Mid-Atlantic Fishery Management Council
800 North State Street, Suite 201, Dover, DE 19901
Phone: 302-674-2331 | FAX: 302-674-5399 | www.mafmc.org
P. Weston Townsend, Chairman | Michael P. Luisi, Vice Chairman
Christopher M. Moore, Ph.D., Executive Director

MEMORANDUM

Date: March 25, 2024

To: Council

From: José Montañez and Brandon Muffley, Staff

Subject: Meeting Materials - Golden Tilefish Research Track Stock Assessment

On Wednesday, April 10, 2024, the Northeast Fisheries Science Center (NEFSC) will provide the Council with an overview of the recently completed 2024 research track stock assessment and peer review for golden tilefish. During the June 2024 management track assessment review, data through 2023 will be included in the modeling and alternative model configurations may be explored. Results of the June management track assessment will be used to inform management and future catch specifications.

The following materials are provided for Council consideration of this agenda item:

- 2024 Report of the Golden Tilefish Research Track Working Group Report Executive Summary
    - The full working group report can be found at: Stock Assessment Support Information (SASINF) Search Tool
- Summary Report of the Golden Tilefish Research Track Stock Assessment Peer Review (to be posted once available)

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
Russell 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 discards 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 $F_{40\%}$ 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 $F_{40\%}$ 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 future 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 U.S. 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 2003-2022 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.