2024 GOLDEN TILEFISH RESEARCH TRACK ASSESSMENT

The 2024 golden tilefish research track working group (RTWG) met 10 times between October 2022 and February 2024. All meetings were held remotely via WebEx.

PARTICIPANTS

Working Group

NAME	AFFILIATION
Jason Boucher	NEFSC
Nikolai Klibansky	SEFSC
Sean Lucey	NEFSC
John Maniscalco	NYSDEC
José Montañez	MAFMC, chair
Paul Nitschke	NEFSC, assessment lead

GARFO = Greater Atlantic Regional Fisheries Office
MAFMC = Mid-Atlantic Fishery Management Council (Council)
NEFSC = Northeast Fisheries Science Center
NMFS = National Marine Fisheries Service
NOAA = National Oceanic and Atmospheric Administration
NYSDEC = New York State Department of Environmental Conservation
SCDNR = South Carolina Department of Natural Resources
SEFSC = Southeast Fisheries Science Center

In addition to the Working Group members, the following individuals participated in some of the meetings:

Chair-invited analytical participants

Daniel Hennen	NEFSC
Kimberly Hyde	NEFSC
Andrew Jones	NEFSC
Anthony Kaufman	NOAA-Affiliate
Adelle Molina	NEFSC
Stephanie Owen	NEFSC
Sarah Salois	NEFSC-Affiliate

Working Group meeting attendees

Charles Adams	NEFSC
Fred Akers	MAFMC Tilefish Advisory Panel Member
Rusell Brown	NEFSC
Kathie Burchard	NEFSC
Greg DiDominico	Lund's Fisheries
Alexander Dunn	NEFSC
Skip Feller	Industry/Council Member
Frank Green	Industry
Homer Hiers	SCDNR

Scott Large	NEFSC
Yong-Woo Lee	NOAA-NMFS
Scott Large	NEFSC
Chris Legault	NEFSC
Laurie Nolan	Industry/ex-Council Member
Douglas Potts	GARFO
Daemian Schreiber	NOAA-Affiliate
Michele Traver	NEFSC
Mark Terceiro	NEFSC
Anthony Wood	NEFSC

ACKNOWLEDGEMENTS

The RTWG thanks: Dan Hennen for assistance with WHAM explorations. We also thank Sarah Salois, Stephanie Owen, Adelle Molina, Andrew Jones, and Kimberly Hyde for developing the ecosystem and socioeconomic profile (ESP) framework. Andy Jones developed the study fleet and observer trawl CPUE. Tony Kaufman provided the RTWG with valuable information and data about the large pelagic survey. Michele Traver and Alexander Dunn provided helpful administrative support. Fred Akers, Skip Feller, Frank Green, and Lauire Nolan offered expert advice regarding the operations of the commercial and recreational golden tilefish fisheries.

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.

The northern stock of golden tilefish are a long-lived, non-migratory demersal species inhabiting the outer continental shelf and slope of the Mid-Atlantic Bight region of the Northwest Atlantic. This species has relatively specific habitat preferences described by soft substrates (for burrowing) and a narrow range in temperatures and salinities. Motivated by the fact that this data-limited stock remains poorly sampled by fishery-independent surveys, this work aims to develop a suite of environmental indicators to better understand geographical distribution and potential drivers of recruitment by utilizing new and under-explored data streams. Quantitative ecosystem indicators were analyzed in relation to in situ larval data, a model-derived recruitment index and a new fishery-dependent catch per unit effort (CPUE) index derived from incidental catch. Linear regressions and generalized additive models (GAM) were used to determine the effects of ecosystem indicators on golden tilefish catch and recruitment. Most principally, there was agreement in bottom temperature and salinity preferences across all analyses and values were consistent with ranges documented in the literature. There was some seasonality to the influence of environmental indicators, such that indicators of habitat condition (bottom temperature and salinity) as well as indicators of food availability (microplankton abundance) in the fall were highly correlated with the presence of larvae and catch of recruitment age (0-1) fish. Analyses suggested physical oceanographic indicators serving as proxies for currents and movement of water masses (shelf water volume, cold pool spatial extent and persistence, Gulf Stream Index) may have important and complex influences on early life history stages. Sources of uncertainty were discussed and our findings informed several research recommendations (TOR 7). In sum, this work highlights the value of the new incidental CPUE index (derived from trawl fisheries) in beginning to make some inferences on drivers of tilefish recruitment and also provides context and support for the further development of ecosystem indicators. Specifically, findings suggest that bottom temperature, salinity at depth, shelf water volume, and microplankton abundance may influence golden tilefish recruitment or mortality and may be of use in as environmental covariates in future stock assessment models.

TOR #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.

Total commercial golden tilefish landings (live weight) increased from less than 125 mt during 1967-1972 to more than 3,900 mt in 1979 during the development of the directed longline fishery. Landings prior to the mid-1960s were landed as a bycatch in the trawl fishery. Annual landings ranged between 454 and 1,838 mt from 1988 to 1998. Landings from 1999 to 2002 were below 900 mt (ranging from 506 to 874 mt). An annual quota of 905 mt was implemented in November of 2001. Landings in 2003 and 2004 were slightly above the quota at 1,130 mt and 1,215 mt, respectively. Landings from 2005 to 2009 were at or below the quota, while landings in 2010 at 922 mt were slightly above the quota (Figure 1). Since 2010 landings have been below the quota and decreased to an estimated 494 mt in 2016. The landings have increased slightly to an average of 695 mt from 2017 to 2022. The Total Allowable Landings (TAL) was reduced for the first time in 2015 to 796 mt from the TAL of 905 mt which was in place from 2001-2014. The TAL in 2016 and 2017 was increased to 856 mt based on projections from the SARC 58 assessment. The TAL was then reduced to 738 mt from 2018 to 2021 based on the 2017 operational assessment and subsequently increased based on the 2021 management track assessment. The top 4 permits hold 80% of the golden tilefish IFQ (individual fishing quota) allocation.

During the development of the directed longline fishery in the late 1970s and early 1980s Barnegat, NJ was the principal tilefish port; more recently Montauk, NY has accounted for most of the landings. Most commercial landings are taken by the directed longline fishery.

The RTWG suggests that a simple scalar assumption of 3.9 mt based on the median estimate from (2014-2021) should be used for the total of all non-directed tilefish fleets (large and small mesh trawl, and gillnet fisheries). The median discards from 2014 to 2021 was estimated to be 2.3 mt in the directed longline tilefish fishery.

The RTWG developed a new recreational catch time series using vessel trip report data, large pelagic survey data, and other historical data available to develop a 1971-2022 time series of recreational catch. Recreational catches have ranged from a low of 3 mt for most years to 100 mt in 1974. More recently, for the last decade (2013-2022), recreational catches have ranged from 14 mt in 2016 to 23 mt in 2015. Based upon the newly developed recreational catch time series, the contribution of recreational golden tilefish landings to total removals for the 2005-2022 period ranged from 0.3% in 2006 to 3.7% in 2015. In 2022, contribution of recreational golden tilefish landings to total removals was 3.2%.

TOR #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.

A time series fishery-independent index of abundance does not exist for tilefish. Effort was considered directed for tilefish when at least 75% of the catch from a trip consisted of tilefish. Three different series of longline effort data were analyzed. The first series was developed by Turner (1986) who used a general linear modeling approach to standardize tilefish effort during 1973-1982 measured in kg per tub (0.9 km of groundline with a hook every 3.7 m) of longline obtained from logbooks of tilefish fishermen. Two additional LPUE series were calculated from the NEFSC weighout (1979-1993) and the VTR logbook data using days absent of the effort metric.

The NEFSC weighout and VTR LPUE series were standardized using a GLM incorporating year and individual vessel effects. Changes in the VTR LPUE can be generally explained with evidence of strong incoming year classes that track through the landings size composition over time (TOR 2). Since the SARC 58 assessment there appear to be increases in

LPUE due to one or two new strong year classes. In general, strong year classes appear to persist longer in the fishery after the FMP and after the constant quota management came into effect which is evident in both the LPUE and size composition data.

The 2024 RTWG developed a method of transitioning from a LPUE index based purely on logbook VTR data to LPUE based on the newly developed CAMS system since the VTR database at the NEFSC will no longer be supported. The CAMS system integrates data collected from dealers with VTRs, observers, electronic monitoring for both landings and discards on a trip by trip basis as a single catch source to be used for assessments and quota monitoring for all managed stocks. The CAMS system is being used for landings and discards in stock assessments starting in 2020. The RTWG developed the most comparable LPUE tilefish index possible within the CAMS system for the transition from the VTR series in 1994 to the CAMS full implementation in 2020. However, the CAMS system has been estimated back in time to 2000. Catch estimates for stocks assessments will likely not use CAMS until the year 2020 and forward into the future. The RTWG did consider linking the VTR and CAMS based LPUE index before 2020 and recommended transitioning the two data series in 2010.

For the 2024 RT assessment the WG also investigated whether other factors could help improve and perhaps better explain the LPUE trends. Reexamination of vessels effects, temporal factors (month), and crew size was examined. None of the available factors reexamined had a large influence on the underlying index. Limiting the index to the top 10 tilefish vessels also did not produce a meaningful difference. Very similar trends are seen in individual vessel LPUE series. The use of crew size also eliminated the data from 1991 to 1993 since that data was not available for that time period which is not desirable. The RTWG agreed to maintain the use of the original LPUE GLM incorporating individual vessel effects for the index.

Past benchmark tilefish assessments concluded that a simple days absent minus one day steam time (DA-1) was the best effort metric from vessel trip report (VTR) data due to data limitations mainly because the data is not collected on a haul by haul basis. Questions remain if landings per unit effort (LPUE) based on data collected at a finer haul basis could provide improvements or provide insights to LPUE indices as an index of biomass. Investigation of the longline study fleet data may help answer questions surrounding the somewhat crude effort metric in the LPUE index and could provide insight for future refinements. To help answer some of these questions the RTWG examined data from a single individual fishing quota (IFQ) tilefish vessel in the study fleet program who has been collecting tilefish catch data on a haul by haul basis since 2010. This analysis seems to support the use of days absent as an effort metric on a trip basis.

Because golden tilefish are poorly sampled by the northeast regions fishery-independent surveys, the assessment is relatively data poor, and additional data sources are vital to better understand trends in abundance. The directed fishery exclusively utilizes longline gear and information from this gear type is the primary source of information underpinning recent assessments. Interestingly, the species is also caught incidentally but with some frequency in trawl gear that is commonly used throughout the region. Despite this being common knowledge, there have been limited explorations of these data to see if they could be useful in understanding abundance patterns. The RTWG examined study fleet and observer data from trawl gear to develop a catch per unit effort (CPUE) index and compare this new index to existing indices from the tilefish assessment. The results suggest that there may be some value in using these data to understand the abundance of fish slightly smaller than those captured in the targeted fishery and the longline landing per unit effort (LPUE) index.

The RTWG estimated the stratified numbers per tow at length indices of relative abundance for the 2017 Tilefish Pilot Longline Survey and the 2020 Golden Tilefish Longline Survey using a standard stratified random mean approach. The 2017 pilot survey used three different offset circle hook sizes (small = 8/0, regular = 12/0, large = 14/0), distributed at a ratio of 20-60-20 and the 2020 survey used two different offset circle hook sizes (small = 8/0, regular = 12/0), distributed at a ratio of 50-50. The pilot survey indicated that small circle hooks (8/0) caught few large golden tilefish and more small individuals relative to regular circle hooks (12/0), and large circle hooks (14/0) caught few individuals overall. Given these findings, the 2020 survey was designed to determine if the small circle hooks (8/0) could provide additional information to a pre-recruit index relative to the regular circle hooks (12/0) as well as inform assessment model selectivity (i.e., domed shaped selectivity), therefore, the large hook (14/0) was dropped from the 2020 survey, as the catchability of large hooks greatly decreases. An adjustment was applied to the hook sizes for 2017 given the difference in the deployment of circle hook sizes between surveys and because of the differences in catchability between hook sizes.

The stratified numbers per haul show a decrease in the abundance index between 2017 and 2020 for both the combined hook indices and for the separate hook size indices. However, the longline stratified survey index at lengths suggests that a relatively large younger year class or perhaps two year classes were present during the 2017 survey (first two modes in the distribution between 35cm and 50cm) in comparison to the 2020 stratified numbers per haul at length index. Three years later in the 2020 survey it can be seen that the stratified numbers per haul between 50 cm and 70 cm is greater than the 2017 survey. This generally follows the expectation of the growth of golden tilefish for the strong year classes seen in the 2017 survey.

Both hook sizes have very similar length distributions but there is some indication that smaller hooks catch a greater amount of smaller, younger fish between 35 and 50 cm relative to regular hooks. The regular hooks appear to catch relatively more large fish greater than 50 cm given that the catchability of regular hooks is about half or that of small hooks. Additional surveys will likely be needed to determine if this data could potentially be used to inform the dome shaped selectivity in the assessment model. This pattern does seem to be consistent with a dome shape selectivity pattern in the fishery in the assessment model.

The survey also provides some indication that as fish age and increase in size they tend to be in deeper strata. However the vast majority of the fish caught in the survey was seen in the core fishing grounds. The combined effects of possible reduction in catchability with larger fish sizes and relatively lower availability of larger/old fish to the fishery remains difficult to quantify at this time.

TOR #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 RTWG goal for TOR 4 was to advance the assessment model from ASAP to the newly developed state-space modeling framework Woods Hole Assessment Model (WHAM). Due the sensitivity of the Tilefish modeling results to random effects the data inputs within TOR 4 were not changed from the last 2021 management track ASAP data input which had a terminal year of 2020. The RTWG goal was to examine model configuration effects in the new modeling framework WHAM without the additional effects of data changes. The RTWG suggests the best configuration to be used in the next management track assessment with the hope that incremental improvement and advancements could be made in future management track assessments as more data can be incorporated from TORs 1-3. A better understanding of random effects influence on model selectivity estimates and biological reference points (BRPs) with this relative data poor stock can then be advanced in future management track assessments once the assessment model is developed in WHAM in this RT assessment.

The RTWG first developed a bridge run which produced similar results to the 2021 ASAP model. The RTWG then investigated configuration changes to improve the model. In general the WHAM model results were similar to ASAP with similar estimates of the dome shaped selectivity in the second block and with the stock rebuilding to roughly SSB_{MSY} after the inception of management in 2001. The WHAM model diagnostics also appears to be acceptable with low retrospective error.

The RTWG developed a base model starting in 1976 using estimated starting numbers at age, self-weighting dirichlet missing 0 for fits to age composition data and shifting the selectivity block to 1976-1986 for the 1st block and 1978 to 2000 for the second block. WHAM model results were sensitive to adding random effects. Adding random effects to the base model NAA appears to allow for additional model flexibility which produces a relatively better fit to the data with improvements in the diagnostics. Most of the change occurs in fitting the 10+ age group while still producing good retrospective diagnostics. Adding numbers at age (NAA) random effects results in a relative flattening of the selectivity curve in the 2nd block, less cryptic biomass, less rebuilding since the inception of management in 2001 and a worse stock status relative to F40% based spawning potential ratio (SPRs) BRP proxies ($F/F_{40\%}$ and $SSB/SSB_{40\%}$ ratios).

Adding additional random effects on selectivity as well as survival continues to improve the relative model diagnostics. In general, it appears that adding additional random effects to the tilefish model seems to result in additional flexibility within the model allowing for further flattening of the selectivity curve which results in lower increase in biomass relative to an F40% based proxies and a relatively poorer stock status.

The RTWG was uncomfortable with the underlying sensitivity of the results even though the diagnostics improved when additional random effects were added. The results became more questionable with additional random effects added to the model given the history of the fishery and management. The perception from industry is that fishing has improved and that increases in biomass have occurred since management was implemented in 2001. The raw data also suggests general improvements in LPUE and size structure after management was put in place. Strong year classes have been entering the fishery relatively consistently every 5-7 years.

While the literature on state space model diagnostics is still developing, some studies have suggested that overfitting may be a concern when data density is relatively low. Liljestrand et al. (2023) demonstrated that low data density may reduce the ability to properly differentiate process and observation errors. Given the relatively low information content of the tilefish data, the RTWG decided to use a less complex model as the basis for continuing model development in the management track.

However, RTWG felt that the WHAM results among models suggests there is considerable uncertainty in the selectivity and stock status. A single model does not seem to capture the true uncertainty in the assessment. The RTWG did not have confidence in the results of the full random effects model as a basis for the assessment and stock status. The RTWG recommends to use the Base model without random effects until more confidence can be gained in future management track that suggests inclusion of some random effects are giving a more accurate depiction of the selectivity and true stock status. However, the RTWG feels that consideration of the random effects model is useful for showing the overall uncertainty and sensitivity of the results in the assessment. Assuming the base model is an accurate depiction of reality also does not account for the true uncertainty in this assessment.

TOR #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.

The RTWG did not recommend a change to the $F_{40\%}$ proxy for F_{MSY} biological reference points (BRP) since a stock-recruit relationship was also not evident in the WHAM base model. There was little difference between using a 10 year or a 5 year recent average for the estimates of the WHAM BRPs. The RTWG suggested using the 10 year average since there can be some variability in the mean weights at ages for the older ages. The recruitment used to estimate the $SSB_{40\%}$ within WHAM was based on the entire time series minus the most recent two years of data (1999 and 2000) since there is limited information to inform recruitment in the last two years of the model. The RTWG recommends the use of the base model configuration for stock status determination (TOR 4). Overfishing ($F/F_{40\%} = 0.55$) was not occurring and the stock was not overfished ($SSB/SSB_{40\%} = 1.29$) according to the base model.

TOR #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.

With the new RTWG base model the projections and biological reference points are integrated within the WHAM framework. The RTWG recommends the use of the base model for $F_{40\%}$ (F_{MSY} proxy) projection for the determination of overfishing limits (OFL) in the next management track assessment. Using the base model would also be consistent with stock status determination. However, the RTWG acknowledges that projections and estimated uncertainty of the base model likely does not capture the true uncertainty in the assessment since the results and status determination were found to be sensitive to changes in selectivity from the use of random effects.

Projections under $F_{40\%}$ show increases in catch in the short-term catch due to a relatively strong recruitment year classes at the end of the time series and because $F_{40\%}$ results in an increase in F within the projection ($F/F_{40\%} = 0.55$). The stock is also estimated to be above $SSB_{40\%}$ ($SSB/SSB_{40\%} = 1.29$) in 2020 for the base model. Therefore the projections become a Fishing down exercise to $SSB_{40\%}$ longer-term in the projections. In the short term, catches at $F_{40\%}$ are higher than the maximum sustainable yield (MSY) when the stock is at $SSB_{40\%}$ (855 mt). The projections for golden tilefish models are also more uncertain because there is limited information to inform recruitment in year t-1 and no information for the terminal year since no survey information for younger smaller fish is available to the assessment model.

TOR #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 RTWG reviewed the status of previous research recommendations and proposed new research ones to address issues raised during the working group meetings. Notable accomplishments relative to past research recommendations include: used survey data to develop a stratified index of relative abundance, examined effort metrics from one longline vessel participating in the study fleet program, variability in recruitment were further investigated using environmental covariates, developed a recreational landings time series, evaluate the reliability of the report of protogynous hermaphroditism in the S. Atlantic stock.

The RTWG proposed new research recommendations that should improve assessing the population through the current or futile models. These include the following: collection of length samples on party/charter trips for potential improvements in recreational time series estimates and evaluate WHAM performance for information poor stocks using simulated tilefish like populations (i.e., only catch data). Do random effects in both survival and selectivity introduce bias?

TOR #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.

Several approaches were considered as potential contingency plans if the proposed assessment model is deemed inappropriate for providing management advice, either as a conclusion of research track peer review or subsequently in the management track process. Many northeast US assessments specify an empirical backup approach based on survey data, either swept-area estimates of stock biomass and a target exploitation rate or survey biomass trends and recent catch. However, due to the current lack of survey data for golden tilefish these approaches are not good options for this stock. The RTWG briefly discussed the use of other data-limited approaches for estimating sustainable yield such as Depletion-Corrected Average Catch (DCAC) and Depletion-Based Stock Reduction Analysis (DB-SRA); however, the RTWG did not pursue these because they heavily rely on assumptions needed to run models and/or they lead to severe retrospective errors in statistical catch-at-age models. In addition, these data-limited methods have been found not to outperform a retrospectively adjusted catch-at-age model over the long-term.

The RTWG recommends that if the proposed assessment approach (WHAM Base model without random effects) does not meet the standards of peer review or is rejected in a future management track assessment, an alternative model be developed to integrate information from catch, age composition and potentially indices (e.g., alternative WHAM configurations).

In addition, the RTWG also proposed an alternative “Plan C” based on historical fishery performance under constant quota strategies. Under Plan C, if modeling fails, management would be based on a commonsense constant catch approach considering the management history since 2001 and response in CPUE and size distribution of fish landed. For example, a constant catch approach using a quota within the range of those implemented in the fishery since 2001 (738 – 905 mt) could be considered when determining an appropriate constant catch if the model fails. Alternatively, using an average of the actual catches (10 year 2013-2022 average catch of 690 mt or 20 year 2021-2022 average catch of 790 mt) may be more justified for the determination of a constant quota catch advice since this is the actual catch that appeared to have a positive effect on recruitment and seemed to allow for strong year classes to persist while supporting the fishery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Golden tilefish, *Lopholatilus chamaeleonticeps*, inhabit the outer continental shelf from Nova Scotia to South America, and are relatively abundant in the Southern New England to Mid-Atlantic region at depths of 80 to 440 m. Tilefish have a narrow temperature preference of 9 to 14 C. Their temperature preference limits their range to a narrow band along the upper slope of the continental shelf where temperatures vary by only a few degrees over the year. They are generally found in and around submarine canyons where they occupy burrows in the sedimentary substrate. Tilefish are relatively slow growing and long-lived, with a maximum observed age of 46 years and a maximum length of 110 cm for females and 39 years and 112 cm for males (Turner 1986). At lengths exceeding 70 cm, the predorsal adipose flap, characteristic of this species, is larger in males and can be used to distinguish the sexes. Tilefish of both sexes are mature at ages between 5 and 7 years (Grimes et al. 1988).

Golden Tilefish was first assessed at SARC 16 in 1992 (NEFSC 1993). The Stock Assessment Review Committee (SARC) accepted a non-equilibrium surplus production model (ASPIC). The ASPIC model estimated biomass-based fishing mortality (F) in 1992 to be 3-times higher than F_{MSY}, and the 1992 total stock biomass to be about 40% of B_{MSY}. The intrinsic rate of increase (r) was estimated at 0.22.

The Science and Statistical Committee reviewed an updated tilefish assessment in 1999. Total biomass in 1998 was estimated to be 2,936 mt, which was 35% of B_{MSY} = 8,448 mt. Fishing mortality was estimated to be 0.45 in 1998, which was about 2-times higher than F_{MSY} = 0.22. The intrinsic rate of increase (r) was estimated to be 0.45. These results were used in the development of the Tilefish Fishery Management Plan (Mid-Atlantic Fishery Management Council 2000). The MAFMC implemented the Tilefish Fishery Management Plan (FMP) in November of 2001. Rebuilding of the tilefish stock to B_{MSY} was based on a ten-year constant harvest quota of 905 mt.

SARC 41 reviewed a benchmark tilefish assessment in 2005. The surplus production model indicated that the tilefish stock biomass in 2005 has improved since the assessment in 1999. Total biomass in 2005 is estimated to be 72% of B_{MSY} and fishing mortality in 2004 is estimated to be 87% of F_{MSY}. Biological reference points did not change greatly from the 1999 assessment. B_{MSY} is estimated to be 9,384 mt and F_{MSY} is estimated to be 0.21. The SARC concluded that the projections are too uncertain to form the basis for evaluating likely biomass recovery schedules relative to B_{MSY}. The total allowable landings (TAL) and reference points were not changed based on the SARC 41 assessment.

Stock status from SARC 48 (2009) was also based on the ASPIC surplus production model which was the basis of the stock assessment for the last three assessments. The model is calibrated with CPUE series, as there are no fishery-independent sources of information on trends in population abundance. While the working group expressed concern about the lack of fit of the model to the VTR CPUE index at the end of the time series, they agreed to accept the estimates of current fishing mortality and biomass and associated reference points. The instability of model results in the scenario projections was also a source of concern. It was noted that the bootstrap uncertainty estimates do not capture the true uncertainty in the assessment. The

ASPIC model indicates that the stock is rebuilt. However, the working group acknowledges that there is high uncertainty on whether the stock is truly rebuilt.

The golden tilefish stock was last assessed at SARC 58 in 2014 with a terminal year of 2012 (https://repository.library.noaa.gov/view/noaa/4719). The golden tilefish stock was not overfished and overfishing was not occurring in 2012 relative to the SARC 58 accepted biological reference points. The stock was declared rebuilt in 2014 by NMFS based on SARC 58 results which indicated that spawning stock biomass (SSB) was at 101% of the accepted SSB_{MSY}. A new model, ASAP, was used in this assessment to incorporate newly available length and age data through the use of pooled age-length key. The ASAP model integrates more realistic life history information on size and growth into a single model framework and better characterizes the population dynamics of the tilefish stock.

A golden tilefish model update was done in 2017 with updated commercial fishery landings, landings size distributions, and CPUE indices of biomass through 2016. The golden tilefish stock was not overfished and overfishing was not occurring in 2016 relative to the newly updated biological reference points.

The last Management Track golden tilefish ASAP model update was done in 2021 with updated commercial fishery landings, landings size distributions, and CPUE indices of biomass through 2020. This assessment began the use of year specific age length keys in the final model configuration. Actual year specific keys were used for 2007, 2009 to 2012, and 2014 to 2020 since improvements in age data become available with efforts made towards production aging for golden tilefish. The golden tilefish stock was not overfished and overfishing was not occurring in 2020 relative to the newly updated biological reference points (https://repository.library.noaa.gov/view/noaa/39406).

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)

TOR 1 was addressed in this research track assessment using an Ecosystem and Socioeconomic Profile (ESP) (Shotwell et al. 2023). The ESP framework began with a comprehensive literature review of ecosystem and socioeconomic information relevant to the stock. Environmental linkages found in the literature are then compiled into a conceptual model outlining key drivers for each life history stage. Ecosystem and socioeconomic indicators relevant to stock performance are proposed, selected, and analyzed relative to stock metrics. Socioeconomic indicators for golden tilefish were not analyzed at this time; however, socioeconomic information from conversations with golden tilefish industry members and details from the 2023 Golden Tilefish Fishery Performance Report (MAFMC 2023) are included in this ESP. Lastly, results are aggregated and reported as advice to inform the stock assessment process. An ESP was used for the 2024 golden tilefish research track assessment in order to gather baseline data and set up a process that can be iterated on by future working groups. Working group members in future assessments will be able to update the indicators presented in this version of TOR1 for inclusion in the golden tilefish stock assessment model.

The Northeast U.S. shelf (NES) golden tilefish ESP includes a detailed literature review of golden tilefish habitat and distribution, size and growth, and ecological linkages for each life history stage. A conceptual model of golden tilefish life history developed from the literature review (Figure 1) was used to identify key ecosystem components of habitat condition, physical oceanography, and food availability. Golden tilefish incidental catch per unit effort (CPUE) from undirected trawl trips from the Northeast Fisheries Science Center's (NEFSC) Study Fleet and Observer programs along with age-1 recruitment data from the previous golden tilefish stock assessment (Nitschke 2021) were analyzed relative to the ecosystem indicators using linear regressions and generalized additive models (GAM) to determine the effects of ecosystem indicators on golden tilefish catch and recruitment. The NEFSC shelf-wide ichthyoplankton and hydrographic survey data (1973-2023) was analyzed to better understand golden tilefish larval geographical distribution and environmental preferences. Relevant results are summarized below and a detailed report is presented in Working Paper 1 (Salois et al. 2024).

1.2 Larval Analysis

The NEFSC ichthyoplankton dataset was used to assess the distribution and pelagic habitat preferences of tilefish larvae. The number of larvae caught has been fairly consistent over time and the locations where larvae were collected were often associated with warmer and saltier waters. Tilefish larvae were most abundant in the summer months (July, August, September), which aligns with the timing of peak spawning months of June and July for this species (Grimes et al. 1988), and rarely-to-never caught in the winter months (January, February, March). Larvae were associated with bottom temperatures between 7-16°C and bottom salinities between 31.47-36.3 psu, which is consistent with observations documented in the literature (Table 1). The newly hatched stage of larvae represented 25% of the sample (n = 22), were only observed in the summer and fall, and occupied a warmer, narrower band of the temperature range (~11-18°C). Older pelagic larvae accounted for 73% of the sample (n = 66), were associated with the widest range of temperatures (7-23°C), and represented the greatest proportion of larvae collected in both the spring and fall (see Salois et al. 2024; Figure 5). For the full analysis see Working Paper 1 (Salois et al. 2024).

1.3 Indicator Analysis

1.3.1 Ecosystem Indicators

1.3.1.1 Habitat condition

Habitat condition was identified as a principal factor for indicator selection on the advice of the RTWG members, who suggested bottom temperature, bottom salinity, and sediment as important variables. Final habitat condition indicators include bottom temperature, salinity at depth (78 m, 92 m, and bottom), cold pool indices, and sediment grain size. All hydrographic data were parsed at spatial and temporal scales relevant to golden tilefish distributions (see Salois et al. 2024 for more details). Adult golden tilefish occupy a narrow band of bottom temperatures (9-14°C) and bottom salinities (33-36 psu) while larvae and early juveniles can tolerate slightly lower temperatures and salinities (Grimes and Turner 1999; Steimle et al. 1999). Bottom temperatures and salinities outside these ranges may impact spawning and recruitment success (Fisher et al. 2014; Grimes et al. 1988). Golden tilefish prefer habitat with small grain size and high malleability (i.e., clay or sand) to construct burrows (Wenner and Barans 2001).

The cold pool index, spatial extent and persistence indices describe the strength, area and duration of the seasonal Mid-Atlantic cold pool (du Pontavice et al. 2022; Chen et al. 2018; Chen and Curchitser 2020). The lower bound of temperature preference for eggs and larvae and juveniles is between 8-13°C (Steimle et al. 1999), thus we suspect years with a larger and more persistent cold pool may not be optimal for early life history stages and recruitment success.

1.3.1.2 Physical oceanography

Physical oceanography indicators included sea surface temperature, shelf water volume, and Gulf Stream index. Monthly sea surface temperatures were spatially cropped to the golden tilefish strata and used as a proxy for water mass movement, which may cause displacement or mortality of eggs and larvae. Shelf water volume is the volume of water inshore of the shelf-slope front, a narrow transition region between masses of cool, low salinity shelf water and warm, high salinity Slope Sea water (Linder and Gawarkiewicz 1998). The position of the shelf-slope front and an increase of shelf water volume brings cold shelf water onto the shelf slope, which may result in mortality (Freeman and Turner 1977). Shelf water temperature and salinity were also included in the analysis. Lastly, the Gulf Stream index (GSI), a measure of the Gulf Stream position relative to the mean position (Pérez-Hernández and Joyce 2014), was used as an indicator to provide context of water mass changes offshore to golden tilefish habitat. A more northerly position of the Gulf Stream (i.e., a positive Gulf Stream Index) is associated with movement of warm water onto the shelf and may create more suitable habitat for golden tilefish.

1.3.1.3 Food availability

Microplankton, the largest phytoplankton size class (> 20 µm), and total chlorophyll-*a* were used as proxies for food availability for golden tilefish. Microplankton are prey for zooplankton and may act as a proxy for zooplankton abundance, a common prey of larval tilefish (Steimle et al. 1999). High microplankton and total chlorophyll-*a* concentrations may indicate periods of high primary productivity that can support strong recruitment year classes.

1.3.2 Socioeconomic Input

Personal conversations with golden tilefish captains, commercial vessel owners, fishers and dealers were conducted to initialize efforts of socioeconomic indicator development. We discussed changes in abundance and distribution of golden tilefish, fishing behavior, and environmental and socioeconomic drivers of golden tilefish.

More than half of the industry members we spoke to reported recent changes in distribution or abundance of golden tilefish, with one reporting that distributions have not changed but densities have shifted. There were multiple reports that tilefish are less available in the east around Atlantis and Veatch Canyons and more concentrated further west near Hudson Canyon in recent years. Industry members had mixed responses as to whether or not golden tilefish congregate in a particular season - some industry members recounted high concentrations of tilefish in the early spring during the 1980s with a shift to late winter months in recent years.

The impacts of weather and climate variability on fishing ability and availability came up in multiple conversations. While stormy conditions have existed for many years, recent changes to storm patterns have made fishability difficult to predict in winter months (MAFMC 2023; Salois et al. 2024; personal communications with industry members, 2024). Consensus was that very cold bottom temperatures potentially push golden tilefish off of the shelf edge and that severe weather conditions in recent winters (specifically 2013-2019) significantly affected tilefish operations.

A key detail in our conversations is the consensus that interannual variability of golden tilefish landings is largely driven by factors outside of environmental variables, including the implementation of the Individual Fishing Quota (IFQ) quota system, participation in other profitable fisheries, increased fuel costs, and economics (i.e., market prices). The cooperative nature of the fishery allows fishers to have some degree of control over the supply, but quotas have reduced the amount of fishing each vessel can do, limiting some vessels to part-time versus year-round fishing.

1.4 Golden Tilefish Datasets

1.4.1 Study fleet/Observer Incidental Trawl CPUE

Two high-resolution fishery-dependent datasets containing catch and effort data from the NEFSC's Study Fleet and Observer programs (1998-2022), were combined using a guild approach to generate haul-level catch per unit effort (CPUE) (NEFSC 2016; Drew 2022; Cheng et al. 2023; Hoyle et al. 2024). This combined dataset captures both incidental golden tilefish catch as well as catch data for a suite of other commonly occurring species from trawl fisheries (Jones and Salois, 2024). A new CPUE index was generated from these undirected trips and expanded using species associations, in order to introduce plausible zeros and reduce bias

(NEFSC 2016; Dettloff 2021; Drew 2022; Jones and Salois 2024). This dataset yielded ~4,600 total catches of golden tilefish and indexes their age around 4 years (Jones and Salois 2024; see TOR3 for more details).

1.4.2 Recruitment Estimate

We used model-derived age-1 recruitment estimates from the 2021 Golden Tilefish Management Track Assessment (Nitschke 2021) Age Structured Assessment Program (ASAP) model (Legault and Restrepo 1998). See Working Paper 1 (Salois et al. 2024) for the full description.

1.4.3 EcoMon Larval Data

Larval and coincident environmental data were acquired from the NEFSC Marine Resources Monitoring, Assessment and Prediction program (MARMAP) (1977-1987) and the NEFSC Ecosystem Monitoring program (EcoMon) (1992-2023) datasets.

1.5 Ecosystem Influences

1.5.1 Ecosystem Influences on Golden Tilefish Abundance

Linear regressions with Pearson's correlation analysis were used to determine relationships between the ecosystem indicators and the incidental CPUE index (derived from the Study Fleet and Observer program data) at time of catch (age 4). For these analyses, each indicator was also lagged by 3 years to examine potential environmental influences for recruitment aged (Age-1) fish.

1.5.1.1 Indicators correlated with CPUE index (Age 3-4, time of catch)

Five ecosystem indicators were significantly correlated with the incidental CPUE index at time of catch (age 4) (Table 1). Catch declined with increases in sea surface temperature, shelf water temperature, and a positive (more northerly) Gulf Stream Index. Increases in CPUE were significantly correlated with increases in microplankton abundance and chlorophyll-*a*. Indicators of bottom temperature and salinity revealed no significant linear trends with CPUE, but catches are clustered in a narrow band between 10-14℃ and 34-36 psu.

Sea surface temperatures can be used as a proxy for water mass movement, which may cause displacement or mortality of eggs and larvae. A more northerly position of the Gulf Stream (positive Gulf Stream Index) pushing warm water onto the shelf slope may increase the spatial range of suitable temperatures for golden tilefish during the winter months. Significant trends between CPUE and indices of food availability highlight the potential value of these data products for future assessment models.

1.5.1.2 Indicators correlated with CPUE index (Age 0-1, as a recruitment proxy)

Eight ecosystem indicators were significantly correlated with golden tilefish CPUE at a lag of 3 years (Age-1) (Table 1). CPUE declined with increasing sea surface temperature, shelf

water volume, and cold pool extent and persistence. Indicators of bottom temperature and salinity revealed no significant linear trends with CPUE, however catches are clustered in a narrow band of temperatures between 10-14℃ and salinities between 34-36 psu.
Golden tilefish CPUE increased with increasing shelf water temperature and salinity, and microplankton during the fall. CPUE increased with a positive (more northerly) position of the Gulf Stream during the winter, but decreased during the summer. These seemingly opposing results may be due to the large spatial extent and seasonal variability of the Gulf Stream. One potential hypothesis of an underlying process driving these trends might be that a more northerly Gulf Stream in the winter may warm waters inhabited by golden tilefish, creating more suitable habitat. Conversely, a northerly position of the Gulf Stream during the summer, however, may result in temperatures exceeding the upper thermal tolerance of golden tilefish.

The observed declines in incidental CPUE with increases in shelf water volume and lower shelf water temperature on the continental shelf edge coincide with hypotheses made after a mass mortality event of golden tilefish in the late 1800s (Freeman and Turner 1977). A positive relationship between microplankton in the fall and CPUE at age 1 indicates the importance of primary productivity during recruitment and may play a role in large recruitment “pulses” observed every 6-7 years. CPUE at age-1 declines when the cold pool is larger and persists longer throughout the year, which may have implications for temperature preferences during early recruitment stages.

1.5.2 Generalized Additive Model (GAM)

Generalized additive model (GAM) results (see Salois et al. 2024; Figure 13) highlighted a seasonal shift in golden tilefish catch from summer to winter/spring, in recent years, with a consistent hotspot of catch concentrated near Hudson Canyon, supporting observations by industry members. Catch increased when bottom temperatures were between 10-14°C and salinity between 34-36 psu, mirroring the regression analysis results. CPUE values increased with lower shelf water volume at time of recruitment (lag 3 years). Catch is variable across the range of shelf water temperatures (10-15°C), with a peak between 11-12°C. There is a unimodal relationship between catch and increased microplankton abundance at time of recruitment. CPUE declined with increased sediment grain size, indicating higher golden tilefish catch in habitats with finer sediment (e.g., mud, very fine sand). The relationship between golden tilefish CPUE and the position of the Gulf Stream differs between time of catch and recruitment (3 year lag). Tilefish are more likely to be caught when the Gulf Stream is in a more northerly position at the time of catch. Conversely, when lagged three years (concurrent with the time of recruitment), a southerly position of the Gulf Stream is associated with higher catch. Though the model was only fed incidental catch data, it was able to capture environmental signals identified as important by both previous studies on golden tilefish as well as industry members, highlighting the value of this trawl-based catch index for the tilefish fishery.

1.5.3 Ecosystem Influences on Golden Tilefish Recruitment

Linear regression analyses with Pearson’s correlation analysis were used to determine relationships between the ecosystem indicators and the modeled recruitment index for age-1 fish.

Each indicator was also lagged by one year to examine potential environmental influences on larval golden tilefish (age 0).

1.5.3.1 Indicators correlated with recruitment index (Age-1)

Recruitment index values were concentrated in a range of bottom temperatures (9-12°C) and salinities (34-36 psu), consistent with what has been documented in the literature (Steimle et al. 1999). Three ecosystem indicators had significant negative correlations with age 1 recruitment: shelf water volume, cold pool spatial extent, and cold pool persistence. Both the cold pool persistence index and spatial extent index were negatively correlated with recruitment at Age-1, as was the mean shelf water volume in the spring. However, there were no discernible trends in shelf water temperature or salinity at Age-1. Further, there were no significant trends between the Gulf Stream Index, chlorophyll-*a*, or microplankton abundance and recruitment index.

1.5.3.2 Indicators correlated with recruitment index (Age 0, time of spawning/hatching)

Three ecosystem indicators were significantly correlated with a recruitment, when lagged by 3 years to approximate conditions at the time of spawning and birth. There was an overall decrease in recruitment with increasing salinity (at 78m and 92m) in the spring and winter, with no discernible trend in the fall and summer. Higher recruitment values were correlated with increased shelf water volume in the spring and increased microplankton abundance in the fall. Recruitment was negatively correlated with higher cold pool spatial extent and persistence, although these trends were not significant.

The significant positive relationship between microplankton in the fall following birth may suggest that microplankton abundance may play a key role in larval survival. Recruitment shows significant, moderate negative correlations with salinity in the winter and with weaker non-significant negative correlations in the spring. This combined with the significant positive correlation with higher shelf water volume may have implications about favorable oceanographic conditions during seasons in which tilefish eggs are spawned. For instance, a cooler, less saline pelagic environment could be a signal of less dynamic shelf conditions, as a higher shelf water volume indicates that the shelf break front is located at the shelf edge and that the shelf is not occupied by slope waters.

1.5.4 Indicator Agreement

Multiple environmental indicators showed consistent trends across datasets and life history stages. The incidental CPUE index (derived from Study Fleet and Observer data sets) declined with increases in sea surface temperature at both time of catch (around age 4) and with a 3 year lag (near age-1 recruitment), however this trend did not match trends between SST and the recruitment index. Bottom temperature and bottom salinity indicators, while exhibiting no clear linear trends, coincided with habitat descriptions across life history with both the incidental CPUE index (derived from Study Fleet and Observer data sets) and recruitment index. Age-1 fish from both the new CPUE index and recruitment index were negatively correlated with shelf water volume, cold pool spatial extent, and cold pool persistence. Golden tilefish were positively correlated with microplankton abundance across three different developmental stages: ages 3-4 (CPUE index: no lag), ages 0-1 (CPUE index: 3 year lag), and as larvae (recruitment index: 1 year lag). Furthermore, the correlation between microplankton abundance in the fall and fish of larval ages across both data sets increases our confidence that this may be a useful indicator of food availability, with potential implications for understanding drivers of growth and development for tilefish early life stages.

1.6 Uncertainty

A major goal of ESPs is to reduce uncertainty by exploring a suite of ecologically relevant ecosystem indicators that could help explain some of the variability in stock dynamics, thus allowing the stock assessment and scientific advice process to more accurately capture ecosystem impacts on the stock. However, there are also several elements of uncertainty inherent in both the tilefish data and ecosystem indicators.

The tilefish abundance data came from fisheries-dependent sources, namely the observer program and the study fleet, which often target 10-15% coverage (Bell et al. 2017, Jones et al. 2020) The observer program attempts to select a random stratified sample (Palmer et al. 2016). Conversely, the fishery-dependent Study Fleet does not follow a statistical design, but rather tracks fishing movement and behavior, which introduces some bias (Cheng et al. 2023; Hoyle et al. 2024). . In an effort to reduce some of the inherent bias and account for the small sample size of the Study Fleet, we focused on undirected catch and determined plausible instances of zero catch using a guild approach (NEFSC 2016; Drew 2022). This approach reduces biases related to fishing behavior but also introduces uncertainty around catch locations. The combined study fleet and observer CPUE index does not represent a random sample, such that data points are not independent and identically distributed. This does not negate the utility of simple correlations and GAMs but does limit the scope of interpretation and conclusions that are drawn from these results.

The recruitment time series is a model-derived product using deviations from the long-term mean and is not based on a stock recruit relationship (Brooks 2024, Nitschke 2021, Miller & Legault 2015). As a model-derived product, recruitment estimates inherit both estimation uncertainty as well as model uncertainty, however as a data-limited stock assessment, this is the only quantification of recruitment that exists. To avoid spurious results, we interpret the preliminary relationships identified here cautiously and acknowledge the uncertainty of the assessment results and the structural assumptions of the model (Brooks & Deroba 2015). Despite the drawbacks and uncertainty in using model derived estimates as a dependent variable, we elected to explore relationships between the timeseries of annual recruitment and environmental indicators. This is consistent with methods used in other studies that evaluate the impacts of environmental drivers on recruitment (Haltuch et al. 2019; Shotwell et al. 2023). Even though environmental covariates were not directly included into the stock assessment model, it is still possible to identify key biotic or abiotic drivers that can inform recruitment estimates qualitatively (Sharma et al. 2019) as well as inform fisheries management, e.g., through a risk table approach (Dorn and Zador 2020).

The bottom temperature and salinity at depth used in the GAM analysis are modeled products (GLORYS12 reanalysis and operational models) with inherent uncertainty. Remotely sensed satellite data products (e.g., sea surface temperature, chlorophyll-a and microplankton abundance) also have inherent uncertainty. This inherent uncertainty in the data is difficult to capture when creating spatially and temporally averaged indicators. In addition, the high temporal and spatial variability of the oceanographic variables increases the uncertainty in the averaged indicators. While environmental data can be highly variable, the remote sensing and in-situ derived products are widely used and commonly accepted for these types of analyses (Phillips et al. 2014, Tommasi et al. 2017) .

1.7 Incorporating Findings into Impacted TORs

TOR 3 (survey): We used the data product (study fleet and observer trawl CPUE) developed in TOR 3 and Working Paper 6 (Jones and Salois 2024) to explore environmental influences on interannual variability in the CPUE index from the trawl fishery. This index provides precise locations of catch and may select for smaller tilefish (Jones and Salois 2024), allowing for analyses to explore insights on habitat use at a finer scale. The work described here provides a unique opportunity to confirm previous insights as well as to explore ecosystem influences on recruitment and fishery-dependent CPUE.

TOR 7 (research recommendations): The ecosystem information compiled for TOR 1 was used to generate a suite of research recommendations under TOR 7. Most notably, we support the need for alternative data streams to better capture the geographic and size distribution of golden tilefish in order to further develop the environmental and socioeconomic indicators highlighted in TOR 1 for testing in future Woods Hole Assessment Model (WHAM) model updates.

2 CATCH

TOR #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.

2.1 Commercial Catch Data

Total commercial golden tilefish landings (live weight) increased from less than 125 mt during 1967-1972 to more than 3,900 mt in 1979 during the development of the directed longline fishery (Figure 2). Landings prior to the mid-1960s were landed as a bycatch in the trawl fishery. Annual landings ranged between 454 and 1,838 mt from 1988 to 1998. Landings from 1999 to 2002 were below 900 mt (ranging from 506 to 874 mt). An annual quota of 905 mt was implemented in November of 2001. Landings in 2003 and 2004 were slightly above the quota at 1,130 mt and 1,215 mt, respectively. Landings from 2005 to 2009 were at or below the quota, while landings in 2010 at 922 mt were slightly above the quota (Figure 2). Since 2010 landings have been below the quota and decreased to an estimated 494 mt in 2016. The landings have increased slightly to an average of 695 mt from 2017 to 2022. The Total Allowable Landings

(TAL) was reduced for the first time in 2015 to 796 mt from the TAL of 905 mt which was in place from 2001-2014. The TAL in 2016 and 2017 was increased to 856 mt based on projections from the SARC 58 assessment. The TAL was then reduced to 738 mt from 2018 to 2021 based on the 2017 operational assessment and subsequently increased based on the 2021 management track assessment (Figure 2). The top 4 permits hold 80% of the golden tilefish IFQ allocation.

Over 75% of the landings came from Statistical Areas 537 and 616 since 1991. In the 1980s a greater proportion of the landings came from 526 (See Working Paper 2 (Nitschke 2024) for additional details). Before 2010, over 85% of the commercial landings of tilefish in the MA-SNE region have been taken in the longline fishery. Since 2010 the percent of the landing coming from longline gear has increased to over 95%. During the development of the directed longline fishery in the late 1970s and early 1980s Barnegat, NJ was the principal tilefish port; more recently Montauk, NY has accounted for most of the landings. In the late 1970s and early 1980s a greater proportion of the landings were taken in the first half of the year. More recently, landings have been relatively evenly distributed throughout the year. See Working Paper 2 (Nitschke 2024) for additional details.

2.1.1 Commercial Market Category and Size Composition Data

Seven market categories exist in the database. From smallest to largest they are: extra small, small, kitten, medium, large/medium, large and extra-large as well as an unclassified category (See Working Paper 2 (Nitschke 2024) for additional details). Evidence of several strong recruitment events can be seen tracking through the market category proportions. More recently, the 2017 model update predicted a strong 2013 year class which began to enter the fishery in 2016. The 2018 data update did show increases in CPUE as the strong year class became more selected by the fishery in 2017. There is also evidence for the 2013 and 2014 year classes with the tracking of the length model in the landings at length. The 2021 management track model update indicates that the 2013 and 2014 year classes were above average.

The RTWG investigation of kept observer length frequencies in 2021 and 2022 suggested a very similar distribution relative to the expanded port sampling landings at length (See Figure 13 from Working Paper 2 (Nitschke 2024)). This could be potentially used as a backup data source if sampling declines continue in the biological port sampling program and expanded length distributions are no longer a good reflection of market category size structure across the year. If temporal market category sampling can no longer characterize the size of a particular market size within a year then the catch at length and therefore catch at age could become biased. However, growth information (age-length-keys) is coming from the biological port sampling program. Observer length samples cannot be used to supplement the port sampling after the market category cull.

2.1.2 Commercial Discard Data

Past tilefish assessments concluded that discards were insignificant and discards were not included as a component of the total catch in the modeling from the limited data that was available at the time.

The observer coverage has improved recently (2014-2021) but still suggests that discards are a relatively minor component of the removals. CVs on discard estimates remain high (Table 2). The recently developed CAMS (Catch Accounting Monitoring System) system will be used for estimating discards of all managed stocks in all fisheries to be used for both monitoring and stock assessments on a trip by trip basis starting in 2020 (See Working Paper 2 (Nitschke 2024) for additional details). The RTWG suggests that a simple scalar assumption of 3.9 mt based on the median estimate from (2014-2021) should be used for the total of all non-directed tilefish fleets (large and small mesh trawl, and gillnet fisheries) since it is likely that some minimal discards occurred in the past. A simple scalar was used instead of hindcasting of the estimates since total effort changes on tilefish grounds for other non-directed fisheries (large and small mesh trawl and gillnet) were not well understood. In addition, the use of tilefish for scaling non-directed fisheries is problematic due to the lack of a golden tilefish species code prior to the tilefish FMP in 2001. For the directed longline fishery, the RTWG suggests using the median discard/kept tilefish ratio from 2014-2021 (0.003386) to hindcast the discard estimates. The median discards from 2014 to 2021 was estimated to be 2.3 mt in the directed longline tilefish fishery. The hindcast estimates a maximum estimate of 12.7 mt with the peak in tilefish landings in 1980 of 3,889 mt. See Working Paper 2 (Nitschke 2024) for additional details.

2.2 Recreational Catch Data

A small recreational fishery occurred briefly in the mid-1970s (< 100 mt annually, Turner 1986) but subsequent recreational catches appear to have been low for the 1981-2022 period, ranging from zero for most years to approximately 200,000 fish in 2010 in the Marine Recreational Information Program (MRIP) (See Working Paper 3 (Montañez et al. 2023) for additional details). The tilefish catch in the MRIP survey is likely below detection levels of the survey judging from the sporadic estimates in the survey and the large Percent Standard Errors associated with reported catches. However there are several party and charter vessels which make a few targeted tilefish trips a year. Party and charter boat vessel trip reports also show relatively low numbers of tilefish being caught although there is an increase in numbers of fish reported towards the end of the 1994-2022 time series. However some of the increase may be more a reflection of recent increases in reporting rate. Most of the reported party and charter boat landings are coming from New Jersey. It appears that a greater proportion of the reported recreational party and charter catch and effort is further south in statistical area 622 relative to the commercial longline fleet that fishes more in 537. Lastly, golden tilefish discards in the recreational fishery appear to be a very minor component of the total removals. See Working Paper 3 (Montañez et al. 2023) for additional details.

The RTWG also reviewed private tilefish recreational data that has been collected since late 2000 (See Working Paper 3 (Montañez et al. 2023) for additional details). Since the new private reporting requirements were implemented, private catch has ranged from 64 fish in 2020 to 298 fish in 2022. Some stakeholders have indicated that the reported private tilefish catch appears to be too low given their observations while on the water. NMFS's GARFO and the MAFMC continue to conduct outreach efforts to ensure that private anglers are aware of the recently implemented permitting and reporting requirements for this fishery. The RTWG also reviewed tilefish landings data collected in the large pelagic survey (LPS). While the LPS was designed as a specialized survey that would focus specifically on the recreational fishery directed at large pelagic species, it also collects information on the quantity of non-LPS species kept (e.g., Atlantic bluefish, king mackerel, black sea bass, spiny dogfish, ocean triggerfish, golden and blueline tilefish) on trips targeting large pelagic species. LPS data estimated that 15,282 and 20,177 golden tilefish have been kept by the charter mode and private mode, respectively for the 2005-200 period combined.

In prior assessments, golden tilefish recreational catches were not included in the modeling as stock assessment working groups were not able to develop a reliable time series for recreational catch. In SAW58th (NEFSC 2014) the working group also concluded that recreational removals were likely a minor component of the catch. The 2024 RTWG believes that recently implemented reporting requirements, improvements in the specialize LPS, and other historical recreational data can now be used to develop a golden tilefish time series for recreational landings which should be considered for inclusion in stock assessment work to better characterized removals in the fishery. The RTWG used recreational landings data developed by Turner (1986), party and charter VTR data, and LPS private mode estimates of tilefish kept to develop a time series of golden tilefish recreational catches. See Working Paper 3 (Montañez et al. 2023) for methodology and assumptions used to develop the time series of golden tilefish recreational catches. Recreational catches have ranged from a low of 3 mt for most years to 100 mt in 1974. More recently, for the last decade (2013-2022), recreational catches have ranged from 14 mt in 2016 to 23 mt in 2015 (Table 3). Based upon the recreational catch time series in Table 3, the contribution of recreational golden tilefish landings to total removals for the 2005-2022 period ranged from 0.3% in 2006 to 3.7% in 2015. In 2022, contribution of recreational golden tilefish landings to total removals was 3.2%.

3 SURVEY DATA

> *TOR #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.*

3.1 Commercial LPUE data

A time series fishery-independent index of abundance does not exist for tilefish. The low catchability in the NEFSC bottom trawl surveys does not produce a reliable index of abundance (Working Paper 14 (Nitschke 2024)). Analyses of catch (landings) and effort data were confined to the longline fishery since directed tilefish effort occurs in this fishery (e.g., the remainder of tilefish landings are taken as bycatch in the trawl fishery). Most longline trips that catch tilefish fall into two categories: (a) trips in which tilefish comprise greater than 90% of the trip catch by weight and (b) trips in which tilefish accounted for less than 10% of the catch. Effort was considered directed for tilefish when at least 75% of the catch from a trip consisted of tilefish.

Three different series of longline effort data were analyzed. The first series was developed by Turner (1986) who used a general linear modeling approach to standardize tilefish effort during 1973-1982 measured in kg per tub (0.9 km of groundline with a hook every 3.7 m) of longline obtained from logbooks of tilefish fishermen. Two additional LPUE series were calculated from the NEFSC weighout (1979-1993) and the VTR logbook data. Effort from the weighout data was derived by port agents' interviews with vessel captains whereas effort from the VTR systems comes directly from mandatory logbook data. In the SARC 58 assessment (2014) and in the 2009, 2005 and 1998 tilefish assessments, Days Absent was used as the best available effort metric. In the 1998 assessment an effort metric based on Days Fished (average hours fished per set / 24 * x number of sets in trip) was not used because effort data were missing in many of the logbooks and the effort data were collected on a trip basis as opposed to a haul by haul basis. In the SARC 58 assessment effort was calculated as:

Effort = days absent (time & date landed - time & date sailed) – 1 day per trip.

For some trips, the reported days absent were calculated to be a single day. This was considered unlikely, as a directed tilefish trip requires time for a vessel to steam to near the edge of the continental shelf, time for fishing, and return trip time. Thus, to produce a realistic effort metric based on days absent, a one day steam time for each trip (or the number of trips) was subtracted from days absent and therefore only trips with days absent greater than one day were used.

The number of vessels targeting tilefish has declined since the 1980s (See Table 1 of Working Paper 4 (Nitschke (2024)); during 1994-2003 and 2005-2015, five vessels accounted for more than 70 percent of the total tilefish landings. The number of vessels targeting tilefish has remained fairly constant since the assessment in 2005. The length of a targeted tilefish trip had been generally increasing until the mid-1990s. At the time of the 2005 assessment trip lengths had shortened to about 5 days. Trip length has increased slightly until 2008 and has subsequently declined until 2011. Trip lengths have been increasing slightly since 2011 to about 8.5 days in 2017 (See Table 1 of Working Paper 4 (Nitschke (2024)). In the weighout data the small number of interviews is a source of concern; very little interview data exists at the beginning of the time series (See Table 1 of Working Paper 4 (Nitschke 2024)). The 5 dominant tilefish vessels make up almost all of the VTR reported landings.

The number of targeted tilefish trips declined in the early 1980s while trip length increased at the time the FMP was being developed in 2000 (See Table 1 of Working Paper 4 (Nitschke (2024)). During the 2005 assessment the number of trips became relatively stable as trip length decreased. The interaction between the number of vessels, the length of a trip and the number of trips can be seen in the total days absent trend in Figure 3. Total days absent remained relatively stable in the early 1980s, but then declined at the end of the weighout series (1979-1994). In the beginning of the VTR series (1994-2004) days absent increased through 1998 but declined to 2005. Days absent increased from 2005 to 2008 but declined until 2010. Again days absent increased from 2010 to 2014 and have subsequently declined. When interpreting total days absent trends, it is important to note with improvements in data collection more recently that the subset of LPUE landings makes up a greater proportion of the total dealer landings (Figure 3).

LPUE trends are very similar for most vessels that target tilefish. A sensitivity test of the general linear model (GLM) using different vessel combinations was done in SARC 41. The SARC 41 GLM was found not to be sensitive to different vessels entering the LPUE series. Very little LPUE data exist for New York vessels in the 1979-1994 weighout series despite the shift in landing from New Jersey to New York before the start of the VTR series in 1994. Splitting the weighout and VTR LPUE series can be justified by the differences in the way effort was measured and difference in the tilefish fleet between the series. In breaking up the series we omitted 1994 because there were very little LPUE data. The sparse 1994 data that existed came mostly from the weighout system in the first quarter of the year. Very similar trends exist in the four years of overlap between Turner (1986) LPUE and the weighout series (Figure 4). At SARC 58 additional logbook data for three New York vessels was collected from New York fishermen from 1991-1994 and added to the VTR series. This was done to provide more information (years of overlap) in the modeling between the weighout and the VTR series.

Since 1979, the tilefish industry has changed from using cotton twine in pre-baited drums to steel cable for the backbone baiting at sea and from J hooks to circle hooks. The gear change to steel cable and snaps started on New York vessels in 1983. In light of possible changes in catchability associated with these changes in fishing gear, past working groups considered that it would be best to use the three available indices separately rather than combined into one or two series. The earliest series (Turner 1986) covered 1973-1982 when gear construction and configuration was thought to be relatively consistent. The weighout series (1979-1993) overlapped the earlier series for four years and showed similar patterns and is based primarily on catch rates from New Jersey vessels. The VTR (1991-2022) series is based primarily on information from New York vessels using steel cable and snaps.

The NEFSC weighout and VTR LPUE series were standardized using a GLM incorporating year and individual vessel effects. The LPUE was standardized to an individual longline vessel and the year 1984; the same year used in the last assessment. For the VTR series the year 2000 was used as the standard. Model coefficients were back-transformed to a linear scale after correcting for transformation bias. The updated GLM model that accounted for individual vessel effects appears to show more of an overall increasing trend in LPUE in comparison to the nominal series (Figure 5).

Changes in the VTR LPUE can be generally explained with evidence of strong incoming year classes that track through the landings size composition over time (Working Paper 2 (Nitschke 2024)). Since the SARC 58 assessment there appear to be increases in LPUE due to one or two new strong year classes. In general, strong year classes appear to persist longer in the fishery after the FMP and after the constant quota management came into effect which is evident in both the LPUE and size composition data.

The 2024 RTWG developed a method of transitioning from a LPUE index based purely on logbook VTR data to LPUE based on the newly developed CAMS system since the VTR database at the NEFSC will no longer be supported. The CAMS system integrates data collected from dealers with VTRs, observers, electronic monitoring for both landings and discards on a trip by trip basis as a single catch source to be used for assessments and quota monitoring for all managed stocks. The CAMS system is being used for landings and discards in stock assessments starting in 2020. The RTWG developed the most comparable LPUE tilefish index possible within the CAMS system for the transition from the VTR series in 1994 to the CAMS full implementation in 2020. However, the CAMS system has been estimated back in time to 2000.

Catch estimates for stocks assessments will likely not use CAMS until the year 2020 and forward into the future. The RTWG did consider linking the VTR and CAMS based LPUE index before 2020.

The RTWG estimated a LPUE time series (2000-2022) that produced very similar trends to the original VTR series used in the past (Figure 6). Theoretically the data source within CAMS is the same (VTR) for the effort metrics on tilefish trips. Some differences can be introduced with the landings source coming from the dealers within CAMS, possible data revisions over time, and data cleanup difference between CAMS and the VTR series when it was developed. Omission of only clear data errors which amounted to only 4 trips (2017, 2020, 2023) were made in the CAMS system to help prevent the introduction of potential biases.

The WG suggested stitching the VTR LPUE to the CAMS LPUE in 2010 (Figure 6). There was some concern that stitching in the CAMS series in 2020 would put a relatively larger abrupt shift into the series at the end of the time series.

For the 2024 RT assessment the WG also investigated whether other factors could help improve and perhaps better explain the LPUE trends (See Table 2 of Working Paper 4 (Nitschke 2024)). Reexamination of vessels effects, temporal factors (month), and crew size was examined. None of the available factors reexamined had a large influence on the underlying index. Limiting the index to the top 10 tilefish vessels also did not produce a meaningful difference. Very similar trends are seen in individual vessel LPUE series (Figure 7). The use of crew size also eliminated the data from 1991 to 1993 since that data was not available for that time period which is not desirable. The RTWG agreed to maintain the use of the original LPUE GLM incorporating individual vessel effects for the index.

3.2 Golden Tilefish Longline Study Fleet LPUE Investigation

Past benchmark tilefish assessments concluded that a simple days absent minus one day steam time (DA-1) was the best effort metric from vessel trip report (VTR) data due to data limitations mainly because the data is not collected on a haul by haul basis. Questions remain if landings per unit effort (LPUE) based on data collected at a finer haul basis could provide improvements or provide insights to LPUE indices as an index of biomass. Investigation of the longline study fleet data may help answer questions surrounding the somewhat crude effort metric in the LPUE index and could provide insight for future refinements. To help answer some of these questions the RTWG examined data from a single individual fishing quota (IFQ) tilefish vessel in the study fleet program who has been collecting tilefish catch data on a haul by haul basis since 2010. Special thanks to Captain Frank Green for his effort in this data collection and for allowing the RTWG to analyze this data source in support of the tilefish RT assessment.

The longline study fleet data examination is described in Working Paper 5 (Nitschke 2024). This analysis concluded that using finer effort metrics on a haul by haul basis supports the use of the day absent effort metric for LPUE indices from VTR data in the directed golden tilefish fishery. Days fished simply based on the number of hooks or miles fished seem to also be a good effort metric instead of days absent. Collection of data at a finer haul by haul resolution using miles fished or hooks fished could provide a more refined LPUE series. However, Days

Absent on a trip level seem to be a good proxy for effort on a directed tilefish longline trip. Using soak time in isolation seems to have a weaker relationship with catch rates. Depth and bycatch could also perhaps help with explaining some of the variability in LPUE if the study fleet program is expanded.

3.3 Golden Tilefish Trawl Study Fleet and Observer LPUE Exploration

To better understand the distribution and abundance of golden tilefish additional data sources are needed. Currently data from the directed fishery is the primary source of information for this stock. Another source of potential information on golden tilefish comes from the fishery-dependent data collected in fisheries where tilefish are not targeted. Specifically, the species is commonly caught in trawl gear that is used throughout the region. While there is a long history of catches in this gear type and some anecdotal information to suggest that catches in trawl gear are linked to recruitment, there have been limited explorations of these data to test for possible associations.

To fill this gap we collected data from the region's two largest data sets of commercial trawl catches. These included the Northeast Fisheries Science Center’s (NEFSC) Observer programs and the NEFSC Study Fleet. These two similar programs collect high-resolution catch and effort information (kept and bycatch weights at the tow level), as well as a suite of other useful information about fishing effort and location. Because of the similar data model shared by the program we were able to pool the records from both programs and then explore catch rates in space in time.

With a combined data set running from 2000 to 2022 we developed nominal and standardized catch per unit effort (CPUE) index from trawl gear. We then compared these trawl indices to two other indices that are available for the species: 1) an index of abundance derived from directed fishery, and 2) a recruitment index derived from the prior assessment. Results from this work suggest that a trawl index is likely sampling fish slightly smaller than those captured in the targeted fishery and the longline landing per unit effort (LPUE) index, but that these indices are tightly coupled when lagged appropriately. Therefore our results suggest that there may be some value in using these data to understand the abundance of fish, however more work is likely needed to test the impact of including such an index in the assessment.

3.4 Golden Tilefish Longline Survey Stratified Numbers per Tow at Length Indices

Stratified numbers per tow at length indices of relative abundance were estimated from the 2017 Tilefish Pilot Survey and 2020 Golden Tilefish Survey using a standard stratified random mean approach. With only two years of data available, this analysis represents a “proof-of-concept” for the estimation of this index which could be considered for inclusion in a Golden Tilefish assessment in the future as more years of data are incorporated.

3.4.1 Tilefish Surveys

This section briefly describes the tilefish surveys that were used to calculate the golden tilefish stratified numbers per tow at length. Catch rates by hook size used to produce these calculations are discussed below.

3.4.1.1 Tilefish pilot survey - 2017

A fisheries-independent pilot survey for golden (*Lopholatilus chamaelonticeps*) and blueline (*Caulolatilus microps*) tilefish throughout the range from Georges Bank to Cape Hatteras was conducted in 2017. This survey was intended as a "proof-of-concept" to establish a comprehensive fishery-independent bottom longline survey for both species along the Atlantic coast. Survey results could be used to standardize the effort across space and time which reduces uncertainty associated with the golden tilefish index of abundance (commercial catch per unit effort). The 2017 longline tilefish pilot survey is described in Working Paper 8 (Frisk et al. 2018) and Working Paper 7 (Boucher et al. 2023).

The 2017 pilot survey was based on a stratified random sampling design. The survey was initially proposed to consist of sampling stations representing the "core" fishing areas of tilefish based on commercial catch and a shallower and deeper "expanded" region to evaluate areas outside of the traditional fishery and better define the species range and abundance. Bottom longlines with one-nautical mile (1,852 m) mainline were deployed with 150 evenly spaced ganglions. Three different offset circle hook sizes (small = 8/0, regular = 12/0, large = 14/0) were deployed, distributed at a ratio of 20-60-20. Small hooks caught 2.2 times more golden tilefish (in numbers) per hook than regular hooks and 4.2 times more golden tilefish (in numbers) than large hooks.

3.4.1.2 Golden tilefish survey - 2020

Given the low incidence of encounters with blueline tilefish during the pilot survey, the Pilot Tilefish Review Committee recommended that the 2020 survey focused on golden tilefish only. The 2020 golden tilefish survey was based on a stratified random sample design consistent with the 2017 pilot survey with a target of 115 stations (a reduction from 206 in 2017). The 2020 survey consisted of sampling stations representing the core fishing areas for the mid-Atlantic golden tilefish population based on commercial catch data and the 2017 pilot survey. The 2020 longline tilefish pilot survey is described in Working Paper 9 (Olin et al. 2020) and Working Paper 7 (Boucher et al. 2023).

To maintain consistency and allow for comparison between surveys, the 2020 survey deployed bottom longlines with one-nautical mile (1,852 m) mainline equipped with 150 evenly spaced ganglions as it was done in the 2017 pilot survey. However, the 2020 survey only deployed two different offset circle hook sizes, distributed at a ratio of 50-50 per each set; these included small hooks (small = 8/0) and those that used by the industry (regular = 12/0), instead of three hook sizes used in the pilot survey (small = 8/0, regular = 12/0, large = 14/0). The 2020 survey was designed to determine if the small circle hooks (8/0) could provide additional information to a pre-recruit index relative to the regular circle hooks (12/0) as well as inform assessment model selectivity (i.e., domed shaped selectivity). Small hooks caught 2.4 times more golden tilefish (in numbers) than regular hooks.

3.4.1.3 Continuation of the golden tilefish survey

The Golden Tilefish Survey was continued in 2023 using the same approach as implemented in 2020. The results were not available in time for calculating the stratified estimates. Future development of the indices will include all additional years of the survey as they become available.

3.4.2 Relative Abundance Indices

Relative abundance indices of stratified mean numbers per haul at length are estimated using design-based methods with the stratified random sampling of the Golden Tilefish survey, following the approach used by the Northeast Fisheries Science Center in the estimation of indices from the Bottom Trawl Survey (Sosebee and Cadrin 2006). Briefly, the survey area was split into strata based on depth and location, reducing the variability of observations within each stratum. The random samples collected from each stratum are used to estimate the mean abundance of individuals per stratum by length, which is then expanded to an annual estimate combining all strata and lengths.

3.4.2.1 Hook selectivity

The 2017 pilot survey used three different offset circle hook sizes (small = 8/0, regular = 12/0, large = 14/0), distributed at a ratio of 20-60-20 and the 2020 survey used two different offset circle hook sizes (small = 8/0, regular = 12/0), distributed at a ratio of 50-50.

The pilot survey indicated that small circle hooks (8/0) caught few large golden tilefish and more small individuals relative to regular circle hooks (12/0), and large circle hooks (14/0) caught few individuals overall. Given these findings, the 2020 survey was designed to determine if the small circle hooks (8/0) could provide additional information to a pre-recruit index relative to the regular circle hooks (12/0) as well as inform assessment model selectivity (i.e., domed shaped selectivity), therefore, the large hook (14/0) was dropped from the 2020 survey, as the catchability of large hooks greatly decreases.

Given the difference in the deployment of circle hook sizes and the change in the ratio of their distribution between the 2017 pilot survey to the 2020 survey, an adjustment was applied to the hook sizes for 2017 (Box 1 below).

Box 1. The adjustment rate for the stratified numbers per haul at length was calculated as the ratio of hooks in 2020 to 2017.

	Small	**Regular**
2017 Pilot Survey	30	90
2020 Survey	75	75
Adjustment	2.50000	0.83333

No adjustment will be necessary between the 2020 and 2023 surveys, as they follow the same survey design focused specifically on golden tilefish and use the same ratio of hook sizes. As additional years of survey data are incorporated, the 2017 survey will be excluded from the index and no hook adjustments will be necessary.

3.4.3 Index Results

The annual stratified indices combining all lengths indicate a slight decrease in relative abundance from 2017 to 2020 (Box 2 below). This trend is consistent across indices calculated with only small hooks, only regular hooks, and the combined small and regular hooks.

Box 2. Annual stratified index values (and CV)

	2017	**2020**
Combined Hooks	13.28 (0.30)	11.06 (0.40)
Regular Hooks	4.80 (0.31)	3.86 (0.41)
Small Hooks	11.35 (0.25)	10.24 (0.28)

The stratified numbers per haul (Box 2 above) show a decrease in the abundance index between 2017 and 2020 for both the combined hook indices and for the separate hook size indices. However, the longline stratified survey index at lengths suggests that a relatively large younger year class or perhaps two year classes were present during the 2017 survey (first two modes in the distribution between 35cm and 50cm) in comparison to the 2020 stratified numbers per haul at length index (Figure 8). Three years later in the 2020 survey it can be seen that the stratified numbers per haul between 50 cm and 70 cm is greater than the 2017 survey. This generally follows the expectation of the growth of golden tilefish for the strong year classes seen in the 2017 survey. This trend can be further examined in the index's age composition data when that data becomes available.

Both hook sizes have very similar length distributions but there is some indication that smaller hooks catch a greater amount of smaller, younger fish between 35 and 50 cm relative to regular hooks. The regular hooks appear to catch relatively more large fish greater than 50 cm given that the catchability of regular hooks is about half or that of small hooks. Additional surveys will likely be needed to determine if this data could potentially be used to inform the dome shaped selectivity in the assessment model. This pattern seems to be consistent with a dome shape selectivity pattern in the fishery in the assessment model.

The survey also provides some indication that as fish age and increase in size they tend to be in deeper strata (shallowest strata zone 3 relative to the deepest strata 4) (Figure 9). However the vast majority of the fish caught in the survey was seen in the core fishery grounds of strata 3. The combined effects of possible reduction in catchability with larger fish sizes and relatively lower availability of larger/old fish to the fishery remains difficult to quantify at this time.

4 ASSESSMENT METHODS

TOR #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 complete history of the golden tilefish assessments can be found in the introduction of the main RT report and Working Paper 13 (Montañez et al. 2023). The current assessment was developed in SARC 58 using the age structured assessment program (ASAP) model (Legault and Restrepo 1998). ASAP is a forward projecting age structured assessment model which can be found on the NOAA Fisheries Toolbox (https://noaa-fisheries-integrated-toolbox.github.io/). Golden tilefish is considered a relatively data poor ASAP model. Production aging started in 2009 and the SARC 58 ASAP model used a pool age length key to derive the catch at age. As production aging continued the ASAP model was advanced with the use of year specific age data when data became more available in the 2021 management track assessment. There are gaps of missing catch at length and age data since the beginning of the longline fishery in the 1970s (Figure 10). The time series mean for weights at age was used for the mean weights at age for years where no catch at length data exists (before 1976, 1983-1994, 2000 and 2001) (Figure 11). A fishery-independent survey index of abundance is not available due to the low catchability of tilefish with trawl gear. Catchability is likely low for trawl gear due to tilefish's burrowing behavior. This assessment therefore relies on commercial LPUE as an index of biomass. In doing so the selectivity that is estimated for removals is also mirrored for the LPUE indices. The data sources that drive this assessment are the total removals, landings at length to estimate the catch at age and the LPUE index.

The RTWG began the research track assessment work with TORs 1-3 with the intention of adding newly developed data streams from these TORs to the final modeling framework. The RTWG also had a goal of advancing the assessment model from ASAP to the newly developed state-space modeling framework Woods Hole Assessment Model (WHAM) (Stock & Miller 2021, https://github.com/timjmiller/wham). WHAM has the ability to incorporate environmental covariates, more options for fitting age compositions and the ability to include process errors.

WHAM can also be configured to make it similar to ASAP. The RTWG used the golden tilefish ASAP data input file from the last 2021 management track ASAP model with a terminal year 2020 to build the bridge run that is similar to the last management track tilefish ASAP model. Once a bridge run was developed the RTWG investigated configuration changes to improve the model. It became apparent that the model results are sensitive to the inclusion of random effects. Due to these challenges under the time constraints the RTWG's goals shifted from the original plan with the assessment modeling TOR4. The RTWG and assessment working group (AWG) decided not to include any data changes within this RT assessment relative to the last management track assessments input data. Our goal for modeling TOR 4 was to better understand the model configuration effects in the new modeling framework WHAM without the additional effects of data changes. The RTWG suggests the best configuration to be used in the next management track assessment with the hope that incremental improvement and advancements could be made in future management track assessments as more data can be incorporated from TORs 1-3. A better understanding of random effects influence on model selectivity estimates and biological reference points (BRPs) with this relative data poor stock can then be advanced in future management track assessments once the assessment model is developed in WHAM in this RT assessment.

The RTWG configured a WHAM run which produced a similar result to the 2021 management track ASAP run assuming a multinomial distribution on the age composition. Some selectivity parameters (ages 6, 7 in block 1 (1971-1983) and ages 1, 2 and 5 in block 2 (1984-2020)) had to be fixed to the ASAP solution in order to get the model to converge (Figure 12). There are some differences between ASAP and WHAM which likely contribute to some of these convergence issues within WHAM. ASAP uses a penalty on recruitment deviations from a mean and initial population estimates were made from a penalty on deviations relative to an input equilibrium population constructed under a zero fishing mortality assumption. WHAM estimates recruitment as a random effect. WHAM starting conditions can be modeled as either from an equilibrium population using two parameters (recruitment and equilibrium fishing mortality) or as estimates of starting numbers at age. The penalty structure within the ASAP model may have made the model more stable to convergence relative to WHAM. During the initial development, WHAM had convergence issues with starting conditions (equilibrium or numbers at age) and with selectivity assumption configuration changes. However, the RTWG was satisfied that ASAP results could be reproduced relatively well within WHAM with some forcing of the parameters to ASAP results (Figure 13).

After constructing a bridge run configuration, the RTWG loosened the constraints on fixing selectivity parameters to ASAP results and explored the use of other self-weighing error structures in fitting the age compositions. Initial WHAM model seemed to have convergence issues with relatively small changes in the configuration. There seemed to be some stability issues caused by the residual pattern in the one-step-ahead (OSA) residuals with age 1 and 2 in the first block and age 1 in the second block using the dirichlet-miss0 (treats missing observations as missing) for age composition error structure. The RTWG fixed the selectivity at zero for these ages (age1 and 2 first block and age 1 in the second block) since there was less than 1% of the removals for these younger ages in each year and because doing so appeared to improve the stability in the convergence (Figure 10). The RTWG did leave the estimation of selectivity for age 2 on in the second block because there was some concern of losing a signal in the data for potential strong incoming year classes at the end of the time series even though a residual pattern also occurred for this age 2 and less than 1% of the catch also came from age 2.

The ASAP model and preliminary runs of WHAM also had relatively high selection of ages 7 through 9 in the first selectivity block (Figure 12). The RTWG fixed full selection in ages 6 through 9 in the first block and kept the full selection at age 5 in the second block to help with the sensitivity of convergence with configuration changes.

Estimation of the initial population sometimes became problematic in WHAM. The 2021 management track ASAP model started in 1971 with the beginning of the development of the directed longline fishery. However age composition data is not available until 1976. Estimation of initial numbers at age seems to cause convergence issues before age composition data was available. Using the equilibrium assumption assuming a low fishing mortality rate (fixing fishing mortality close to zero) seems to produce reasonable starting conditions given the history of the fishery. Configuring the model to start in 1976 while estimating the starting numbers at age can produce unrealistic high population estimates under different model configurations relative to a population based on an equilibrium low F assumption. Model results will differ among the different starting conditions but the overall trends and results were not very sensitive to the starting conditions (Figure 14). Starting the model much later in 1995 with the availability of port sampling length data did not produce a large difference in the population estimates. However the population scaling in retrospective peels seemed more problematic when starting the model in 1995 (Figure 15).

There is also limited information on the best time frame for blocking the two assumed selectivity blocks since there is no age composition data between 1983 and 1994. However model results did not seem very sensitive to changes in the block within this 10 year period. To be more pragmatic the RTWG decided to change the first block to the middle this 10 year period with no age composition data (1^{st} block 1976-1986 and 2^{nd} block from 1987 to 2020).

In general the WHAM model results were similar to ASAP with similar estimates of the dome shaped selectivity in the second block and with the stock rebuilding to roughly SSB_{MSY} after the inception of management in 2001. The WHAM model diagnostics also appears to be acceptable with low retrospective error. Similar to past ASAP assessments, due to questions surrounding the cryptic biomass the manual forcing of a flatter selectivity curve in the second block (fixing 10+ age group to 0.5) tends to produce retrospective issues (Mohn's Rho on SSB = 0.51).

The RTWG developed a base model starting in 1976 using estimated starting numbers at age, self-weighting dirichlet missing 0 for fits to age composition data and shifting the selectivity block to 1976-1986 for the 1^{st} block and 1978 to 2000 for the second block. WHAM model results were sensitive to adding random effects. Adding random effects to the base model NAA (iid or survival) appears to allow for additional model flexibility which produces a relatively better fit to the data with improvements in the diagnostics. Most of the change occurs in fitting the 10+ age group while still producing good retrospective diagnostics. Adding numbers at age (NAA) random effects results in a relative flattening of the selectivity curve in the 2^{nd} block, less cryptic biomass, less rebuilding since the inception of management in 2001 and a worse stock status relative to $F_{40\%}$ based spawning potential ratio (SPRs) BRP proxies ($F/F_{40\%}$ and $SSB/SSB_{40\%}$ ratios). The results with random effects on survival were similar (except for starting condition) among runs whether using iid (uncorrelated), ar1 (correlated by age), ar1_y (correlated by year) or 2dar1 (correlated by year and age). The improvements in the diagnostic and comparison of results between the base model and the base model with added NAA iid random effects (base_NAAiid) is described in Working Paper 11 (Nitschke 2024).

The RTWG continues to work on adding additional random effects on selectivity as well as survival which continues to improve the relative model diagnostics. The full random effects exploration is described in Working Paper 12 (Hennen 2024). In general, it appears that adding additional random effects to the tilefish model seems to result in additional flexibility within the model allowing for further flattening of the selectivity curve which results in lower increase in biomass relative to an $F_{40\%}$ based proxies and a relatively poorer stock status

.

Figures 16 to 19 compare the Base model (ASAP like) to the base configuration with added iid random effects on NAA (Base_NAAiid) and full random effects model which had random effects on both NAA and selectivity (Full_RE). The relative model diagnostics which is also reflected in the AIC improved as random effects are added to the tilefish WHAM model (Working Paper 12, Hennen 2024). The Stock status based on SPR BRPs are highly sensitive to the inclusion of random effects within WHAM due to the change in perception in the relative improvements in the stock since the inception of management in 2001 with the estimated lower cryptic biomass from a flatter selectivity curve. There are less 10+ fish in the catch at age data over the last two decades than expected for a long lived species if the selectively is more flat-topped for a more positive stock status. Therefore, the model can either estimate a dome shaped selectivity pattern that produces cryptic biomass in the 10+ group like in the ASAP and non-random effect WHAM model runs or the model can produce flatter selectivity with the added flexibility allowed for fitting the data through random effects with relative improvements in the diagnostics. However, if a flatter selectivity is estimated then the biomass has to be low under high Fs since the proportion of 10+ fish in the catch at age is low.

The RTWG was uncomfortable with the underlying sensitivity of the results even though the diagnostics improved when additional random effects were added. The results became more questionable with additional random effects added (Full_RE) to the model given the history of the fishery and management. The perception from industry is that fishing has improved and that increases in biomass have occurred since management was implemented in 2001. The raw data also suggests general improvements in LPUE and size structure after management was put in place. Strong year classes have been entering the fishery relatively consistently every 5-7 years. This did not seem to be completely consistent with overfishing over the entire time series (close to 5 decades).

The SSB_{MSY} estimates are relatively stable among models and all of the models indicate some level of stock increase since 2001 but because the full random effect model has a flatter selectivity curve it also estimates a higher maximum sustainable yield (MSY, TOR 5) when the stock is rebuilt. The projection in the full random effects model seems unrealistic in terms of the ability for the stock to quickly rebuild with relatively small reductions in the catch from the recent averages. Random effects likely quickly dissipated in the projections within WHAM so the stock quickly rebuilds the 10+ age group and catch increases quickly to approach MSY (1,075 mt). This did not seem realistic given the response observed in the stock under the catches over the last 10 to 20 years. A comparison of the actual LPUE index through data updated through 2022 also does not correspond well with the projection from the 2020 terminal year assessment since the LPUE index suggests a decrease in biomass occurs similar to what has occurred in the past as strong year classes age (Figure 20). It seemed unrealistic that this would change in the projection with the stock continuing to rebuild. The RTWG has concerns that if the full random effect model was used for management, then the biomass would continue to be re-estimated lower as the assessment gets updated with additional years of data and that rebuilding may only occur in theory within the projections.

Although the full random effect model (Full_RE) provided the best fit to the data and perhaps had the most desirable diagnostic properties as a whole, the RTWG ultimately decided against further development of the random effects model in this research track. The RTWG was unable to support the very different impression of fishery selectivity implied by the model results (Figure 21). The incorporation of random effects on selectivity and survival led to much higher selection of older tilefish, which in turn reduced implied recruitment and SSB over the time series (Figure 16). The WG believed that reduced selection of older tilefish coupled was better supported by expert opinion and ongoing research. Of most concern to the WG was the perception that the stock had far higher MSY than previously thought, which seemed unrealistically high given the performance of the fishery since the inception of management in 2001.

While the literature on state space model diagnostics is still developing, some studies have suggested that overfitting may be a concern when data density is relatively low. Liljestrand et al. (2023) demonstrated that low data density may reduce the ability to properly differentiate process and observation errors. Additionally, Li et al. (In press) indicated caution when including random effects on processes that have strong effects on reference point calculations. Given the relatively low information content of the tilefish data, the RTWG decided to use a less complex model as the basis for continuing model development in the management track.

However, RTWG felt that the WHAM results among models suggests there is considerable uncertainty in the selectivity and stock status. A single model does not seem to capture the true uncertainty in the assessment. The RTWG did not have confidence in the results of the full random effects model as a basis for the assessment and stock status. The RTWG recommends to use the Base model without random effects until more confidence can be gained in future management track that suggests inclusion of some random effects are giving a more accurate depiction of the selectivity and true stock status. However, the RTWG feels that consideration of the random effects model is useful for showing the overall uncertainty and sensitivity of the results in the assessment. Assuming the base model is an accurate depiction of reality also does not account for the true uncertainty in this assessment.

5 STATUS DETERMINATION CRITERIA

> *TOR #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 the 2021 ASAP management track assessment maximum sustainable yield (MSY) reference points were based on spawning potential ratio (SPR) of 40% due to a lack of a stock recruit relationship. SSB_{MSY} was estimated from stochastic long-term (100 year) projections fishing at $F_{40\%}$ and the overfished threshold is defined as one half of $SSB_{40\%}$. Stochastic projections resampled from a cumulative density function (CDF) of empirical age-1 recruitment for the entire time series (1971-2020). The recent 5 year average was used to account for prevailing conditions with a natural mortality assumption of 0.15 used for the estimation of $F_{40\%}$ and $SSB_{40\%}$. The 2021 management track assessment concluded that overfishing was not occurring and the stock was not overfished.

The RTWG did not recommend a change to the $F_{40\%}$ proxy for F_{MSY} biological reference points (BRP) since a stock-recruit relationship was also not evident in the WHAM base model (Figures 22 and 23). Examination of the weight at age did not suggest large changes in trend at the end of the times series (Figure 23). There was little difference between using a 10 year or a 5 year recent average for the estimates of the WHAM BRPs (Table 4). The RTWG suggested using the 10 year average since there can be some variability in the mean weights at ages for the older ages (Figure 24). The number of years to use for the average could be changed if a trend in the mean weights emerges in the future management track assessments. The recruitment used to estimate the $SSB_{40\%}$ within WHAM was based on the entire time series minus the most recent two years of data (1999 and 2000) since there is limited information to inform recruitment in the last two years of the model.

The dynamic estimates of SSB at $F_{40\%}$, $F_{40\%}$, and yield at $F_{40\%}$ for the base model can also be seen in Figure 25. The time series of $SSB/SSB_{40\%}$ and $F/F_{40\%}$ ratios for the base model along with the estimated WHAM model uncertainty is shown in Figure 26. The Kobe stock status plot is shown in Figure 27 for the base model.

The RTWG recommends that stock status should be based on the model without random effects (Base model). However, Table 5 also compares the results for the two random effects models (Base_NAAiid and Full_RE) to show the selectivity effects on the biological reference points. $F_{40\%}$ decreases as selectivity flattens with random effects (Table 4, Figure 21). Mean recruitment and therefore $SSB_{40\%}$ are similar among the three models but MSY will differ due to the selectivity differences among models and the stock status also differs among models due to the difference in the recent biomass estimates in 2020. Therefore the stock status among models also differs (Table 5). The RTWG recommends the use of the base model configuration for stock status determination (TOR 4). Overfishing ($F/F_{40\%} = 0.55$) was not occurring and the stock was not overfished ($SSB/SSB_{40\%} = 1.29$) according to the base model.

6 PROJECTION METHODS

TOR #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

The 2020 ASAP management track assessment projections were based on 1000 iterations from the MCMC of the final ASAP model as input to a separate AGEPRO projection program which can be found on the NOAA Fisheries toolbox (https://noaa-fisheries-integrated-toolbox.github.io/). With the new RTWG base model the projections and biological reference points are integrated within the WHAM framework and the estimated uncertainty from the model are projected forward. The RTWG recommends the use of the base model for $F_{40\%}$ (F_{MSY} proxy) projection for the determination of overfishing limits (OFL) in the next management track assessment. Using the base model would also be consistent with stock status determination. However, the RTWG acknowledges that projections and estimated uncertainty of the base model likely does not capture the true uncertainty in the assessment since the results and status determination were found to be sensitive to changes in selectivity from the use of random effects.

$F_{40\%}$ 10 year projections were made for illustrative purposes for this RT assessment since the model will be updated with additional years of data in the next management track assessment (Figures 28 to 32). The base model with terminal year 2020 estimated relatively high catches in the projections under $F_{40\%}$ (2022 to 2023) (Table 6, Figure 32). This increase in the short-term catch is due to a relatively strong recruitment year classes at the end of the time series and because $F_{40\%}$ results in an increase in F within the projection ($F/F_{40\%}$ = 0.55) (Figure 31). The stock is also estimated to be above $SSB_{40\%}$ ($SSB/SSB_{40\%}$ =1.29) in 2020 for the base model. Therefore the projections become a Fishing down exercise to $SSB_{40\%}$ longer-term in the projections. In the short term, catches at $F_{40\%}$ are higher than the MSY when the stock is at $SSB_{40\%}$ (855 mt).

To a certain extent the RTWG was not confident in the full random effect model because it seemed unrealistic that the stock would suddenly rebuild under catches that were slightly lower than what was observed in the last 10 years (TOR 4). These projections would suggest a change in the proportion of larger older fish caught by the fishery with higher catches as the stock rebuilds (Figures 33 to 35). This model and projection was considered a sensitivity run due to the remaining questions with the random effects and the results. However, this sensitivity does propose a different perception of the stock and projection trends which suggests that the true uncertainty in the assessment is likely not captured by the base model projection in isolation.

The projections for golden tilefish models are also more uncertain because there is limited information to inform recruitment in year t-1 and no information for the terminal year since no survey information for younger smaller fish is available to the assessment model. This lack of information is reflected in the low estimated selectivity for younger ages (age 1 and 2) from the WHAM model (Figure 36). For example, with this research track assessment the terminal year is 2020 and if we assume a bridge year for short-term projections then the first year of the catch advice would be 2022. Therefore, for the first year that is used for quotas, the 2018 year class will be 4 years old and the 2019 year class will be 3 years old. By the 2nd year (2023), most of the catch advice will not be based on data to inform year class strength since the 2018 year class will be 5 years old and the 2019 will be 4 years old.

7 RESEARCH RECOMMENDATIONS

TOR #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

Research recommendations from previous stock assessments of Golden Tilefish were compiled, and the status of each recommendation was evaluated by the RTWG. Some research recommendations were repeated in various documents (e.g., 58 SAW, SSC documents); for brevity, where the same, or substantially similar recommendations were made, we consolidated them under a single heading (e.g., 58 SAW), but noted the additional documents in which the recommendation was raised.

7.1.1 1999 Science and Statistical Committee Review Assessment (NEFSC 2006)

• 'Ensure that market category distributions accurately reflect the landings.' – Considered completed, a large/medium category was added in 2013. Sampling of the commercial lengths has improved over the last decade or so. Small, kitten, and medium market category distributions can shift from one year to the next due to the growth of a strong year class. Intensive length sampling of the landings by market categories is needed to account for possible shifts in the distribution within a market category over time. Similar landings distributions were seen among the observer, study fleet, and commercial port sampling data sources. In addition, a large-medium market category code was added in 2013 which appears to have resulted in a decrease in the amount of unclassified fish landed. This suggests that unclassified landings were a combination of true unclassified fish and large-medium fish due to the lack of a large-medium code prior to 2013. The development of large-medium market category code should reduce this source of error in the landings at length and landings at age after 2013.

• 'Ensure that length frequency sampling is proportional to landings by market category.' – This 2024 research track looked at observer kept length frequencies as possible backup for the port sampling program if it gets worse. Commercial length sampling has been sporadic during the beginning of the time series. In particular, length samples from the large market category have been lacking. However commercial length sampling has greatly improved over the last decade or so with a higher proportion of the sampling coming from Montauk where most of the fish are landed. In addition, a large-medium market category code was added in 2013 which appears to have resulted in a decrease in the amount of unclassified fish landed. This suggests that unclassified landings were a combination of true unclassified fish and large-medium fish due to the lack of a large-medium code prior to 2013. The development of large-medium market category code should reduce this source of error in the landings at length and landings at age after 2013. Since 2020, biological port sampling of market category lengths and age has decreased due to cuts in funding. Recommend that sampling is increased or remain at least at current levels in the future.

• 'Increase and ensure adequate length sampling coverage of the fishery.' – Considered completed, superseded by new SARC 58 research recommendations under bullets 1 and 2.

• ‘Update age- and length- weight relationships.’ – Considered completed for SARC 58.
• ‘Update the maturity-at-age, weight-at-age, and partial recruitment patterns.’ – Considered completed for SARC 58.

• ‘Develop fork length to total length conversion factors for the estimation of total length to weight relationships.’ – Considered completed in SARC 41.

• ‘Incorporate auxiliary data to estimate r [intrinsic rate of population increase] independent of the ASPIC model.’ – No longer applicable. The 2005 SARC 41 questioned if this can be done or should be done. However the 2009 SARC 48 SCALE results suggest that r is overestimated in the ASPIC model. The SARC 58 working group did not consider the ASPIC model to be sufficient to evaluate the status of the stock and explored other models in SARC 58 assessment. This research recommendation is no longer relevant with the use of the ASAP model or WHAM model.

7.1.2 41st SAW Assessment Report (NEFSC 2006)

• ‘Conduct a hook selectivity study to determine partial recruitment changes with hook size. Determine catch rates by hook size. Update data on growth, maturity, size structure, and sex ratios at length.’ – Updated growth, maturity, and size structure studies were completed during the 2009 SARC 48 assessment. Hook selectivity data was collected in the 2017 fishery-independent pilot bottom longline survey for tilefish (for both golden and blueline tilefish) and the fishery-independent bottom longline surveys for golden tilefish that were conducted in 2020 and 2023. In this 2024 research track assessment, data collected in the 2017 pilot survey and the 2020 survey were used to examine hook selectivity. More specifically, stratified numbers per tow at length indices of relative abundance were estimated using a standard stratified random mean approach.

• ‘Collect data on spatial distribution and population size structure. This can help answer the question of the existence of a possible dome[-]shaped partial recruitment pattern where larger fish are less vulnerable to the fishery due to spatial segregation by size.’ – This research recommendation was initially addressed in the study fleet data during the 2009 SARC 48 assessment. The surveys discussed under the first bullet of SAW 58th research recommendations have helped further address this research recommendation.
• ‘Continue to develop the forward projecting catch-length model as additional length data becomes available. Investigate the influence of adding a tuning index of abundance and model estimated partial recruitment (logistic) to the catch-length model.’ – This research recommendation was completed during the 2009 SARC 48 assessment. The improved catch-length model was renamed as the SCALE model. An ASAP model was developed in SARC 58. However, in this 2024 research track assessment, the model is developed in WHAM.

• ‘Collect appropriate effort metrics (number and size of hooks, length of main line, soak time, time of day, area fished) on a haul basis to estimate commercial CPUE.’ – This research recommendation was completed with the study fleet analysis during the 2009 SARC 48 assessment. In this 2024 research track assessment we examined effort metrics for a single longline vessel participating in the study fleet program.

• ‘Initiate a study to examine the effects of density dependence on life history parameters between the 1978-82 period and present.’ – This research recommendation was completed with the updated growth and maturity study during the 2009 SARC 48 assessment.

• ‘Increased observer coverage in the tilefish fishery to obtain additional length data.’ – Considered completed due to increased port sampling to obtain sufficient lengths from the landings. Discards in the fishery are relatively small and adequately sampled. In this 2024 research track assessment, kept/length and estimated discards from observer program data were reevaluated.

• ‘Develop a bioeconomic model to calculate maximum economic yield per recruit.’ – No progress.

7.1.3 48th SAW Assessment Report (NEFSC 2009)

• ‘Continue the development of an improved haul based fishery-dependent CPUE index (i.e., continue the current study fleet project) or design a tilefish longline survey as a semi fishery-independent index of abundance that could be conducted by an existing longline vessel and the study fleet platform. If a tilefish longline survey is developed then size information should be incorporated into the survey design for the estimation of a recruitment and size specific index of abundance which could improve the tilefish assessment.’ – Considered completed, superseded by SARC 58 research recommendation under first bullet.

• ‘For the study fleet project and any potential semi fishery independent survey, include additional information on conflicts with lobster and trawl gear, the possibility of unknown effects on tilefish CPUE due to competition/interference from an increased abundance of dogfish, the unknown effects of bait type on tilefish CPUE (e.g., substitutes for the preferred squid).’ – No progress.

• ‘Develop protocols to ensure consistency between dealer, VTR, and IVR reports of the tilefish landings.’ – The IVR (Interactive Voice Response) requirements were first implemented when the FMP was initiated in 2001 as a way to track quota landings in the fishery in a timely fashion. However, with the implementation of electronic dealer reporting in 2004 and improved VTR (Vessel Trip Report) reporting processing by the agency, the information provided by fishermen using the IVR system has become redundant. The IVR requirements were eliminated in 2018.

• ‘Develop protocols to ensure consistency in market category designation among fishing ports.’ – The fishing industry has implemented protocols to ensure consistency in market categories among fishing ports. In addition, a large-medium market category code was added in 2013 which appears to have resulted in a decrease in the amount of unclassified fish landed.

• ‘Explore the influence of water temperature and other environmental factors on trends in the commercial fishery CPUE index of stock abundance.’ – Work in progress, but note that extremely limited catch and temperature data are available to address this research recommendation. Available data was examined in the SARC 58 assessment in TOR 3. The 2024

RTWG noted that more recently, temperature plus other variables (DO, salinity, etc.) were collected in the 2017 pilot survey for tilefish and the 2020 and 2023 golden tilefish surveys. While these data were not used to develop CPUE values used in the assessment (due to shortness of the time series), it is expected that as future golden tilefish longline surveys are conducted and more information is collected, this could be evaluated. Continuing to collect water temperature and other environmental factors in future longline surveys could be useful.

7.1.4 2012 Scientific and Statistical Committee Discussion from 48th SAW Review Assessment

• 'Understand the role of tilefish in creating secondary habitats through their burrowing activity, thereby increasing diversity and the extent to which this diversity is compromised by the removal of these ecosystem engineers by the fishery.' – No progress.

• 'Understand the causes in the pattern and variability in recruitment.' – A 2021 study indicated several potential environmental drivers for the golden tilefish fishery and stock dynamics (Nesslage et al. 2021). In this 2024 research track assessment, variability in recruitment patterns were further evaluated using environmental covariates.

• 'Quantify and understand the spatial dynamics of the stock and the fishery (specifically, assess historical changes in the distribution of fishing effort, develop haul-by-haul information on the spatial and temporal distribution of catch, and evaluate the potential of a rigorously-designed study fleet program).' – Work in progress, through an initial examination of the 2008 study fleet data and ongoing use of the VTR as the source of information for the fishery-dependent CPUE index of stock abundance. Expansion of the study fleet program to more vessels on directed tilefish trips might be beneficial. In this 2024 research track assessment we examined effort metrics for a single longline vessel participating in the study fleet program.

• 'Assess the potential for and extent of local population structure.' – No recent progress. The work of Katz et al. (1983) used significant differences in allelic frequencies to identify distinct stocks between mid-Atlantic and South Atlantic tilefish. Those authors also felt that certain aspects of golden tilefish distribution, life history and ocean circulation patterns supported their two stock hypothesis for the United States Atlantic. However, tissue samples (liver and muscle) were collected on the 2017 pilot survey for tilefish and the 2020 and 2023 surveys for golden tilefish which can be used to conduct future genetic work.

• 'Assess coherence between north and south Atlantic stocks and evaluate the effects of climate indices in driving stock dynamics.' – Nesslage et al. (2021) examined long-term, low-frequency climate drivers on golden tilefish in an attempt to explain recruitment.

• 'Evaluate the potential effect of time-varying catchability on assessment models that rely on commercial CPUE data.' – Work in progress, through examination of catchability trends in SCALE and ASAP models developed for SAW 58th.

• 'Evaluate the potential for a stakeholder survey to assess extent of population outside of normal fishing area.' – The 2017 fishery-independent pilot bottom longline survey for tilefish evaluated the spatial distribution of golden tilefish.

7.1.5 58th SAW Assessment Report (NEFSC 2014)

• 'Develop an industry based survey using two or three designated fishing trips per year. Industry based survey trips would follow a design similar to a fishery independent survey and collect more intensive size and catch information on a haul-by-haul basis. However, a reduction in catch rates likely occur on these survey trips relative to normal fishing operation. The benefits of a survey design to the stock assessment will likely surpass a more intensive and burdensome haul-by-haul data collection on trips during normal fishing operation. The SAW 58 working group suggests this science could be funded through the Cooperative Research Program, the habitat assessment improvement plan, or MAFMC research set-aside (RSA)' [also recommended in 2017 Scientific and Statistical Committee Discussion from Assessment Update Through 2016; and 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment] – Work in progress. An industry based survey as described above was not implemented. However, the 2024 RTWG indicated that a pilot fishery-independent tilefish longline survey was conducted in 2017 (for both golden and blueline tilefish) and fishery-independent golden tilefish longline surveys were conducted in 2020 and 2023. These surveys have been used to collect more intensive size and catch information on a haul-by-haul basis. Survey data indicates that there is a dome-shape selectivity pattern from spatial effects and from possible gear hook size selection. Uncertainty remains with the ability to quantify the degree of doming in the fishery.

• 'Increase the sampling of maturity at size and age and commercial landings at size and age.' – Commercial length sampling has greatly improved over the last decade or so with a higher proportion of the sampling coming from Montauk where most of the fish are landed. However, since 2020, biological port sampling of market category lengths and age has decreased due to cuts in funding. Recommend that sampling is increased or remain at least at current levels in the future.

7.1.6 2014 Scientific and Statistical Committee Discussion from 58th SAW Review Assessment

• 'Explore methods to estimate the abundance and distribution of burrows as a forerunner of a fishery-independent survey.' – No progress.

• 'Perform exploratory analyses of fish distributions to assess whether the dome-shaped selectivity curve used in the assessment reflects fishery selectivity or availability, or both' [also recommended in 2017 Scientific and Statistical Committee Discussion from Assessment Update Through 2016; and 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment]. – Work in progress. See first bullet under SAW 58th research recommendations for additional narrative regarding recent survey efforts.

• 'Expand observer coverage to improve index standardization of fishery-dependent data' [also recommended in 2017 Scientific and Statistical Committee Discussion from Assessment Update Through 2016; and 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment]. – Considered completed due to increased port sampling to obtain sufficient lengths from the landings. Discards in the fishery are relatively small and adequately sampled.

• 'Leverage large pelagic recreational fishing activity to improve life history information.' – The 2024 RTWG noted that there is no biological information collected for tilefish in the large pelagic survey that could be used to improve life history information. However, the RTWG used the estimated large pelagic private mode golden tilefish recreational landings to develop a time series of recreational catch.

• 'Assess the accuracy and reliability of aging techniques' [also recommended in 2017 Scientific and Statistical Committee Discussion from Assessment Update Through 2016; and 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment]. – Work in progress. The NEFSC does QA/QC exercise in the production of aging of each species https://www.fisheries.noaa.gov/new-england-mid-atlantic/science-data/golden-tilefish-qa-qc-exercise-results.

• 'Consider genetic approaches to assess possible stock structure.' – No recent progress. See fourth bullet under 2012 Scientific and Statistical Committee from 48th SAW Review Assessment for additional narrative about this research recommendation. However, tissue samples (liver and muscle) were collected on the 2017 pilot survey for tilefish and the 2020 and 2023 surveys for golden tilefish which can be used to conduct future genetic work.

• 'Evaluate the reliability of the report of protogynous hermaphroditism in the S. Atlantic stock.' – In a study conducted by McBride at al. (2013) the gonochoristic sexual pattern of the golden tilefish northern stock was confirmed. The RTWG spoke with South Carolina Department of Natural Resources technical staff who are familiar with recent research of protogynous hermaphroditism in the South Atlantic stock, and it was indicated that, SEDAR 89 South Atlantic Tilefish Working Group discussions have noted reservations of recent research showing protogynous hermaphroditism in the S. Atlantic stock.

7.1.7 2017 Scientific and Statistical Committee Discussion from Assessment Update Through 2016

• 'Leverage existing fishing activity to provide samples to improve life history and distribution information' [also recommended in 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment]. – No progress.

• 'Evaluate the role of sanctuaries on the Golden Tilefish stock and its fisheries' [also recommended in 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment]. – Some progress. The 2017 fishery-independent pilot survey for tilefish can be used to evaluate the role of sanctuaries and justification for a dome-shaped selectivity pattern in the commercial fishery.

7.1.8 2020 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2021 Management Track Assessment

• 'New survey results will be incorporated into assessment.' – Golden tilefish longline survey results are considered preliminary. Survey results will be considered for the 2024 Research Track Assessment. The 2024 RTWG developed a stratified index of abundance from the 2017 pilot survey for tilefish and the 2020 survey for golden tilefish for possible inclusion in future management track assessments.

• 'Use of an aggregate age length key should be reconsidered. Perhaps consider an age and length-based model. (It was noted that this often requires a full benchmark assessment).' – This recommendation is completed. The age structure tilefish model is now based on production aging.

• 'In the meantime, continue use of contemporary age length keys and enhance use, if possible.' – Considered completed for the 2021 Management Track Assessment.

• 'Review new data on recreational data derived from mandatory permitting and reports.' – Work in progress. The 2024 RTWG reviewed collected private recreational data and concluded that at this point it is not reliable. However, it is possible that as future improvements in reporting occur, this data may be more useful.

• 'Consider adding MRIP and recreational VTR data to assessment. Comprehensive review of all sources of estimated removals (e.g., discards, too).' – Considered completed for the 2024 research track assessment. The 2024 RTWG developed a time series of recreational catch.

7.1.9 2021 Scientific and Statistical Committee Discussion Regarding Factor for Consideration in the 2024 Research Track Assessment

• 'Given the results of the assessment update, it seems reasonable to change the overfishing definition to $F_{40\%}$'. – This was done and completed.

• 'Continuation of adequate age sampling is critical to the switch from the use of pooled age-length-key to year specific age-length-keys for more appropriate characterization of age structure and better tracking of year classes.' – This was done and completed.

• 'There is a significant concern with reductions in the biological port sampling that may negatively affect future assessments, including the next RT assessment model in 2024.' – The 2024 RTWG discussed that while the recent reductions in the biological port sampling are troublesome, it is recommended that sampling be increased or remain at least at current levels in the future.

• 'Due to the lack of information on incoming recruitment at the end of the time series (no fishery independent surveys that capture young fish), alternatives to the TAL calculations based on projections that rely on uncertain indications of year class strength should be considered. A conservative approach to changes in the TAL over time appears to have resulted in overall benefits for both the Golden Tilefish stock and for the fishery.' – No progress.

7.2 New Research Recommendations

1. Collection of length samples on party/charter trips for potential improvements in recreational time series estimates for golden tilefish.

2. Evaluate WHAM performance for information poor stocks using simulated tilefish like populations (i.e., only catch data). Do random effects in both survival and selectivity introduce bias?

In addition to the new research recombination listed above, the sub-working group that developed the golden tilefish ecosystem and socioeconomic profile discussed in Working Paper 1 (Salois et al. 2024), also made several research recommendations. The research recommendations listed in Working Paper 1 (Salois et al. 2024) may help improve the tilefish ecosystem and socioeconomic profile and are related to golden tilefish size distribution, availability, recruitment, size distribution and early life history. However, these recommendations were not included above as the RTWG did not have time to review these research recommendations.

8 BACKUP ASSESSMENT APPROACH

> *TOR #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.*

The Northeast US stock assessment process requires an accepted stock assessment method to provide the best scientific information available for fishery management, including a contingency plan if the proposed assessment method fails peer review in the research track process or subsequently fails peer review in the routine management track process. Many northeast US assessments specify an empirical backup approach based on survey data, either swept-area estimates of stock biomass and a target exploitation rate or survey biomass trends and recent catch. However, due to the current lack of survey data for golden tilefish these approaches are not good options for this stock.

The RTWG briefly discussed the use of other data-limited approaches for estimating sustainable yield such as Depletion-Corrected Average Catch (DCAC; MacCall 2009) and Depletion-Based Stock Reduction Analysis (DB-SRA; Dick and MacCall 2011). However, the RTWG did not pursue these because they heavily rely on assumptions needed to run models. Lastly, DCAC and DB-SRA were two of several data-limited approaches examined by the NEFSC Index Based Methods Working Group (NEFSC 2020, Legault et al. 2023). The primary focus of that working group was to quantify the performance of data-limited approaches in circumstances that led to severe retrospective errors in statistical catch-at-age models. That group found that none of the data-limited methods outperformed a retrospectively adjusted catch-at-age model over the long-term.

The RTWG recommends that if the proposed assessment approach (WHAM Base model without random effects) does not meet the standards of peer review or is rejected in a future management track assessment, an alternative model be developed to integrate information from catch, age composition and potentially indices (e.g., alternative WHAM configurations).

The RTWG also proposed an alternative "Plan C" based on historical fishery performance under constant quota strategies. Under Plan C, if modeling fails, management would be based on a common sense constant catch approach considering the management history since 2001 and response in CPUE and size distribution of fish landed.

Total commercial landings (live weight) increased from less than 125 metric tons (mt) during 1967-1972 to more than 3,900 mt in 1979 during the development of the directed longline fishery (Figure 2). Landings prior to the mid-1960s were landed as a bycatch in the trawl fishery. Annual landings ranged between 454 and 1,838 mt from 1988 to 1998. Landings from 1999 to 2002 were below 900 mt (ranging from 506 to 874 mt). An annual quota of 905 mt was implemented in November of 2001. Landings in 2003 and 2004 were slightly above the quota at 1,130 mt and 1,215 mt, respectively. Landings from 2005 to 2009 were at or below the quota, while landings in 2010 at 922 mt were slightly above the quota (Figure 2). Since 2010 landings have been below the quota and decreased to an estimated 494 mt in 2016. The landings have increased slightly to an average of 693 mt from 2017 to 2022. The Total Allowable Landings (TAL) was reduced for the first time in 2015 to 796 mt from the TAL of 905 mt which was in place from 2001-2014 (Figure 2). The TAL in 2016 and 2017 was increased to 856 mt based on projections from the SARC 58 assessment. The TAL was then reduced to 738 mt from 2018 to 2021 based on the 2017 operational assessment and then increased to 834 mt in 2022 based on the 2021 management track assessment.

The top 4 permits hold 80% of the golden tilefish IFQ (individual fishing quota) allocation. As discussed in the working paper describing commercial landings (working paper 2), since the IFQ system was put in place late 2009, golden tilefish landings are more evenly spread through the year as fishermen try to avoid market gluts and promote ample supply of fish during the course of the year to support market development and promote price stability. On a temporal basis, for the last decade or so, the bulk of the landings (approximately 70%) occurred during the April through October period, with the remaining landings (approximately 30%) occurring during the November through March period (Working Paper 2). However, IFQ stakeholders have indicated that sometimes, a vessel may underharvest its quota allocation due to fear of overfishing (exceeding their individual quota allocation). Consequently, some MAFMC's Tilefish Advisory Panel Members have asked for the MAFMC to consider a carry-over of unused portions (with a small proportion of the cap) to the next fishing year as done in the scallop fishery.[1] According to stakeholders, this would not only benefit fishermen that may underharvest due to fears of exceeding their IFQ allocation, but would also benefit vessels that may not be able to land their entire allocation in one fishing year due to repairs and maintenance (MAFMC, 2014, 2015, 2016, 2017). Stakeholders have indicated that the large underage that

[1] Note: the Council has not considered an action to implement a quota carry-over.

occurred in 2015 and 2016 were due to several factors; including, inactive vessels (some IFQ allocations were not fished at all), some vessels with large allocations were out of the water for an extended period due to repairs and maintenance (about 2 months each), some vessels that leased quota did not get a chance to fish, and severe winter conditions (MAFMC, 2016, 2017). However, it is not clear if the underharvest of quota is due to management practices, fishing behavior, or poor fish availability.

Changes in the commercial CPUE can be generally explained by the impact of strong incoming year classes that track through the landings size composition over time. Since 2000, there appears to be increases in CPUE due to three strong year classes (Figures 4 and 37). In general, strong year classes and the proportion of larger fish in the catch appears to persist longer in the fishery after the FMP's quota based management was implemented in 2001. The fishery has been managed as an IFQ with a relatively small number of participants that have repeatedly expressed a desire for stability over maximized access as it benefits their markets and pricing. For all of these reasons, a constant catch approach using a quota within the range of those implemented in the fishery since 2001 (738 – 905mt) could be considered when determining an appropriate constant catch if the model fails. However, using an average of the actual catches (10 year 2013-2022 average catch of 690 mt or 20 year 2021-2022 average catch of 790 mt) may be more justified for the determination of a constant quota catch advice since this is the actual catch that appeared to have a positive effect on recruitment and seemed to allow for strong year classes to persist while supporting the fishery.

9 REFERENCES

Bell, R.J., Gervelis, B., Chamberlain, G., and Hoey, J. (2017). Discard estimates from self-reported catch data: An example from the US northeast shelf. North American Journal of Fisheries Management, 37(5), pp.1130-1144.Brooks, E.N., 2024. Pragmatic approaches to modeling recruitment in fisheries stock assessment: A perspective. Fisheries Research, 270, p.106896.

Brooks, E.N. and Deroba, J.J. (2015). When "data" are not data: the pitfalls of post hoc analyses that use stock assessment model output. Canadian Journal of Fisheries and Aquatic Sciences, 72(4), pp.634-641.

Cheng, M.L.H., Rodgveller, C.J., Langan, J.A., and Cunningham, C.J. (2023). Standardizing fishery-dependent catch-rate information across gears and data collection programs for Alaska sablefish (*Anoplopoma fimbria*). ICES Journal of Marine Science, 80(4), 1028–1042. https://doi.org/10.1093/icesjms/fsad037

Chen, Z., Curchitser, E., Chant, R., and Kang, D. (2018). Seasonal variability of the cold pool over the Mid-Atlantic Bight continental shelf. Journal of Geophysical Research: Oceans, 123(11), 8203–8226.

Chen, Z. and Curchitser, E.N. (2020). Interannual Variability of the Mid-Atlantic Bight Cold Pool. Journal of Geophysical Research: Oceans, 125(8), e2020JC016445. https://doi.org/10.1029/2020JC016445

Dettloff, K. (2021). Improvements to the Stephens-MacCall approach for calculating CPUE from multispecies fisheries logbook data. Fisheries Research, 242, 106038. https://doi.org/10.1016/j.fishres.2021.106038

Dick, E.J. and MacCall, A.D. (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.

Dorn, M.W. and Zador, S.G. (2020). A risk table to address concerns external to stock assessments when developing fisheries harvest recommendations. Ecosystem Health and Sustainability, 6(1), p.1813634.

Drew, K. (2022). Fishery-dependent CPUE index for bluefish derived from MRIP data. Working paper for the 2022 Bluefish research track assessment. https://static1.squarespace.com/static/511cdc7fe4b00307a2628ac6/t/64501a9ee391c8439cbc3c04/1682971295614/c_MRIP+Index_Drew+2022d.pdf

du Pontavice, H., Miller, T.J., Stock, B.C., Chen, Z., and Saba, V.S. (2022). Ocean model-based covariates improve a marine fish stock assessment when observations are limited. ICES Journal of Marine Science, 79(4), 1259–1273.

Fisher, J.A.D., Frank, K.T., Petrie, B., and Leggett, W.C. (2014). Life on the edge: Environmental determinants of tilefish (*Lopholatilus chamaeleonticeps*) abundance since its virtual extinction in 1882. ICES Journal of Marine Science, 71(9), 2371–2378. https://doi.org/10.1093/icesjms/fsu053

Freeman, B.L. and Turner, S.C. (1977). Biological and fisheries data on tilefish, *Lopholatilus chamaeleonticeps.* Goode and Bean (Technical Series) [Report No. 5].

Grimes, C.B., Idelberger, C.F., Able, K.W., and Turner, S.C. (1988). The reproductive biology of tilefish, Lopholatilus chamaeleonticeps Goode and Bean, from the United States Mid-Atlantic Bight, and the effects of fishing on the breeding system.

Grimes, C.B. and Turner, S.C. (1999). The Complex Life History of Tilefish *Lopholatilus chamaeleonticeps* and Vulnerability to Exploitation. American Fisheries Society Symposium, 23, 17–26.

Haltuch, M.A., Brooks, E.N., Brodziak, J., Devine, J.A., Johnson, K.F., Klibansky, N., Nash, R.D.M., Payne, M.R., Shertzer, K.W., and Subbey, S. (2019). Unraveling the recruitment problem: A review of environmentally-informed forecasting and management strategy evaluation. Fisheries Research, 217, 198–216.

Hoyle, S.D., Campbell, R.A., Ducharme-Barth, N.D., Grüss, A., Moore, B.R., Thorson, J.T., Tremblay-Boyer, L., Winker, H., Zhou, S., and Maunder, M.N. (2024). Catch per unit effort modelling for stock assessment: A summary of good practices. Fisheries Research, 269, 106860. https://doi.org/10.1016/j.fishres.2023.106860

Jones, A.W., Lowman, B.L., Manderson, J.M., and Mercer, A. (2020). An investigation of fine-scale CPUE for northern shortfin squid (*Illex illecebrosus*) using NEFSC Study Fleet data. Working Paper for the MAFMC *Illex* Working Group and Science and Statistical Committee.

Jones, A.W., Salois, S.L. (2024). Exploring a CPUE index from trawl gear using high-resolution catch data. Working paper for the 2024 Golden Tilefish Research Track Assessment.

Legault, C.M. and Restrepo, V.R. (1998). A flexible forward age-structured assessment program. ICCAT. Col. Vol. Sci. Pap, 49, 246-253.

Legault, C.M., Wiedenmann, J., Deroba, J.J., , Fay, G., Miller, T.J., Brooks, E.N., Bell, R.J., Langan, J.A., Cournane, J.M., Jones, A.W., and Muffley, B. (2023). Data-rich but model-resistant: an evaluation of data-limited methods to manage fisheries with failed age-based stock assessments. Canadian Journal of Fisheries and Aquatic Sciences 80 (1): 27-42.

Li, C., Deroba, J., Miller, T., Legault, C., and Peretti, C. (2024). “An evaluation of common stock assessment diagnostic tools for choosing among state-space models with multiple random effects processes.” Fisheries Research. In press.

Liljestrand, E.M., James R.B., and Deroba J.J. (2023). "Applying a novel state-space stock assessment framework using a fisheries-dependent index of fishing mortality." Fisheries Research 264: 106707.

Linder, C.A. and Gawarkiewicz, G. (1998). A climatology of the shelfbreak front in the Middle Atlantic Bight. Journal of Geophysical Research: Oceans, 103(C9), 18405–18423. https://doi.org/10.1029/98JC01438

MAFMC. (2000). Tilefish Fishery Management Plan. Dover, DE. https://www.mafmc.org/tilefish

MAFMC. (2014). 2014 Tilefish Advisory Panel Fishery Performance Report. Dover, DE. https://www.mafmc.org/fishery-performance-reports

MAFMC. (2015). 2015 Tilefish Advisory Panel Fishery Performance Report. Dover, DE. https://www.mafmc.org/fishery-performance-reports

MAFMC. (2016). 2016 Golden Tilefish Advisory Panel Fishery Performance Report. Dover, DE. https://www.mafmc.org/fishery-performance-reports

MAFMC. (2017). 2017 Advisory Panel Golden Tilefish Fishery Performance Report. Dover, DE. https://www.mafmc.org/fishery-performance-reports

MAFMC. (2023). Golden Tilefish Fishery Performance Report. https://www.mafmc.org/fishery-performance-reports

Marshall, K.N., Koehn, L.E., Levin, P.S., Essington, T.E., and Jensen, O.P. (2019). Inclusion of ecosystem information in US fish stock assessments suggests progress toward ecosystem-based fisheries management. ICES Journal of Marine Science, 76(1), pp.1-9.

McBride, R.S., Vidal, T.E., and Cadrin, S.X. (2013). Changes in size and age at maturity of the northern stock of Tilefish (*Lopholatilus chamaeleonticeps*) after a period of overfishing. Fish. Bull., 111, pp. 161-174.

Miller, T.J. and Legault, C.M. (2015). Technical details for ASAP version 4. US Dept Commer, Northeast. Fish Sci Cent Ref Doc. 15-17; 136 p. https://doi.org/10.7289/V57W695G

NEFSC. (1993). Report of the 16th Northeast Regional Stock Assessment Workshop (16th SAW) Stock Assessment Review Committee (SARC) Consensus Summary of Assessments. NEFSC Ref Doc. 93-18. 108 p.

NEFSC. (2006) (Revised and Reissued). 41st Northeast Regional Stock Assessment Workshop (41st SAW). US Dept Commer, Northeast Fish Sci Cent Ref Doc. 05-14; 237 p. https://apps-nefsc.fisheries.noaa.gov/rcb/publications/assessment-documents.html

NEFSC. (2009). 48th Northeast Regional Stock Assessment Workshop (48th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-15; 834 p. https://apps-nefsc.fisheries.noaa.gov/rcb/publications/assessment-documents.html

NEFSC. (2014). 58th Northeast Regional Stock Assessment Workshop (58th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 14-04; 784 p. https://apps-nefsc.fisheries.noaa.gov/rcb/publications/assessment-documents.html

NEFSC. (2016). Black Sea Bass Benchmark Stock Assessment Review Term of Reference 1 – Spatial issues Proposed partitioning of northern black sea bass stock for purposes of developing spatial stock assessment models. https://static1.squarespace.com/static/511cdc7fe4b00307a2628ac6/t/56cb73b5e707ebc39

cec6bf3/1456174007277/Spatial+report+to+MAFMC+SSC_draft.pdf
NEFSC. (2020). Draft report of the Index Based Methods Working Group, submitted for peer review November 20th 2020. 128 pp. https://static1.squarespace.com/static/511cdc7fe4b00307a2628ac6/t/60391d7c0a8d8429d50f72b2/1614355837553/Draft+Index+Based+Methods+Peer+Review+Report.pdf

Nesslage G., Lyubchich V., Nitschke P., Williams E., Grimes C., and Wiedenmann J. (2021). Environmental drivers of golden tilefish (*Lopholatilus chamaeleonticeps*) commercial landings and catch-per-unit-effort. Fish Oceanogr. 2021;00:1–15. https://doi.org/10.1111/fog.12540

Nitschke, P. (2021). Golden Tilefish, *Lopholatilus chamaeleonticeps,* Management Track Assessment through 2020 in the Middle Atlantic-Southern New England Region.

Palmer, M.C., Hersey, P., Marotta, H., Shield, G.R., and Cierpich, S.B. (2016). The design and performance of an automated observer deployment system for the Northeastern United States groundfish fishery. Fisheries Research, 179, pp.33-46.

Pérez-Hernández, M.D. and Joyce, T.M. (2014). Two Modes of Gulf Stream Variability Revealed in the Last Two Decades of Satellite Altimeter Data. *Journal of Physical Oceanography*, *44*(1), 149–163. https://doi.org/10.1175/JPO-D-13-0136.1

Phillips, A.J., Ciannelli, L., Brodeur, R.D., Pearcy, W.G., and Childers, J. (2014). Spatio-temporal associations of albacore CPUEs in the Northeastern Pacific with regional SST and climate environmental variables. ICES Journal of Marine Science, 71(7), pp.1717-1727.

Shotwell, S.K., Blackhart K., Cunningham C., Fedewa E., Hanselman D., Aydin K., Doyle M., Fissel B., Lynch P., Ormseth O., Spencer P., and Zador S. (2023). Introducing the Ecosystem and Socioeconomic Profile, a Proving Ground for Next Generation Stock Assessments. Coastal Management, 51(5–6), 319–352. https://doi.org/10.1080/08920753.2023.2291858

Salois, S.L., Owen S., Molina A., Jones A.W., and Hyde K.J.W. (2024). Golden Tilefish Ecosystem and Socioeconomic Profile. Working paper for the 2024 Golden Tilefish Research Track Assessment.

Sharma, R., Porch C.E., Babcock E.A., Maunder M., and Punt A.E., editors. (2019). Recruitment: Theory, Estimation, and Application in Fishery Stock Assessment Models. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-148. https://repository.library.noaa.gov/view/noaa/19834

Sosebee, K.A. and Cadrin, S.X. (2006). A historical perspective on the abundance and biomass of Northeast complex stocks from NMFS and Massachusetts inshore bottom trawl surveys, 1963-2002. U.S. Dep. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 06-05; 200 p.

Steimle, F.W., Zetlin, C.A., Berrien, P.L., Johnson, D.L., and Chang, S. (1999). Essential Fish Habitat Source Document: Tilefish, *Lopholatilus chamaeleonticeps*, Life History and Habitat Characteristics (NOAA Technical Memorandum NMFS-NE-152).

Tommasi, D., Stock, C.A., Hobday, A.J., Methot, R., Kaplan, I.C., Eveson, J.P., Holsman K., Miller, T.J., Gaichas, S., Gehlen, M., and Pershing, A. (2017). Managing living marine resources in a dynamic environment: the role of seasonal to decadal climate forecasts. Progress in Oceanography, 152, pp.15-49.

Turner S.C. (1986). Population dynamics of and, impact of fishing on tilefish, *Lopholatilus chamaeleonticeps*, in the Middle Atlantic-Southern New England region during the

1970's and early 1980's. New Brunswick, N.J.: Rutgers University. Ph.D. dissertation. 289 p.
Wenner, E.L. and Barans, C.A. (2001). Benthic habitats and associated fauna of the upper- and middle-continental slope near the Charleston Bump. American Fisheries Society Symposium, 25, 161–175.

10 TABLES

Table 1. Environmental indicator trends with study fleet/observer CPUE and model recruitment estimates with lags. Arrows represent direction of trend with negative trends indicated by a downward facing arrow and positive trends indicated by an upwards facing arrow. If the trend was significant, the arrows are red (negative) or blue (positive). Black arrows represent trends that were non-significant. Dashed lines indicate no trend.

	CPUE No Lag (Age 4)	CPUE 3y Lag (Age 1)	Rec. Estimate (Age 1)	Rec. 1y Lag (Age 0)
SST	↓	↓	—	—
BT	—	—	—	—
Salinity	—	—	—	↓
SW Volume	—	↓	↓	↑ spring
SW Temp.	↓	↑	—	—
SW Salinity	—	↑	—	—
GSI	↓	↑ winter ↓ summer	—	—
CP Extent	—	↓	↓	↓
CP Persistence	—	↓	↓	↓
CP Index	—	—	↑	—
Microplankton	↑	↑ fall	—	↑ fall
CHL-a	↑	—	—	—

	No trend, but matches literature	↑	Positive significant trend	—	No trend
	Not tested	↓	Negative significant trend	↑↓	Non-significant trend (+ or -)

Table 2. Number of observed trips, discard ratios (discard/ sum all species kept), estimated CVs, and estimated discards in metric tons for large and small mesh trawl and gillnet gear.

	longline		trawl				gillnet		
			LG		SM				
year	mt	cv	mt	cv	mt	cv	mt	cv	total Mt
1989			13.7	0.46	15.1	0.53			28.7
1990			0.0	0.00	1.3	0.57			1.3
1991			1.6	1.46	15.3	0.56			17.0
1992			4.7	0.95	19.8	0.77			24.5
1993			0.4	1.08	7.4	0.32			7.8
1994			1.2	0.64	2.6	0.55			3.8
1995			2.1	0.99	0.7	1.95	0.2	1.19	3.0
1996			0.5	0.73	5.2	0.92			5.7
1997			0.1	2.31	4.7	1.98	0.6	0.84	5.4
1998			4.5	1.65	3.8	0.69			8.3
1999			7.5	0.65	0.6	0.97			8.1
2000			4.2	1.19	2.7	0.69			6.8
2001			78.5	0.40	0.7	0.76			79.1
2002			0.0	0.93	1.4	1.07			1.4
2003			0.6	0.78	25.7	0.69	0.5	0.90	26.8
2004	3.3	1.40	6.3	0.56	0.9	0.70	4.8	0.28	15.4
2005			0.2	0.83	2.5	0.56	2.3	0.70	5.0
2006	2.3	0.48	1.0	0.66	2.6	0.76			6.0
2007			0.0	0.87	0.1	1.49	4.3	0.96	4.4
2008			0.2	0.52	3.8	0.83	0.5	0.91	4.5
2009			2.5	0.48	1.2	0.44	2.5	0.55	6.2
2010			0.5	0.37	2.7	0.50	0.5	0.85	3.8
2011			2.2	0.59	1.2	0.39	3.3	0.37	6.7
2012			1.0	0.78	1.1	0.67	1.6	0.43	3.7
2013			0.9	0.42	0.5	0.55	4.0	0.42	5.4
2014	3.6	0.63	0.0	0.67	1.2	0.45	2.1	0.44	6.9
2015	4.2	1.71	0.2	0.53	3.6	0.47	0.6	0.53	8.6
2016	0.5	0.84	0.5	0.43	4.7	0.31			5.6
2017	2.6	0.45	3.8	0.37	2.6	0.39	0.2	0.67	9.2
2018	0.5	0.93	0.3	0.69	2.3	0.27	0.6	0.47	3.6
2019	3.4	0.48	0.1	0.49	2.7	0.28	0.3	0.51	6.5
2020	2.0	0.98	0.7	0.38	6.9	0.49			9.6
2021	2.0	0.57	0.4	0.76	1.1	0.49	0.9	0.57	4.4

Table 3. Recreational catch time series, 1971-2022.

Year	mt	Year	mt		
1971	3	1997	3		
1972	39	1998	3		
1973	75	1999	3		
1974	100	2000	3		
1975	60	2001	3		
1976	50	2002	3		
1977	25	2003	3	Estimated proportion of landings	
1978	5	2004	3	Recreational	Commercial
1979	5	2005	3	0.4%	99.6%
1980	3	2006	3	0.3%	99.7%
1981	3	2007	5	0.7%	99.3%
1982	3	2008	3	0.4%	99.6%
1983	3	2009	3	0.4%	99.6%
1984	3	2010	4	0.4%	99.6%
1985	3	2011	11	1.2%	98.8%
1986	3	2012	21	2.5%	97.5%
1987	3	2013	18	2.0%	98.0%
1988	3	2014	17	2.0%	98.0%
1989	3	2015	23	3.7%	96.3%
1990	3	2016	14	2.7%	97.3%
1991	3	2017	16	2.2%	97.8%
1992	3	2018	16	2.1%	97.9%
1993	3	2019	17	2.4%	97.6%
1994	3	2020	15	2.3%	97.7%
1995	3	2021	22	2.9%	97.1%
1996	3	2022	22	3.2%	96.8%

Table 4. Estimated biological reference points from the Base model, Base_NAAiid and Full_RE WHAM models using a recent 10 year average (top) and a recent 5 year average (bottom).

10 Year Average (2011-2020)

Model	$F_{40\%}$	$SSB_{40\%}$	MSY at $SSB_{40\%}$	Mean Recruitment
Base	0.265	9,314	855	1,339
Base_NAAiid	0.238	8,014	791	1,148
Full_RE	0.138	8,195	1,075	1,181

5 Year Average (2016-2020)

model	$F_{40\%}$	$SSB_{40\%}$	MSY at $SSB_{40\%}$	Mean Recruitment
Base	0.264	9,925	868	1,339
Base_NAAiid	0.237	8,539	808	1,148
Full_RE	0.137	8,733	1,127	1,181

Table 5. Fishing mortality and SSB estimates for 2020 and stock status shown with $F/F_{40\%}$ and $SSB/SSB_{40\%}$ ratios for the Base model, Base_NAAiid and Full_RE WHAM models using a recent 10 year average.

Model	F_{2020}	SSB_{2020}	$F/F_{40\%}$	$SSB/SSB_{40\%}$
Base	0.146	11,980	0.55	1.29
Base_NAAiid	0.190	5,246	0.80	0.65
Full_RE	0.223	2,567	1.61	0.31

Table 6. Catch from $F_{40\%}$ 10 year projections for the base model ($F_{40\%} = 0.265$) and the Full_RE ($F_{40\%} = 0.138$) model shown here as a sensitivity.

Year	Base	Full_RE
2021	1,023	475
2022	1,096	616
2023	1,068	765
2024	981	883
2025	939	935
2026	925	983
2027	905	1,038
2028	895	1,072
2029	889	1,080
2030	883	1,080

11 FIGURES

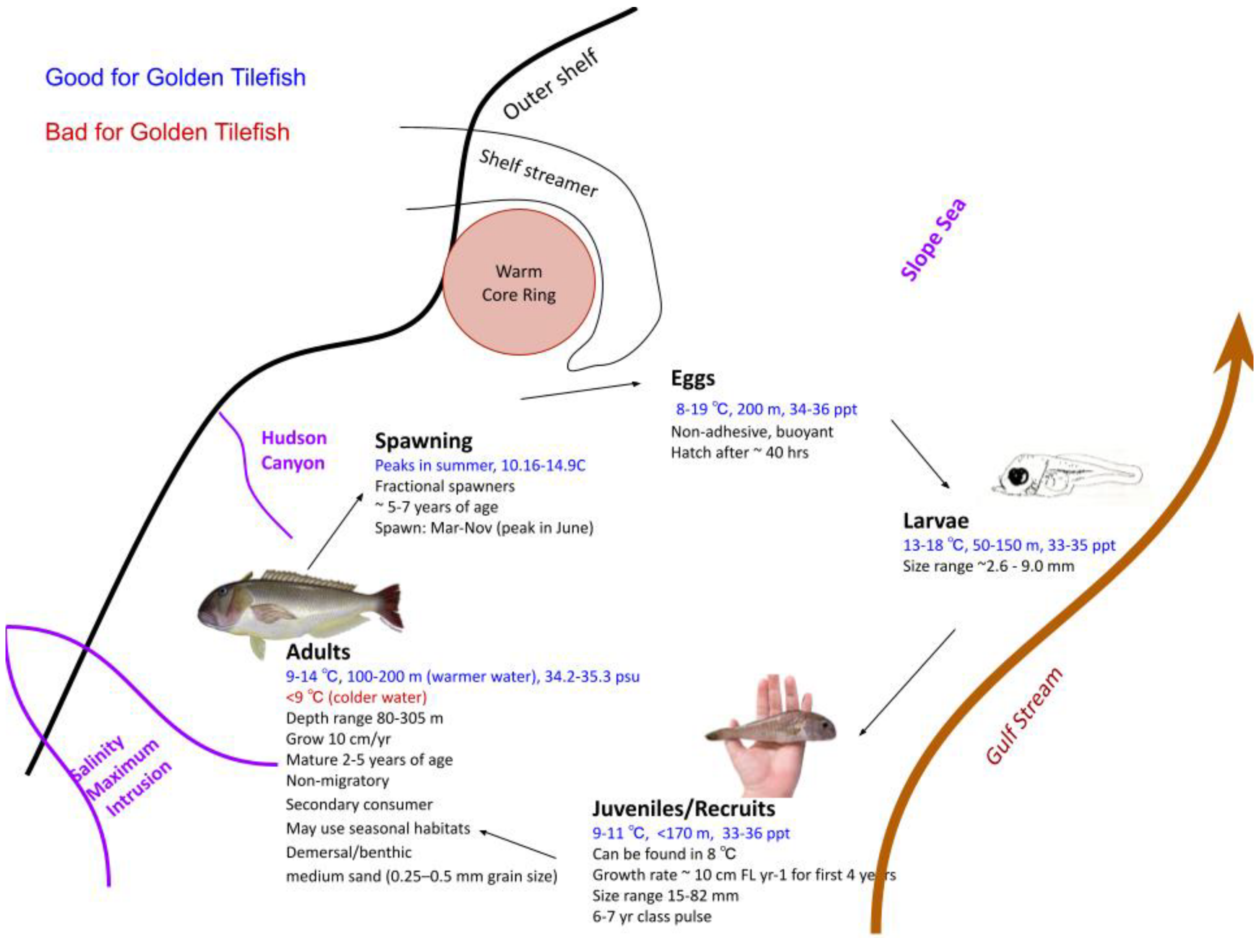


Figure 1. Conceptual model of golden tilefish life history stages and potential ecosystem impacts.

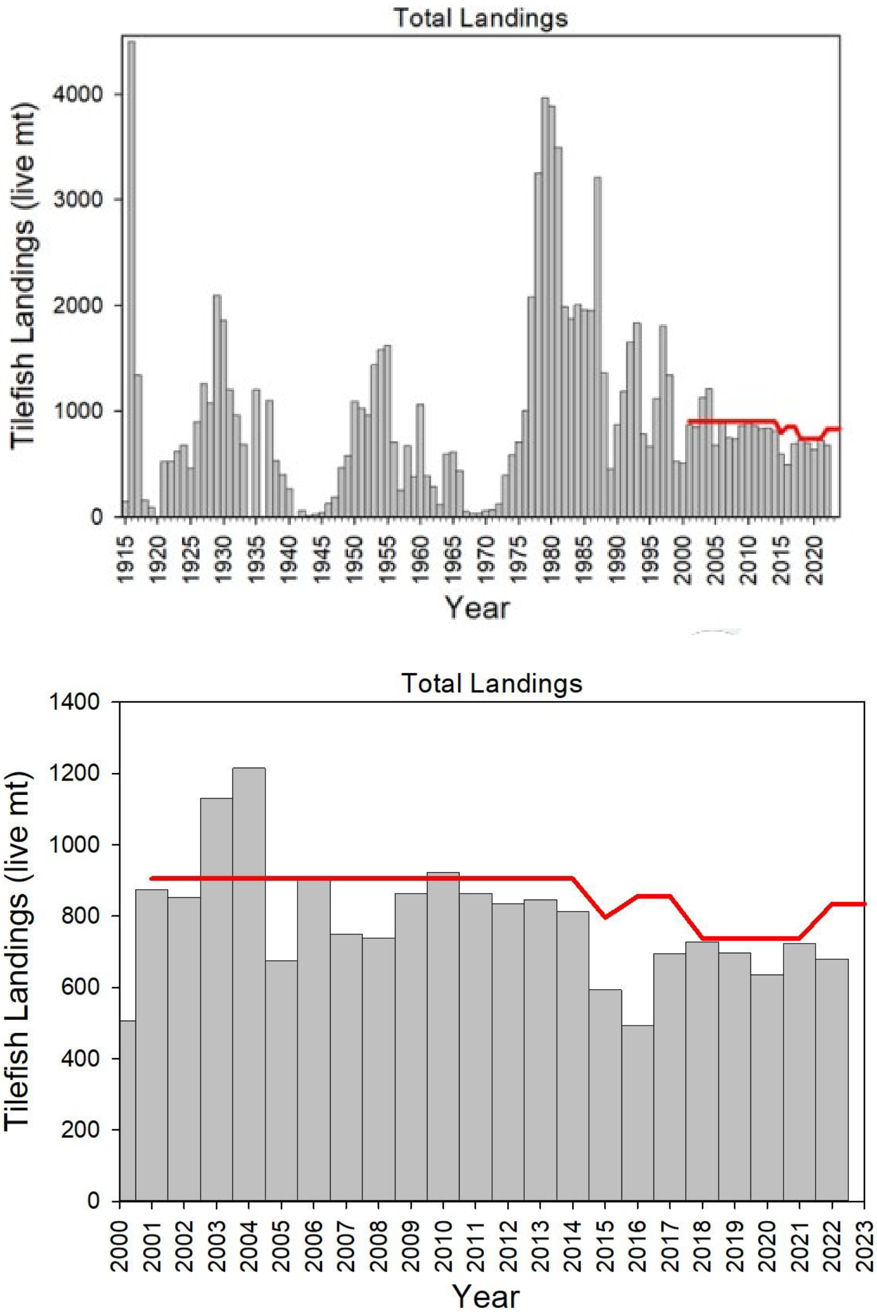


Figure 2. Landings of tilefish in metric tons from 1915-2022 (top) and from 2000-2022 (bottom). Landings in 1915-1972 are from Freeman and Turner (1977), 1973-1989 are from the general canvas data, 1990-1993 are from the weighout system, 1994-2003 are from the dealer reported data, and 2004-2020 are from dealer electronic reporting. Red line is the Total Allowable Landings (TAL) from 2001-2022.

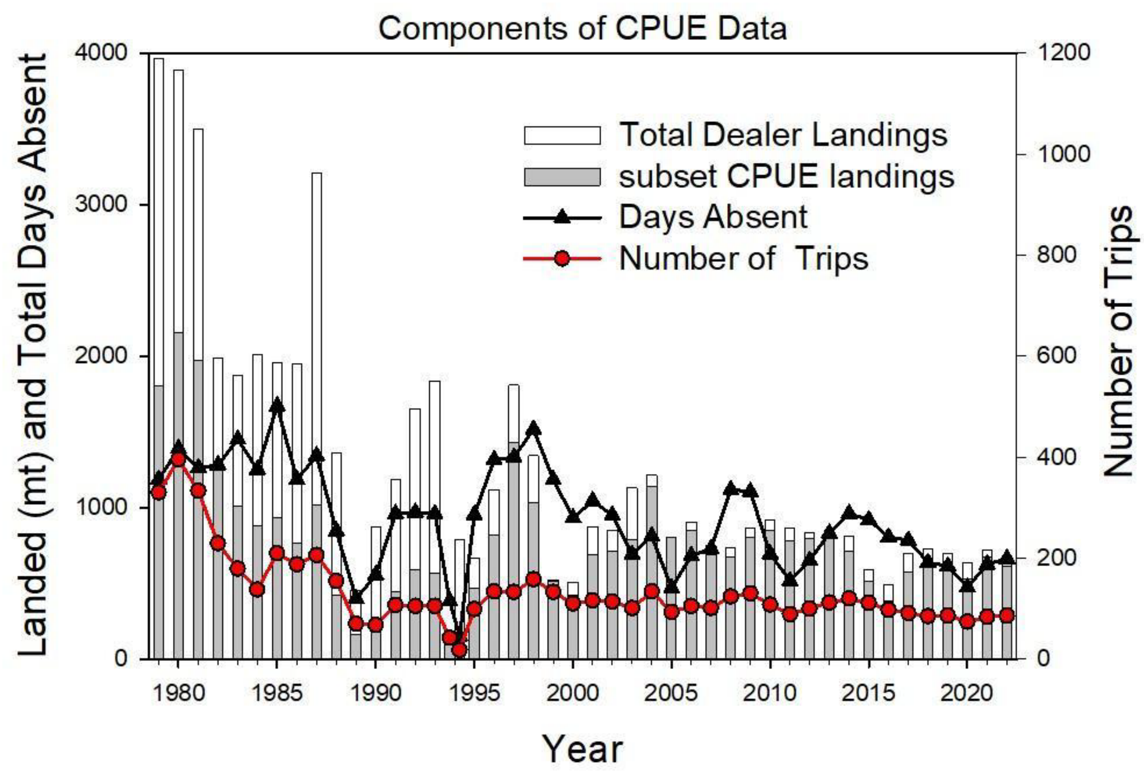


Figure 3. Total number of trips and days absent for trips targeting tilefish (= or >75% tilefish) from 1979-2022. Total Dealer and LPUE subset landings are also shown.

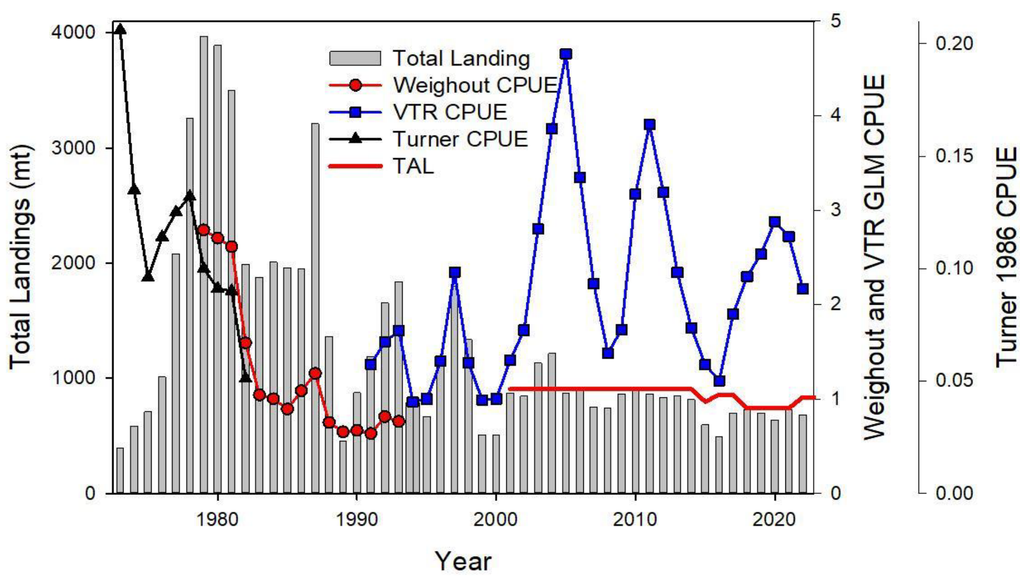


Figure 4. GLM LPUE for the weighout and VTR data split into two series with additional New York logbook LPUE data from three vessels (1991-1994) added to the VTR series. Four years of overlap between Turner's and the weighout LPUE series can also be seen. ASAP relative changes in qs among LPUE series were not incorporated into the plot. Assumed total landings are also shown. Landings in 2005 were taken from the IVR system. Red line is the TAL.

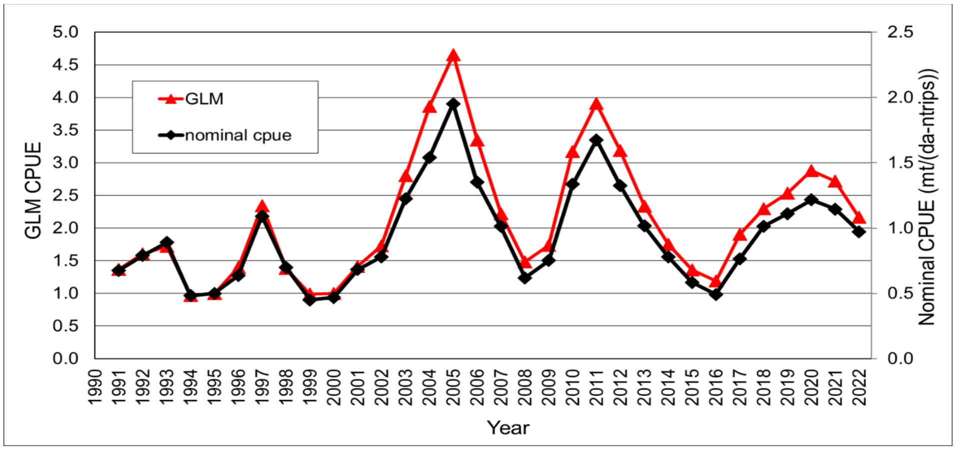


Figure 5. Comparison of nominal and vessel standardized GLM LPUE indices.

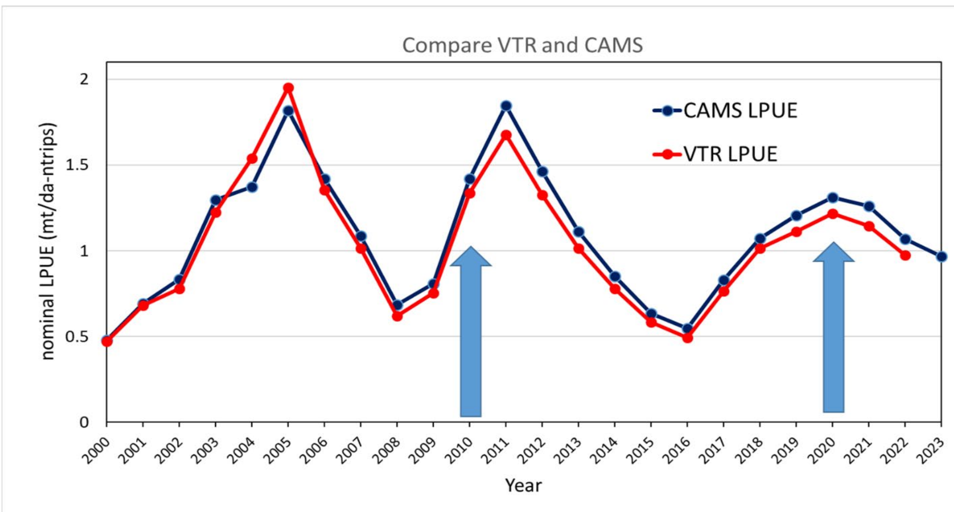


Figure 6. Comparison of nominal VTR and CAMS based LPUE indices. Arrows indicate years the RTWG considered for stitching the two series together.

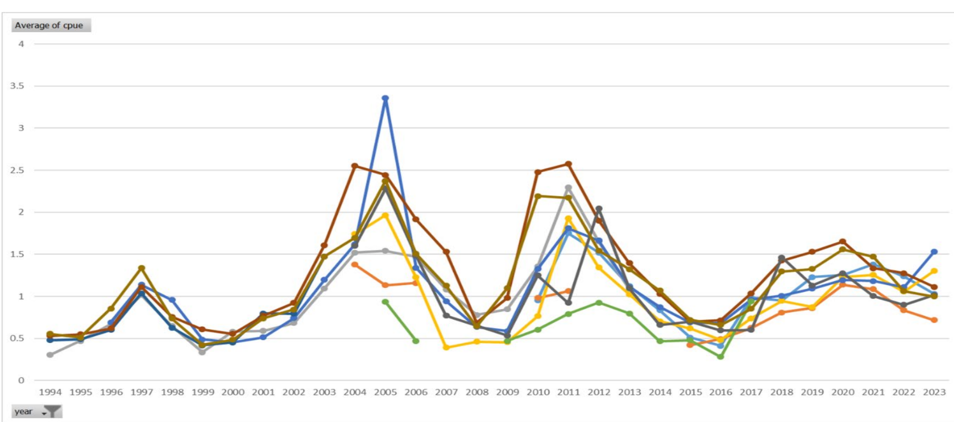


Figure 7. Comparison of individual tilefish vessel specific nominal LPUE series.

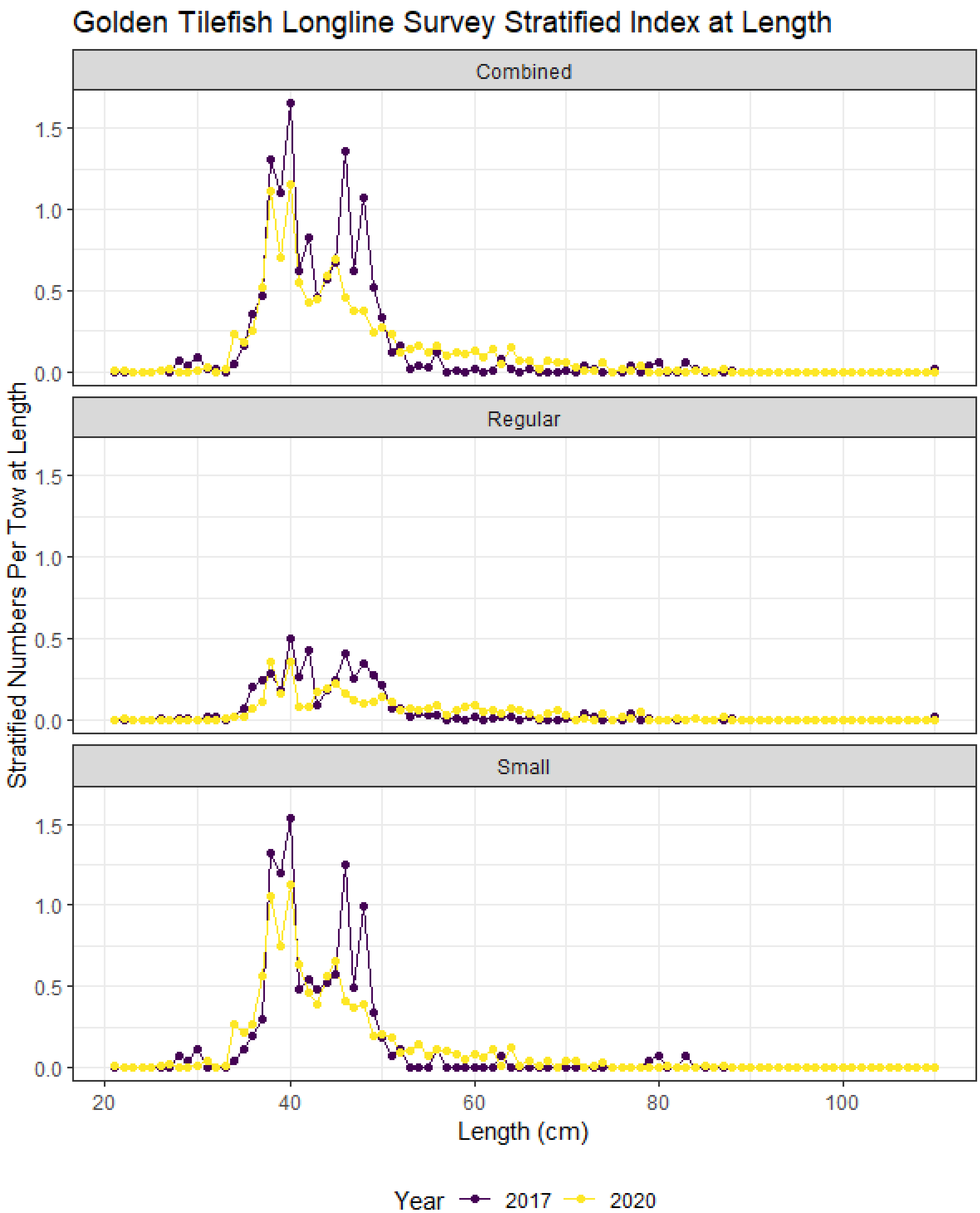


Figure 8. Stratified numbers per tow at length indices of golden tilefish in 2017 and 2020 using small and regular hooks combined (top), only regular hooks (middle), and only small hooks (bottom).

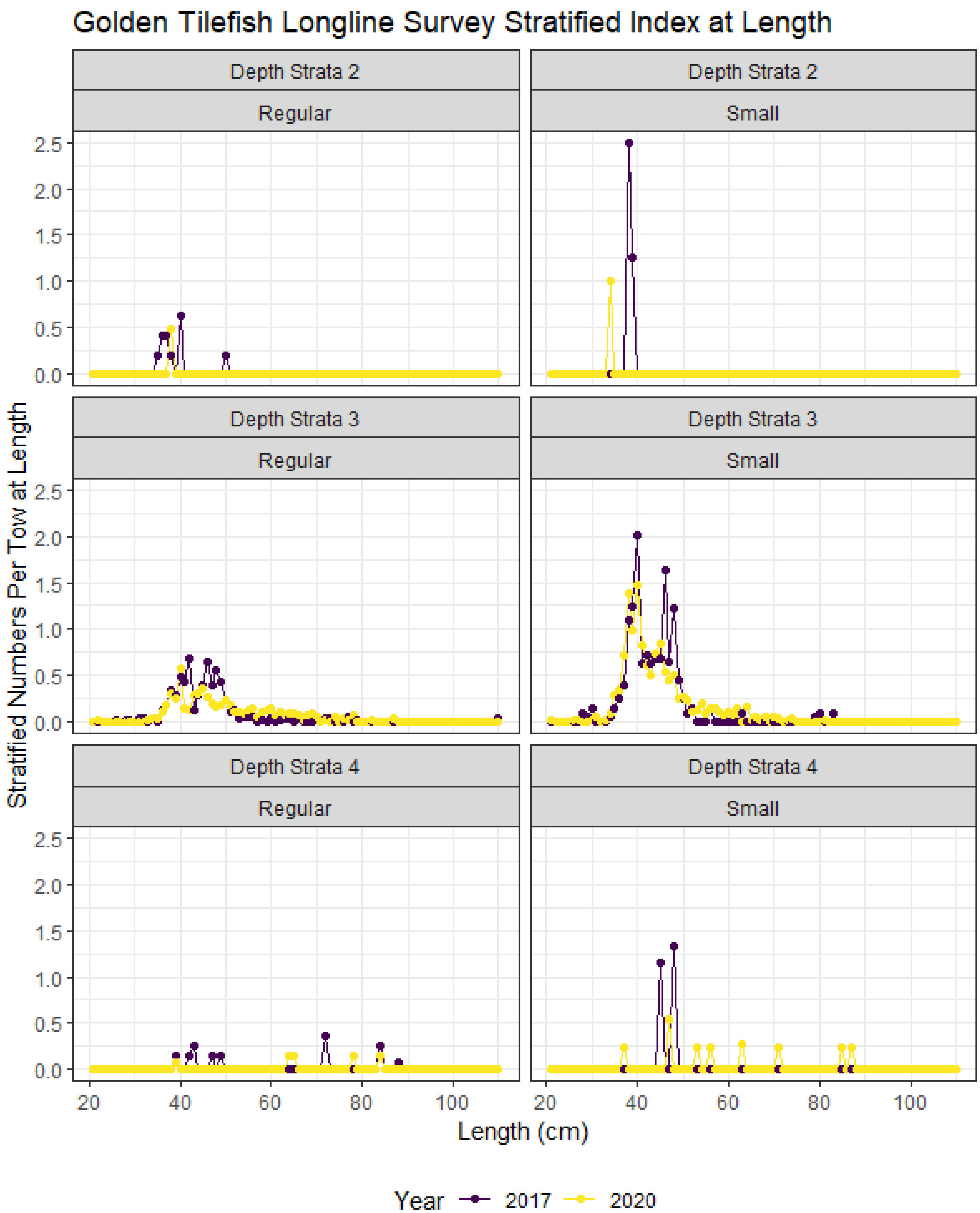


Figure 9. Stratified numbers per tow at length indices of golden tilefish in 2017 and 2020 by depth strata using only regular hooks (left) and only small hooks (right). Depth stratum ranges (in meters): 2 = 82.3-98.6, 3 = 98.8-252.2 and 4 = 252.4-303.6. A weak effect of hook size was observed, with smaller hooks having a slight shift to smaller fish.

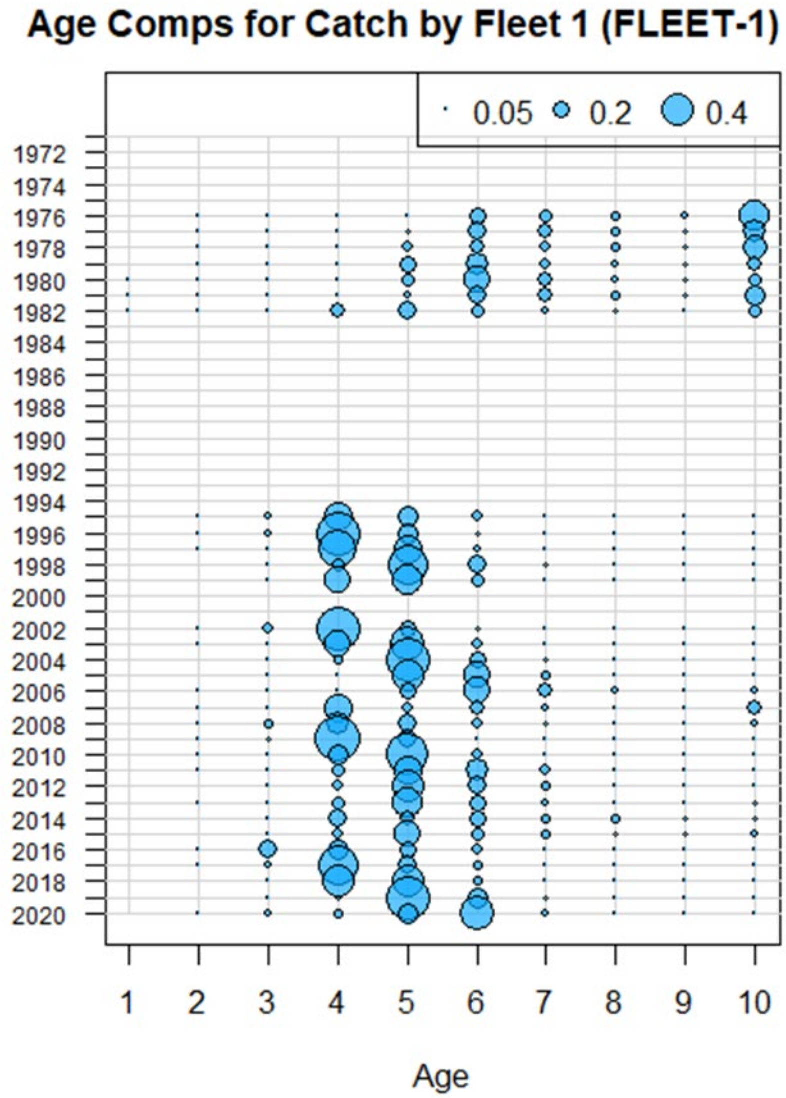


Figure 10. Catch at age input from the 2021 management track assessment used for ASAP and WHAM.

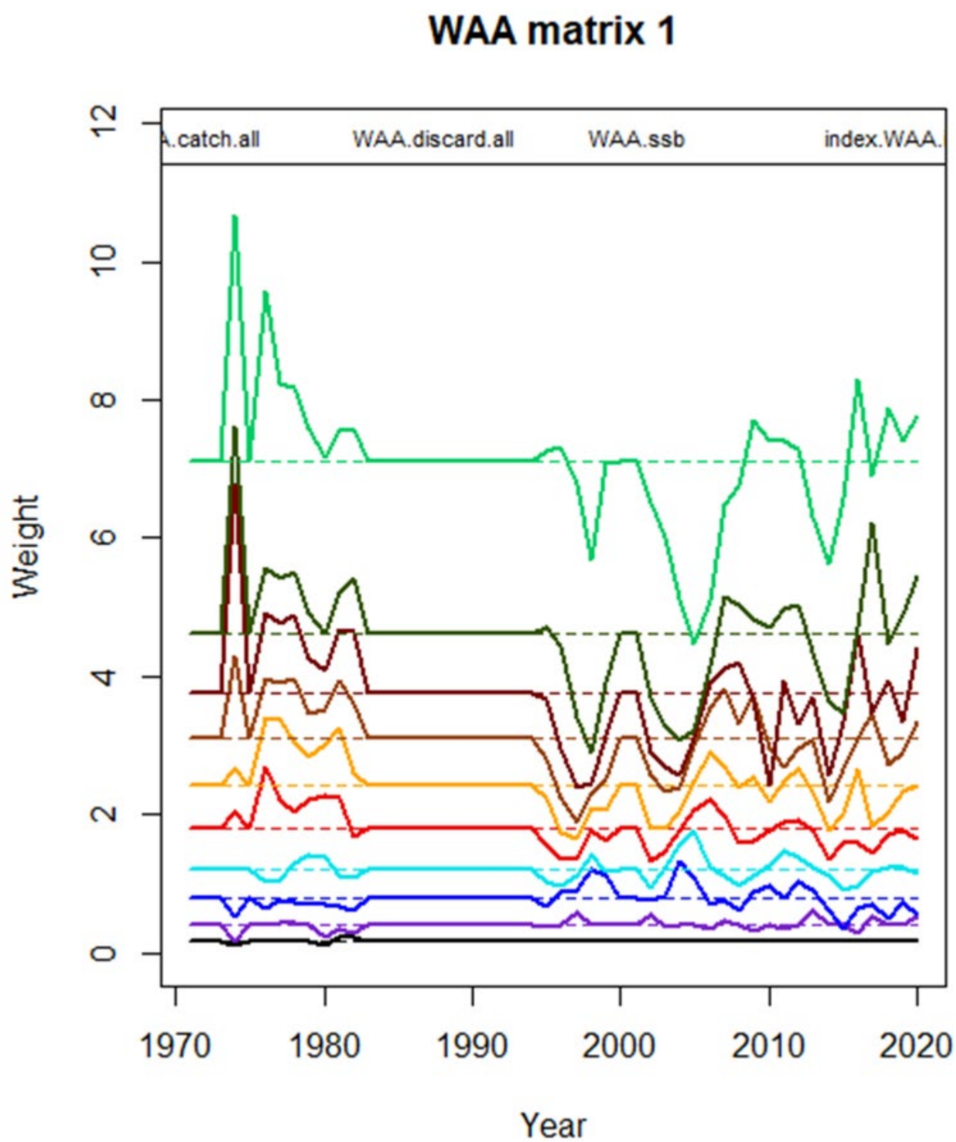


Figure 11. Mean weight at age input from the 2021 management track assessment used for ASAP and WHAM with terminal year 2020.

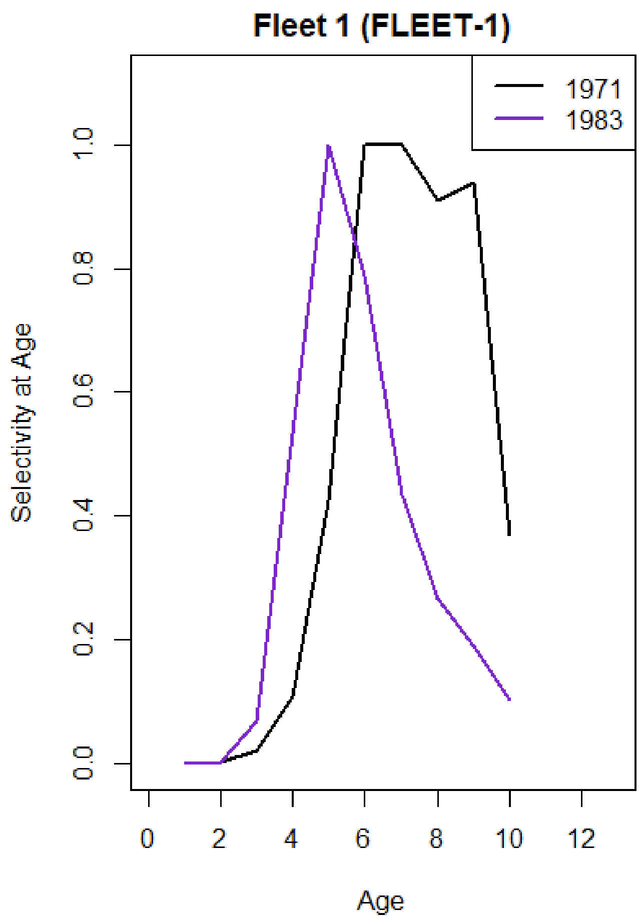


Figure 12. Estimated selectivity from the 2021 management track ASAP model for two selectivity blocks (1971 to 1982 in black and 1983 to 2020 in purple).

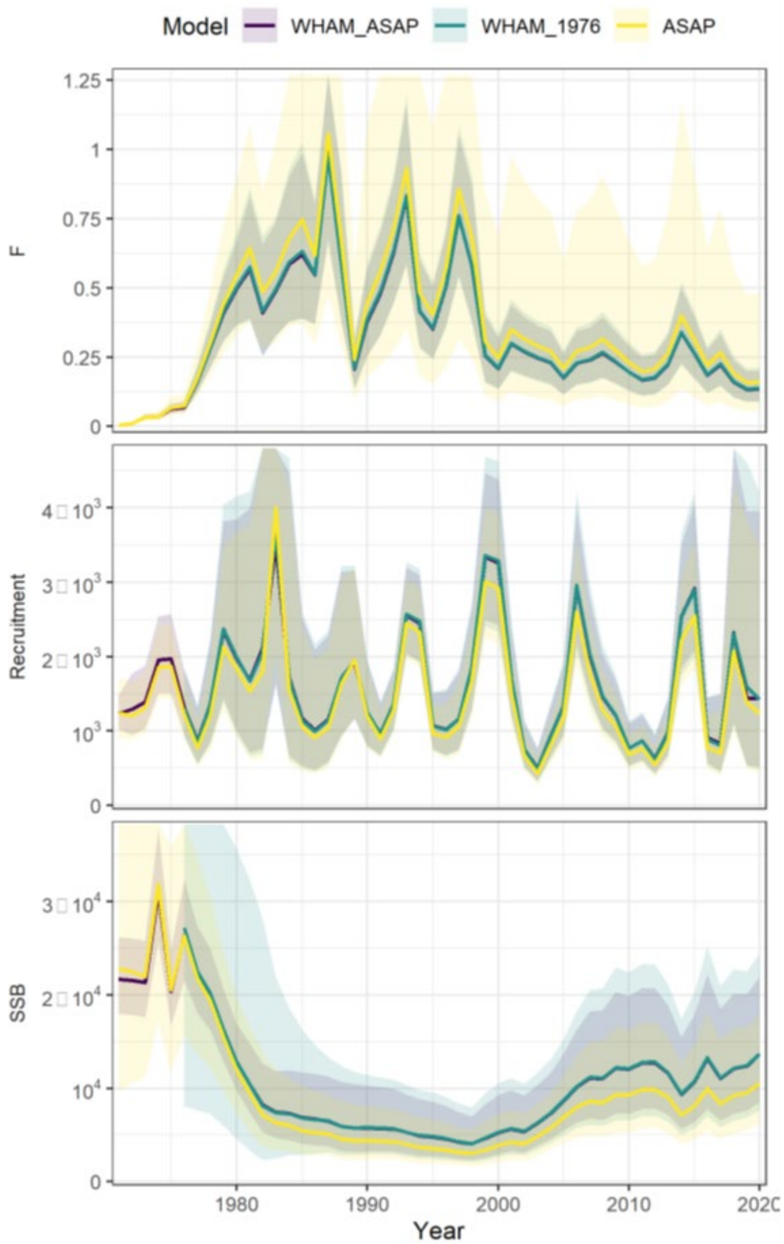


Figure 13. Comparison the 2021 management track ASAP run (yellow), to WHAM model runs (purple 1971 start & turquoise 1976 start) for fishing mortality, recruitment (age 1) and SSB.

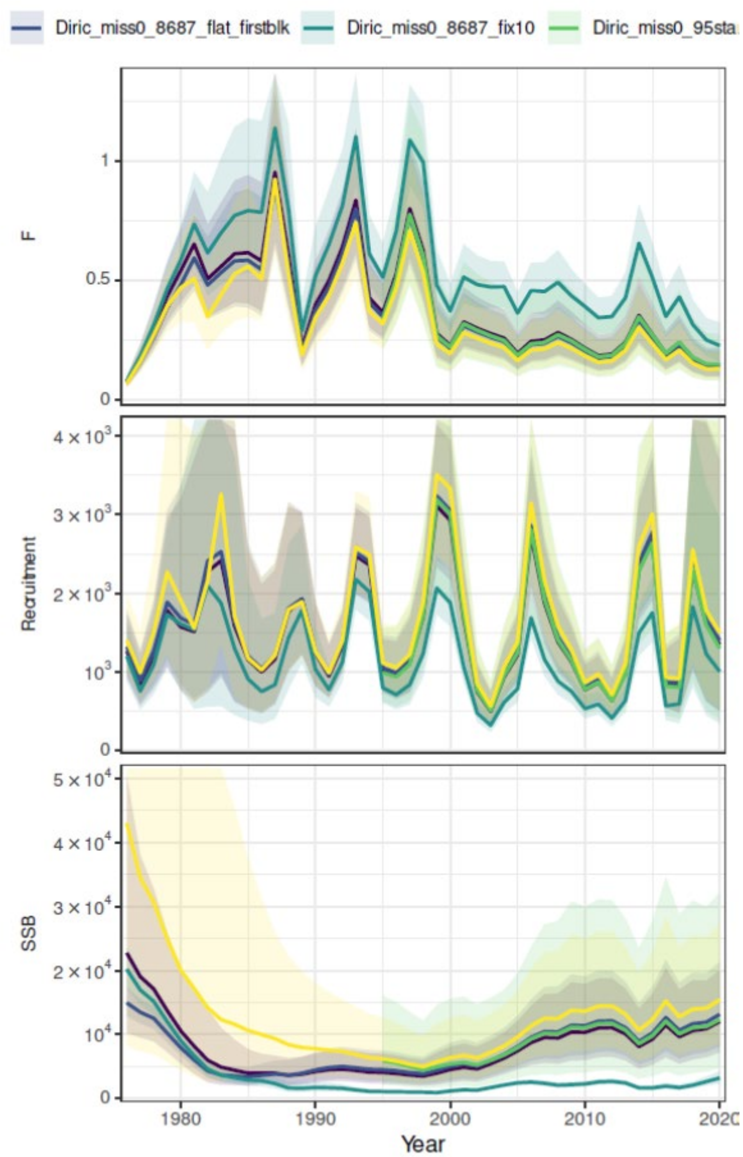


Figure 14. Comparison of WHAM runs results for fishing mortality, recruitment (age 1) and SSB. The model in yellow is an example of a run which starts at unrealistic high biomass. Fixing the 10+ selectivity at 0.5 produces the low SSB run from 1990 to 2020 (turquoise). Other runs all produce similar results with a run starting in 1995 in green.

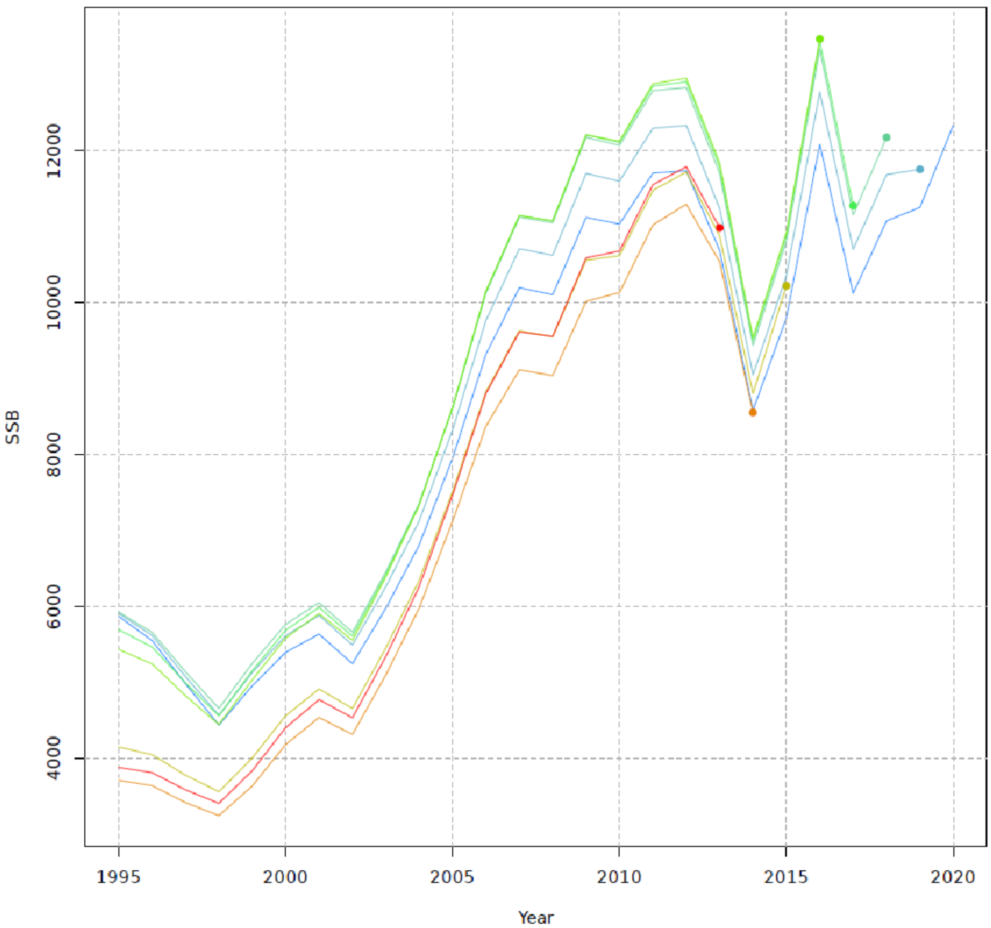


Figure 15. SSB retrospective plot for the WHAM run which starts in 1995.

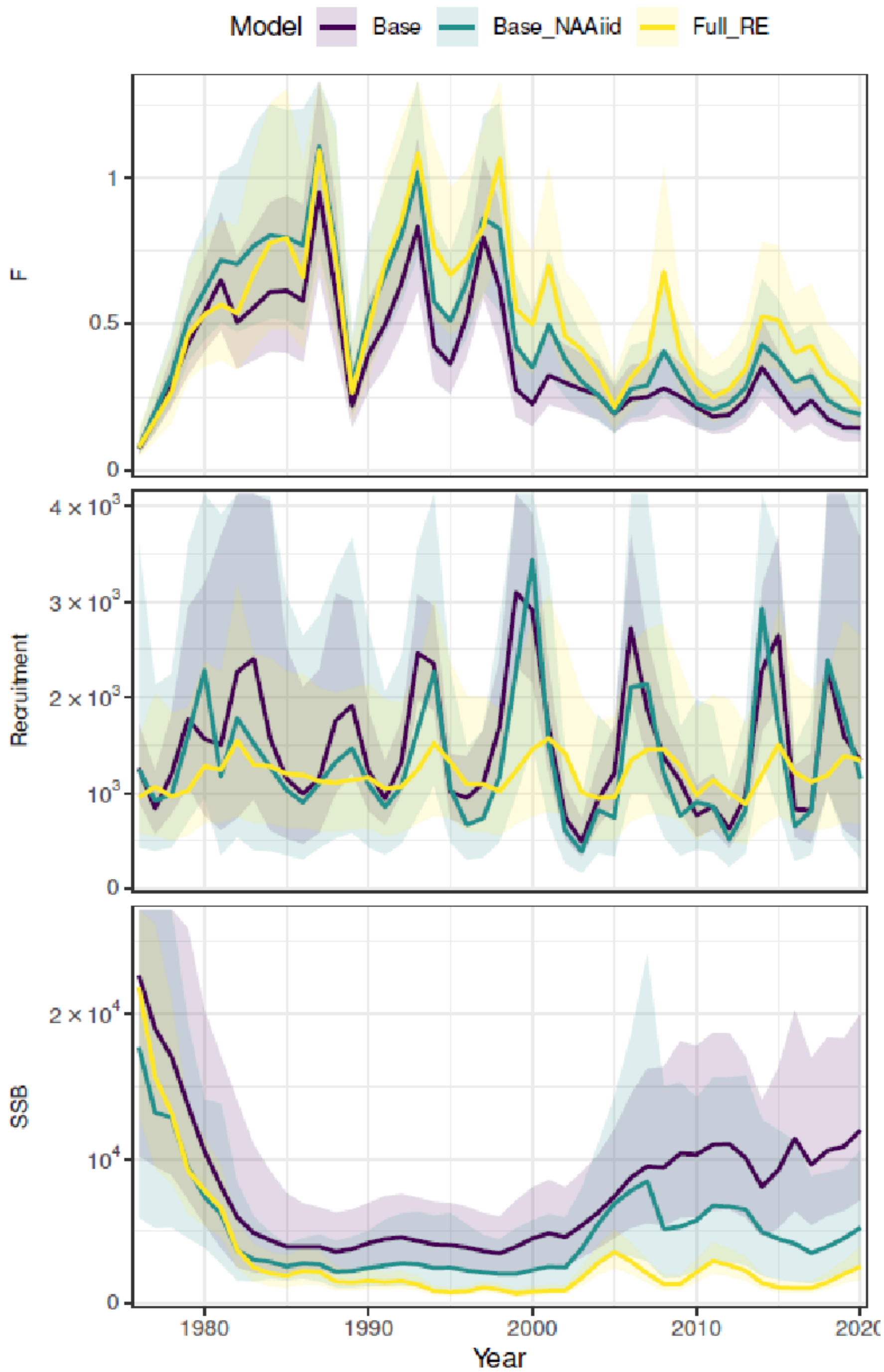


Figure 16. Comparison of WHAM Base (purple), Base with NAA iid random effects (turquoise) and the full random effects (selectivity & NAA) run (yellow) for fishing mortality, recruitment (age 1) and SSB.

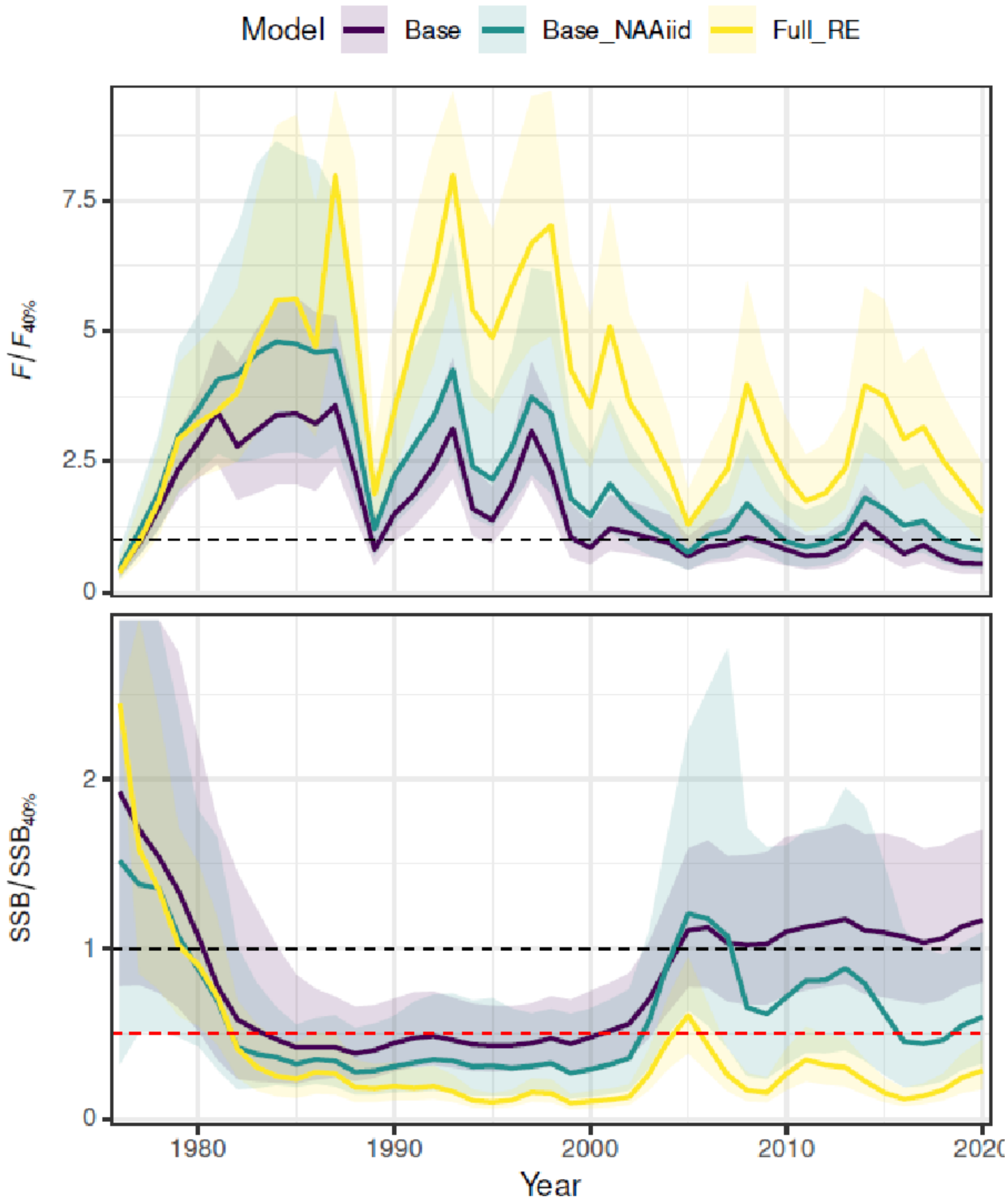


Figure 17. Comparison of WHAM Base (purple), Base with NAA iid random effects (turquoise) and the full random effects (selectivity & NAA) run (yellow) for $F/F_{40\%}$ and $SSB/SSB_{40\%}$ ratios.

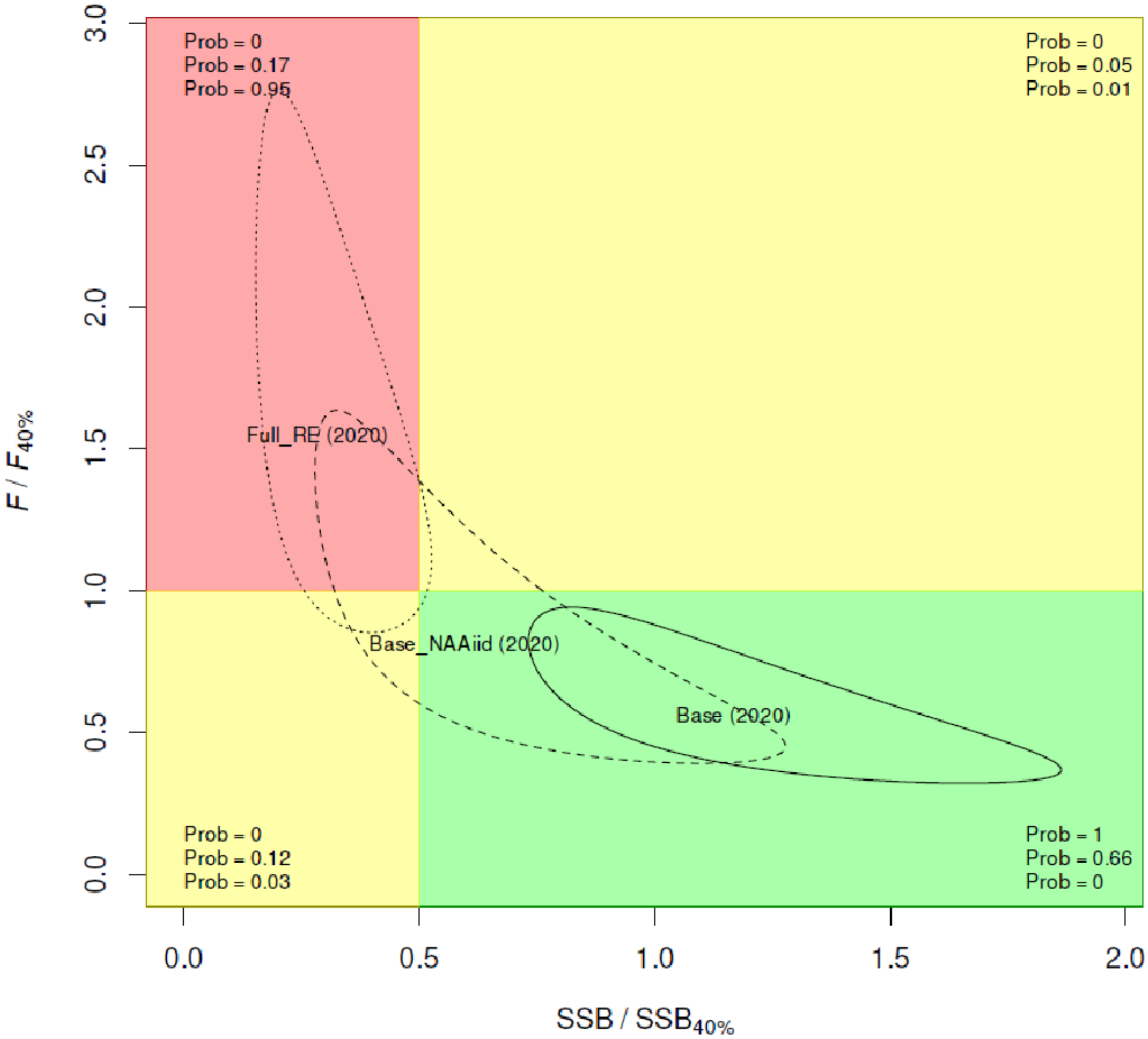


Figure 18. Comparison of WHAM Base, Base run with NAA iid random effects and the full random effects (selectivity & NAA) status Kobe plots.

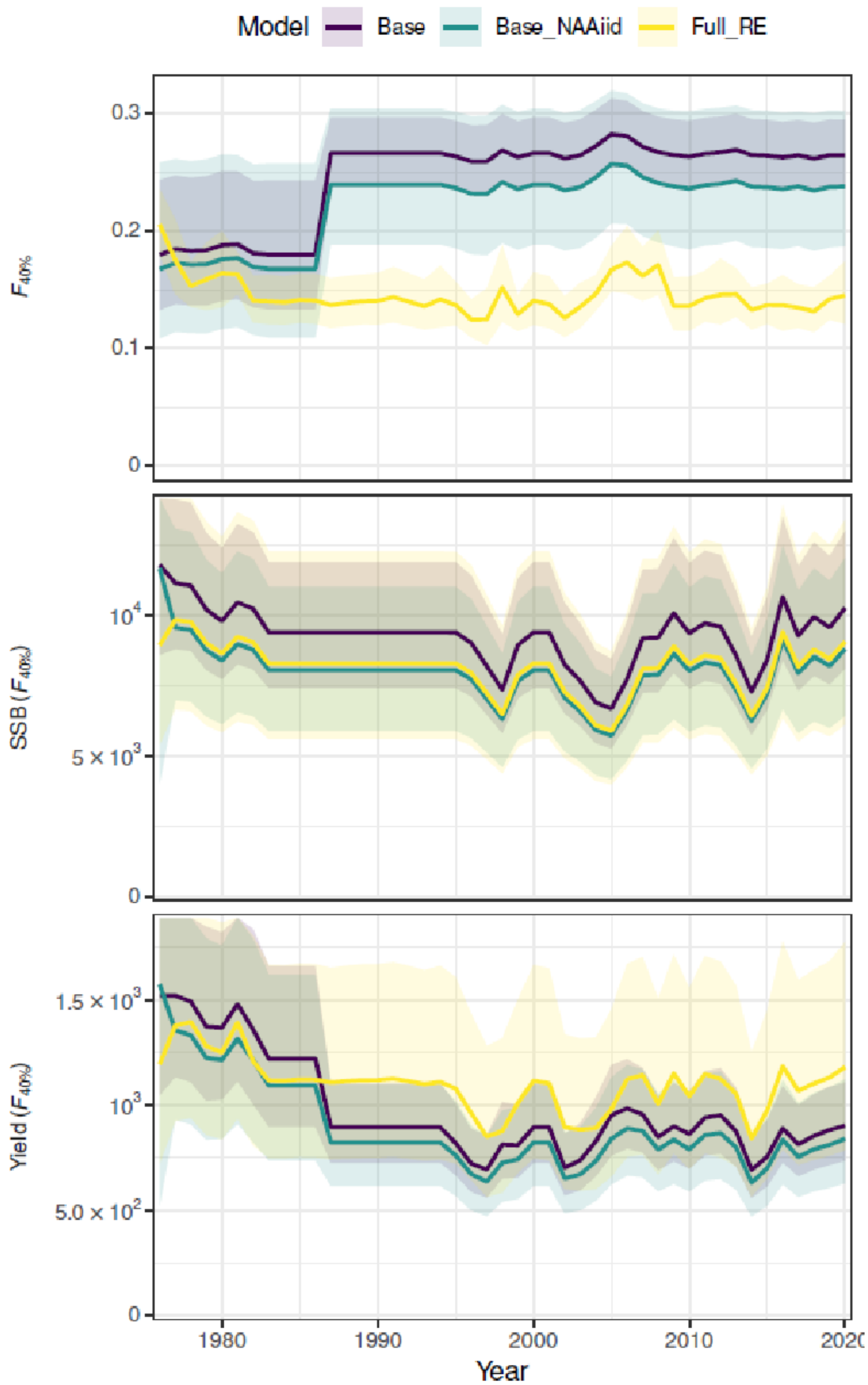


Figure 19. Comparison of WHAM Base (purple), Base run with NAA iid random effects (turquoise) and the full random effects (selectivity & NAA) run (yellow) for dynamic estimates of $F_{40\%}$, $SSB_{40\%}$ and yield at $F_{40\%}$.

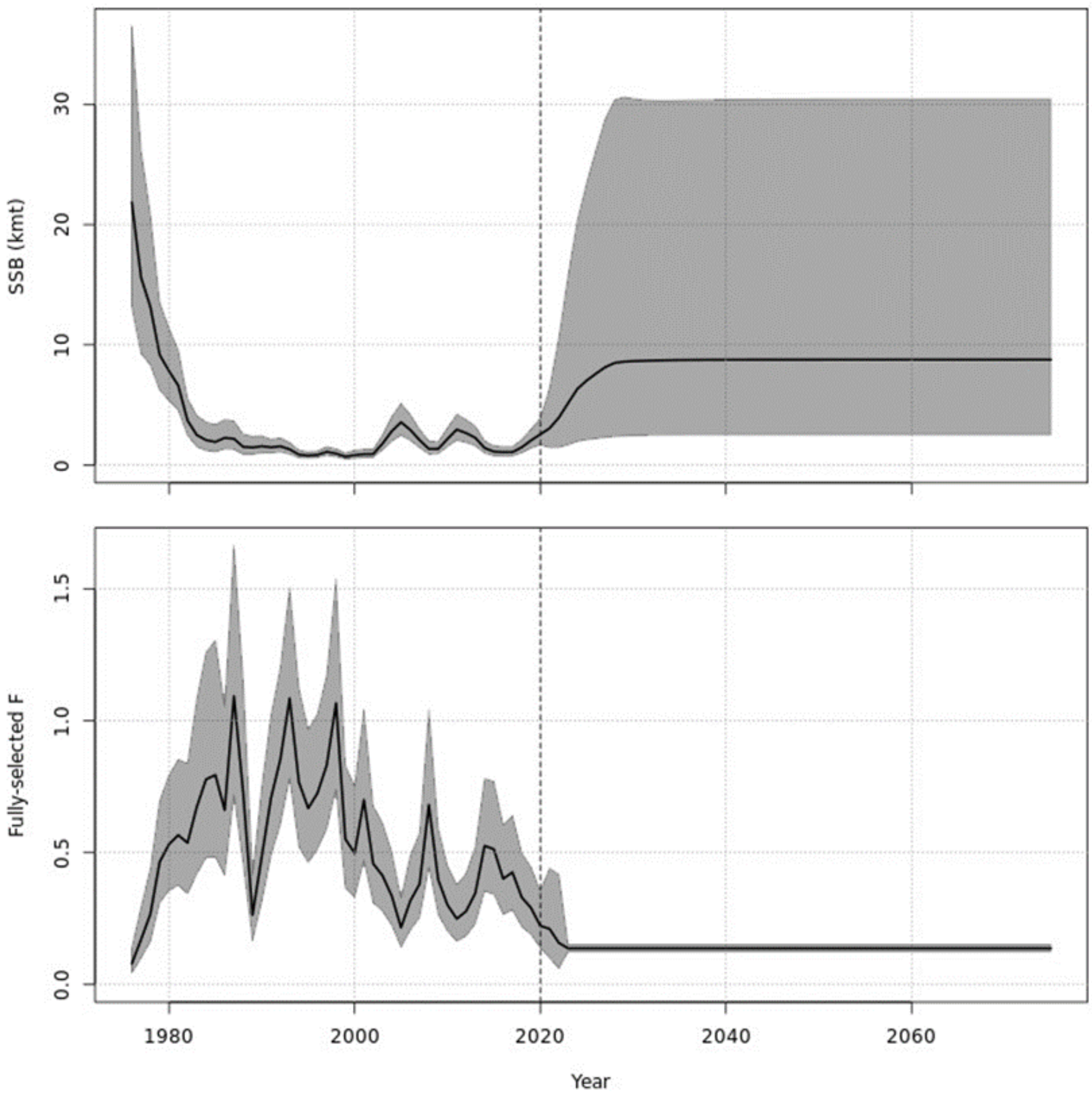


Figure 20. The full random effect model with random effects on selectivity and NAA for SSB and fishing mortality. After the dotted line are projections which assume the catch for 2021 (723 mt) and 2022 (680) and $F_{40\%}$ after 2022.

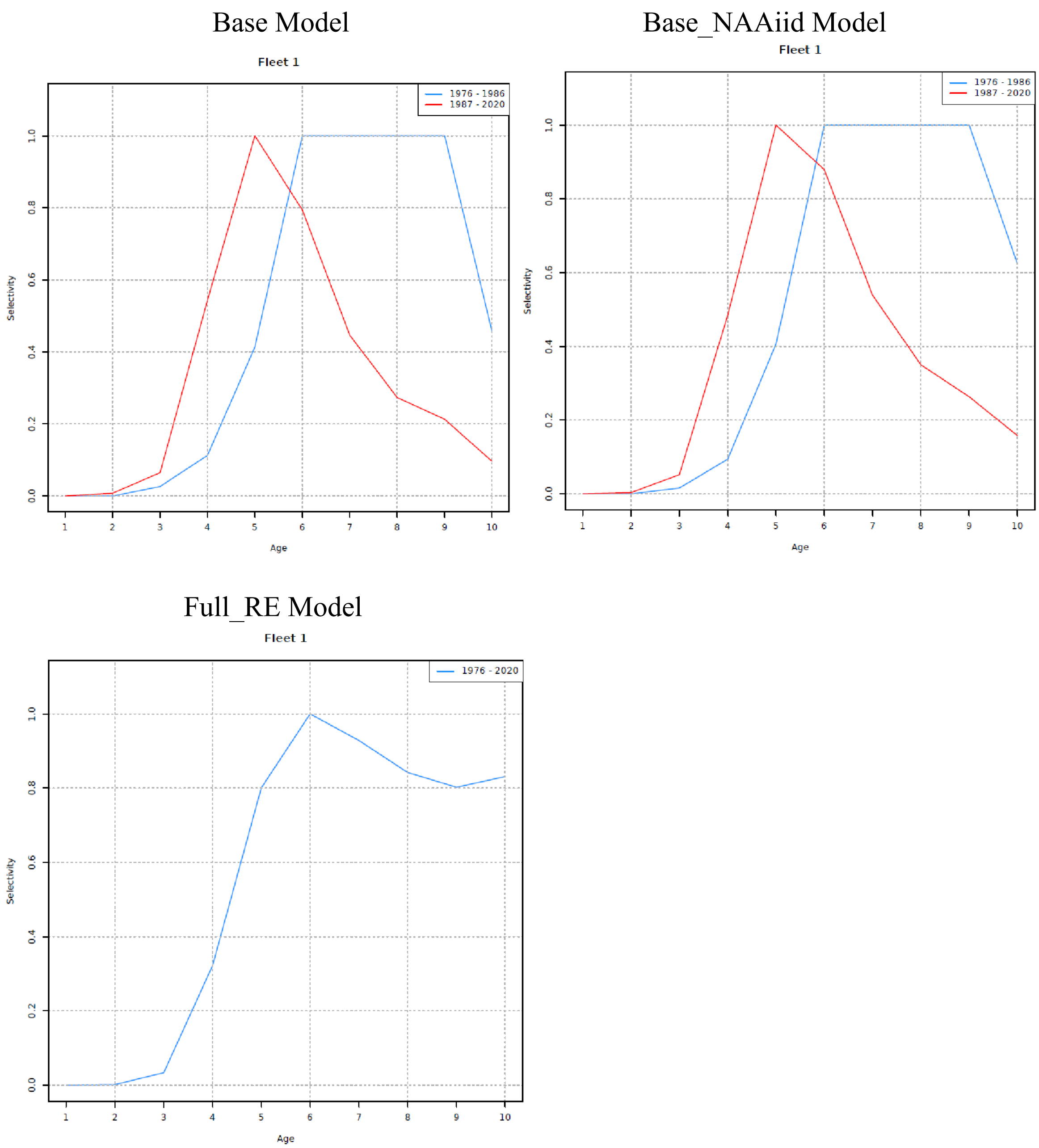


Figure 21. Comparison of selectivity from the Base, Base_NAAiid and Full_RE WHAM models. The average selectivity over the time series is shown for the Full_RE model since the Full_RE model incorporates random effects on selectivity.

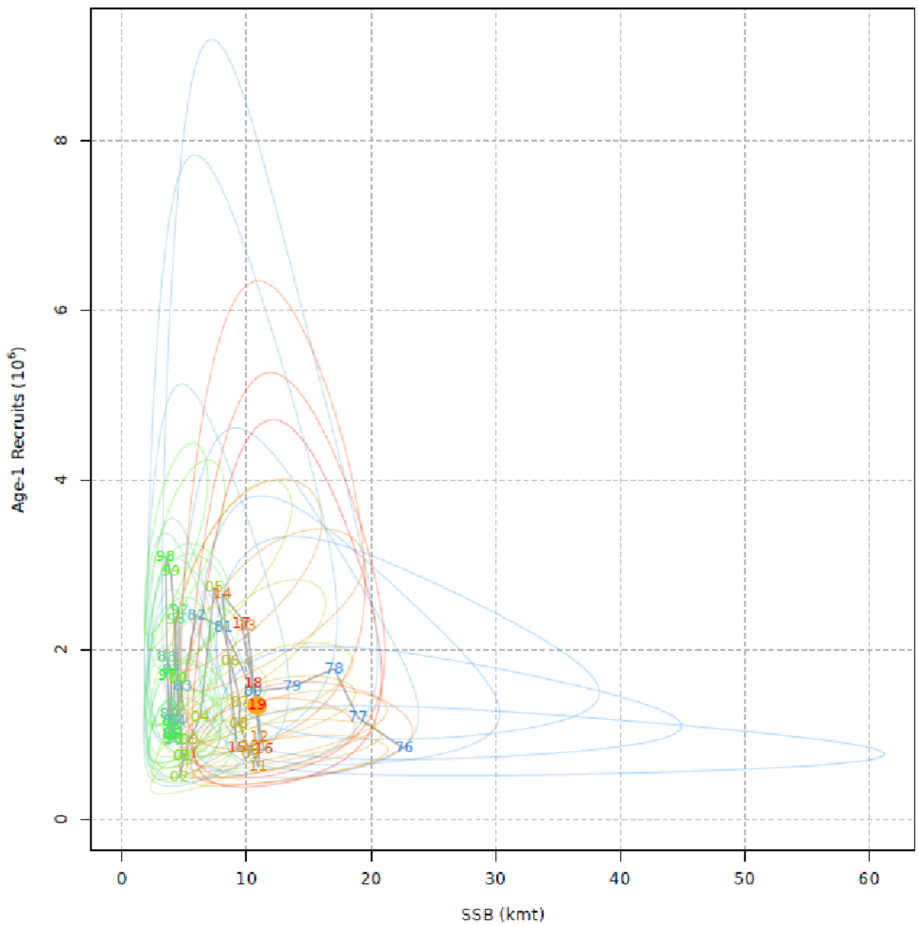


Figure 22. Stock (SSB) and age-1 recruit estimates with joint confidence bounds from the WHAM base model.

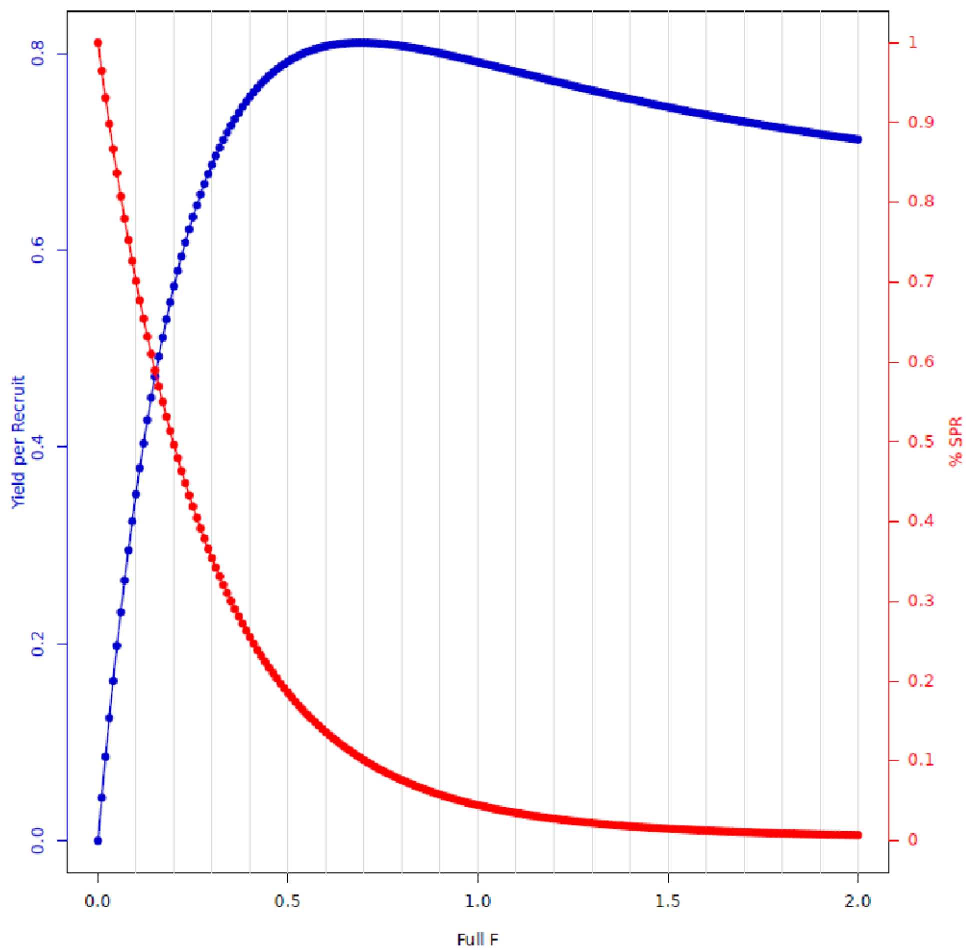


Figure 23. Yield per recruit and percent spawning potential ratio for base WHAM model.

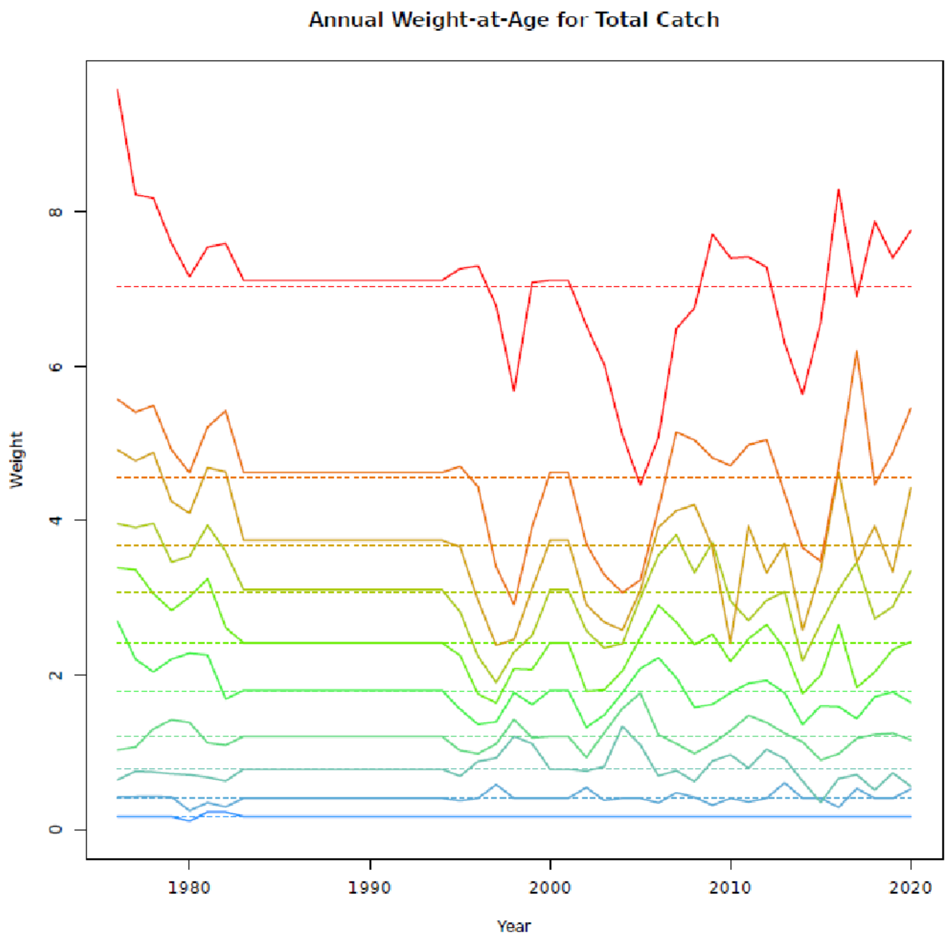


Figure 24. Catch mean weights at age input for golden tilefish.

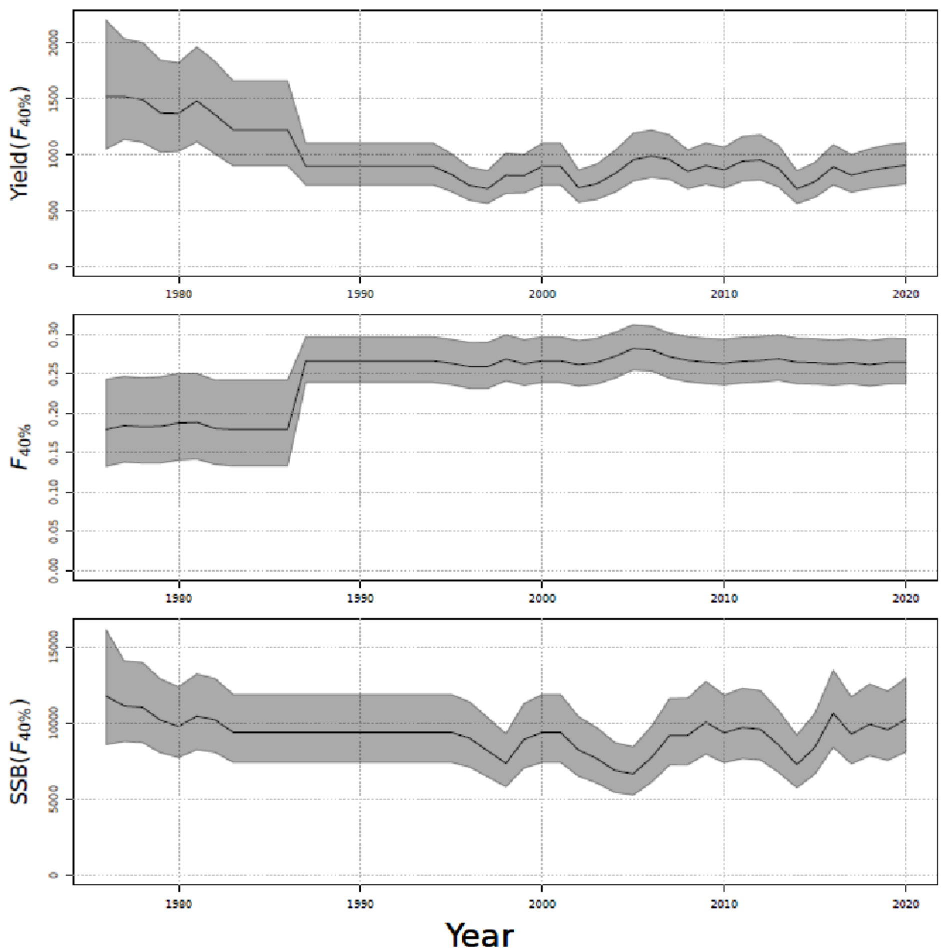


Figure 25. Dynamic estimates of SSB at $F_{40\%}$, $F_{40\%}$, and yield at $F_{40\%}$ for the base model.

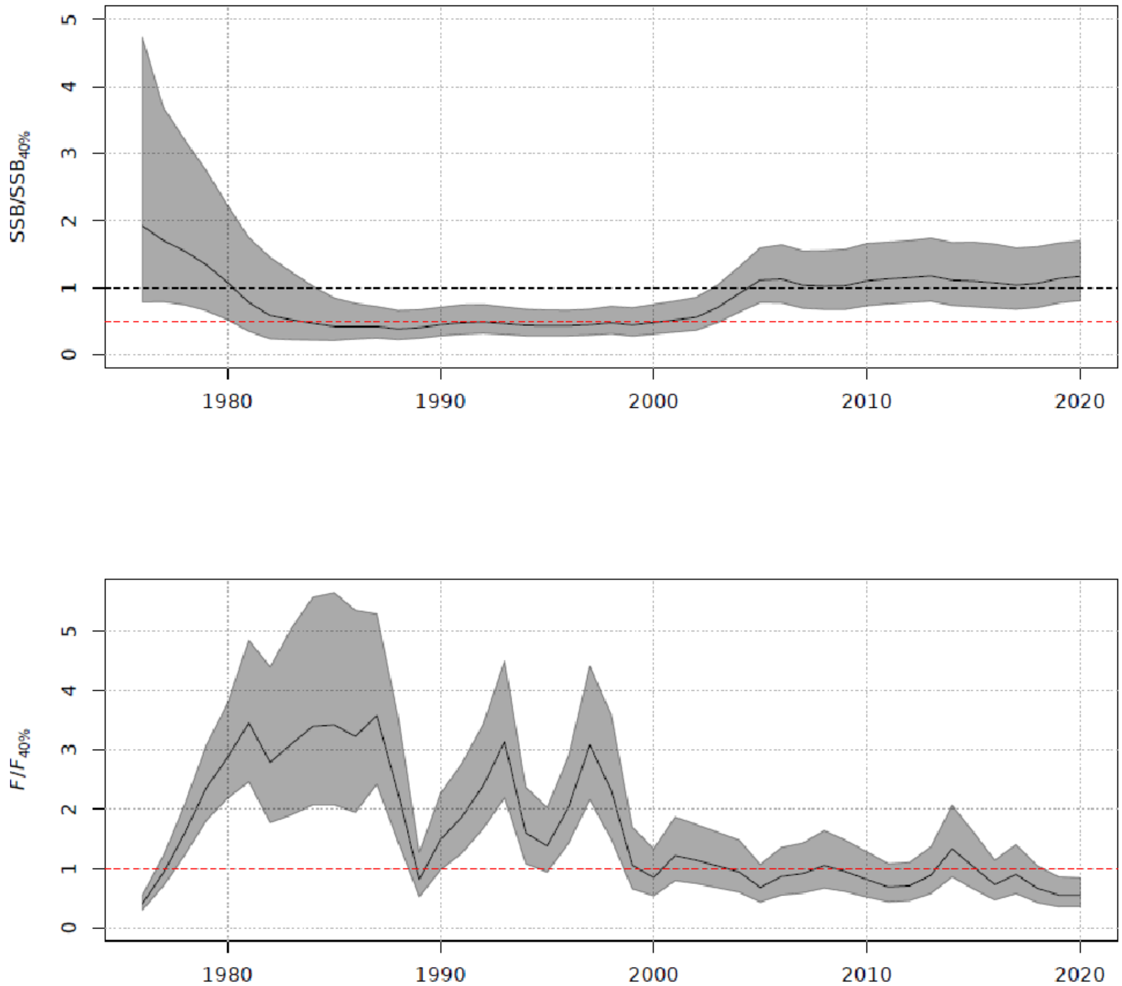


Figure 26. $SSB/SSB_{40\%}$ and $F/F_{40\%}$ ratios for the base model.

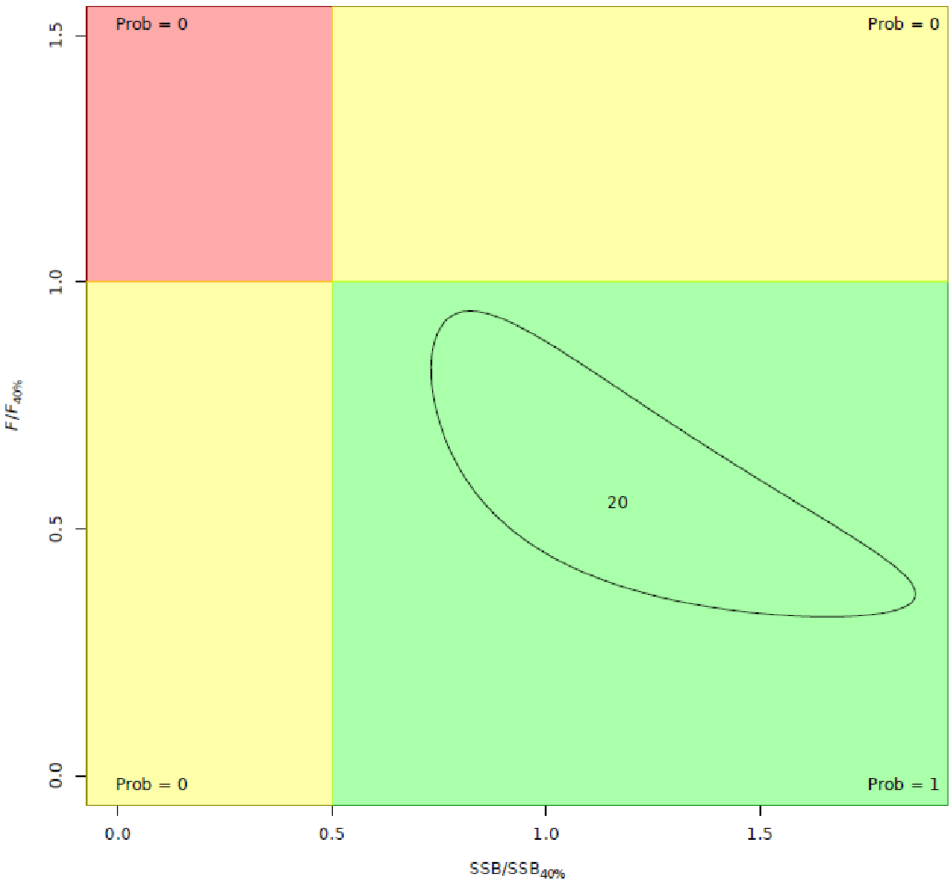


Figure 27. Kobe status plot for the base WHAM model.

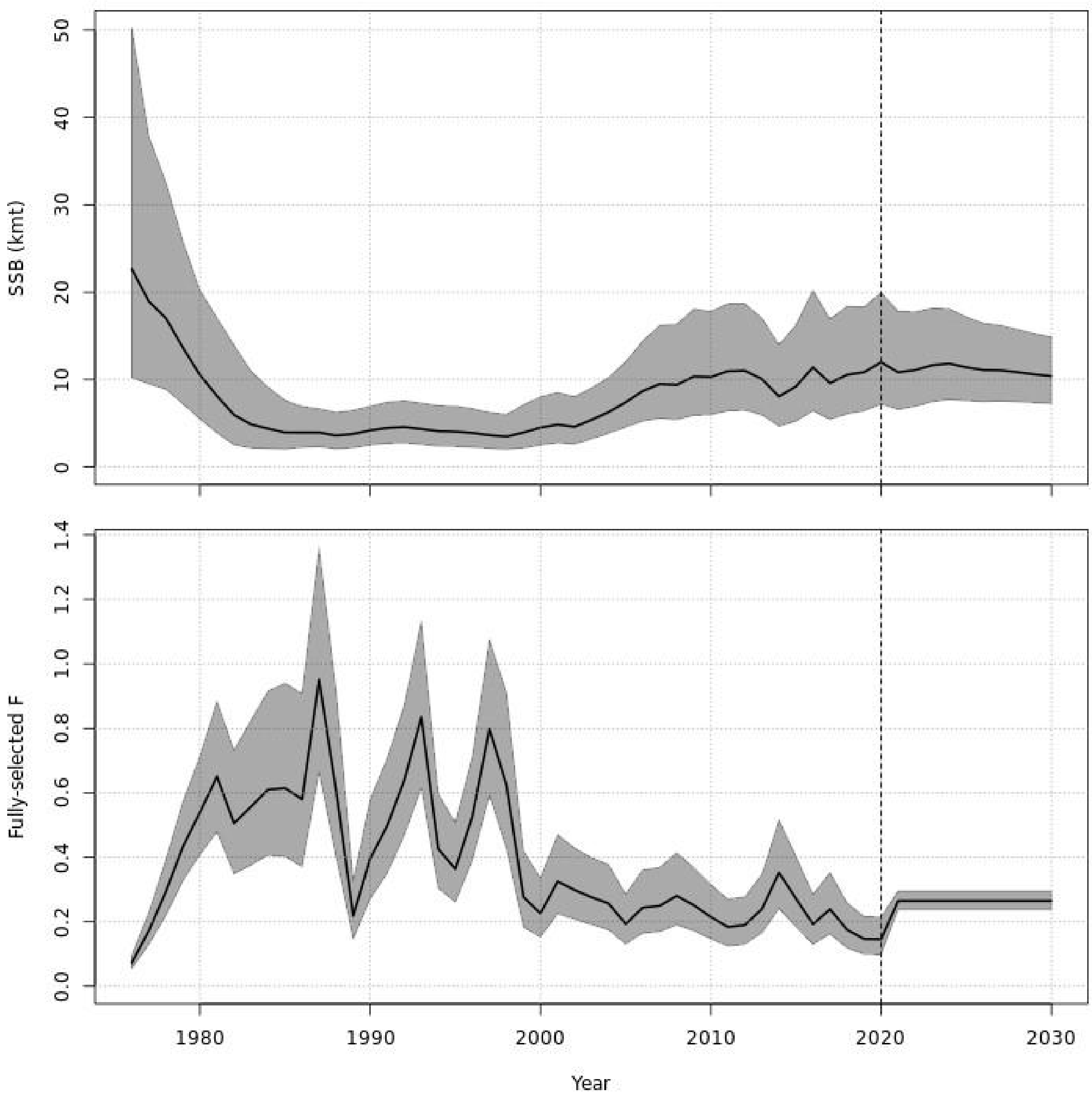


Figure 28. WHAM Base model SSB and fishing mortality with 10 year projections at $F_{40\%}$. The vertical dotted line shows the beginning of the projection under $F_{40\%}$.

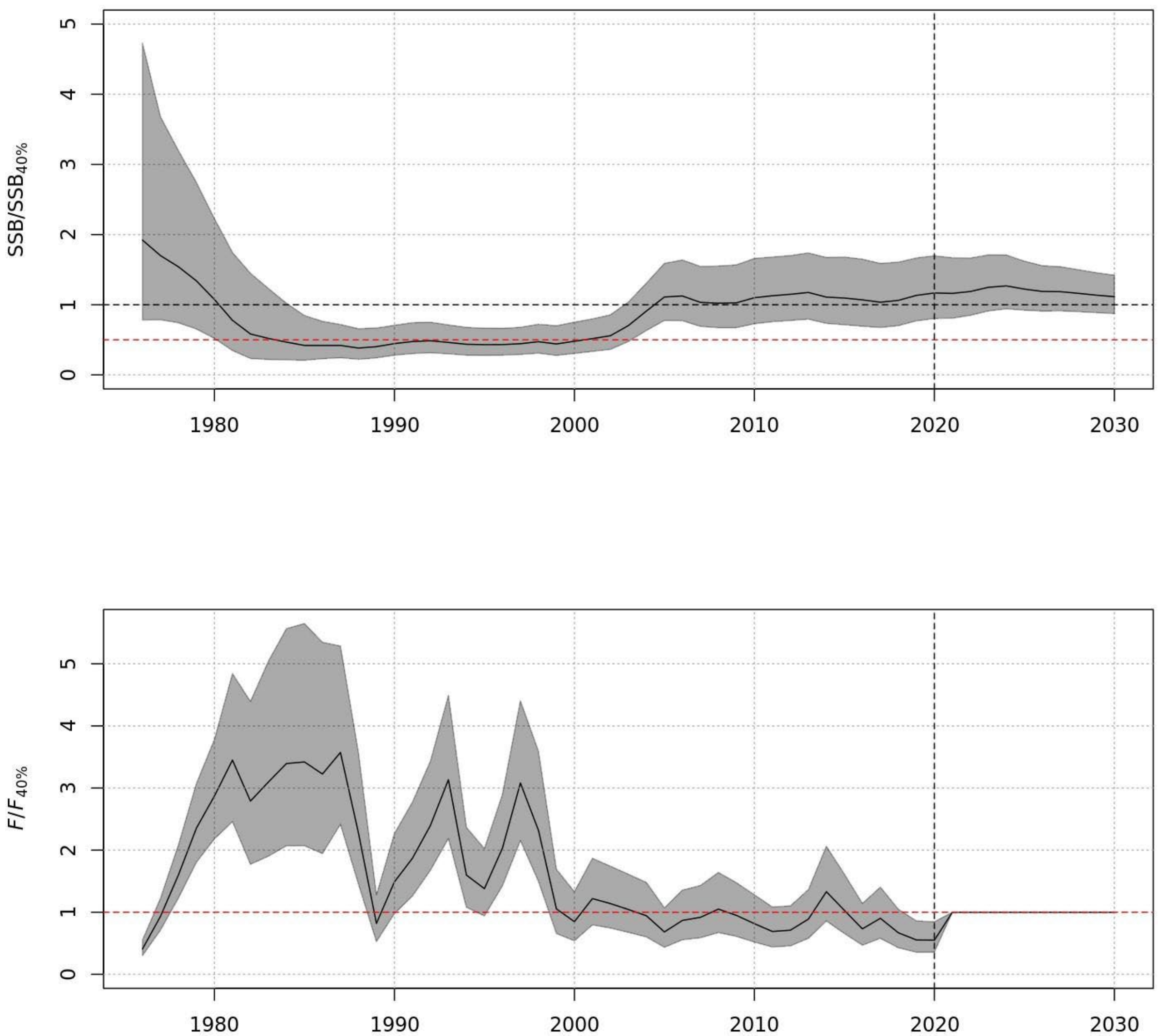


Figure 29. WHAM Base model $SSB/SSB_{40\%}$ and $F/F_{40\%}$ ratios with 10 year projections at $F_{40\%}$. The vertical dotted line shows the beginning of the projection under $F_{40\%}$. Red dotted line signifies the overfished and overfishing levels.

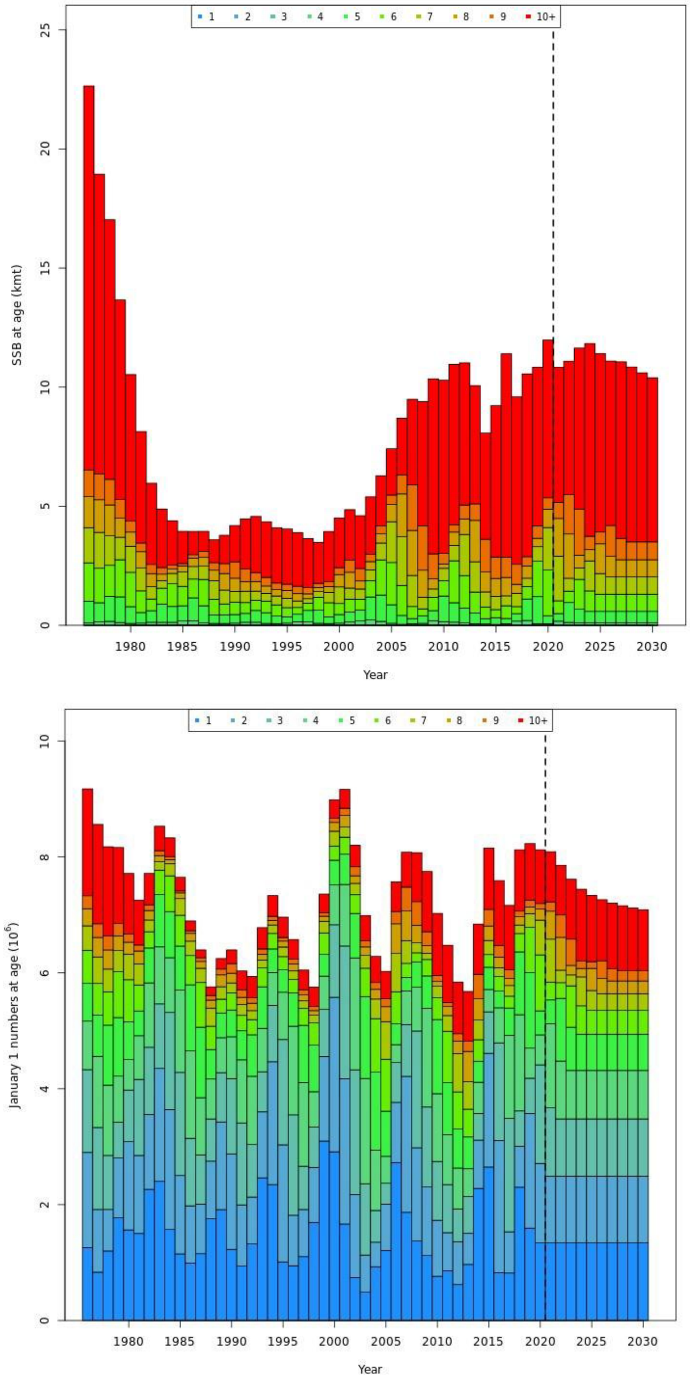


Figure 30. WHAM Base model SSB at age (top) and numbers at age (bottom). The vertical dotted line shows the beginning of the projection under $F_{40\%}$.

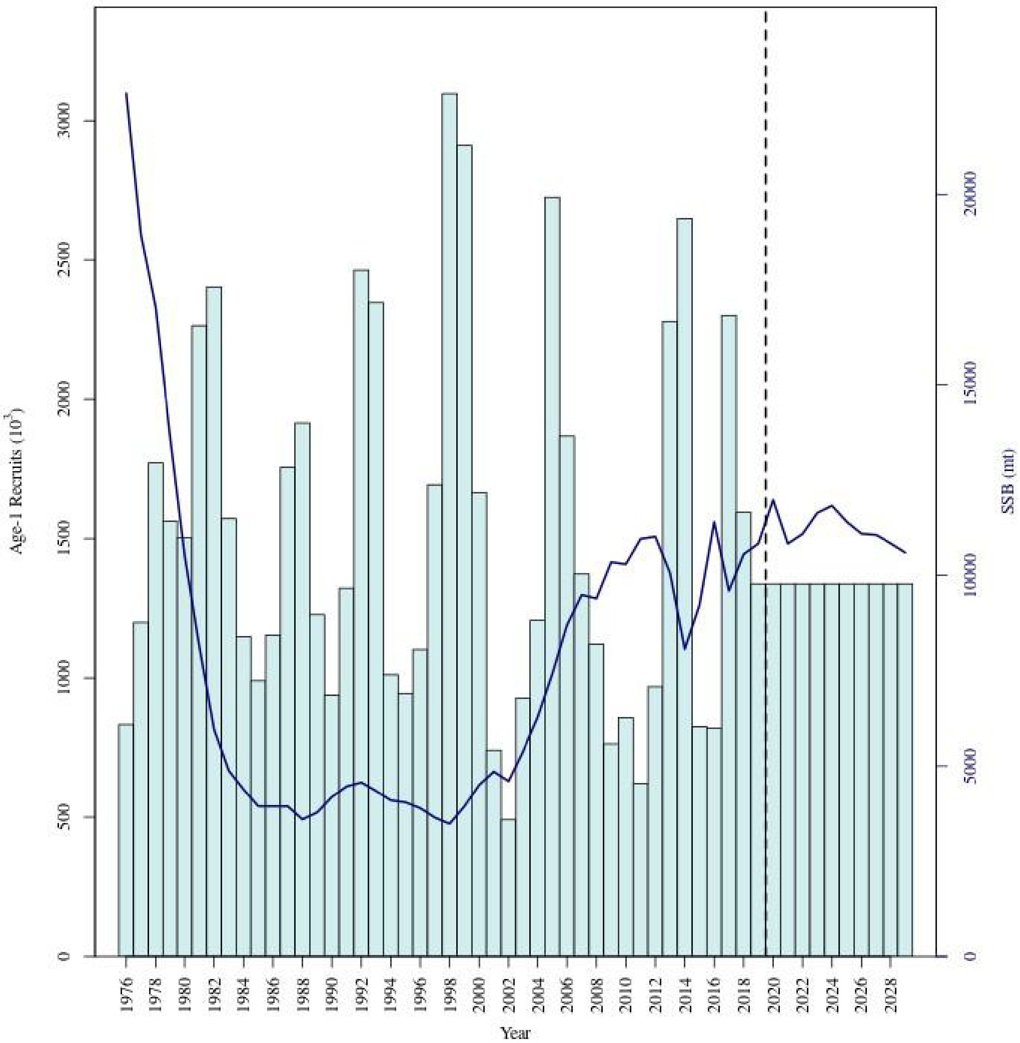


Figure 31. WHAM base model comparison of SSB and age-1 recruitment. The dotted line shows the beginning of the projection fishing at $F_{40\%}$.

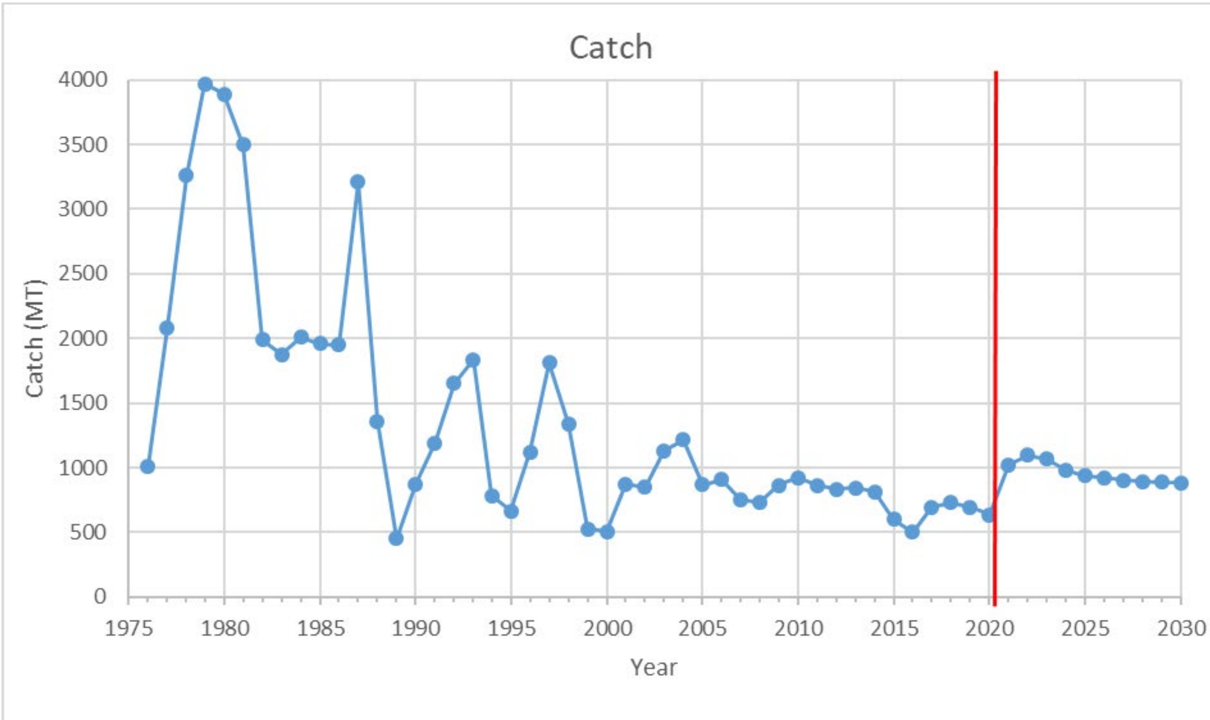
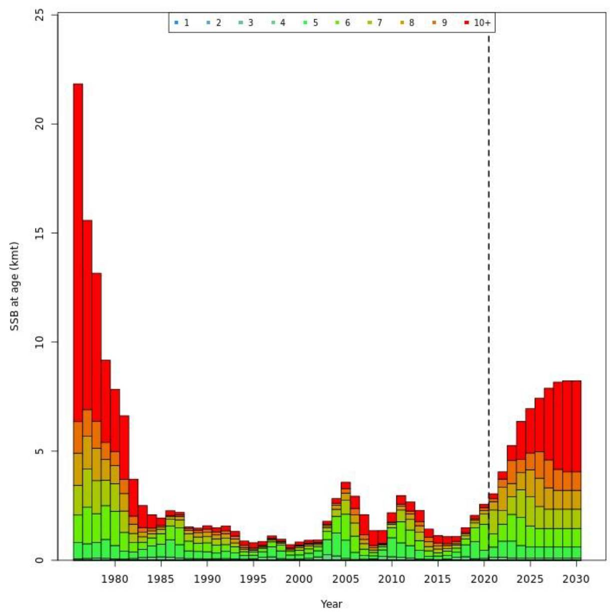


Figure 32. WHAM base model time series of catch and 10 year projected catch. The red line shows the beginning of the projection fishing at $F_{40\%}$.



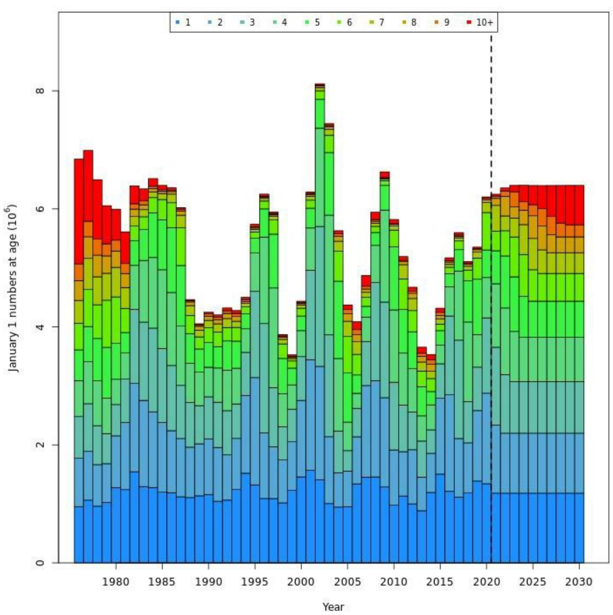


Figure 33. WHAM Full_RE sensitivity model SSB at age (top) and numbers at age (bottom). The vertical dotted line shows the beginning of the projection under $F_{40\%}$.

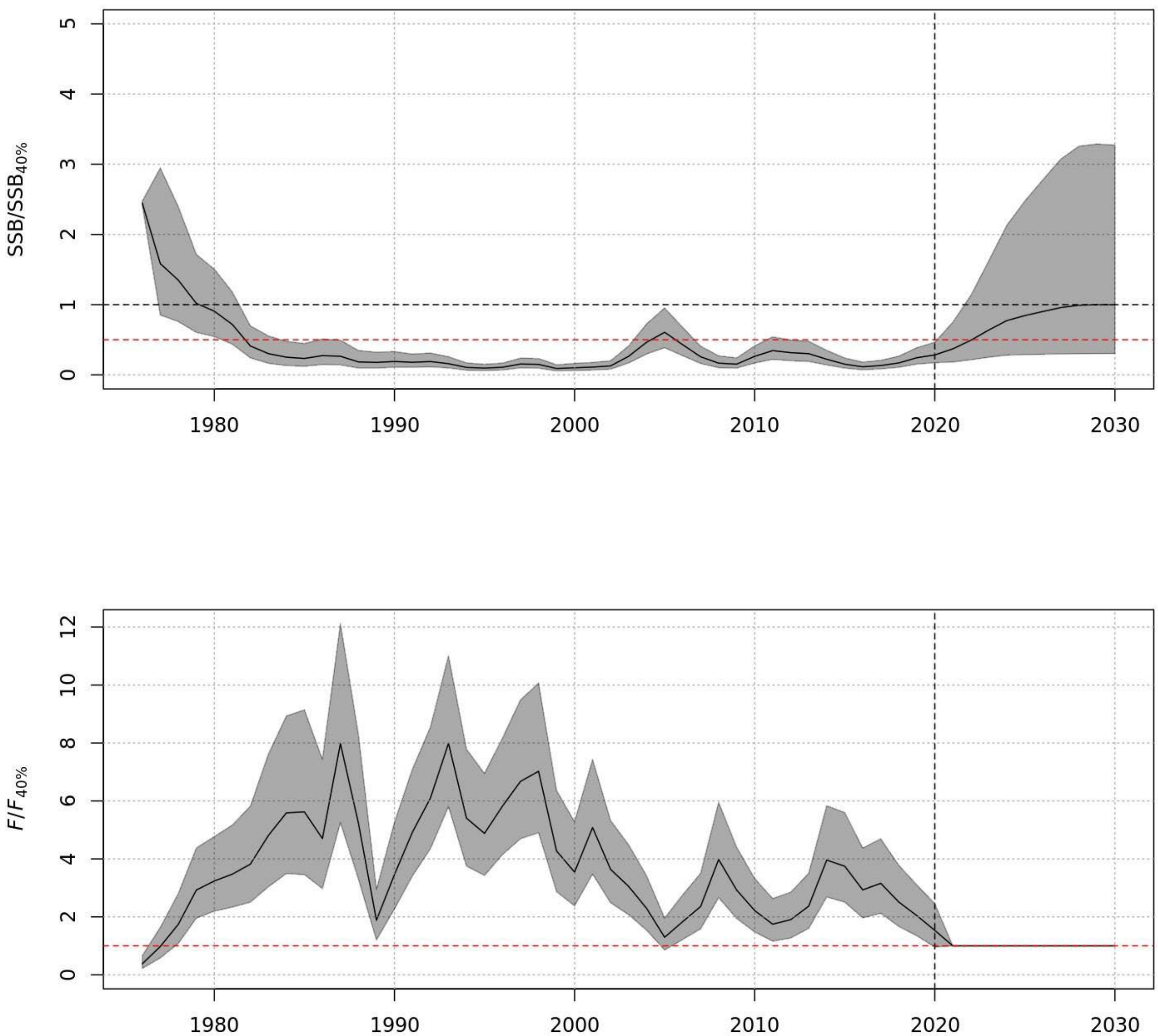


Figure 34. WHAM Full_RE sensitivity model $SSB/SSB_{40\%}$ and $F/F_{40\%}$ ratios with 10 year projections at $F_{40\%}$. The vertical dotted line shows the beginning of the projection. Red dotted line signifies the overfished and overfishing levels.

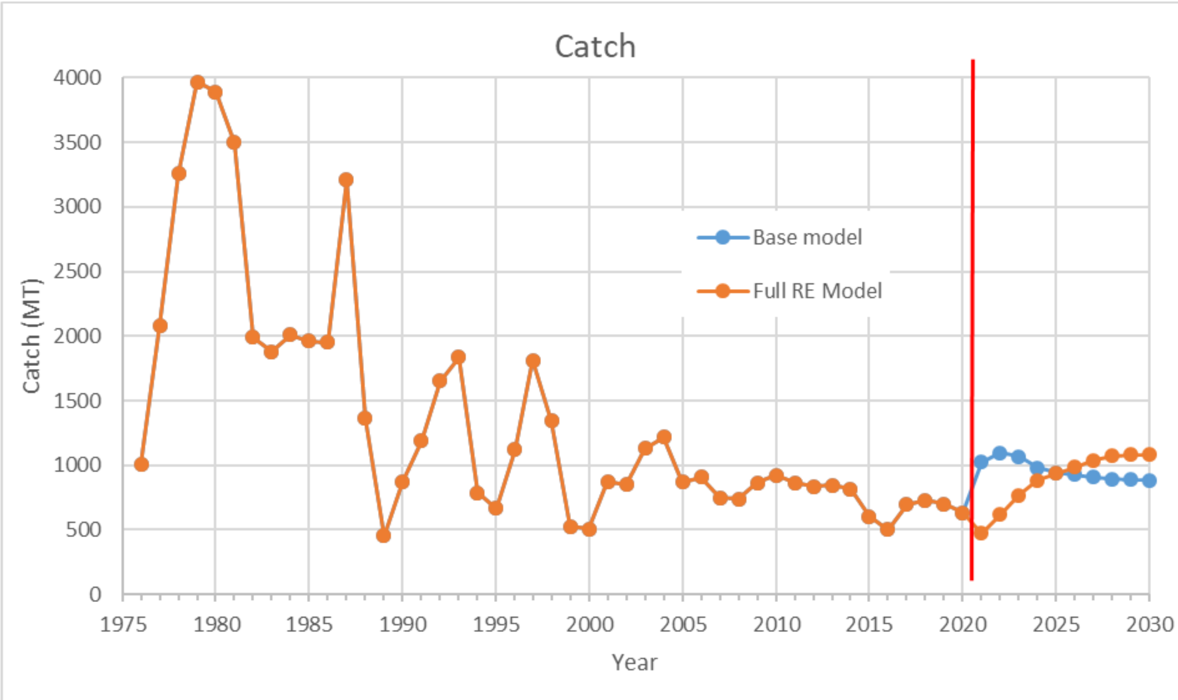


Figure 35. Time series of catch and comparison of 10 year projected catch between the Base and Full_RE sensitivity model. The red line shows the beginning of the projection fishing at $F_{40\%}$. The Base model projection shows a fishing down trend to MSY while the Full_RE projection shows rebuilding towards a relatively high MSY.

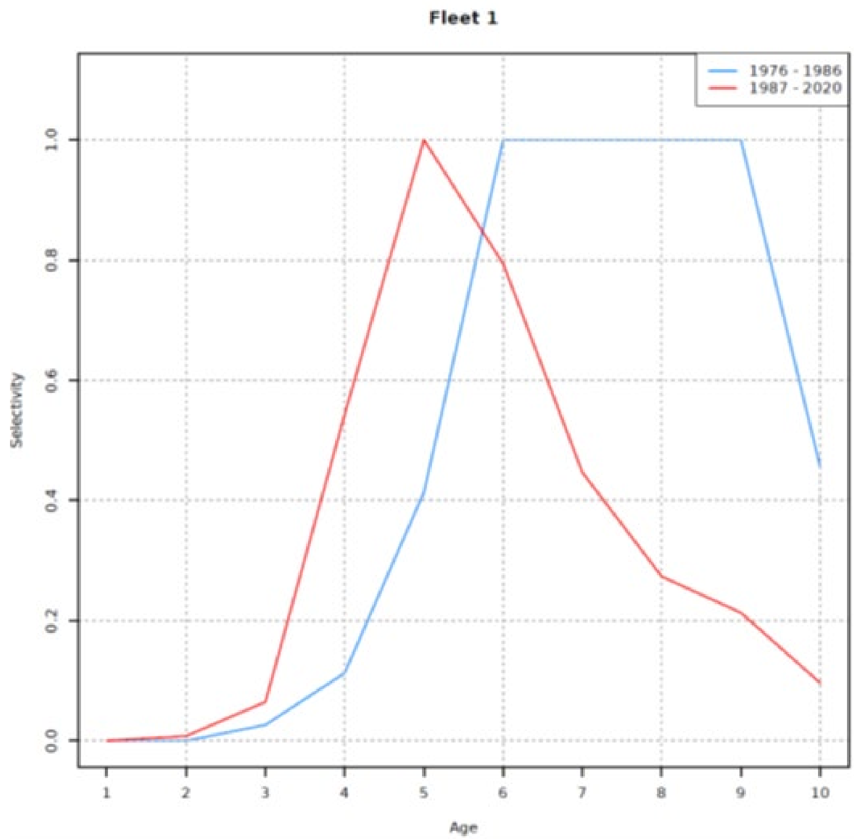


Figure 36. Estimated selectivity for the two selectivity blocks for the WHAM base model.

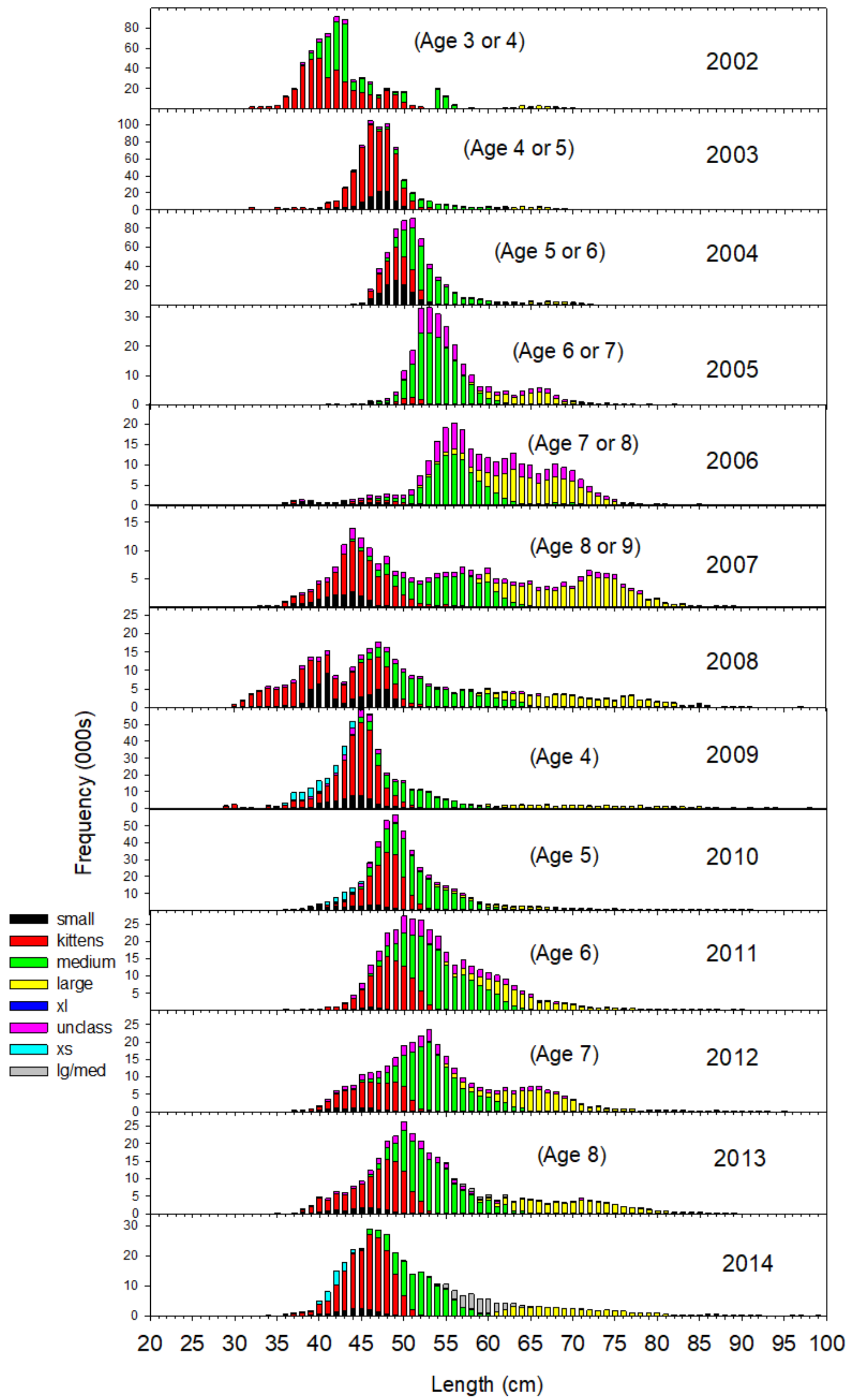


Figure 37. Expanded length frequency distributions from 2002 to 2022. Kittens lengths were used to characterize the extra small category in 2013. **Y-axis is allowed to rescale**.

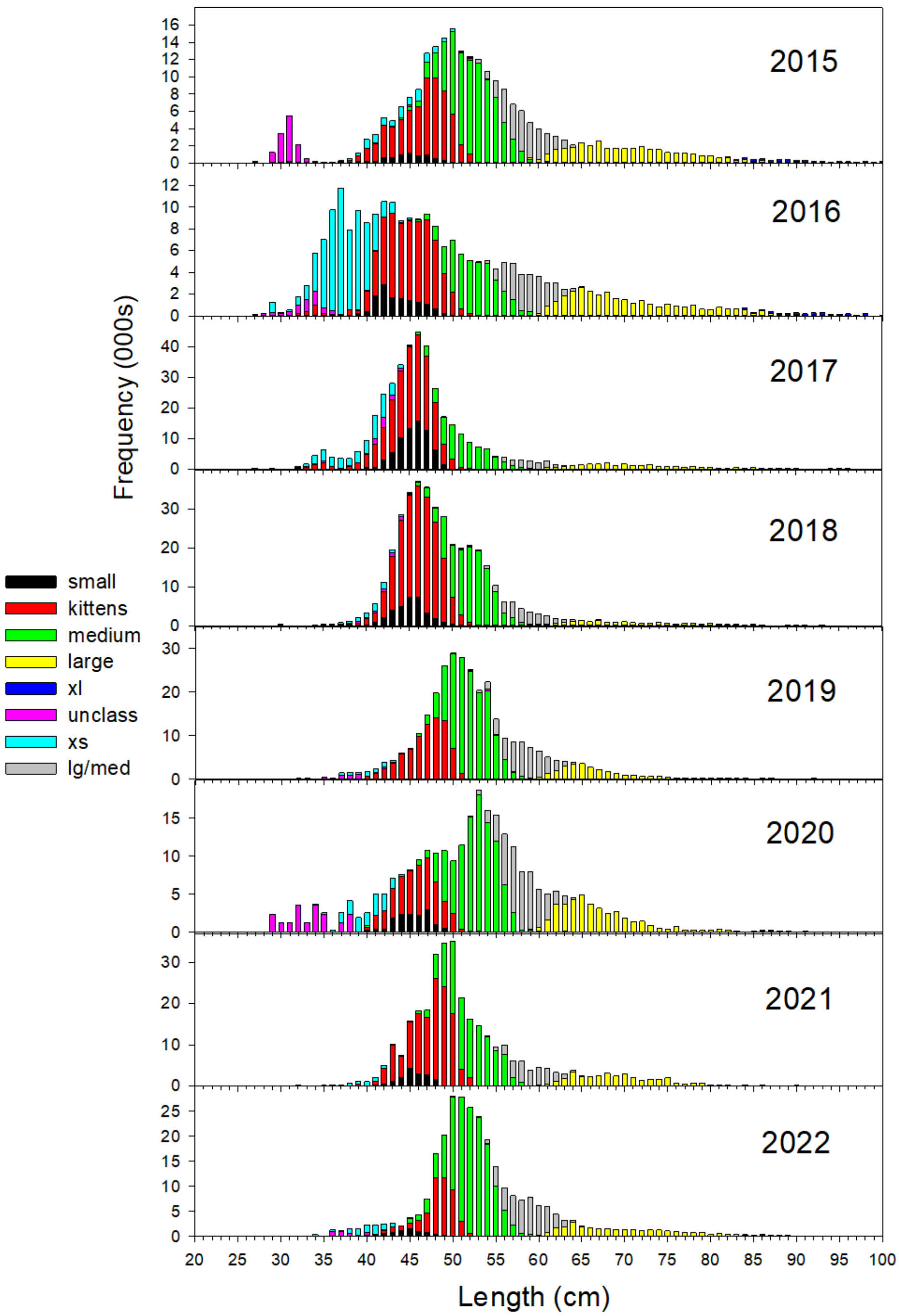


Figure 37 (continued). Expanded length frequency distributions from 2002 to 2022. Kittens lengths were used to characterize the extra small category in 2013. **Y-axis is allowed to rescale.**

12 APPENDIX A: LIST OF RELEVANT WORKING PAPERS

Working papers are available on the NEFSC data portal in the background folder for this assessment.

1. Salois et al 2024. Golden Tilefish Ecosystem and Socioeconomic Profile.
2. Nitschke 2024. Golden Tilefish Commercial Removals.
3. Montañez et al. 2023. Golden Tilefish Recreational Data Collection and Analysis.
4. Nitschke 2024. LPUE and CAMS LPUE Transition.
5. Nitschke 2024. Tilefish Longline Study Fleet.
6. Jones and Salois 2024. Exploring a CPUE index from trawl gear using high-resolution catch data.
7. Boucher et al. 2023. Estimation of Stratified Numbers Per Tow.
8. Frisk et al. 2018. Fisheries-independent pilot survey for Golden (*Lopholatilus chamaelonticeps*) & Blueline (*Caulolatilus microps*) Tilefish throughout the range from Georges Bank to Cape Hatteras. Final Report.
9. Olin et al. 2020. Fishery-independent 2020 bottom longline survey for the Mid-Atlantic Golden Tilefish (*Lopholatilus chamaeleonticeps*) stock. Final Report.
10. Tilefish Survey Review Committee 2018. Report of the 2017 Pilot Tilefish Survey Review.
11. Nitschke 2024. Golden Tilefish Stock Assessment Modeling.
12. Hennen 2024. Golden Tilefish Full Random Effects WHAM Model.
13. Montañez and Nitschke 2023. Summary of the Assessment Work Conducted for Golden Tilefish.
14. Nitschke 2024. Tilefish NEFSC Bottom Trawl Survey Plots.