	Trawl Conn not used
CT	Long Island Sound Trawl Survey	Yes	No	Use again in 2022	Remove; the index is dominated by age-0 fish, covers a limited spatial area, and was poorly fit by the model
NY	Peconic Bay Trawl	No	No	Explore additional standardization; consider as a YOY index if Conn dataset is expanded	Trawl Conn not used
NJ	NJ Ocean Trawl	Yes	No	Revise strata choice, standardization; consider as YOY index	Trawl Conn not used
DE	30’ Trawl	No	No	Explore additional standardization	Limited spatial coverage for recruitment index; trawl Conn not used
MD	Coastal Bays Juvenile Trawl Survey	No	No	Explore as part of trawl composite YOY survey	Trawl Conn not used
NC	IGNS	No	No	Consider River Regions data to expand spatial extent of PSIGNS	Trends in other regions the same as PSIGNS; not worth shortening the time series
NC	P195	No	No	Explore additional standardization; consider as a YOY index if Conn dataset is expanded	Trawl Conn not used

Table 5. Model outputs and diagnostics from ASAP runs using various temporal and spatial levels of ALK and data resolution. ρ = Mohn's rho

ALK	2021 SSB (mt)	Recruitment (millions of fish)	F	SSB ρ	R ρ	F ρ	# at initial objective function (out of 200)	# unique objective function solutions	# Not converged (out of 200)
Traditional-Seasonal	59,540	28.7	0.19	0.341	0.080	-0.23	190	2	4
Multinomial-Seasonal	51,562	27.4	0.19	0.215	0.024	-0.18	193	2	3
Multinomial-Season-Region	43,916	27.1	0.21	0.222	0.033	-0.19	192	2	4

Table 6. Model table showing linear steps in the ASAP and WHAM model bridge building process. R is recruitment (in millions of age-0 fish). "W" in the model names indicates WHAM model runs. ~ indicates jitter analysis was not performed for that run. P is Mohn's rho measure of retrospective patterning.

Model	Description	2021 SSB (MT)	2021 R (mil)	2021 F	SSB ρ	R ρ	F ρ	# at OG OBFunc	Jitter Sol	Not conv
BF00	BLF 2021 MT model	95,742	27.9	0.172	0.226	0.192	-0.221	~	~	~
BF01	BLF RT Continuity Run	91,745	39.4	0.160	0.294	0.170	-0.277	132	51	18
BF03	Update all new data	85,975	39.2	0.172	0.364	0.174	-0.323	142	38	21
BF04	New LW parameters	86,581	39.1	0.172	0.359	0.174	-0.320	~	~	~
BF05	New Rec discard mortality	82,103	35.9	0.159	0.380	0.186	-0.334	~	~	~
BF07	Add commercial discards	82,018	36.2	0.160	0.378	0.185	-0.332	140	28	31
BF08	New Indices: MRIP Continuity	88,424	35.3	0.158	0.319	0.123	-0.313	129	17	52
BF09	New Indices: MRIP Continuity, multinomial ALKs	70,336	26.3	0.158	0.352	0.042	-0.321	193	2	6
BF10	New Indices: MRIP Continuity, multinomial ALKs, Rec discard length by season/region	67,029	26.7	0.138	0.405	0.051	-0.361	197	2	2
BF11	New Indices: MRIP Guild, multinomial ALKs, Rec discard length by season/region	47,734	25.8	0.172	0.253	0.033	-0.214	191	2	8
BF12	New M: Lorenzen based on empirical WAA	65,946	97.3	0.110	0.346	0.113	-0.266	188	3	9
BF18	5 Sel blocks	79,849	97.9	0.116	0.293	0.035	-0.220	183	2	9
BF19	Fix bounded selectivities F2to3	82,858	98.6	0.113	0.288	0.014	-0.221	194	2	5
BF20	MRIP PSE for fleet 2	91,149	101.7	0.107	0.257	0.023	-0.205	194	4	2
BF21	MRIP index input CV (from 0.3)	84,212	85.3	0.116	0.193	0.000	-0.223	191	3	6
BF22	No CT survey	88,051	93.1	0.111	0.187	-0.001	-0.229	193	2	6
BF23	Adjust MRIP CV to reduce RMSE (x1.6)	94,886	102.0	0.102	0.225	-0.014	-0.209	199	1	1
BF24	Adjust fixed selectivity at age 2 for some blocks	63,320	94.6	0.159	0.326	0.017	-0.277	192	4	4
BF26W	Reduce fleet 1 CV	63,606	95.7	0.160	0.270	-0.066	-0.215	~	~	~
BF27W	Reduce Fleet 2 CV	68,546	96.4	0.152	0.249	-0.062	-0.198	~	~	~
BF28W	Reduce both fleets CV	68,631	96.4	0.152	0.248	-0.063	-0.197	~	~	~

Table 7. Results and diagnostics for different state-space model variations of the WHAM model BF28W examining different options for treating the yearly transitions (survival) in recruitment and number-at-age. R is recruitment (in millions of age-0 fish). Ρ is Mohn's rho measure of retrospective patterning.

Model	Description	dAIC	AIC	2021 SSB (MT)	2021 R (mil)	2021 F	R ρ	SSB ρ	F ρ	Con-verged?	Positive definite Hessian?
BF28W	Base model: traditional statistical catch-at-age	~	~	68,631	96.4	0.152	-0.063	0.248	-0.197	TRUE	TRUE
m7	All NAA transitions are random effects correlated by year and age	0	3229	55,344	86.5	0.166	0.010	0.130	-0.096	TRUE	TRUE
m5	All NAA transitions are random effects correlated by year	3	3232	55,070	82.3	0.167	0.019	0.126	-0.097	TRUE	TRUE
m4	All NAA transitions are random effects independent, identically distributed	46.2	3275	58,114	98.6	0.160	-0.008	0.172	-0.144	TRUE	TRUE
m6	All NAA transitions are random effects correlated by age	46.9	3276	58,786	99.9	0.159	-0.004	0.177	-0.148	TRUE	TRUE
m2	Recruitment transitions are random effects independent, identically distributed	111	3340	73,843	104.1	0.144	-0.022	0.236	-0.195	TRUE	TRUE
m3	Recruitment transitions are random effects correlated by year	111	3340	72,329	101.3	0.146	-0.020	0.245	-0.198	TRUE	TRUE

Table 8. Results, retrospective, and convergence properties of the final model sensitivity runs for the WHAM final model (BF28W_m7, bolded row). R is recruitment (in millions of age-0 fish). ρ is Mohn's rho measure of retrospective patterning.

Model	2021 R (mil)	2021 SSB (MT)	2021 F	R ρ	SSB ρ	F ρ	Con-verged?	Positive Definite Hessian?
BF28W_m7	**86.5**	**55,344**	**0.166**	**0.010**	**0.130**	**-0.096**	**TRUE**	**TRUE**
rmALB	83.1	53,880	0.171	0.0075	0.127	-0.0924	TRUE	TRUE
rmBIG	86.6	56,327	0.163	0.0281	0.1222	-0.0885	TRUE	TRUE
rmMRIP	101.4	64,964	0.142	-0.0233	0.176	-0.0993	TRUE	TRUE
rmNEA	81.3	57,488	0.162	0.0071	0.1326	-0.0995	TRUE	TRUE
rmSEA0	90.3	55,826	0.165	0.0045	0.1266	-0.0932	TRUE	TRUE
rmPSIGN	75.1	38,725	0.236	0.0635	0.25	-0.1689	TRUE	TRUE
rmYOY	77.4	53,209	0.175	0.0278	0.1473	-0.1118	TRUE	TRUE
rmCHES	86.0	54,749	0.168	0.0048	0.1256	-0.0908	TRUE	TRUE
rmSEA1	86.9	55,633	0.165	0.0116	0.1316	-0.0983	TRUE	TRUE
NEFSC offshore	97.6	59,020	0.169	0.046	0.128	-0.094	TRUE	TRUE
GLM indices	90.5	57,758	0.158	0.0535	0.1513	-0.1155	TRUE	TRUE
MRIP direct	101.4	71,334	0.131	0.004	0.130	-0.096	TRUE	TRUE
MRIP SAA	86.8	60,378	0.165	0.007	0.121	-0.093	TRUE	**FALSE**
Borrow Region	83.4	95,775	0.130	0.090	0.204	-0.132	TRUE	**FALSE**
Borrow AB1	83.6	55,473	0.172	0.0066	0.141	-0.1023	TRUE	TRUE
Use AB1 for B2	80.9	53,674	0.186	0.0071	0.1326	-0.0921	TRUE	TRUE
B2 15% DM	93.9	58,842	0.172	0.0103	0.1244	-0.0935	TRUE	TRUE
M Lorenzen Low	31.4	42,296	0.226	0.0007	0.1316	-0.1003	TRUE	**FALSE**
M Lorenzen High	480.3	205,189	0.045	0.2398	0.4017	-0.2348	TRUE	TRUE

Table 9. Biological reference points from the 2015 benchmark, 2021 operational assessment, the continuity run (BF01), and the final model (BF28W_m7).

Reference Point	SAW60	OA2019	BF01: Cont Run	BF28W_m7: Final Model
$F_{35\%}$	0.190	0.181	0.176	0.248
SSB_{TARGET}	101,343 MT	201,729 MT	190,771 MT	91,897 MT
$SSB_{THRESHOLD}$	50,672 MT	100,865 MT	93,386 MT	45,949 MT

Table 10. Short-term (2022-2025) projections of SSB and the probability of being above $B_{THRESHOLD}$ in 2025 for bluefish under 3 different F scenarios.

Projection scenario	2022	2023	2024	2025	P (2025) > $B_{threshold}$
F_{MSY} = 0.248	65,805 (39,305-110,170)	66,340 (37,604-117,034)	64,083 (35,017-117,275)	61,784 (32,086-118,971)	0.84
F_0 = 0	65,805 (39,305-110,170)	72,637 (41,394-127,462)	83,806 (46,270-151,792)	94,956 (49,788-181,098)	0.99
F_{status_quo} = 0.166	65,805 (39,305-110,170)	68,357 (38,820-120,367)	70,009 (38,411-127,601)	71,150 (37,110-136,412)	0.93
MAFMC risk policy (60% CV)	65,805 (39,305-110,170)	67,891 (37,217-123,847)	68,583 (33,654-139,765)	68,804 (29,551-160,198)	0.85
MAFMC risk policy (100% CV)	65,805 (39,305-110,170)	68,514 (37,767-124,295)	70,385 (35,116-141,078)	71,553 (31,586-162,089)	0.88

Table 11. Short term (2022-2025) projections of total catch for bluefish under 3 different F scenarios.

Projection scenario	2022	2023	2024	2025
F_{MSY} = 0.248	11,460	13,909 (8,098-23,889)	13,957 (7,784-25,022)	13,584 (7,157-25,784)
F_0 = 0	11,460	0	0	0
F_{status_quo} = 0.166	11,460	9,569 (5,564-16,458)	10,127 (5,628-18,223)	10,292 (5,399-19,623)
MAFMC risk policy (60% CV)	11,460	10,581 (P* = 0.311)	11,118 (P* = 0.314)	11,202 (P* = 0.316)
MAFMC risk policy (100% CV)	11,460	9,225 (P* = 0.311)	10,027 (P* = 0.321)	10,357 (P* = 0.327)

11 FIGURES

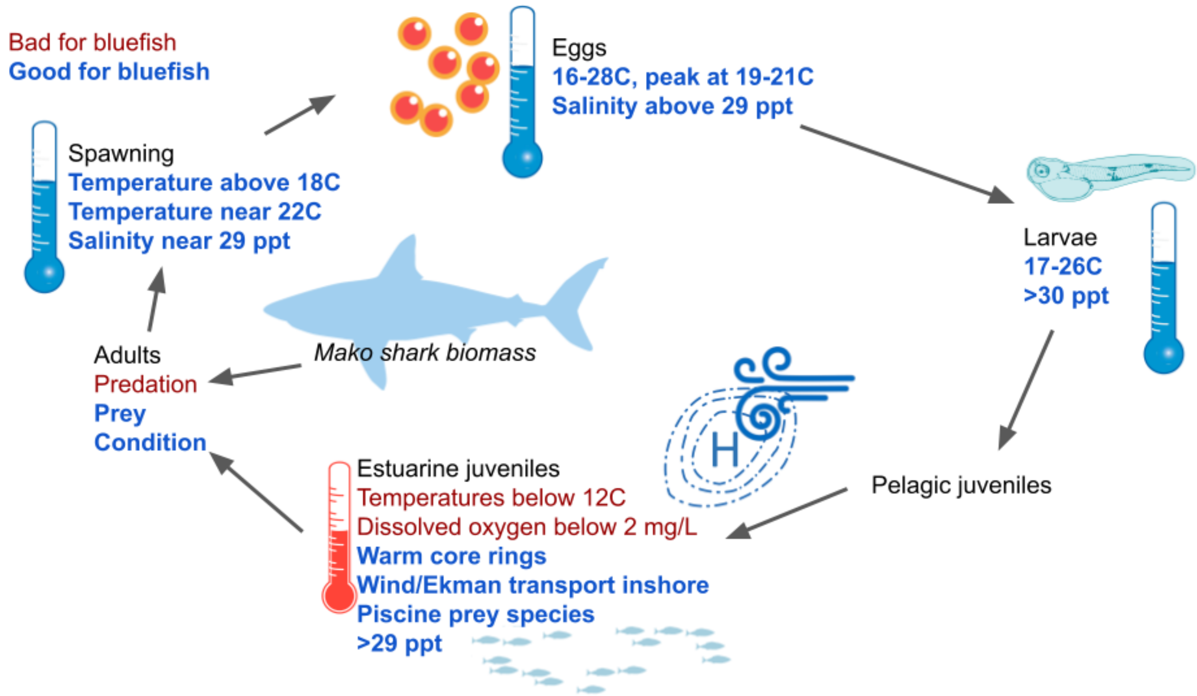

Figure 1. Life history conceptual model of bluefish identifying environmental factors with positive (blue text) or negative (red text) effects on different life stages of bluefish.

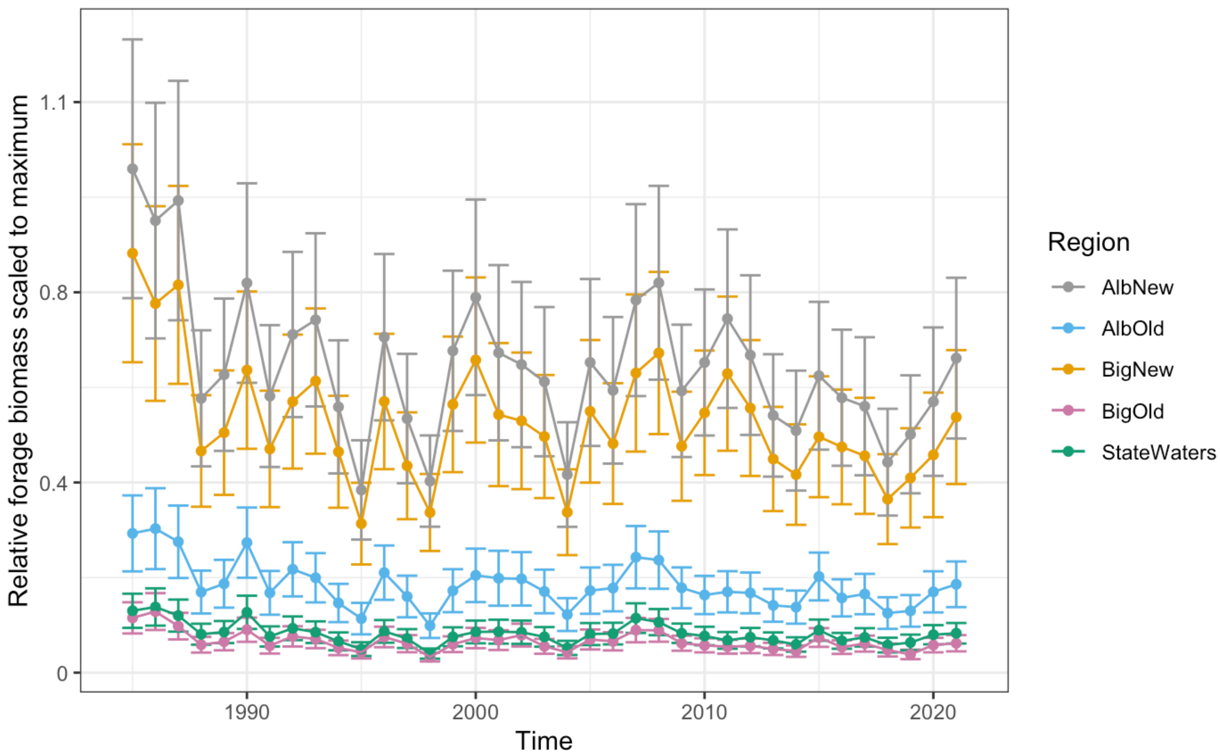

Figure 2. Forage fish indices for Fall 1985-2021 in Mid-Atlantic and Georges Bank nearshore and offshore areas. AlbNew= Albatross New, all inshore and new offshore survey strata (largest area); AlbOld= Albatross Old, includes all inshore survey strata; BigNew= Bigelow New, includes the subset of inshore survey strata that can be sampled by the R/V Henry Bigelow plus new offshore strata; BigOld = Bigelow Old, includes the subset of inshore survey strata that can be sampled by the R/V Henry Bigelow; StateWaters includes the coastline to 3 nautical miles offshore (smallest area)

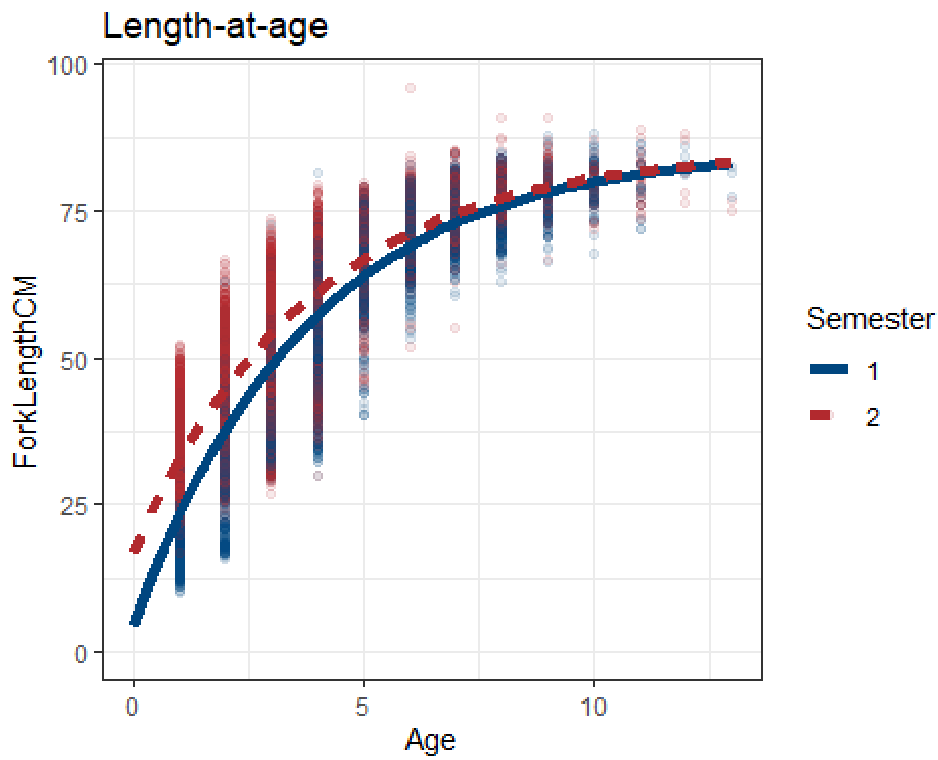

Figure 3. Fitted von Bertalanffy relationship by season using age-length data from state and federal fishery-dependent and fishery-independent sources. Jan-Jun=Semester 1, July-Dec=Semester 2.

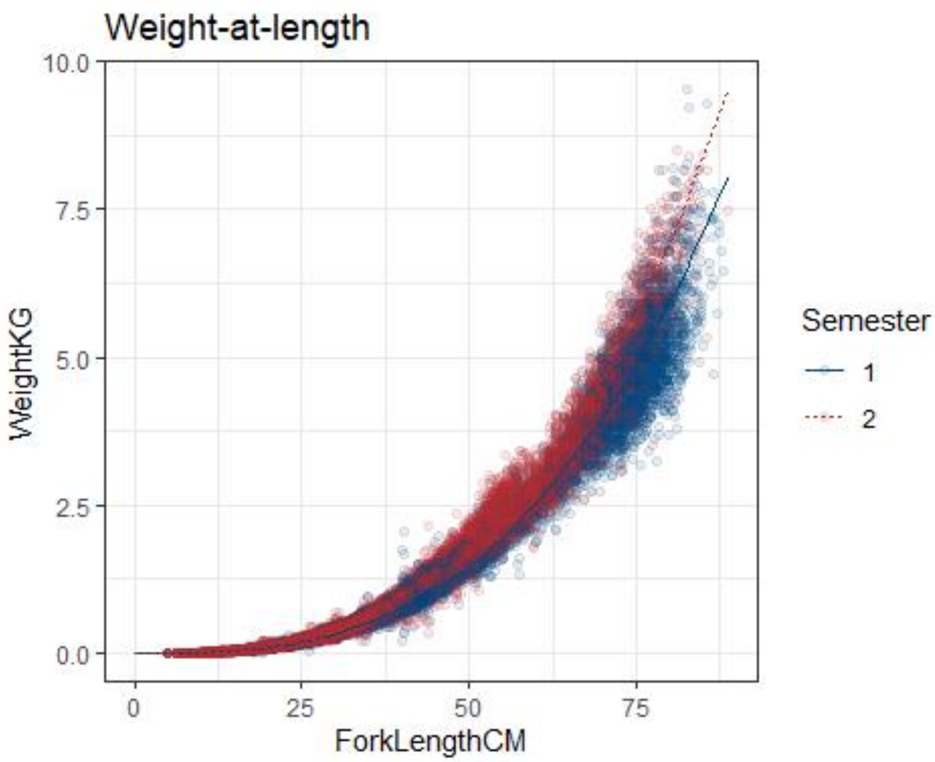

Figure 4. Fitted length-weight relationship by season using age-length data from state and federal fishery-dependent and fishery-independent sources. Jan-Jun=Semester 1, July-Dec=Semester 2.

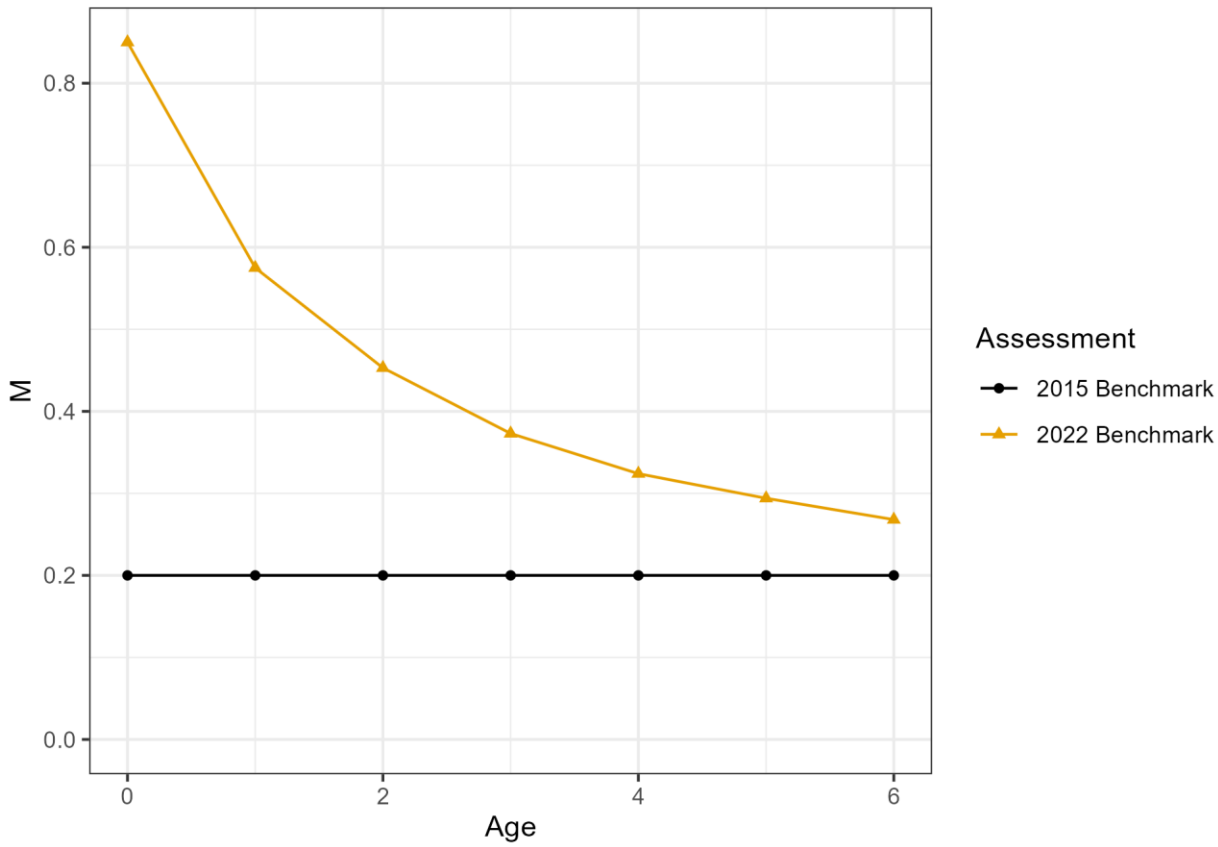

Figure 5. Comparison of the estimates of M-at-age used in the 2015 benchmark assessment and this assessment.

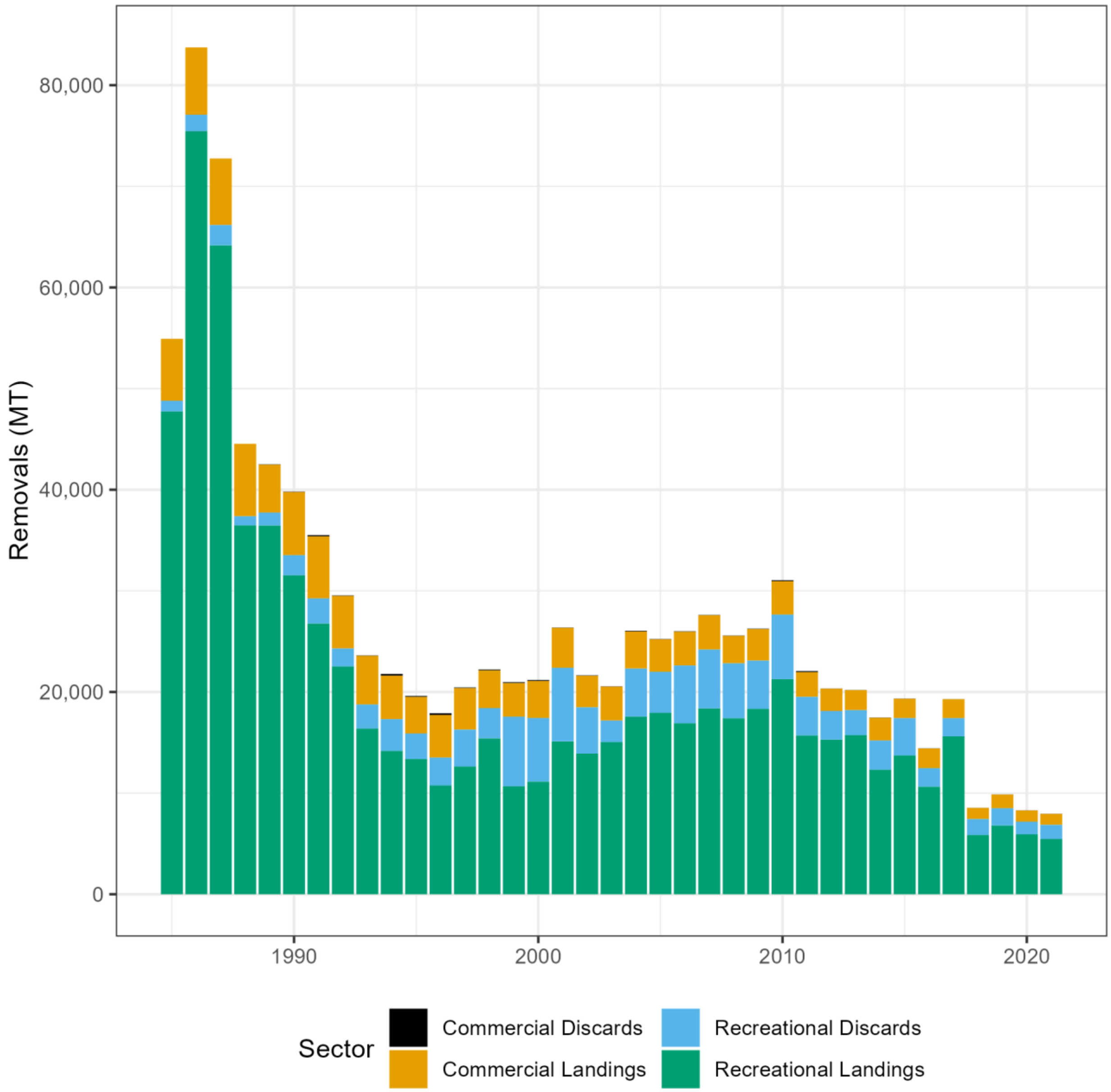

Figure 6. Total removals of bluefish on the Atlantic coast by sector, 1985-2021.

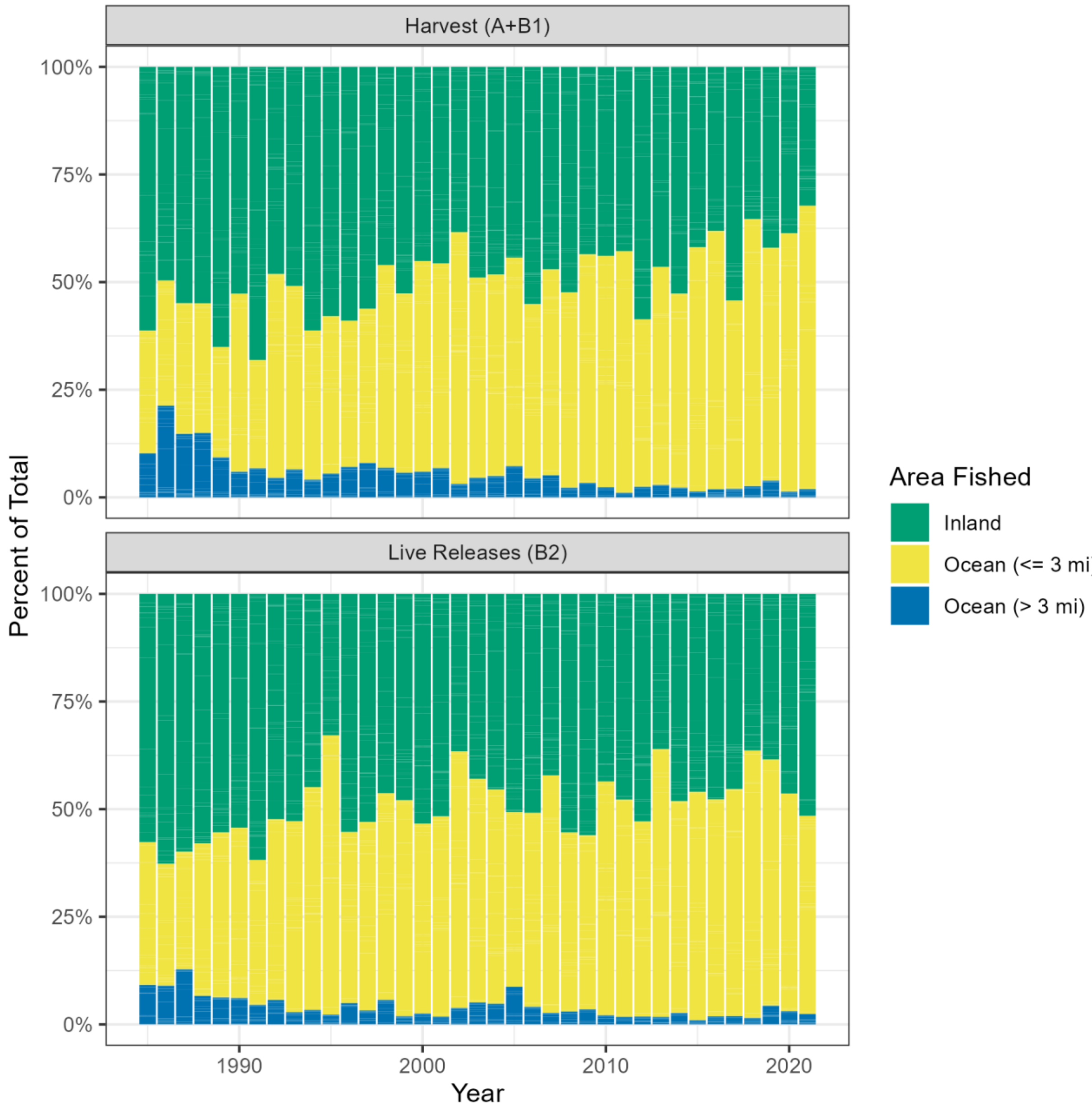

Figure 7. Percent of recreational harvest (top) and live releases (bottom) in numbers by area fished over time.

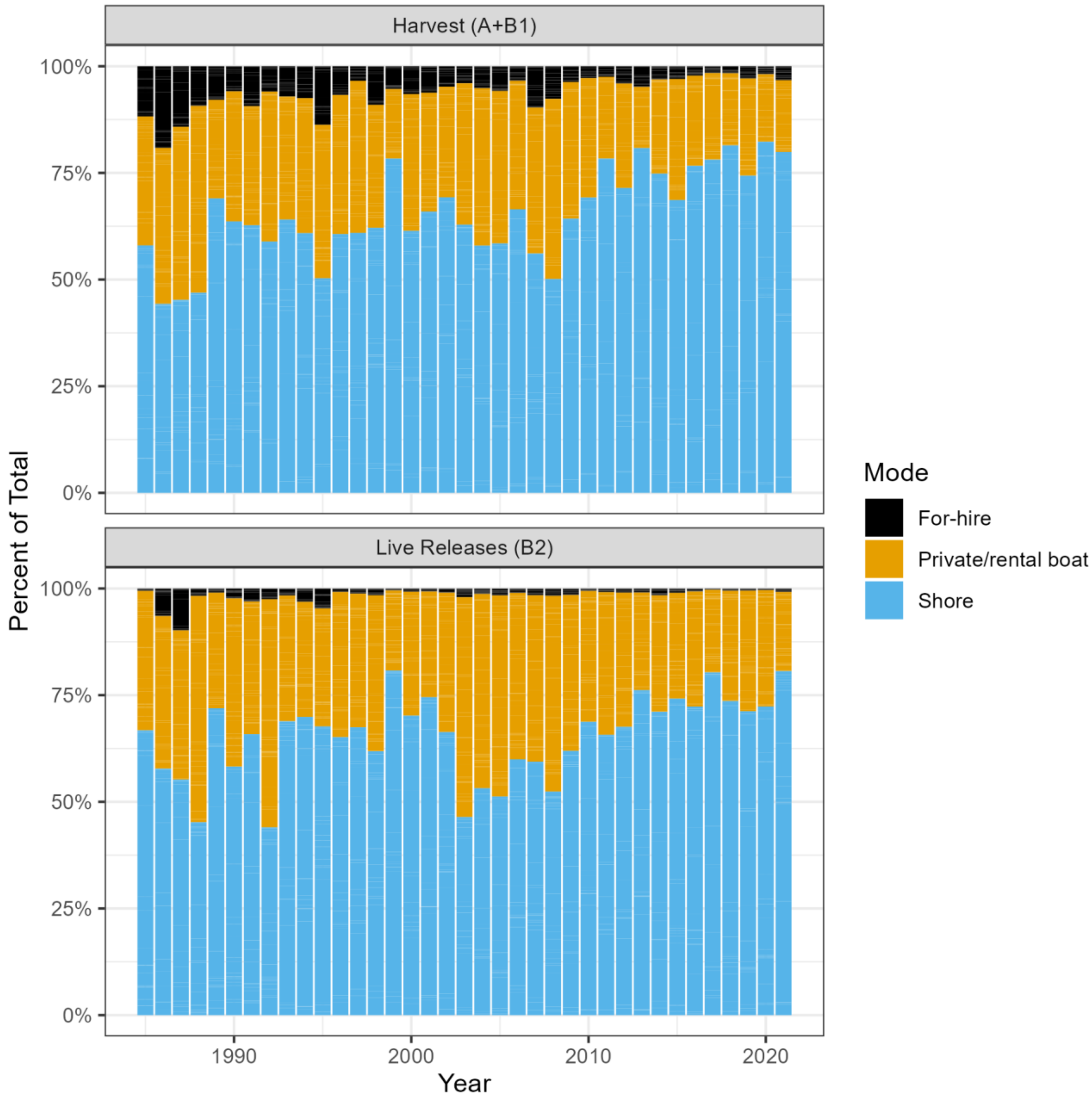

Figure 8. Percent of recreational harvest (top) and live releases (bottom) in numbers by mode of fishing over time.

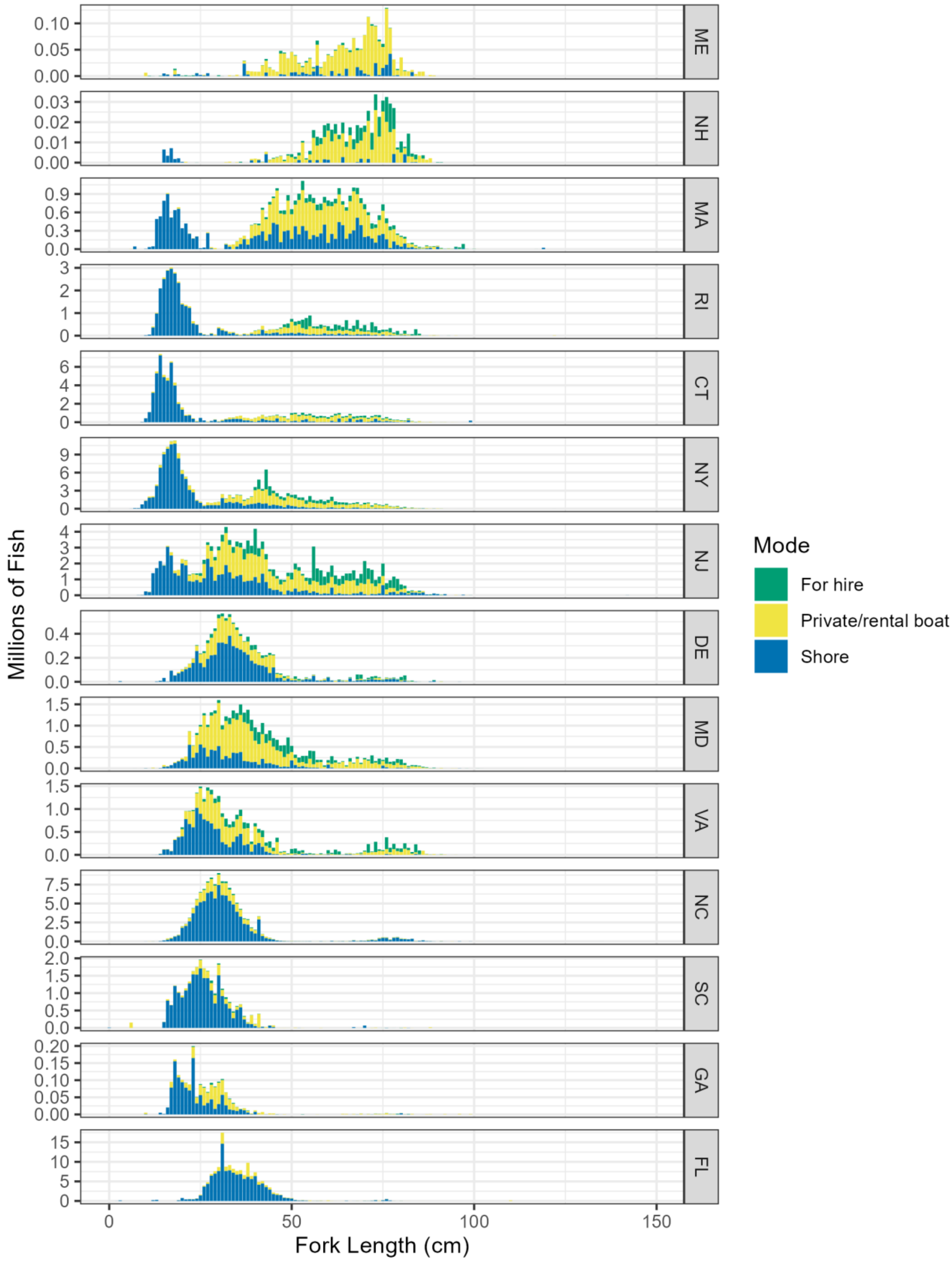

Figure 9. Length frequency of recreationally harvested bluefish by state and mode of fishing, 1982-2020 pooled.

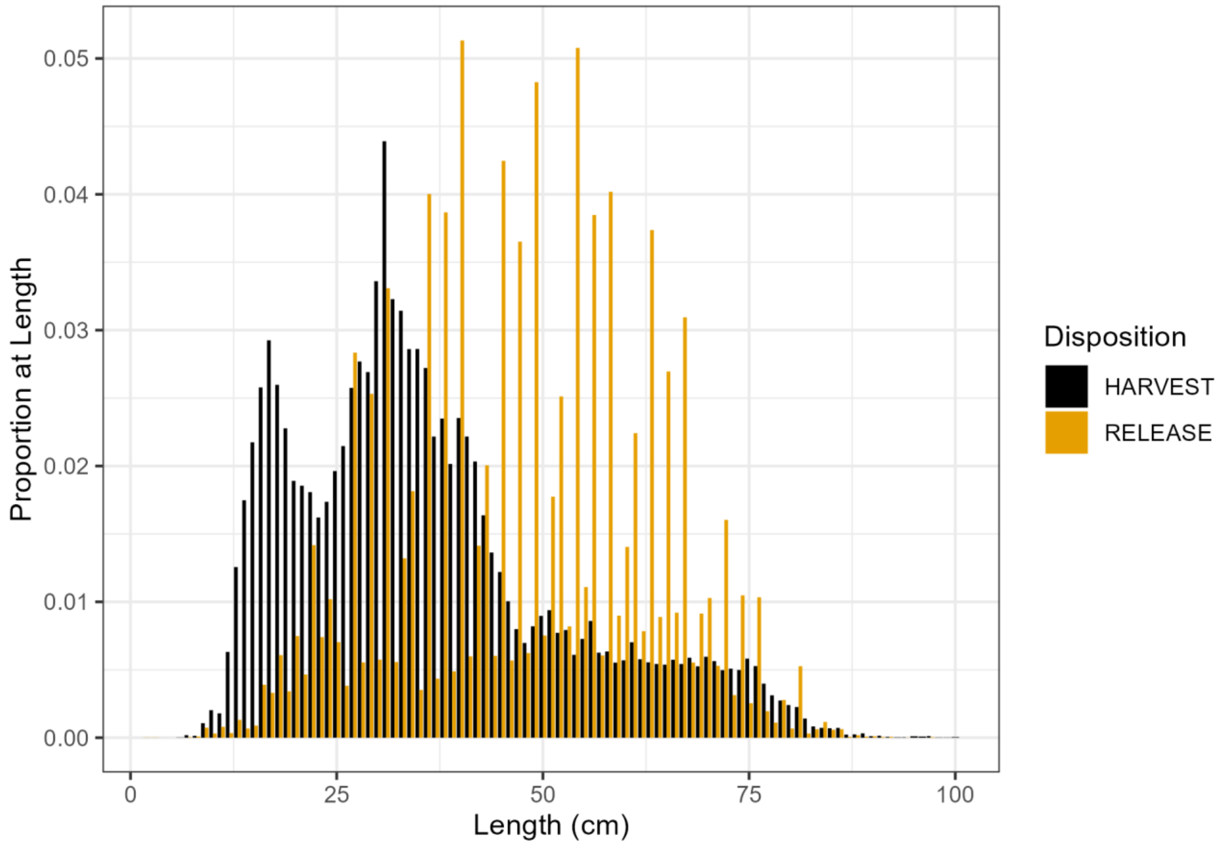

Figure 10. Proportion at length for harvested and released bluefish sampled, pooled over region, season, and years.

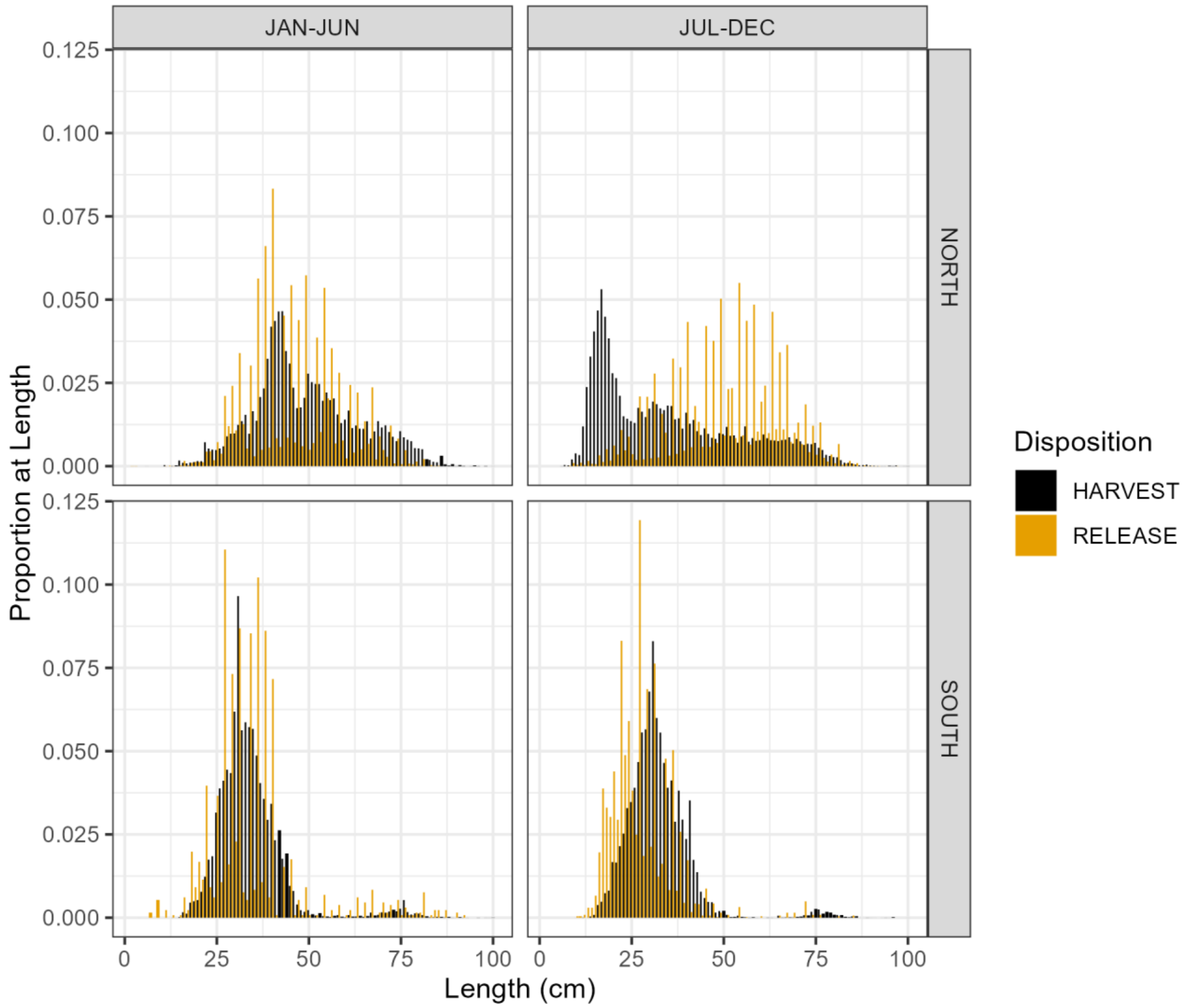

Figure 11. Proportions of length of harvested and released bluefish by region and season. Data pooled over 1985-2021.

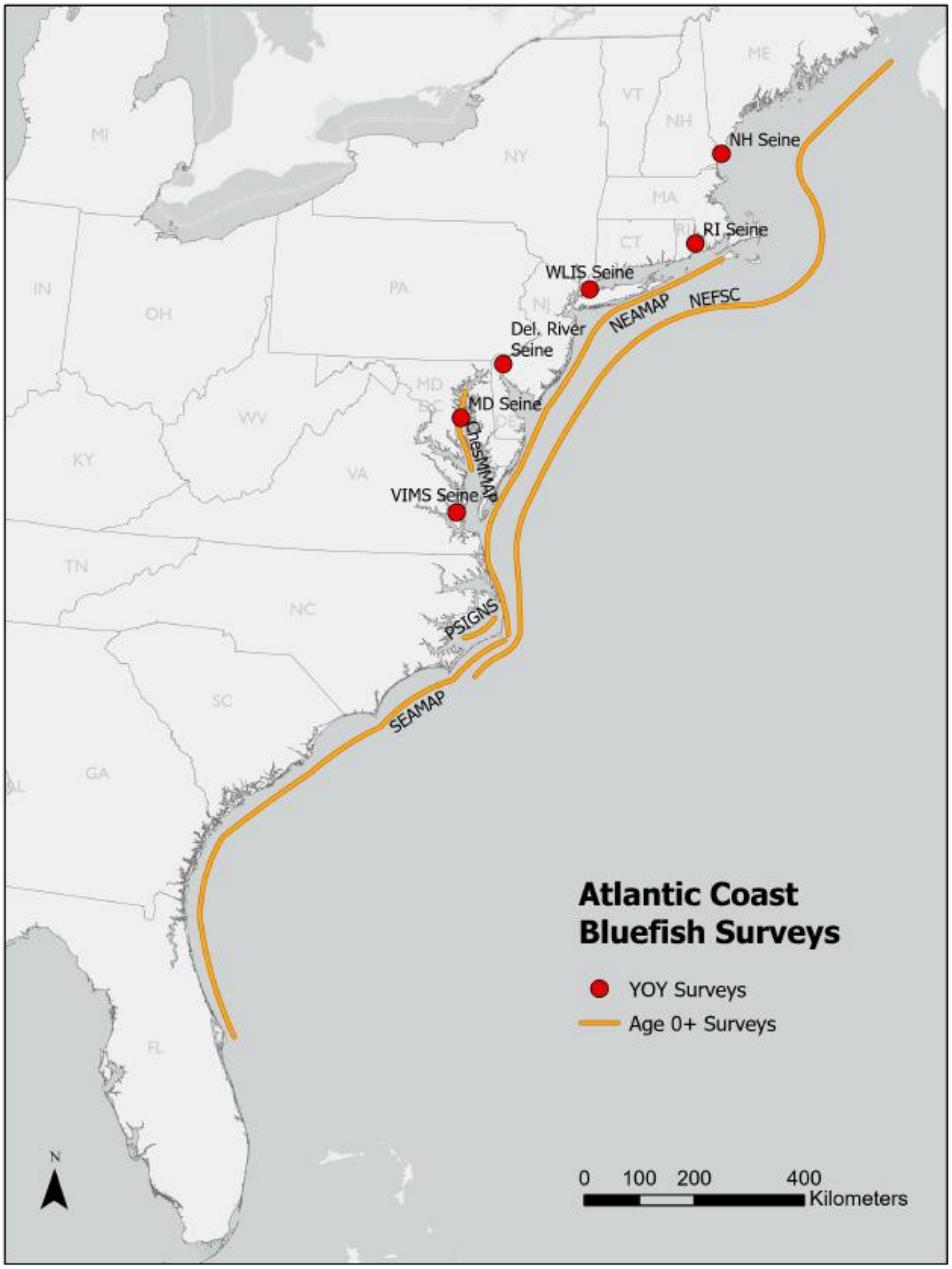

Figure 12. Map of the east coast of the United States showing the approximate locations of each of the surveys used in the final assessment model.

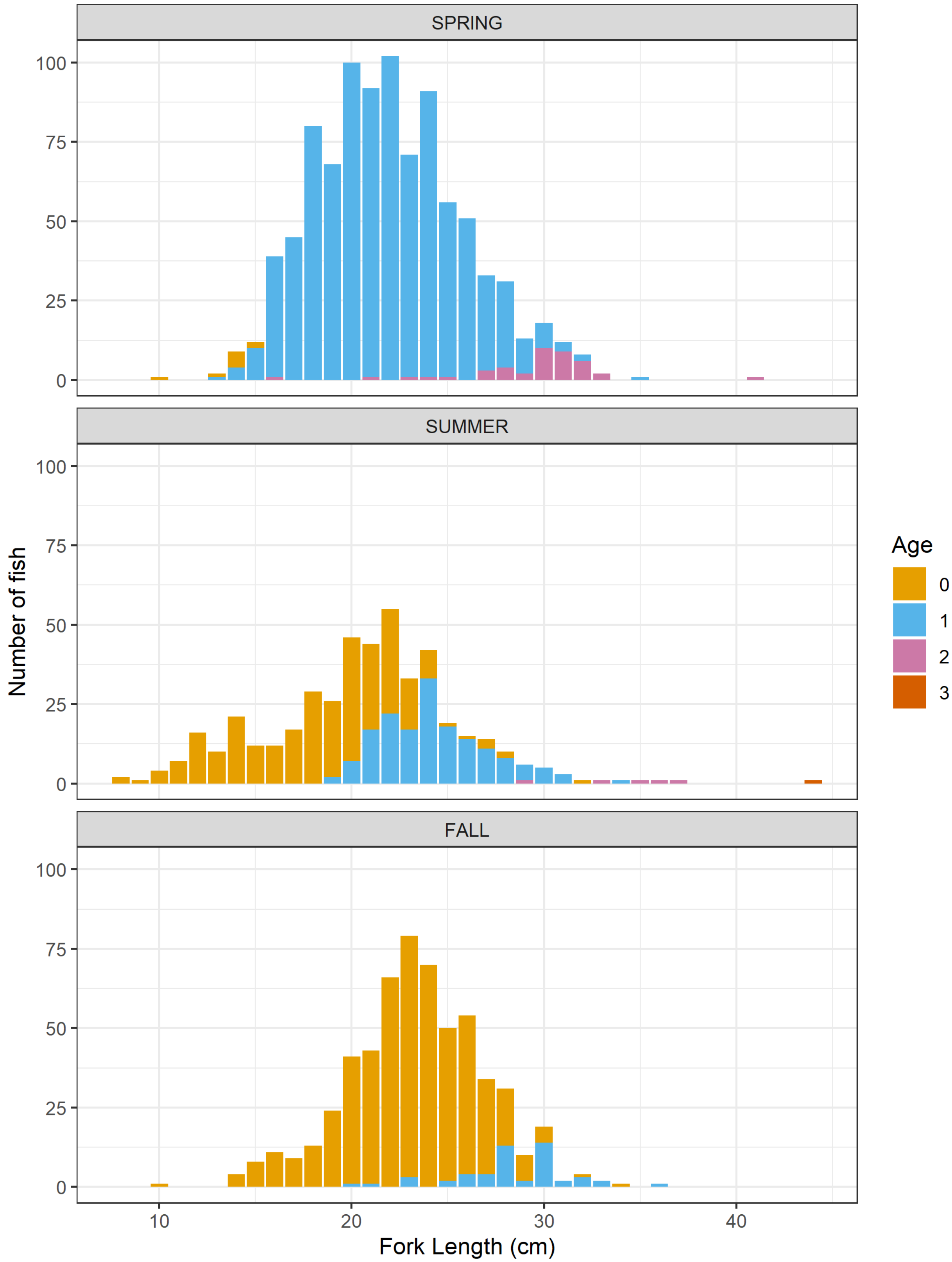

Figure 13. Length and age frequency of the SEAMAP survey by season, pooled over all years.

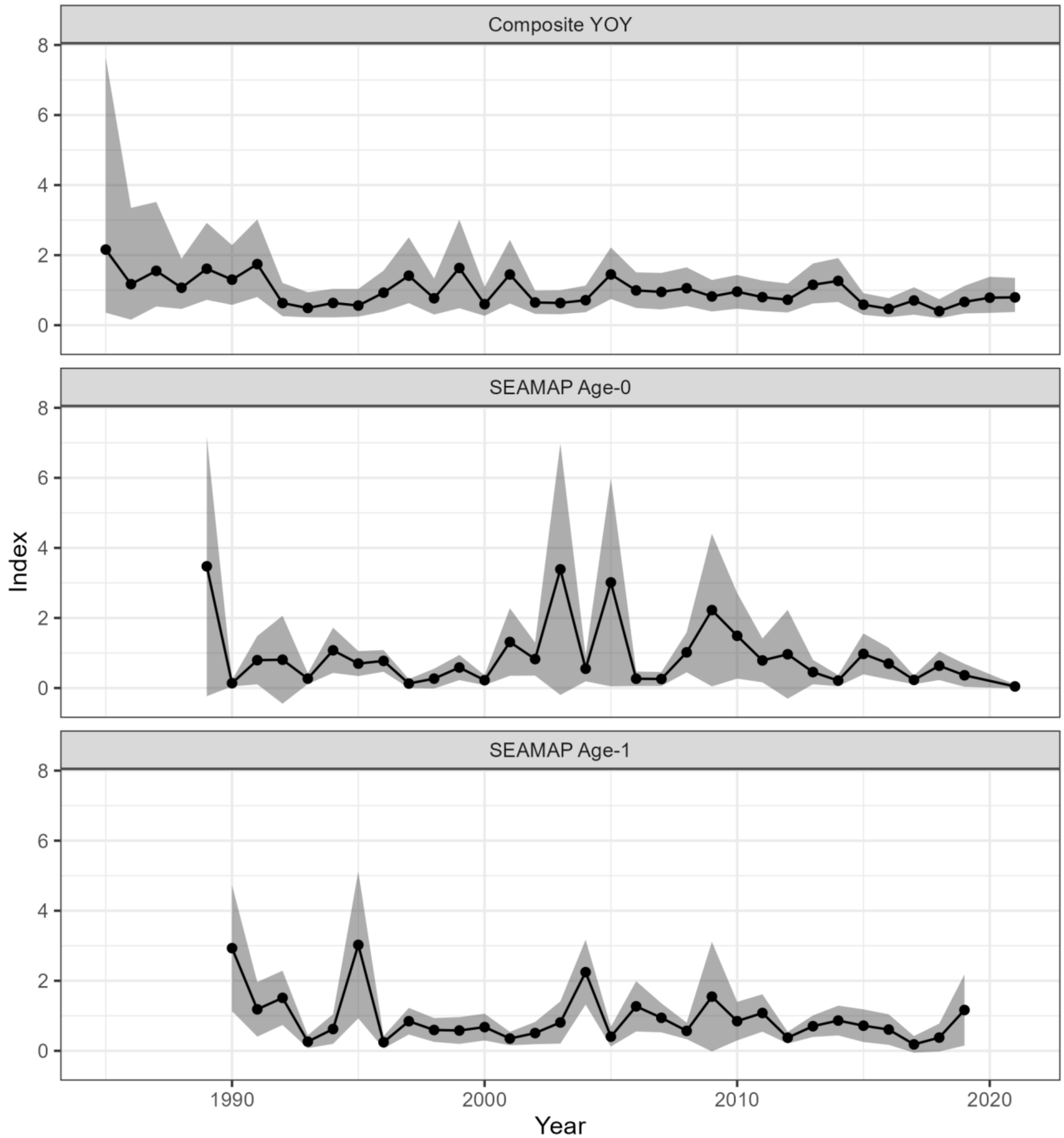

Figure 14. Indices of bluefish recruitment (i.e, age-0 and age-1 only) used in the ASAP and WHAM models.

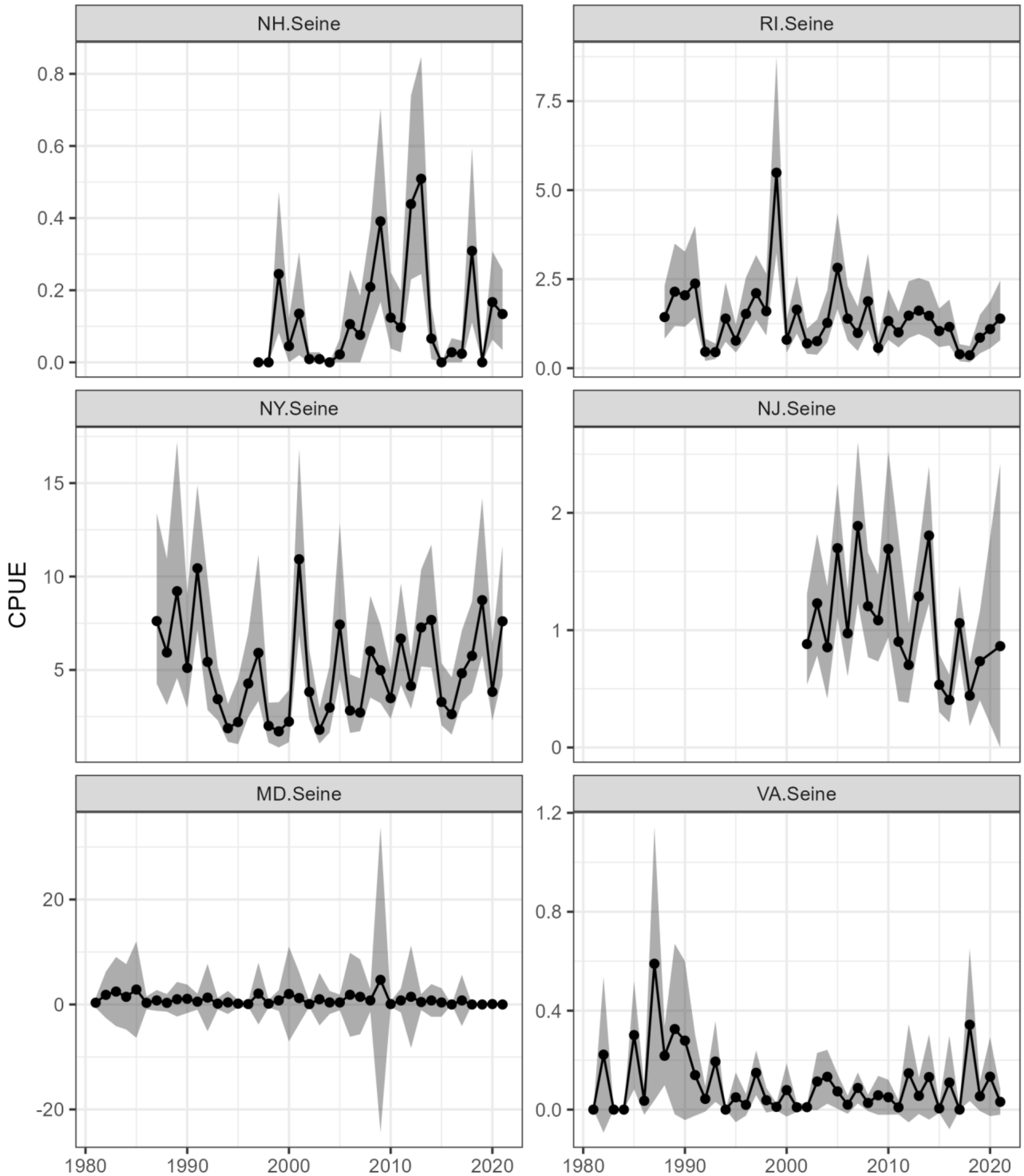

Figure 15. Indices of young-of-year abundance from state seine surveys used in the composite YOY index. Shaded area indicates 95% confidence interval.

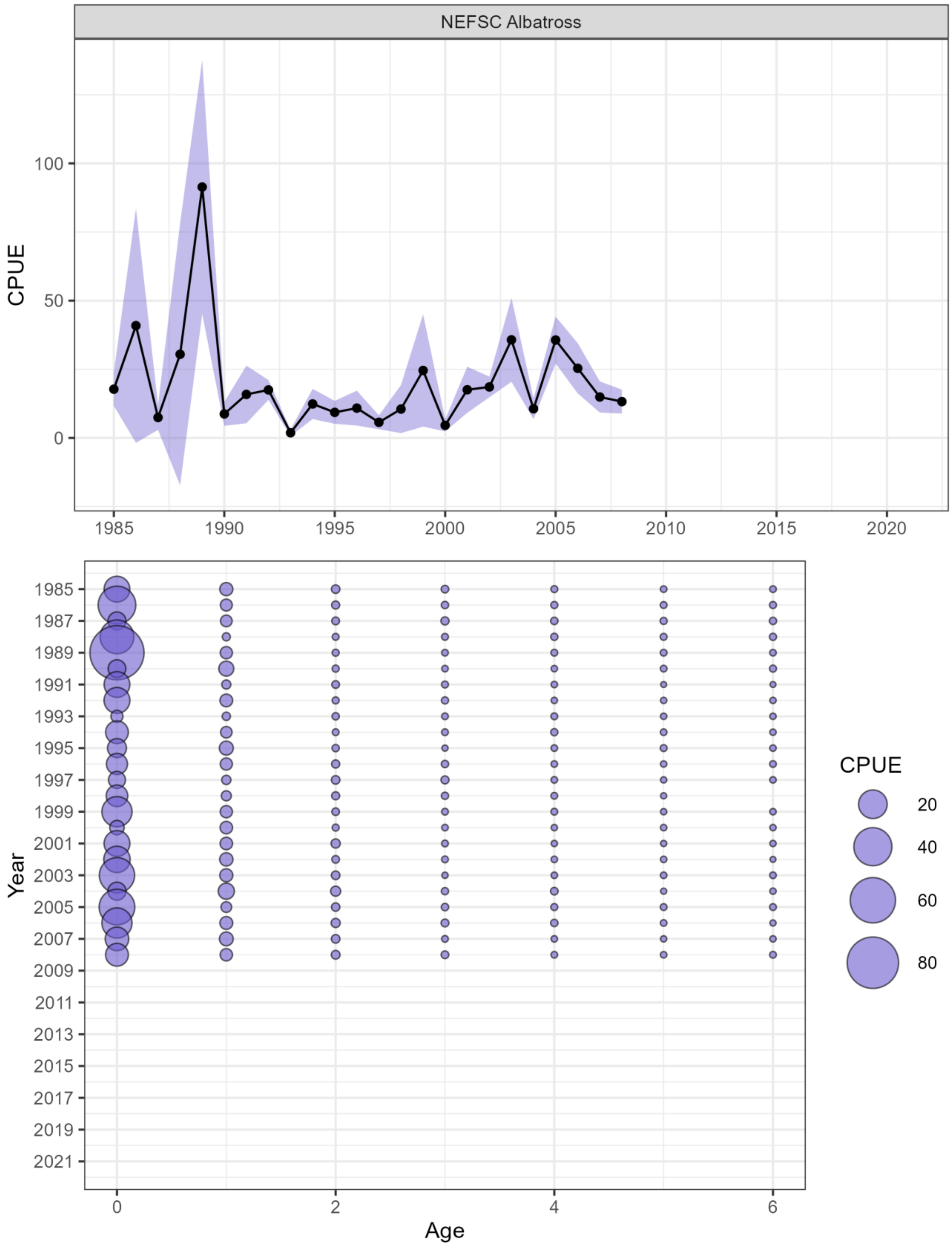

Figure 16. Index of abundance (top) and age composition (bottom) from the NEFSC fall trawl survey (Albatross years). Shaded area indicates 95% confidence interval.

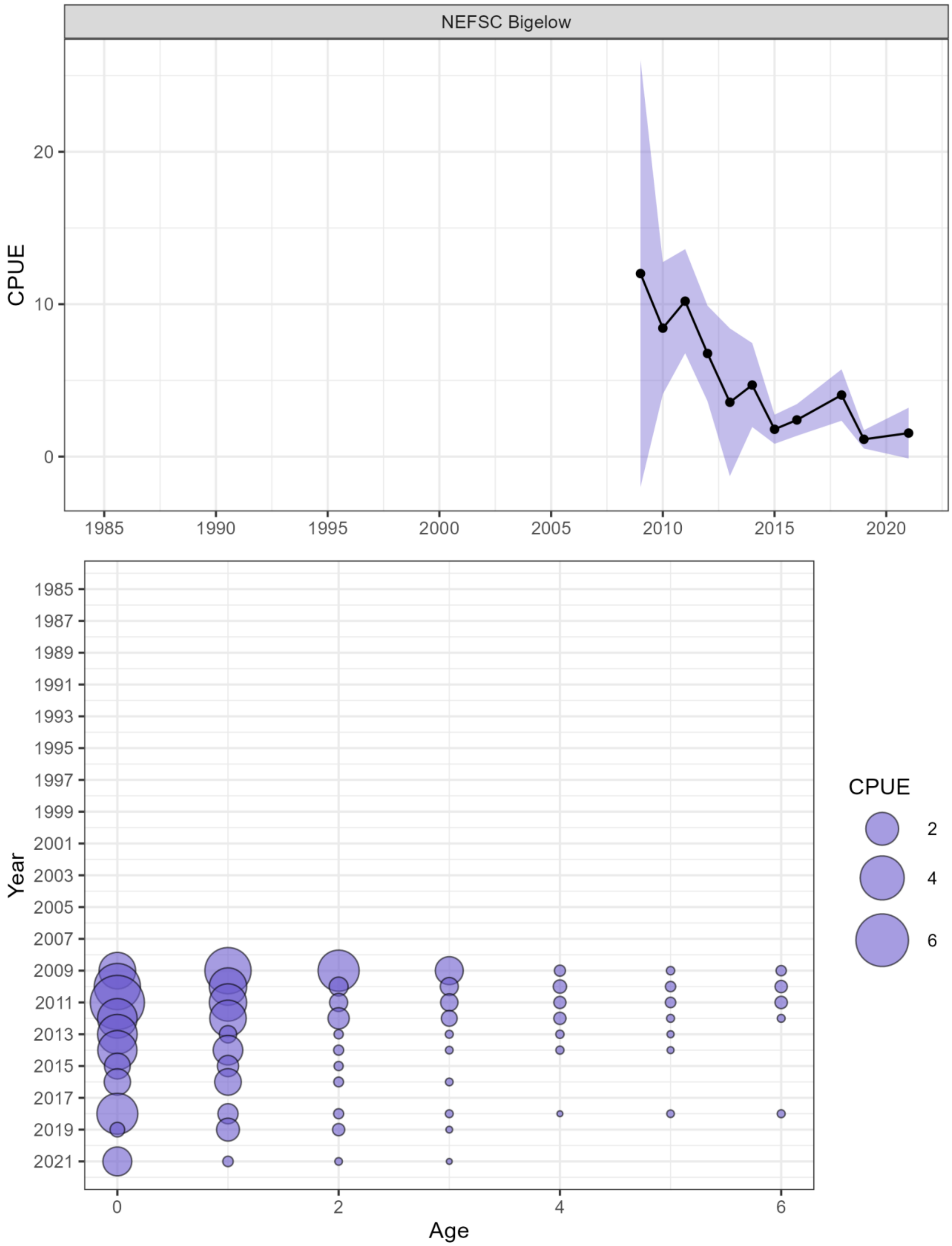

Figure 17. Index of abundance (top) and age composition (bottom) of the NEFSC fall trawl survey (Bigelow years). Shaded area indicates 95% confidence interval.

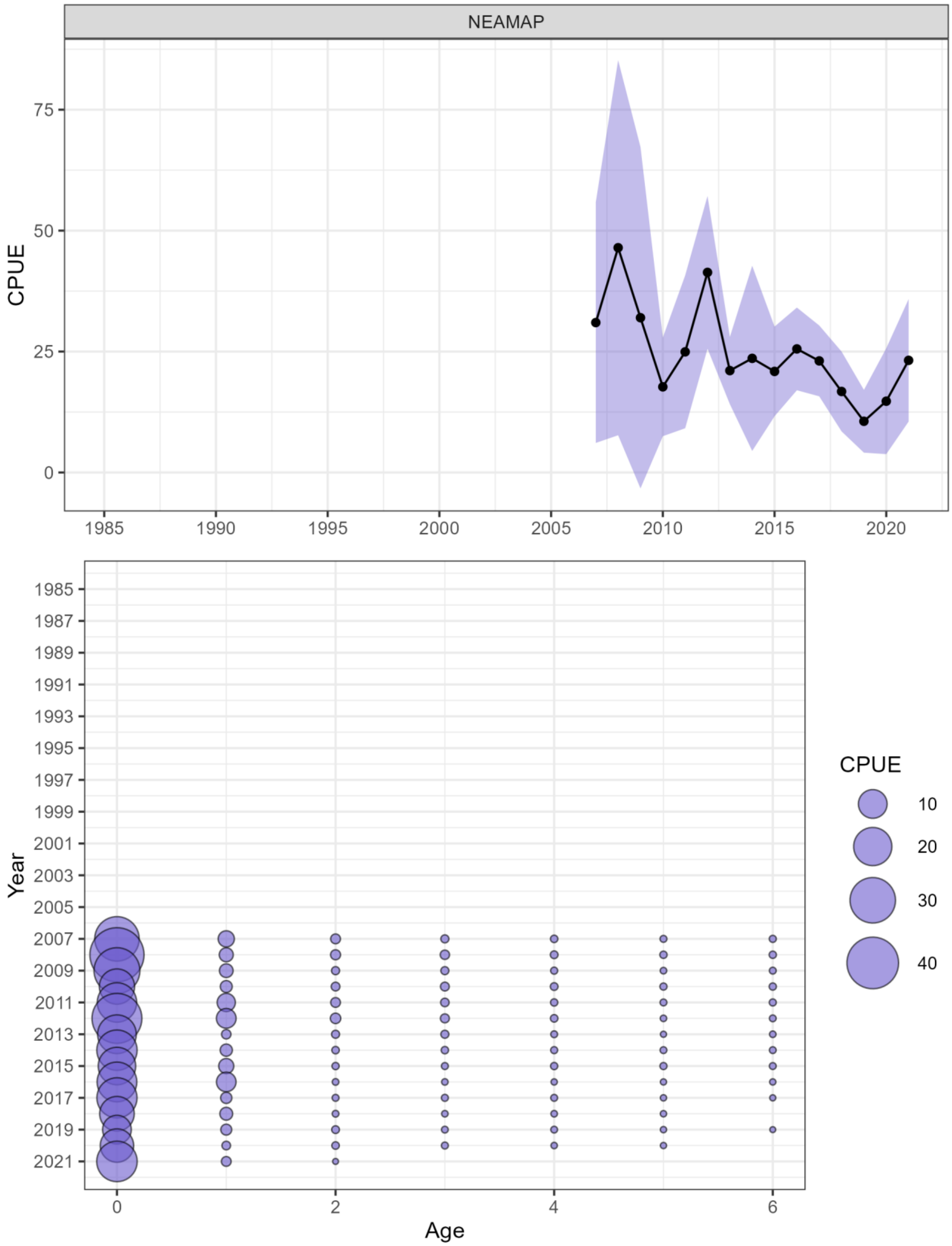

Figure 18. Index of abundance (top) and age composition (bottom) of the NEAMAP survey. Shaded area indicates 95% confidence interval.

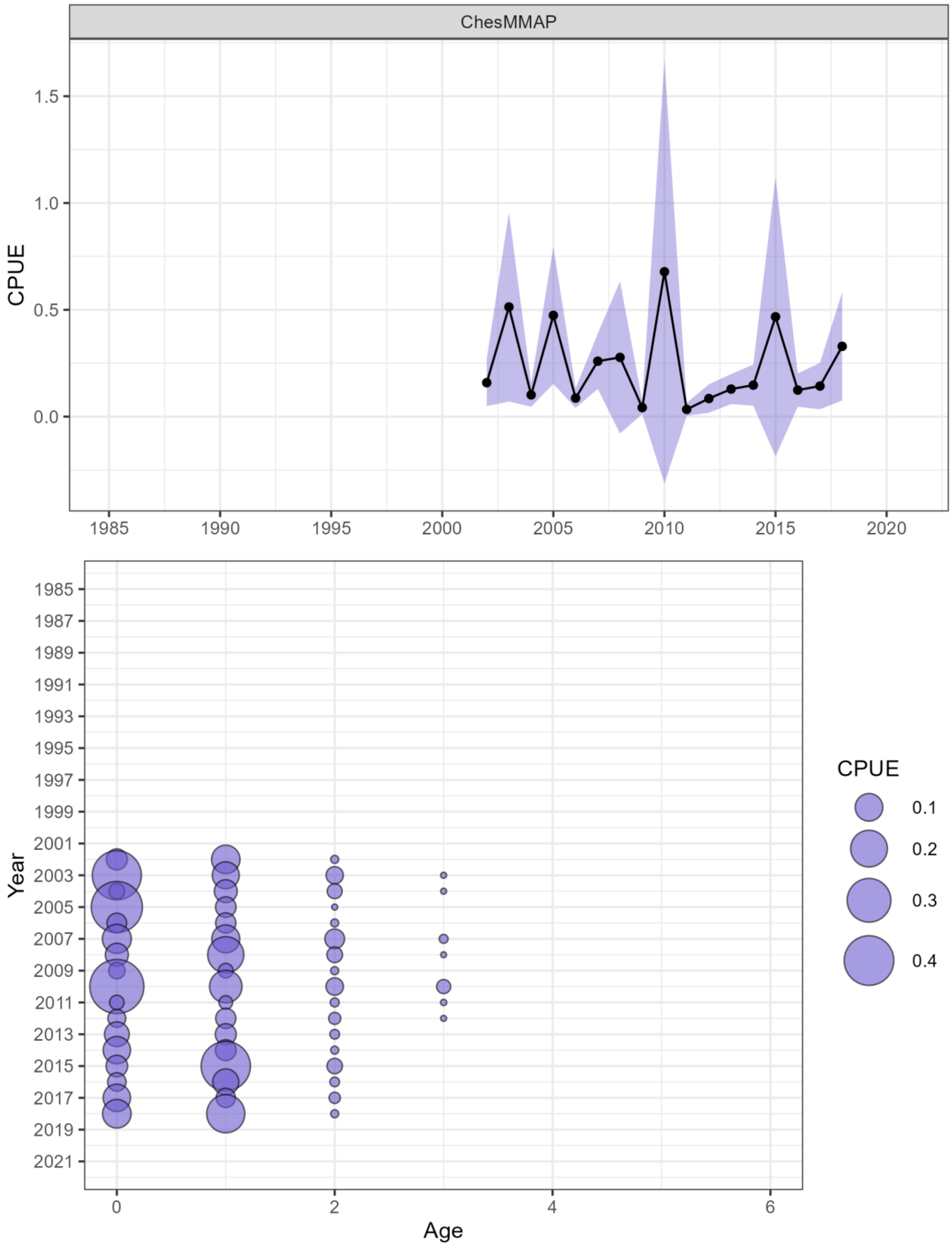

Figure 19. Index of abundance (top) and age composition (bottom) of the ChesMMAP survey. Shaded area indicates 95% confidence interval.

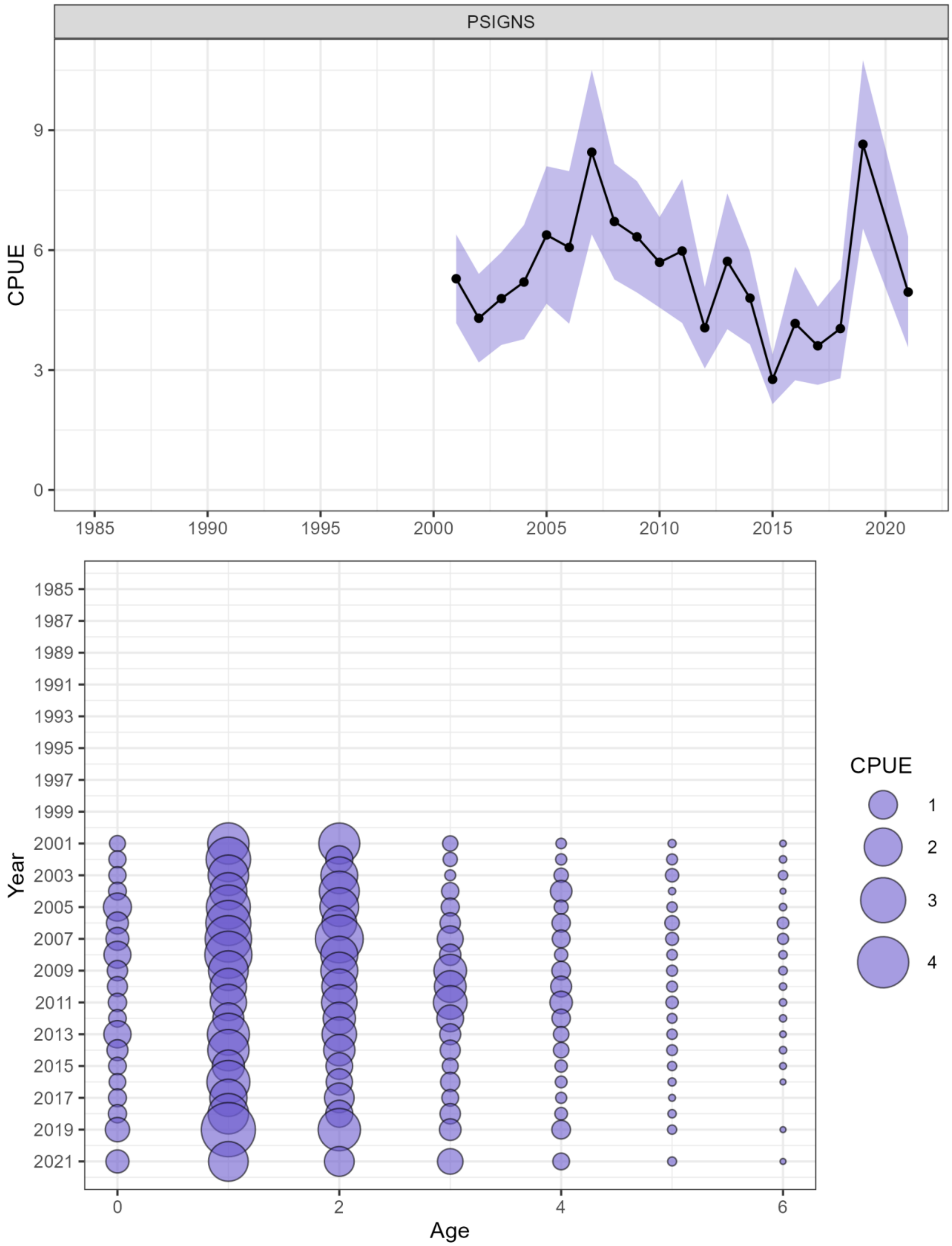

Figure 20. Index of abundance (top) and age composition (bottom) of the NC PSIGN survey. Shaded area indicates 95% confidence interval.

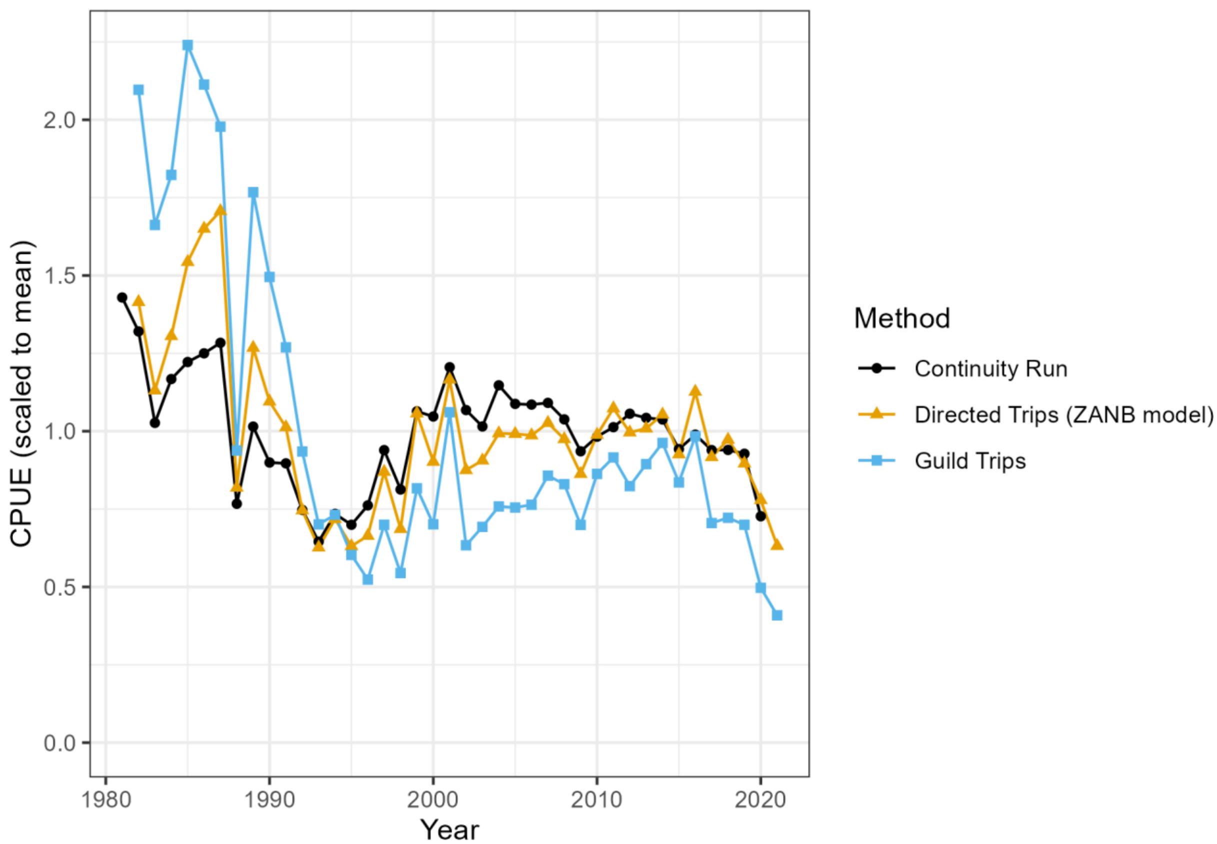

Figure 21. Comparison of trends in the MRIP CPUE index developed using different trip selection and standardization methods.

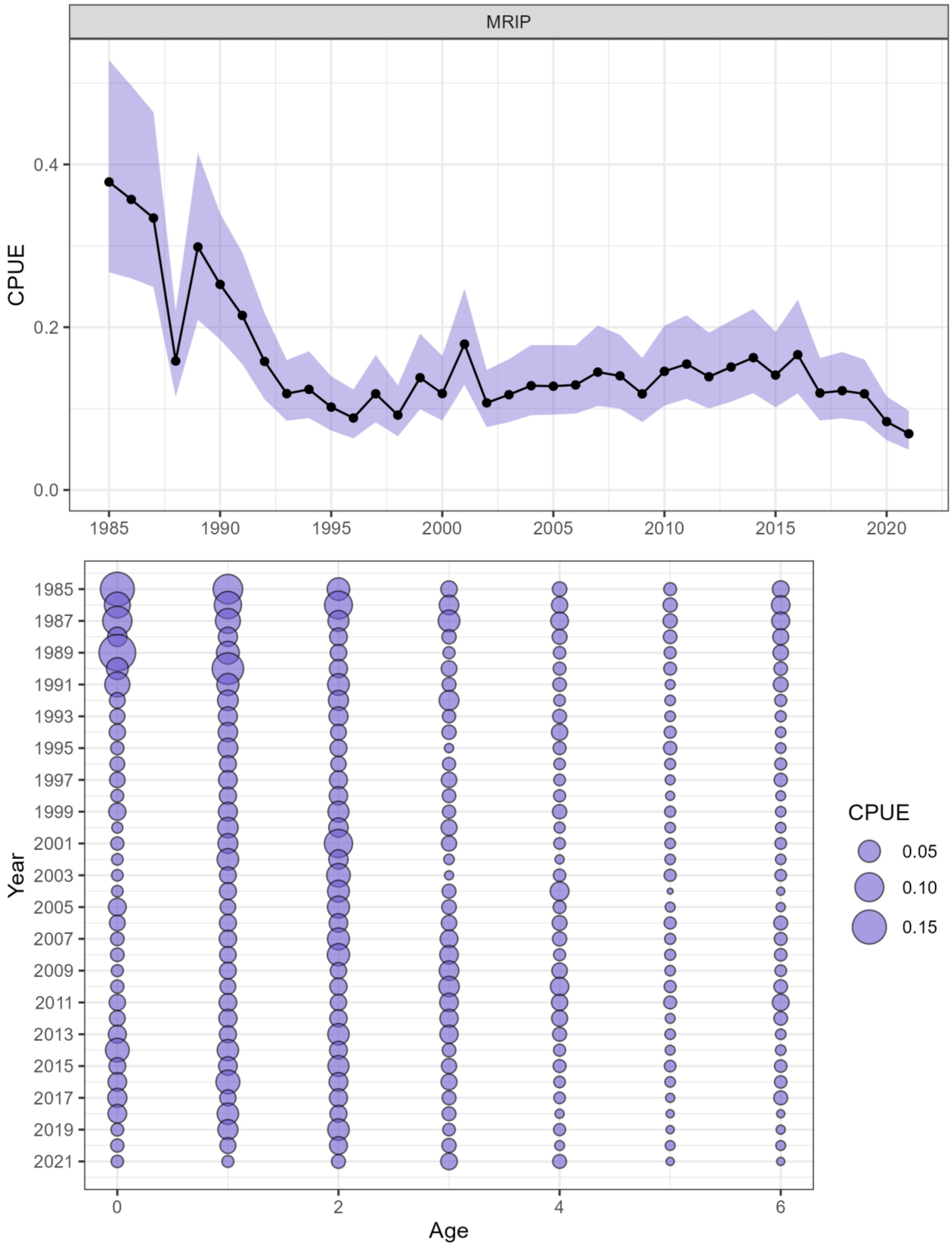

Figure 22. MRIP CPUE (top) and age composition (bottom).Shaded area indicates 95% confidence interval.

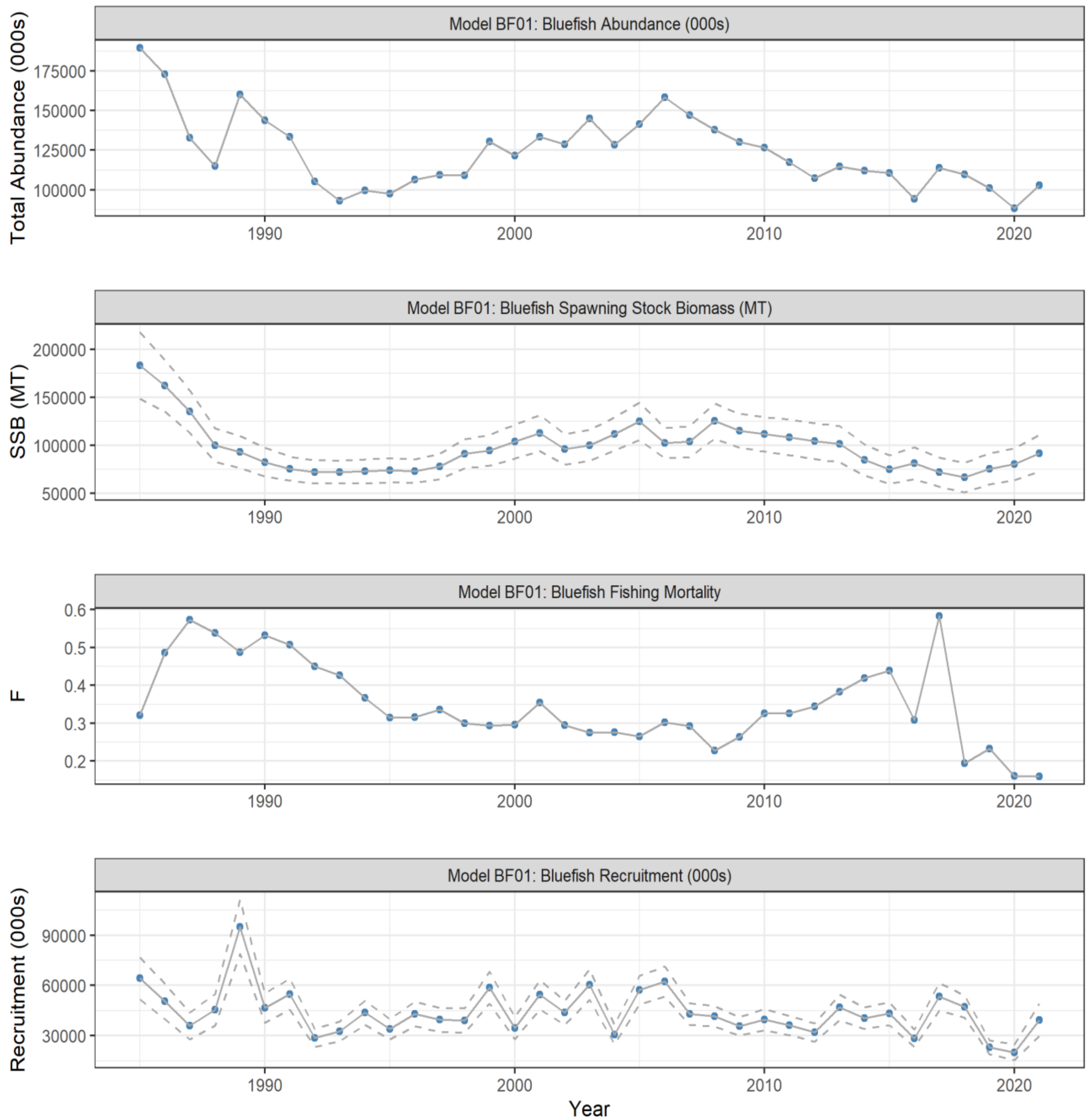

Figure 23. Model results of abundance, SSB, fishing mortality, and recruitment for the continuity run (Model BF01).

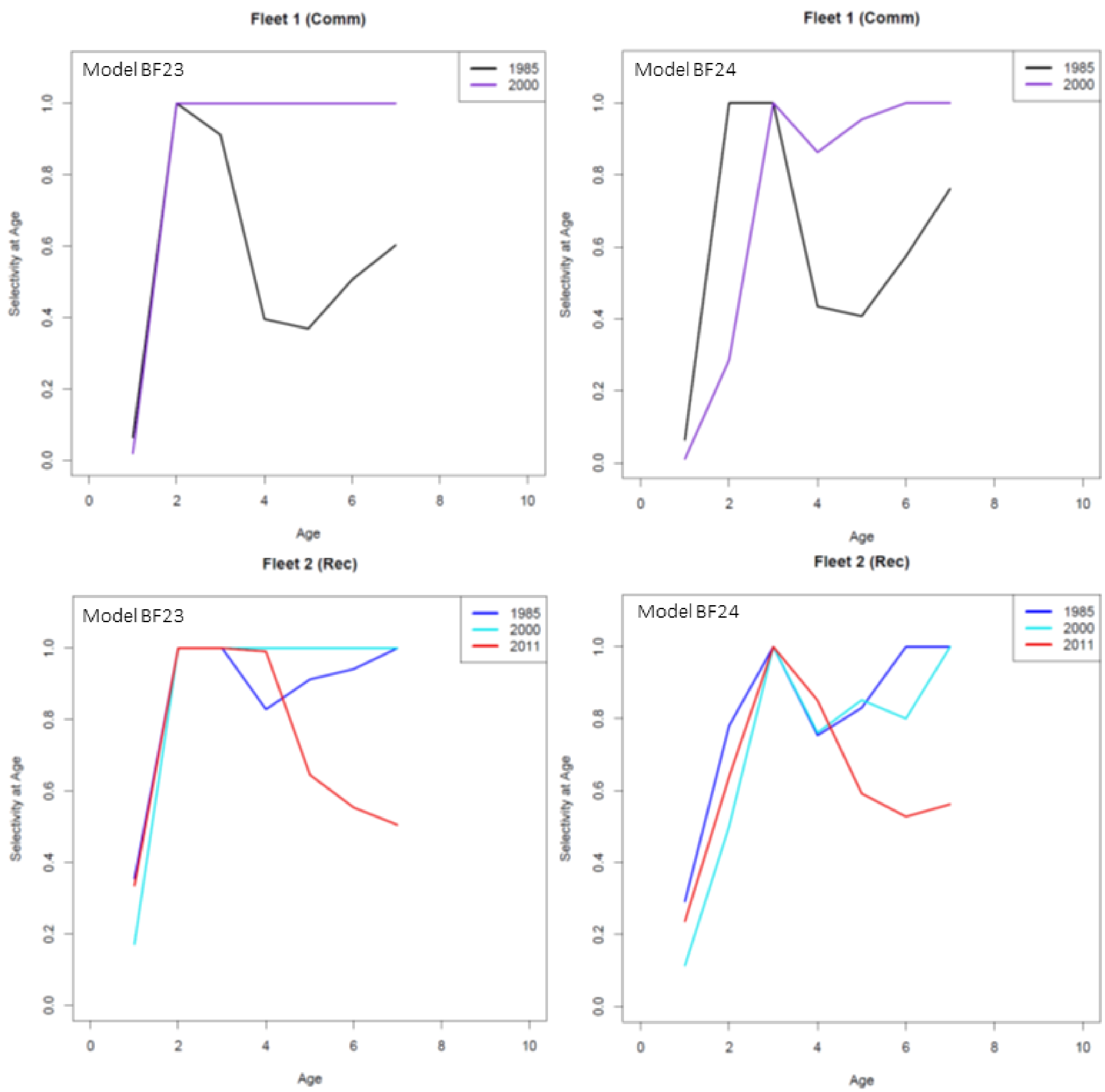

Figure 24. Fleet selectivity block comparison between model BF23 (left) and final model BF24 (right) after addressing poor age composition residual blocks in BF23.

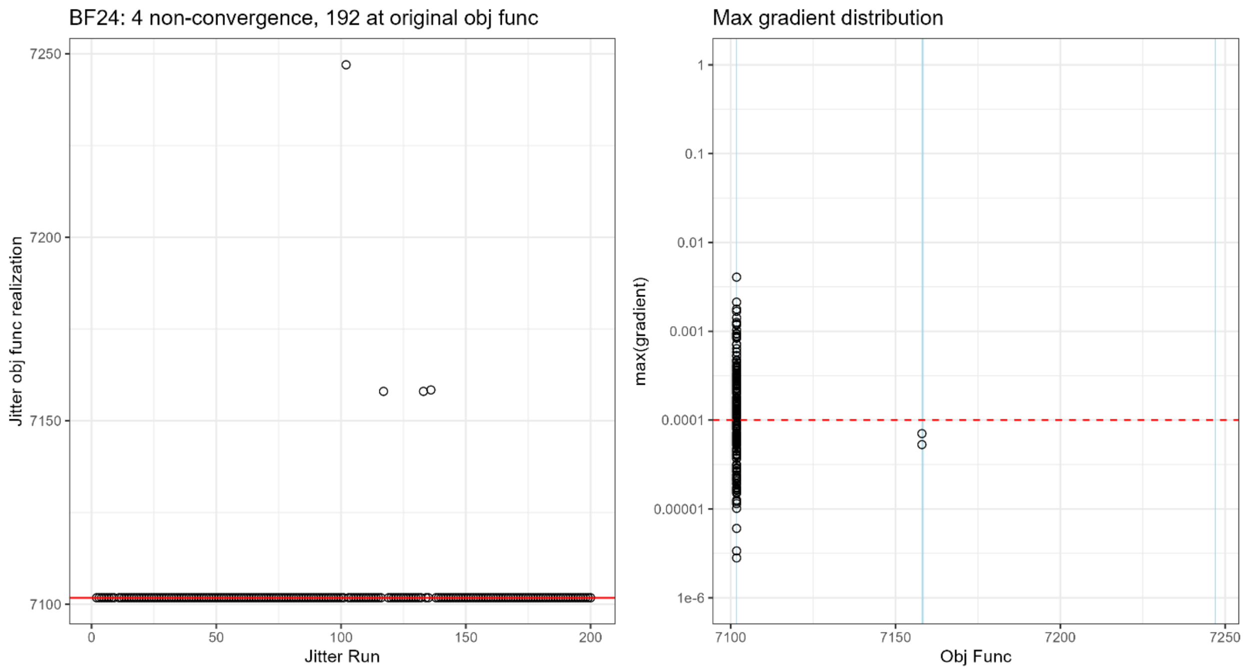

Figure 25. Convergence diagnostics for the final ASAP model, model BF24. The left plot shows objective function results for 200 random sets of starting values, with the original objective function designated by the horizontal red line. The right plot shows gradient values from each model distributed around a 0.0001 criterion for 'good' convergence (blue vertical lines represent other objective function solutions, some where the gradient result is above or below the y-axis range of the plot).

Figure 26. Comparison between model results from ASAP model BF24 (blue lines and points) and the same model run as a traditional statistical CAA model in WHAM (orange lines).

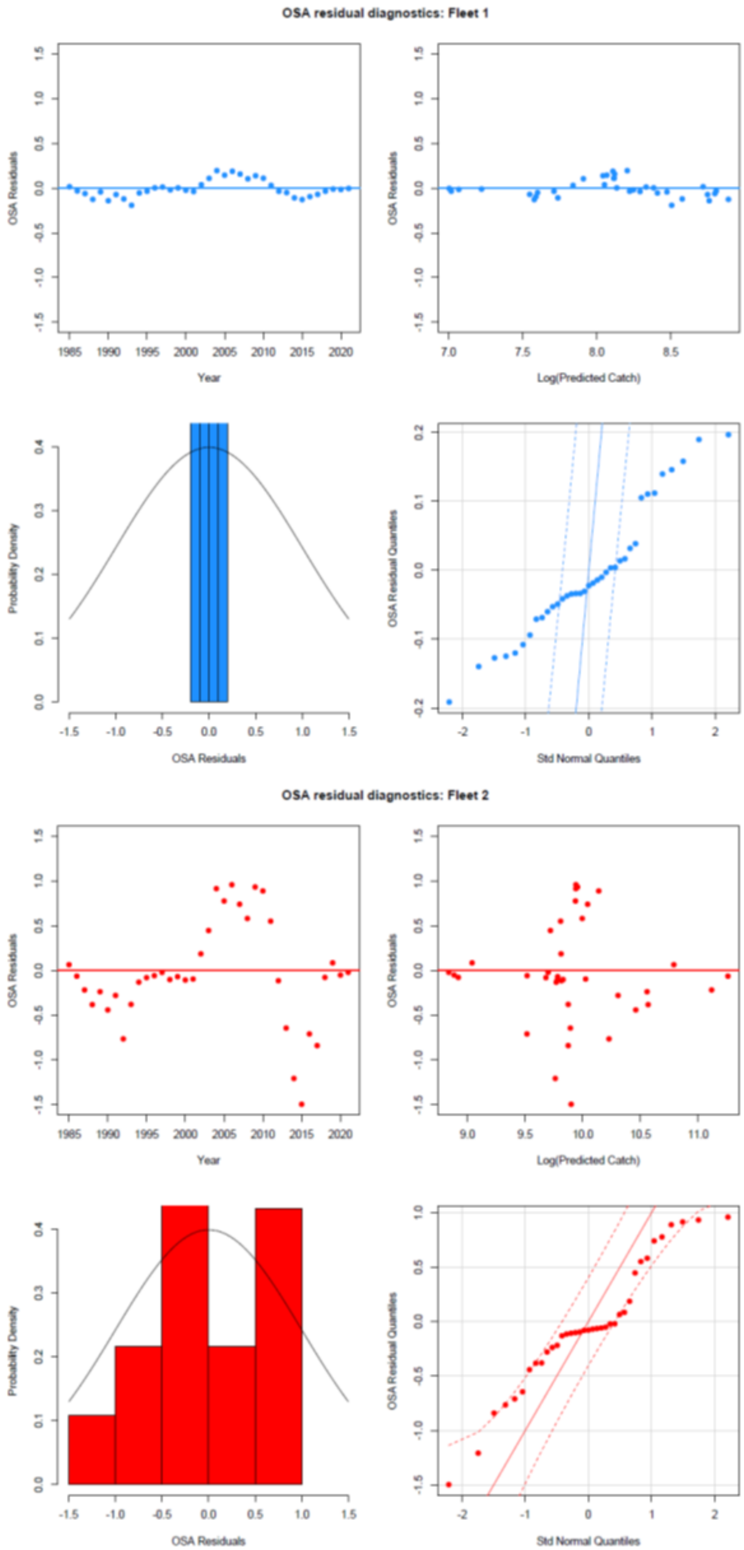

Figure 27. One-step ahead residual diagnostics for the 2 fleets. Patterns in the diagnostics for both fleets led to a reduction in the fleet input CVs.

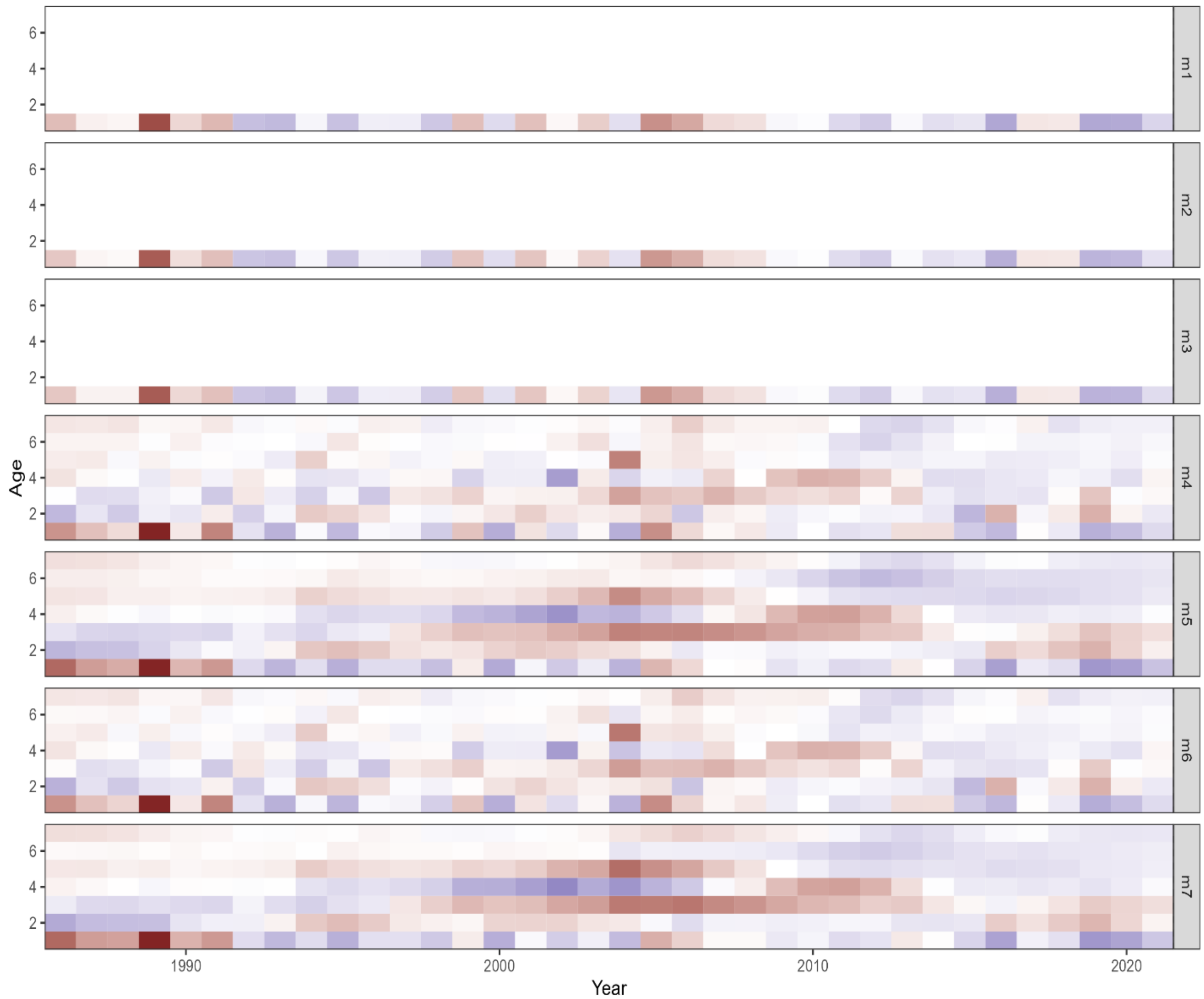

Figure 28. Number-at-age deviations for the models BF28W_m1 through BF28W_m7. Red indicates positive deviations and blue indicates negative deviations.

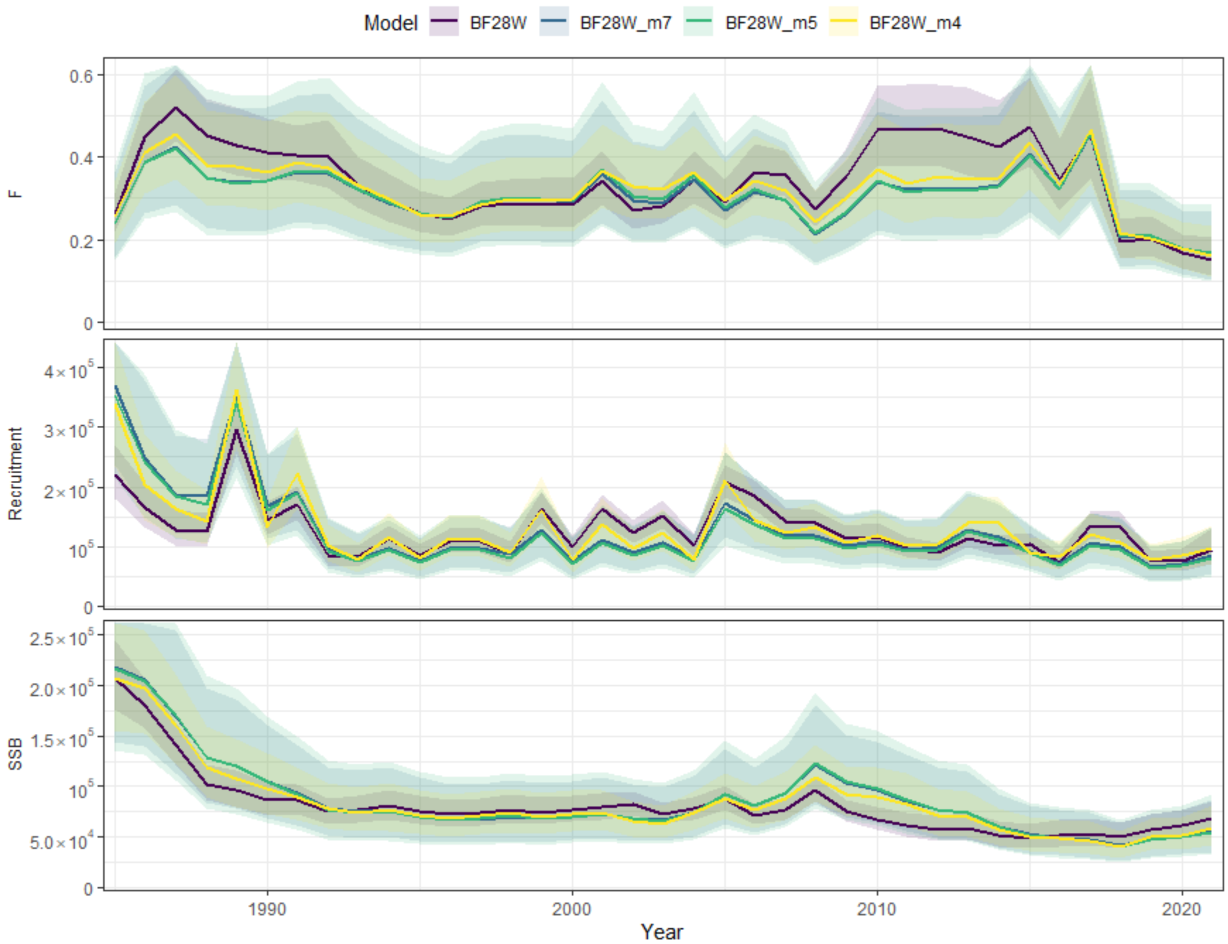

Figure 29. A comparison of SSB, F, and recruitment between the final bluefish model (BF28W_m7), the base statistical catch-at-age model (BF28W), and the top two closest models chosen by AIC.

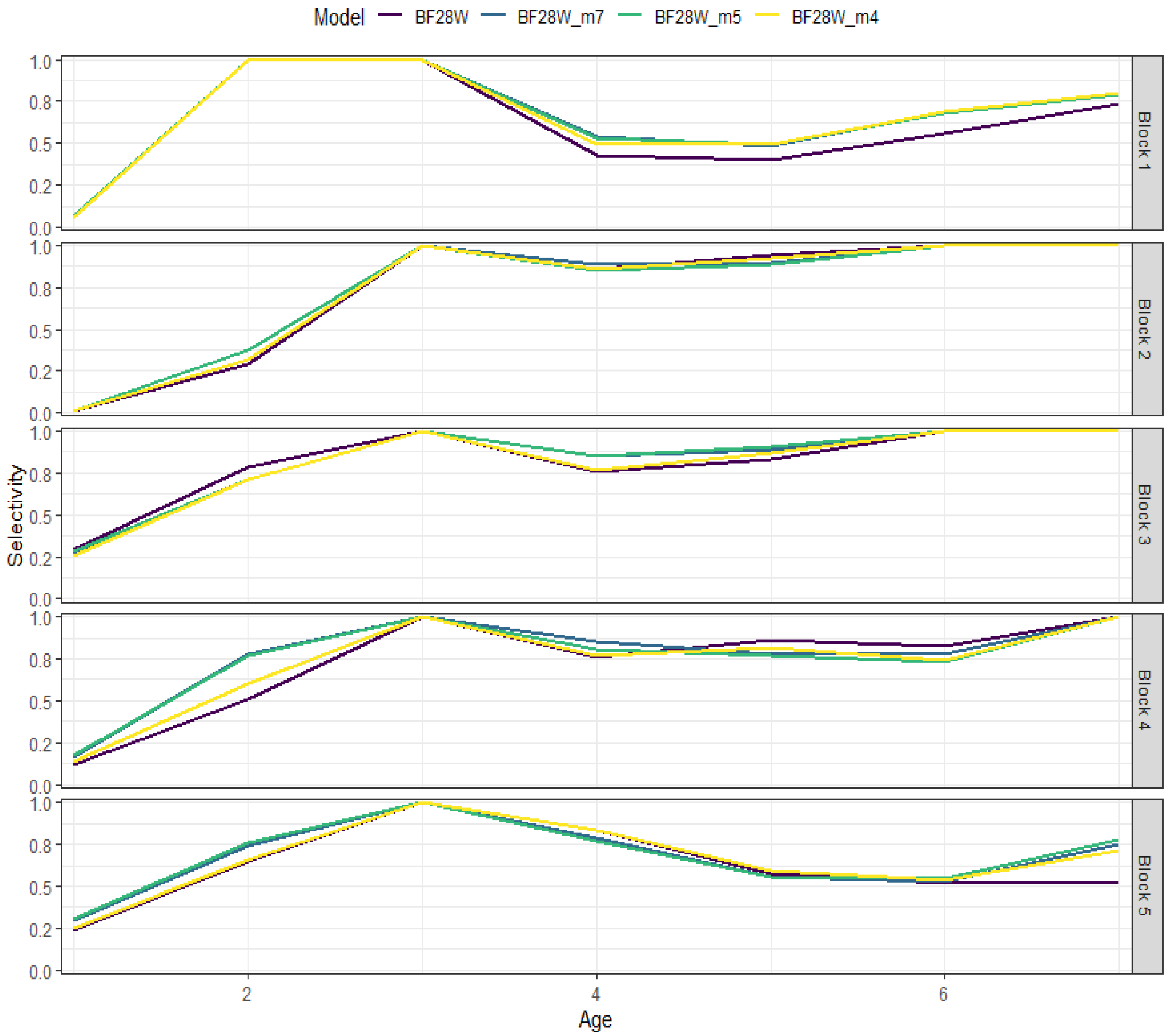

Figure 30. A comparison of the fleet selectivity block estimates between the final bluefish model (BF28W_m7), the base statistical catch-at-age model (BF28W), and the top two closest models chosen by AIC. Block 1: commercial fleet 1985-1999, Block 2: commercial fleet 2000-2021, Block 3: recreational fleet 1985-1999, Block 4: recreational fleet 2000-2010, Block 5: recreational fleet 2011-2021.

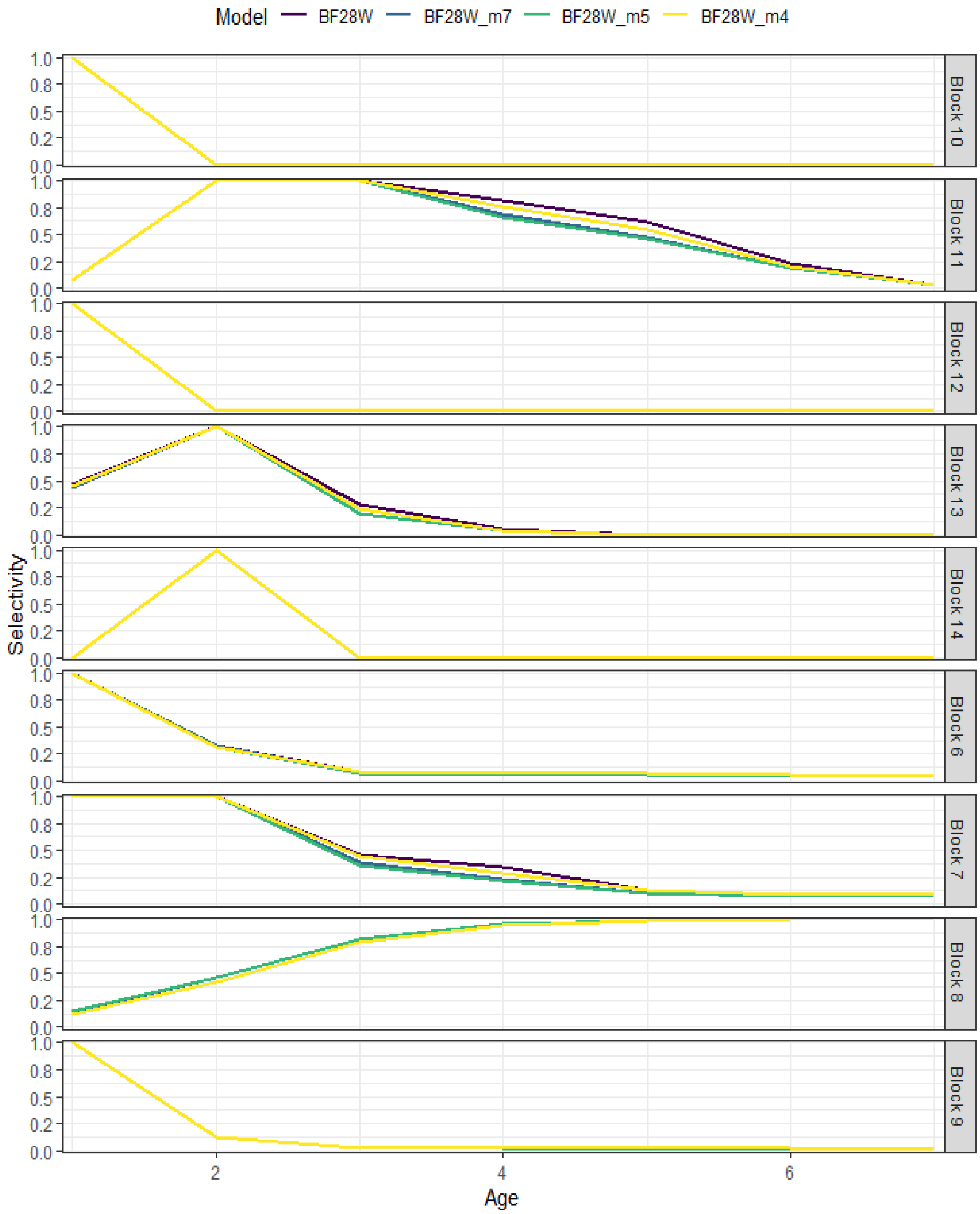

Figure 31. A comparison of the Index selectivity estimates between the final bluefish model (BF28W_m7), the base statistical catch-at-age model (BF28W), and the top two closest models chosen by AIC. Block 6: NEFSC Albatross, Block 7: NEFSC Bigelow, Block 8: MRIP CPA, Block 9: NEAMAP, Block 10: SEAMAP Age 0, Block 11: PSIGNS, Block 12: Conn YoY, Block 13: ChesMMAP, Block 14: SEAMAP Age1.

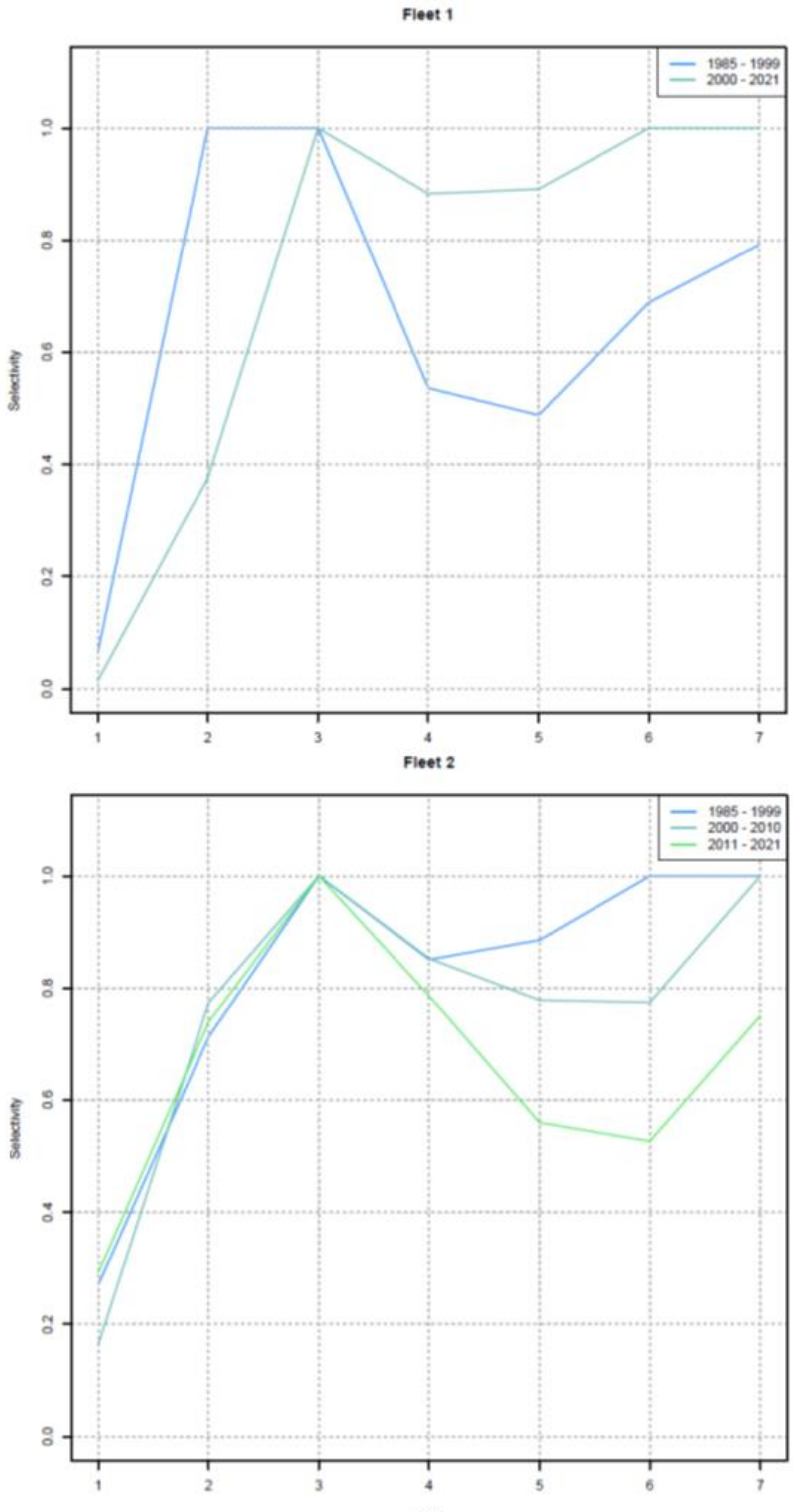

Figure 32. Selectivity estimates for the commercial (top) and recreational (bottom) fleets from the final model BF28W_m7.

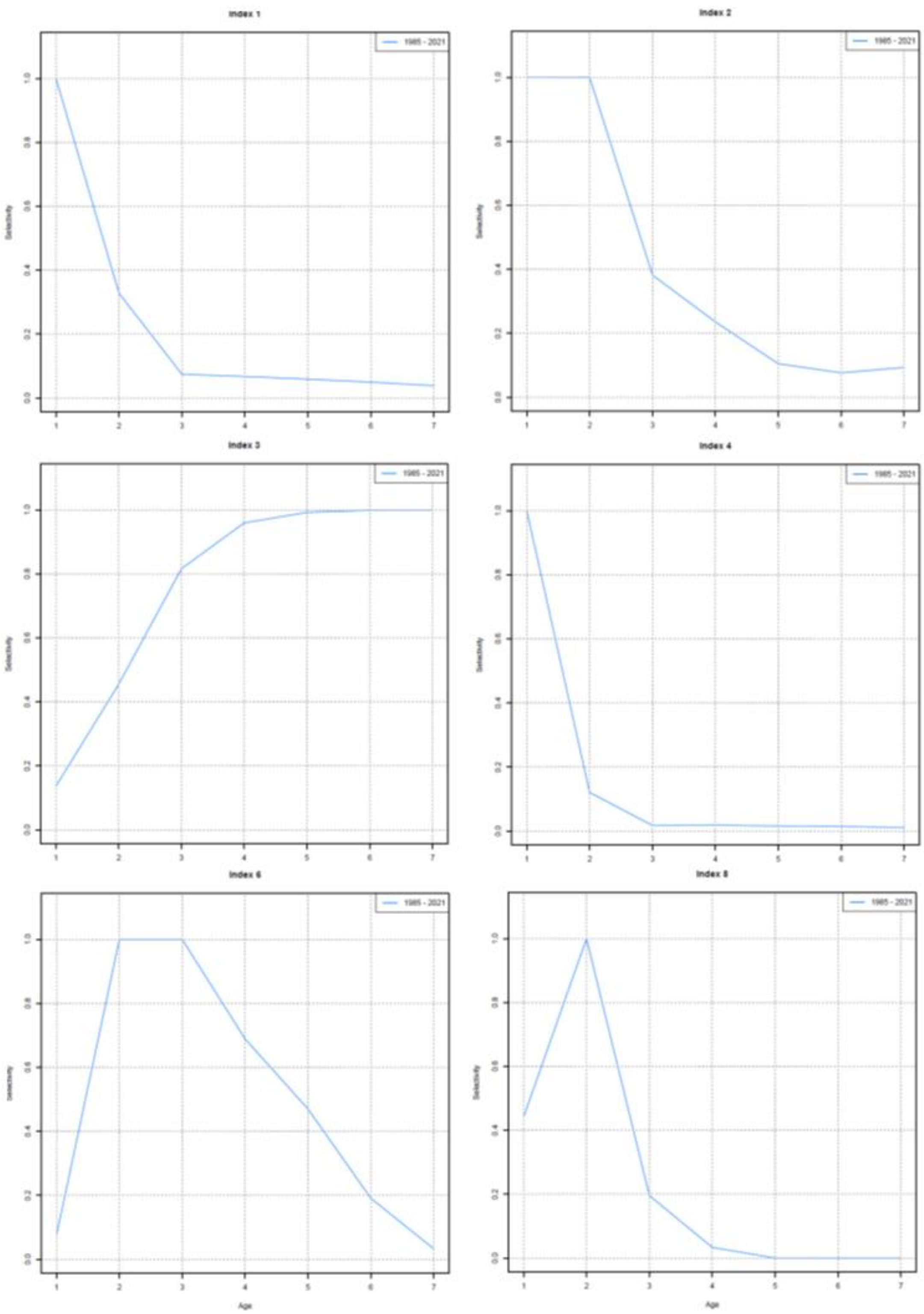

Figure 33. Final index selectivity estimates for all indices with age comps in the final model BF28W_m7. Index 1: NEFSC Alb, Index 2: NEFSC Big, Index 3: MRIP CPA, Index 4: NEAMAP, Index 6: PSIGNS, Index 8: ChesMMAP.

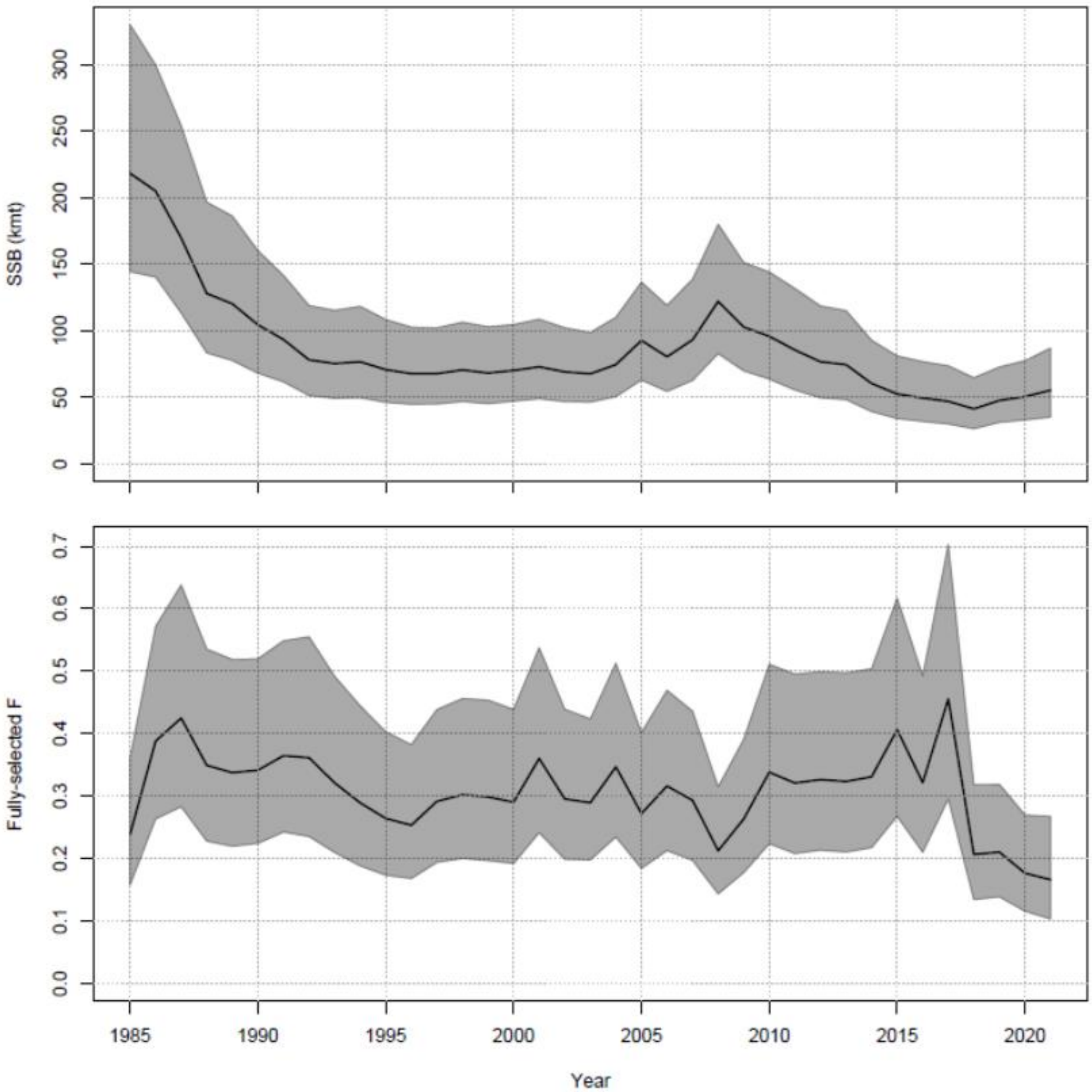

Figure 34. Spawning stock biomass (top) and fully selected fishing mortality (bottom) results from the final model BF28W_m7 from 1985-2021.

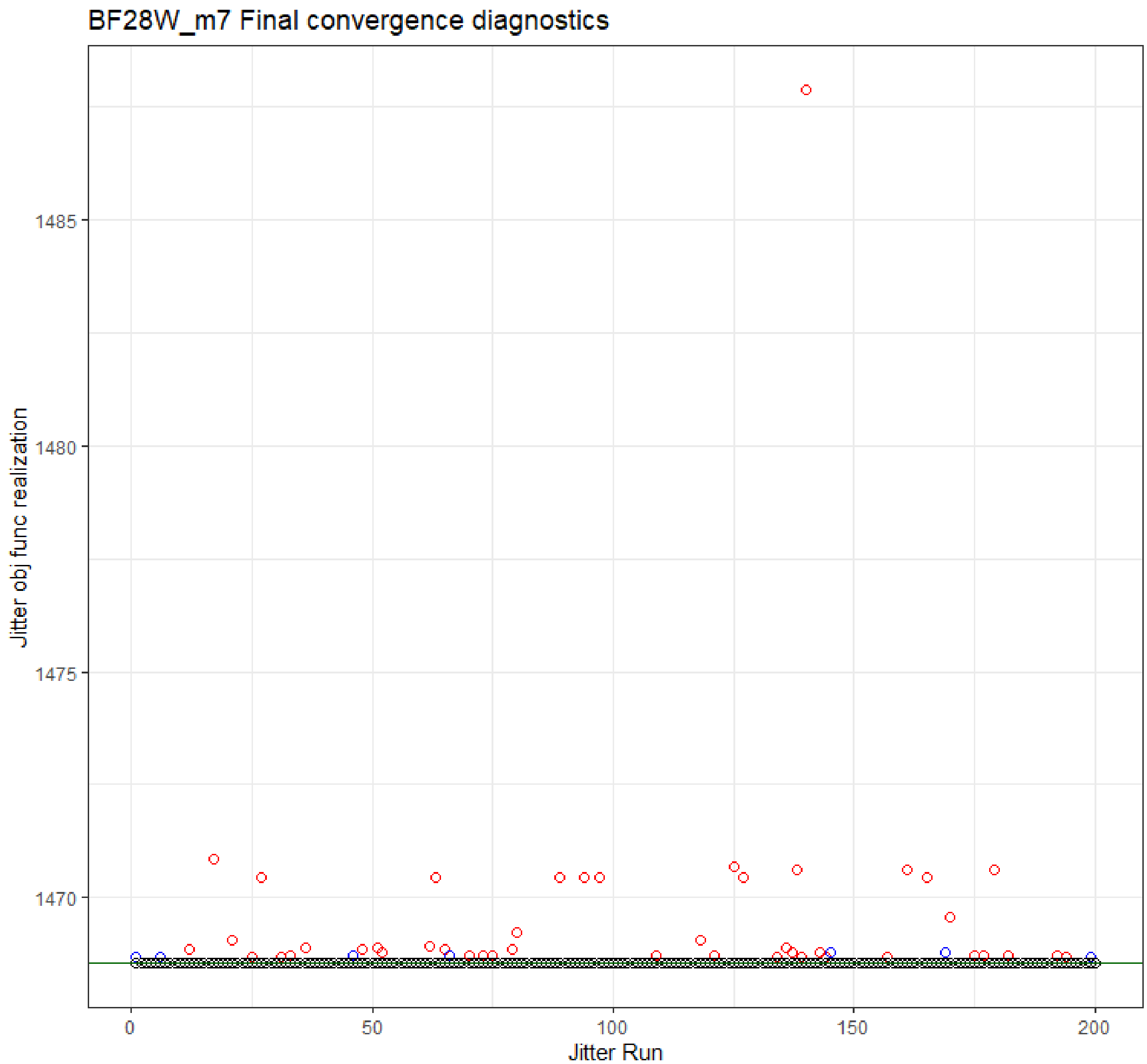

Figure 35. Jitter analysis to investigate convergence properties of final model BF28W_m7. 200 jitter runs at 3 different variance scales were run with convergence results shown (Scale 1 = black, Scale 2 = blue, Scale 3 = red). The green line indicates the original model objective function.

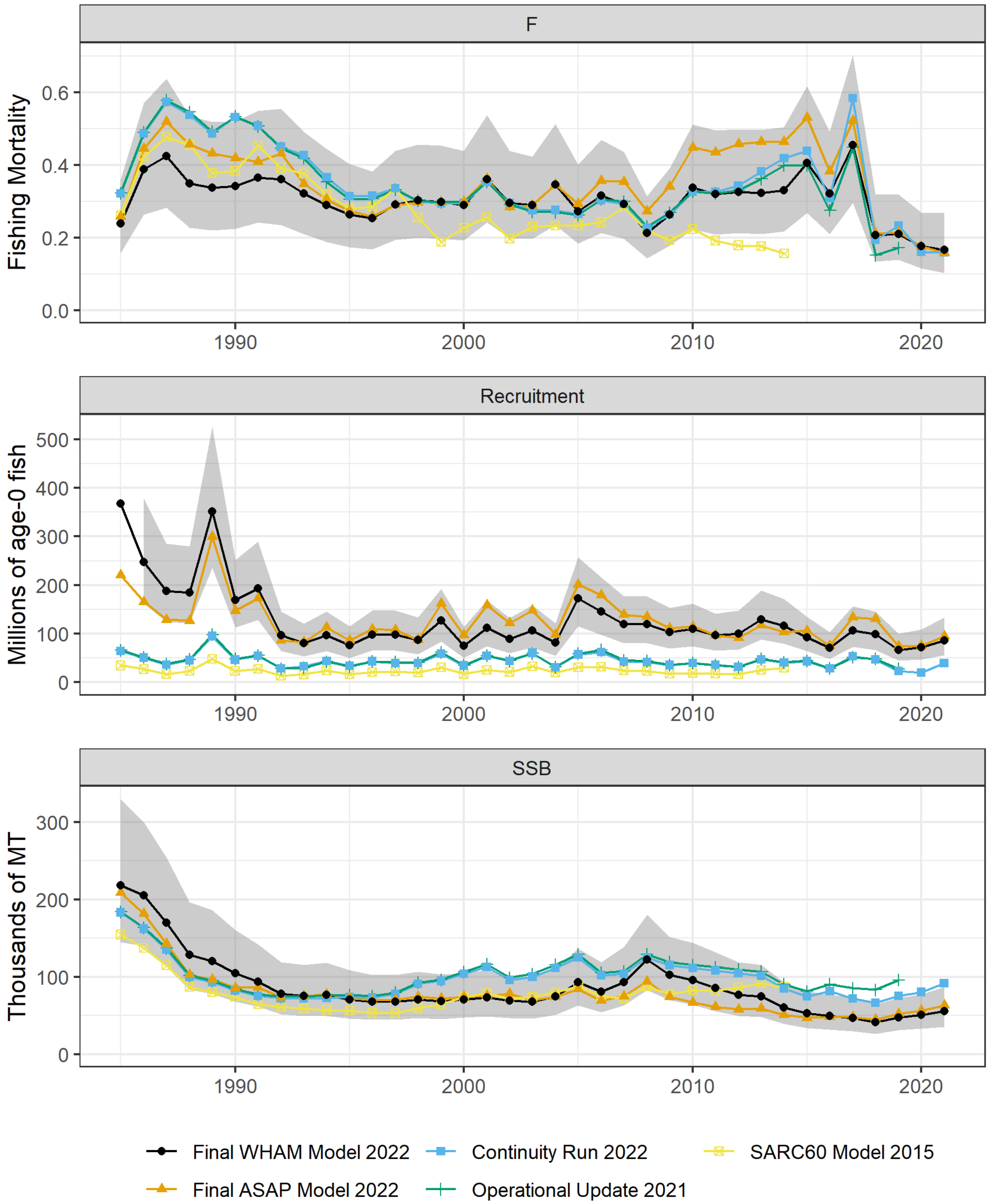

Figure 36. Historical retrospective of model results from the final WHAM model, the final ASAP final model, the continuity run update of the SAW60 model, the operational assessment in 2021, and the SARC60 benchmark model. The shaded area indicates the 95% confidence intervals for the final WHAM model estimates.

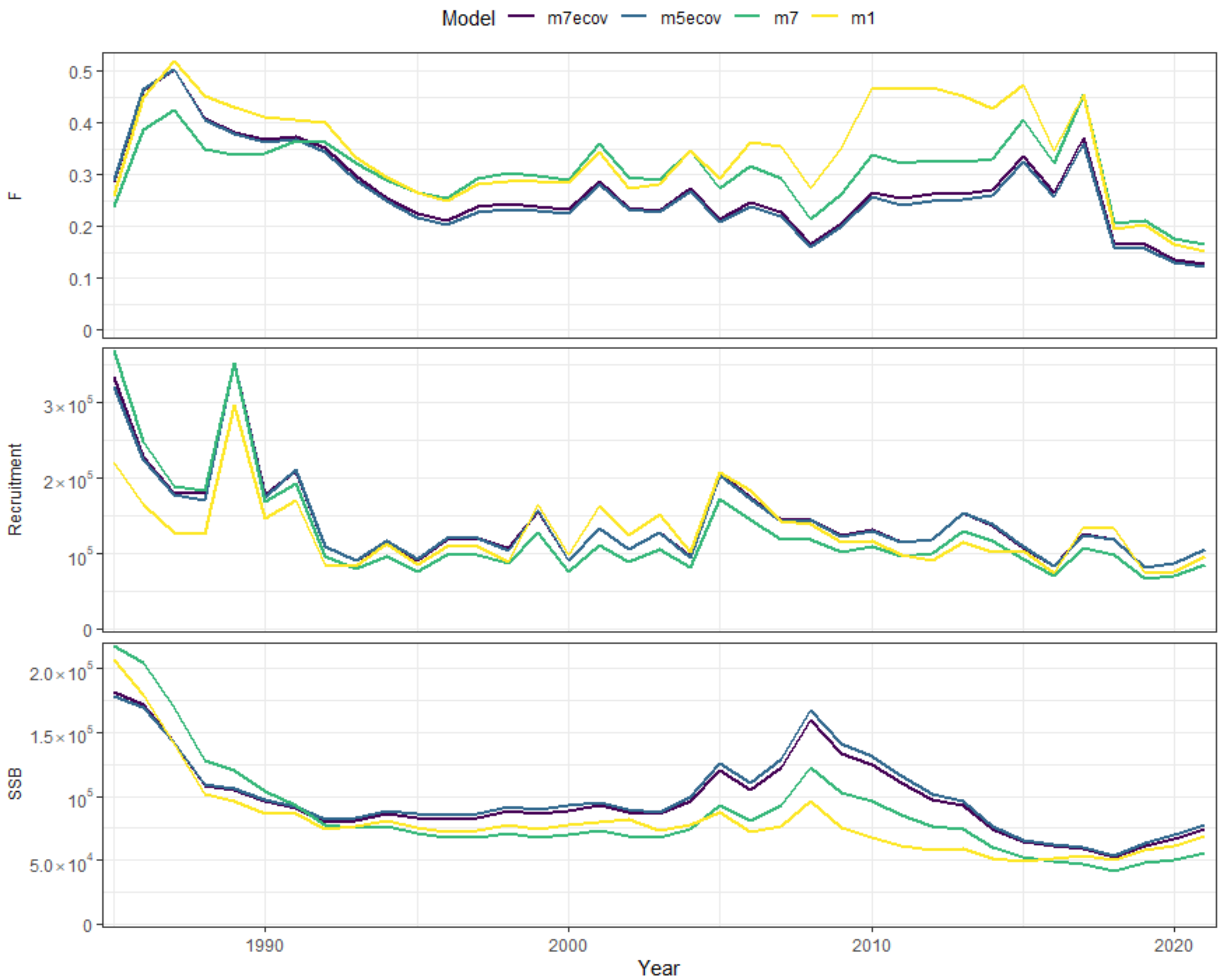

Figure 37. A comparison of the results from the base model (BF28W) and the top 3 models that include the environmental covariate on the MRIP index catchability.

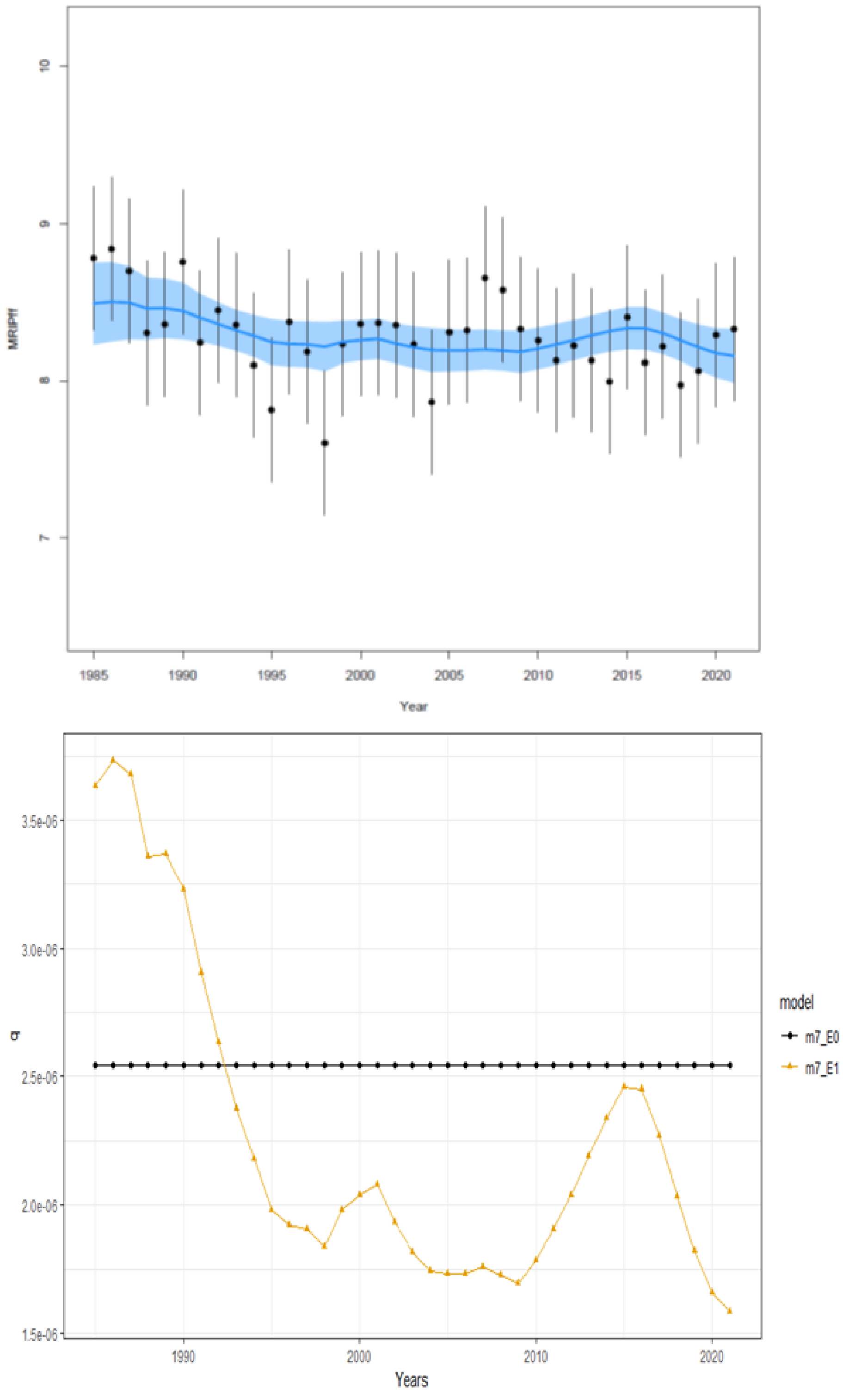

Figure 38. Fit to the forage fish index used as a covariate on the catchability (availability) of the MRIP index (top) and resulting trend in estimated catchability over time for the MRIP index (bottom).

Figure 39. Results from the sensitivity analyses testing the impact of removing each index on the model results.

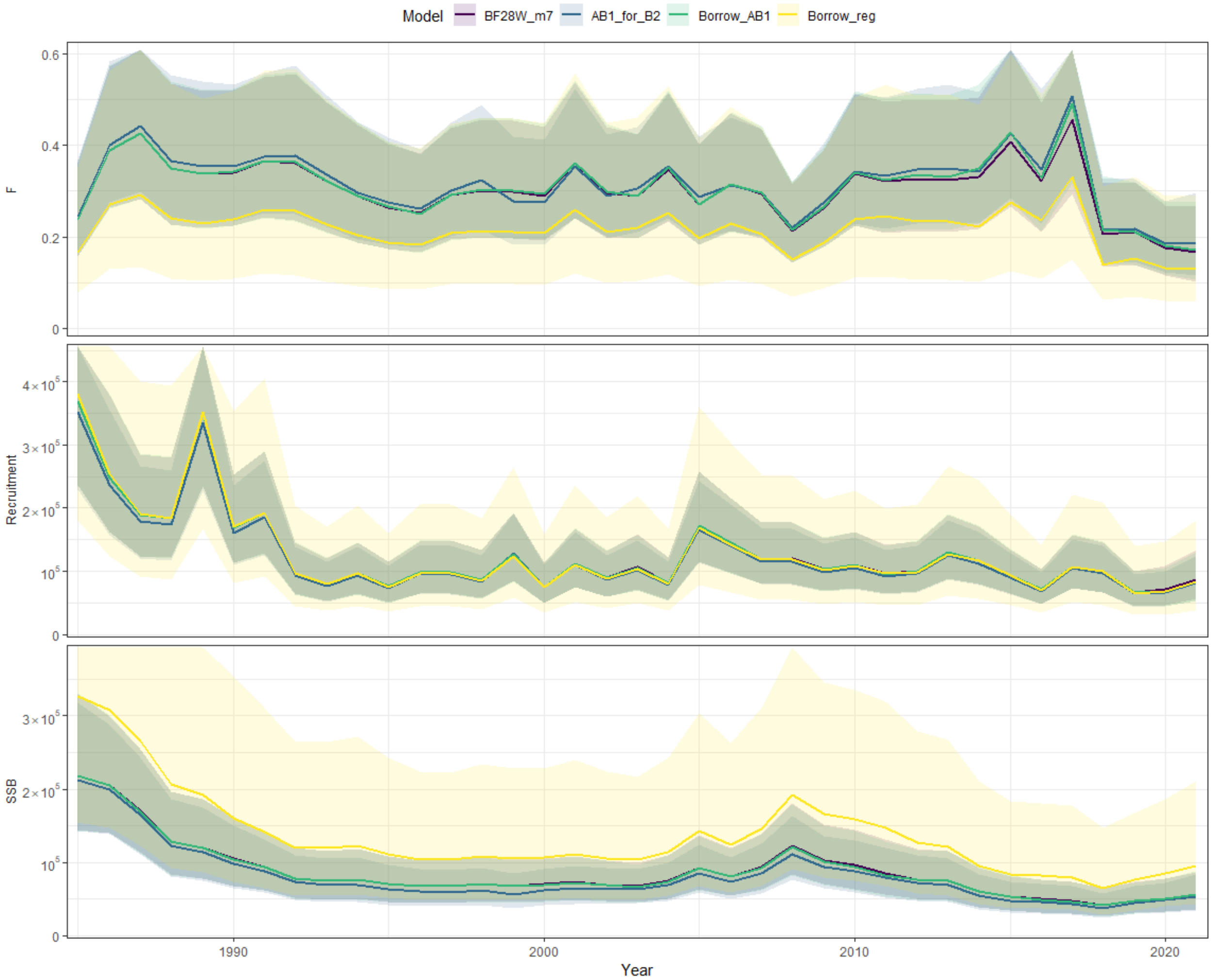

Figure 40. Results from the sensitivity analyses testing different borrowing rules for the recreational discard lengths. The borrowing from region model ("Borrow_reg") had poor convergence properties with bounded parameter estimates.

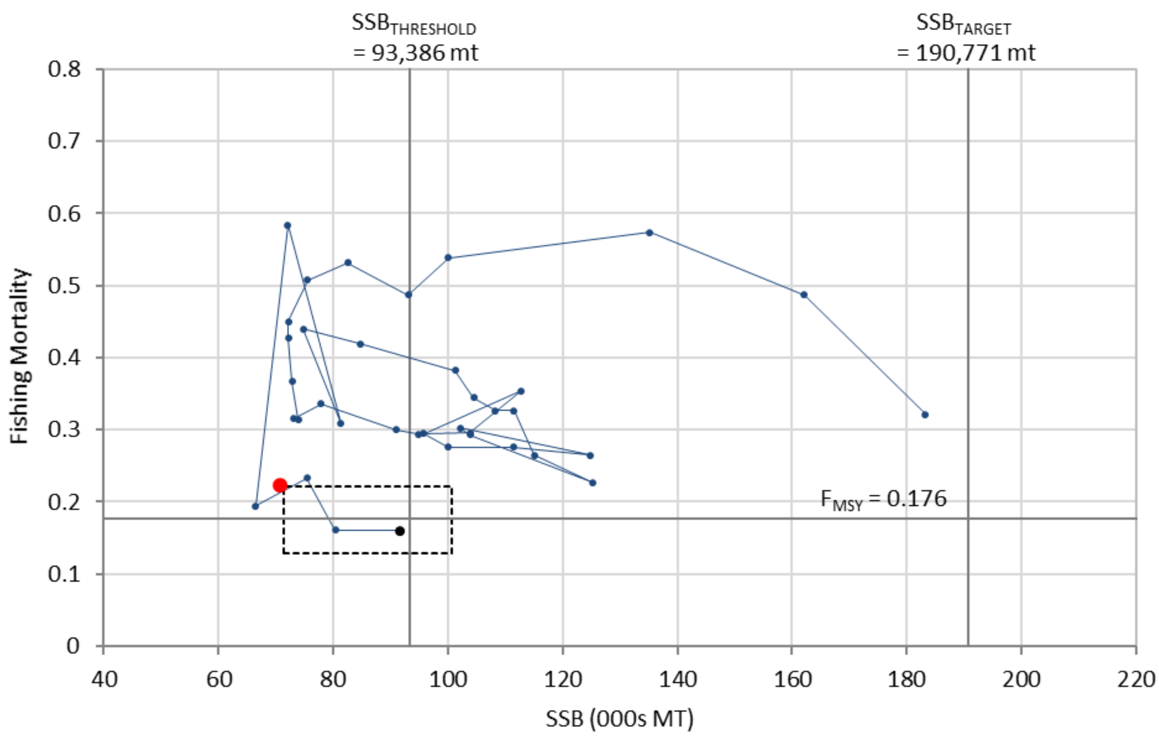

Figure 41. Stock status plot with status determination criteria for the model BF01, the continuity run. Dashed line indicates the 90% confidence region around the terminal year estimates of F and SSB. Red dot shows the retrospective adjusted terminal year values, which fall outside the confidence region and indicate that status should be determined using the retrospectively adjusted values.

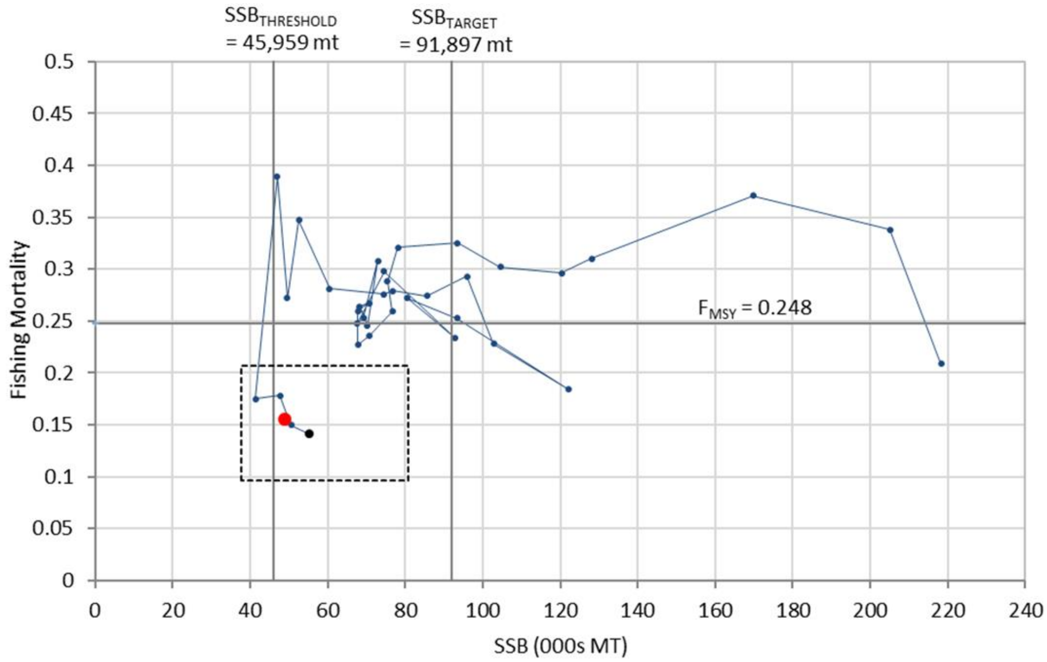

Figure 42. Stock status plot with status determination criteria for the final WHAM model for this assessment. Dashed line indicates the 90% confidence region around the terminal year estimates of F and SSB. Red dot shows the retrospective adjusted terminal year values, which fall within the confidence region and indicate that a retrospective adjustment is not necessary to determine stock status.

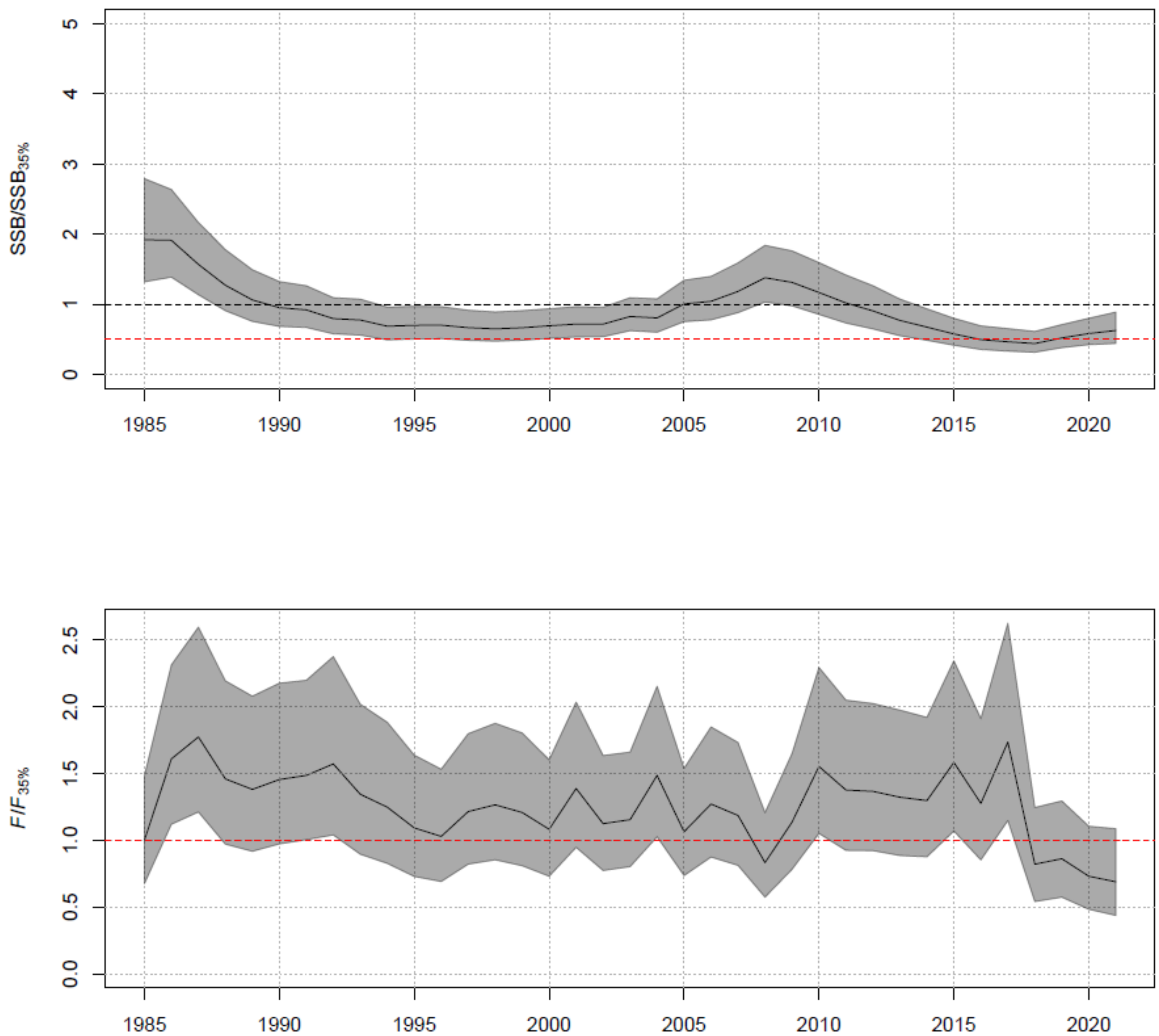

Figure 43. Final model SSB and fishing mortality in relation to SSB35% and F35%, the status determination criteria. The current bluefish stock is not-overfished and over-fishing is not occurring.

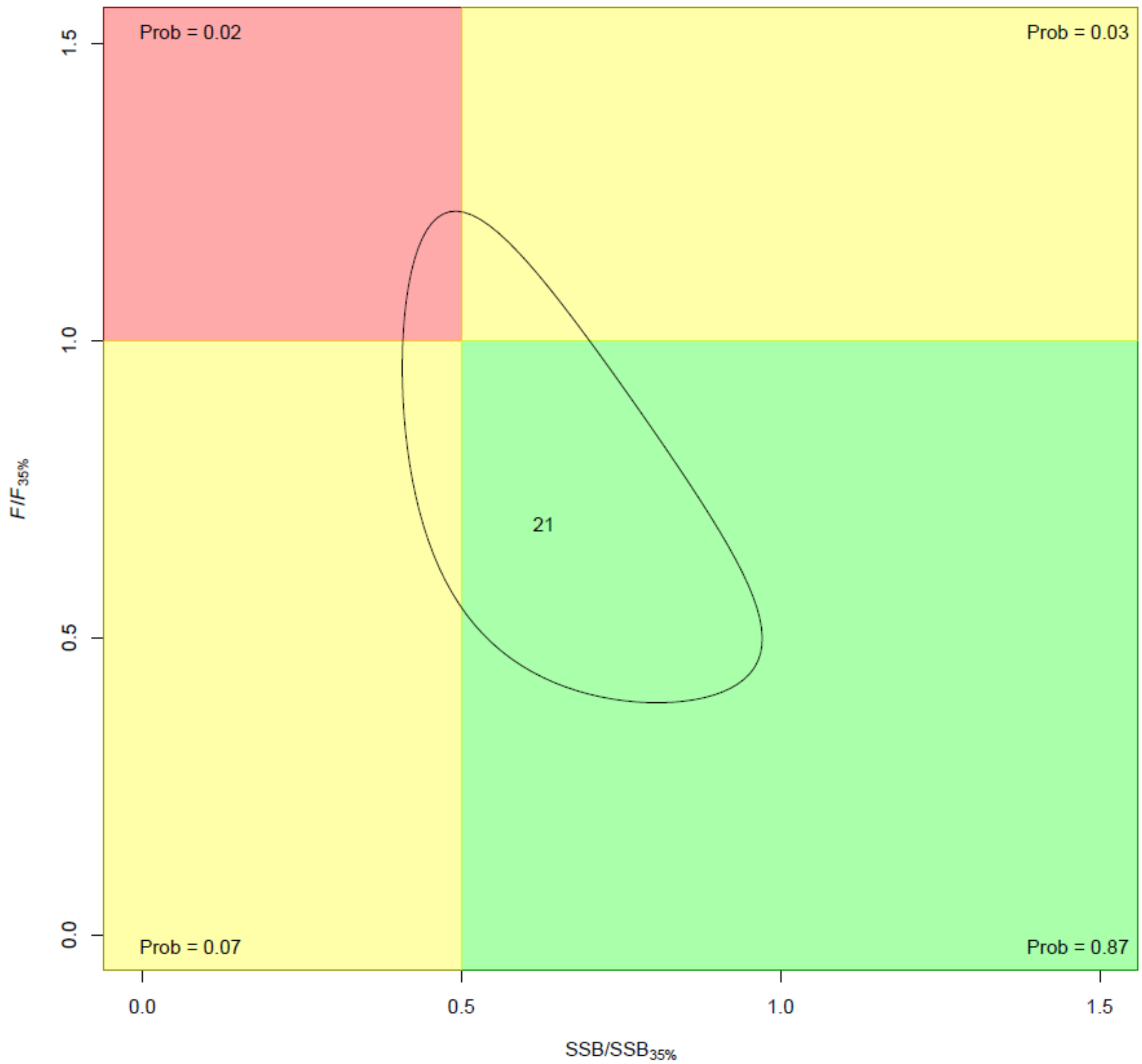

Figure 44. Kobe plot from final model (BF28W_m7) with stock status probability ellipse showing an 87% probability the bluefish stock is not overfished and over-fishing is not occurring. "21" indicates the 2021 point estimates of $SSB/SSB_{35\%}$ and $F/F_{35\%}$.

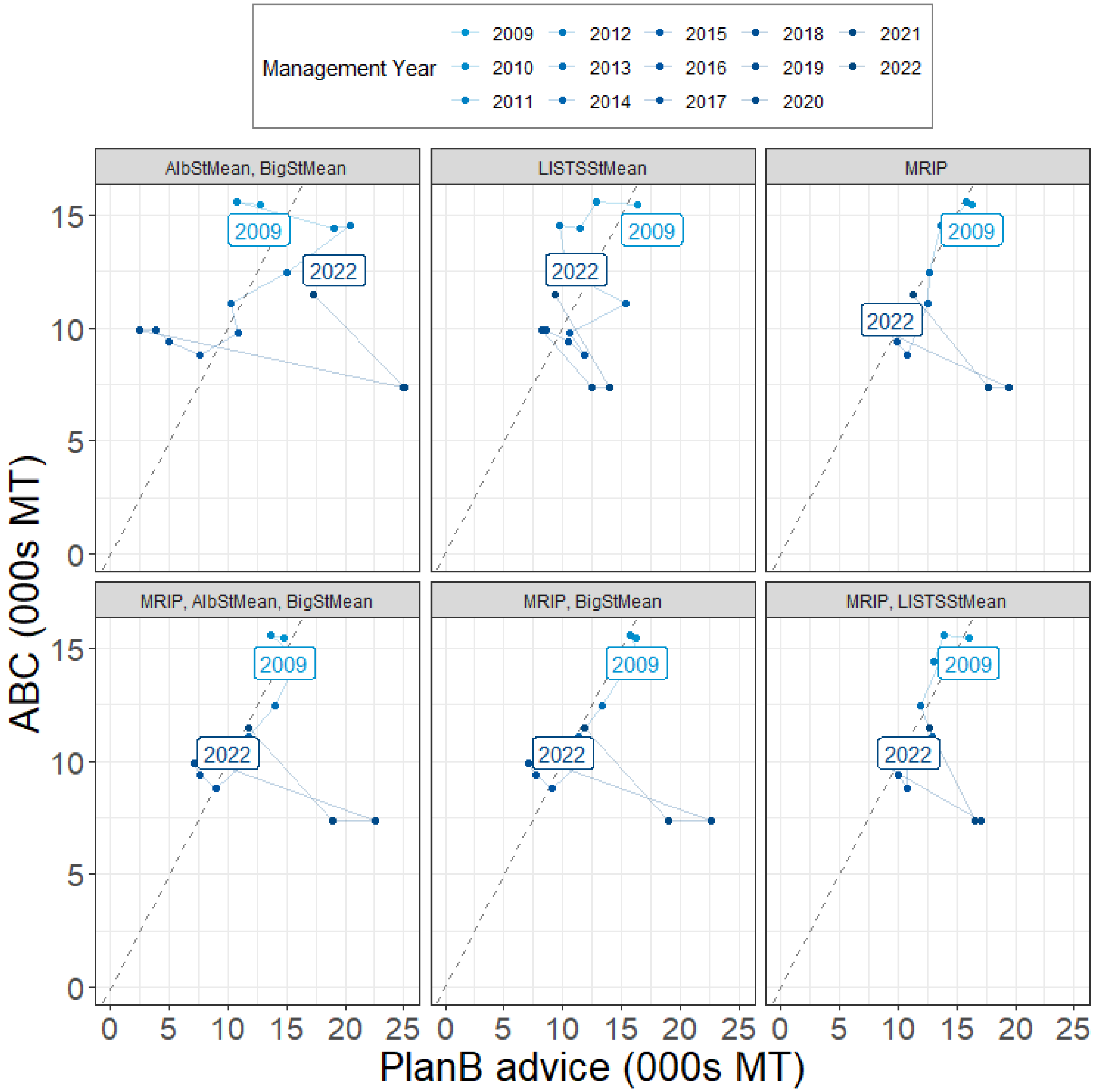

Figure 45. Results of applying a hypothetical Ismooth analysis compared to the actual ABC recommended for use in management. The dashed line is the 1-1 line.

12 LIST OF RELEVANT WORKING PAPERS

Working papers are available on the NEFSC data portal for this assessment, and at the hyperlinks below.

1. Tyrell et al. 2022. Bluefish Ecosystem and Socioeconomic Profile.
2. Valenti 2022a. The Spatial Distribution of Bluefish (*Pomatomus saltatrix*): Insights from American Littoral Society Fish Tagging Data
3. Tyrell 2022. Bluefish VAST Index Exploration.
4. Gaichas et al. 2022. Vector Autoregressive Spatio-Temporal (VAST) modeling of piscivore stomach contents, 1985-2021.
5. Truesdell et al. 2022. Life History Analyses for Bluefish.
6. Tyrell and Truesdell 2022. Natural mortality of bluefish.
7. Celestino et al. 2022a. Index of abundance exploration and development by the Bluefish Working Group's Fishery Independent Data Group.
8. Wood 2022a. Commercial and Recreational Data Collection and Analysis.
9. Drew 2022a. Recreational Data Changes for Bluefish, 2012-2021.
10. Drew 2022b. The Spatial Distribution of Bluefish (*Pomatomus saltatrix*): Insights from MRIP Data.
11. Valenti 2022b. Catch-and-Release Recreational Angling Mortality of Bluefish (*Pomatomus saltatrix*): Updated Analysis for 2022
12. Drew 2022c. Development of the Composite YOY Index for Bluefish.
13. Drew 2022d. A Fishery-dependent CPUE index for bluefish derived from MRIP data.
14. Celestino et al. 2022b. Development of Bluefish Age-Length Keys.
15. Wood 2022b. Bluefish Model Bridge-Building in ASAP.
16. Wood 2022c. ASAP diagnostic plots.
17. Wood 2022d. WHAM diagnostic plots.
18. Truesdell 2022. Alternative assessment plan.