Summary Report of the Golden Tilefish Research Track Stock Assessment Peer Review

March 11-14, 2024
Northeast Fisheries Science Center, Woods Hole, Massachusetts

Report prepared by Panel Members:
Michael Wilberg (Chair), MAFMC SSC
Matt Cieri, independent contractor with CIE
Samuel Subbey, independent contractor with CIE
Joe Powers, independent contractor with CIE

Introduction

The Northeast Region Coordinating Council (NRCC)[1] has developed an enhanced stock assessment process to improve the quality of assessments. The process involves two tracks of assessment work: 1) a management track that includes routine updates of previously approved assessment methods to support regular management actions (e.g., annual catch limits), and 2) a research track that allows comprehensive research and development of improved assessments on a stock-by-stock or topical basis. The research track assessment process allows for a more thorough review of information available and for the evaluation of different assessment approaches than would be possible in a standard stock assessment process where the results are immediately used for management advice. This Panel reviewed the Research Track Assessment for the golden tilefish stock.

The most recent stock assessment for golden tilefish was a management track assessment in 2021 based on the benchmark assessment from SARC 58 in 2014 (https://repository.library.noaa.gov/view/noaa/4719). The previous stock assessment used the Age-Structured Assessment Program (ASAP) with estimates of fishery landings at age and multiple landings per unit effort (LPUE) time series. There were no fishery independent indices of abundance included in the assessment model because the existing surveys rarely encountered golden tilefish. This stock assessment represented golden tilefish from North Carolina to Georges Bank as a single stock. The Golden Tilefish Research Track Working Group (WG) developed multiple Woods Hole Assessment Model (WHAM) models as candidates for use in a management track assessment. The WG was chaired by José Montañez and included staff from NOAA Fisheries, the MAFMC, and state management agencies. Terms of Reference (ToRs) for the WG are provided in Appendix 1.

[1] Atlantic States Marine Fisheries Commission (ASFMC), Greater Atlantic Regional Fisheries Office (GARFO), Mid-Atlantic Fishery Management Council (MAFMC), New England Fishery Management Council (NEFMC), and Northeast Fisheries Science Center (NEFSC).

The Golden Tilefish Research Track Assessment Peer Review took place in Woods Hole, MA during March 11-14, 2024 for a hybrid meeting. The Panel included three independent scientists selected by the Center for Independent Experts (CIE): Matt Cieri, Joe Powers, and Sam Subbey. The Panel was chaired by Mike Wilberg (Member of the MAFMC SSC).

The Research Track Assessment Report and 14 supporting working papers were provided to the Panel on the NEFSC data portal (https://apps-nefsc.fisheries.noaa.gov/saw/sasi.php) on February 21, 2024. The panel was also provided with the presentations and model diagnostics for the WHAM models. Panel members drafted portions of the Panel Summary Report individually, but the entire Panel edited and reviewed the report such that it represents the consensus views of the Panel. Prior to the meeting, members of the Panel met with Michele Traver (NEFSC's Stock Assessment Workshop Process Lead) and Kristan Blackhart (Chief, NEFSC Population Dynamics Branch) on March 7, 2024 to review and discuss the meeting agenda, reporting requirements, meeting logistics and the overall process.

Presentations made by the WG members are listed in the meeting agenda (Appendix 2) and are available as pdfs on the data portal. WG members were present during the meeting to answer questions from the Panel. Charles Peretti, Toni Chute, Amanda Hart, Chris Legault, Liz Brooks, and Jessica Blaylock served as rapporteurs (see Appendix 4 for the materials provided and Appendix 5 for meeting attendees).

Panel members and the Chair drafted this Summary Report in a Google Doc. The Panel Chair compiled and edited this Summary Report with assistance from the CIE Panelists before submission of a draft report to the WG. The scope of the WG review of the draft was limited to suggesting corrections for errors of fact or requesting that Panel recommendations be clarified. Additionally, each of the CIE Panelists will submit separate reviewer reports to the CIE.

The Panel agreed that TORs 1-3, 5, 6, and 9 were fully met and ToRs 4, 7, and 8 were partially met. At the time of the peer review, many of the models were still showing either diagnostic issues or estimates that the Panel considered unlikely. The Panel agrees that WHAM may be an appropriate tool for the golden tilefish assessment, but that additional work needs to be done prior to selecting a base model for the management track assessment. In particular, the Panel makes recommendations in ToRs 4 and 8 on ways to potentially proceed. The Panel commended the WG for a substantial amount of work and for the collegial nature of the review.

The Panel's evaluation of the WG's response to the nine Terms of Reference (ToRs) is provided below.

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 Panel agreed that this ToR has been **fully met**, and welcomed the use of Ecosystem and Socioeconomic Profiles (ESP) to address this ToR.

Significant progress has been made in the identification of pertinent ecosystem and climate variables that influence the stock. Using incidental CPUE indices of abundance and model estimates of recruitment, the Golden Tilefish Research Track Assessment (2024) results show that: (i) bottom temperature and salinity preferences across life stages were consistent with the literature, (ii) indicators of habitat condition and food availability highly correlated with the presence of larvae and recruitment-age (0-1 years) fish, and physical oceanographic indicators may have complex influences on early life stages and fish recruiting to the fishery (□ age 4).

The results are promising and show a clear potential for incorporating environmental variables into the index standardization and recruitment variability into the stock assessment. This potential notwithstanding, the findings of this TOR have not been incorporated into the models for assessing the golden tilefish stock. It remains unclear when environmental patterns would be convincing enough to include in WHAM, and thus address the recommendations under TOR 3 of the Applied State Space Models Research Track Stock Assessment Peer Review. The current methodological approach should be reevaluated in terms of developing the generalized additive model (GAM) that links environmental variables to CPUE and recruitment, and consideration of other modeling approaches that can potentially provide insight into the spatio-temporal dynamics of the golden tilefish.

The methodological approach adopted in this study comprises two main steps.
1. Identification of pairwise relationships: Initially, pairwise relationships between variables were determined. This involved using Pearson's correlation analysis to assess the strength and direction of the relationships between CPUE/recruitment and each environmental variable individually and between pairs of environmental variables.
2. GAM development: following the identification of significant pairwise relationships, a GAM was constructed, which linked CPUE with all the environmental variables that were found to be significant in the first step.

In principle, this methodological approach should allow for a comprehensive analysis of the relationships between CPUE and environmental variables, incorporating both linear and non-linear effects. However, a thorough revision of the methodological approach for establishing the linkages between these variables and stock dynamics may be warranted because non-monotonic nonlinear relationships (e.g., dome shapes) may not have significant linear correlations. Furthermore, the need to quantify uncertainty in the identified variables and their potential impacts stock dynamics may necessitate the exploration of additional tools and methodologies not covered in the report. Understanding the relationship between ecosystem variables and the spatio-temporal dynamics of the golden tilefish that may be relevant to other ToRs in this Assessment may also require the development of new models.

Revision of methodological approach:
When exploring relationships, Pearson's correlation coefficient may not be suitable due to the possibility of non-linear relationships or other complexities in the data. The analysis results (see Table 1) show that (with few exceptions) the null hypothesis (there is no correlation between the

two variables) cannot be rejected at conventional p-values (0.05 or 0.01). However, non-detection of a linear relationship may either be due to the nature of the data, or that the relationship being investigated is non-linear.

Exploring other methodologies may be useful. For instance:

1. Spearman's Rank Correlation: a non-parametric measure of association that assesses the monotonic relationship between two variables. It does not assume linearity and is based on the ranks of the data rather than the actual values. Spearman's correlation can be more robust to outliers and is suitable for ordinal or ranked data.

2. Kendall's Tau: Similar to Spearman's correlation, Kendall's Tau is a non-parametric measure that assesses the ordinal association between two variables. It measures the similarity of the ordering of data pairs and is particularly useful when dealing with small sample sizes or tied ranks.

3. Partial Correlation: Partial correlation measures the strength and direction of the linear relationship between two variables while controlling for the effects of one or more additional variables. It helps assess the relationship between two variables while holding other variables constant, thus providing insights into direct associations.

Variables	**R and p values**	**Inference**
Monthly CPUE/hr vs. Bottom Temp	R=0.0072, p=0.95	Cannot reject NH
Recruitment vs Mean salinity	R=-0.14, p=0.54 (Fall) R=-0.14, p=0.54 (Summer) R=-0.29, p=0.19 (S) R=-0.48, p=0.022 (W)	Reject NH
Recruitment vs Cold pool extent Recruitment vs 3yr-lag Cold pool extent	R=-0.37, p=0.059 R=-0.58, p=0.03	Cannot Reject NH moderately strong, Reject NH
CPU vs Weighted mean SST CPUE vs 3yr-lag Weighted mean SSt	SP, Summer, Winter All R and P	Weak Significant Cannot reject NH

Table 1: Example results – Pearson correlation analysis

The GAM analyses appear to be comprehensive, integrating various environmental and spatiotemporal predictors to explain the variation in the response variable, CPUE. However, like any statistical model, there are potential drawbacks, which warrant consideration. Since the methodological details about how the GAM was implemented are lacking in the report, the Panel provides the following cautions for consideration:

1. Collinearity: Including multiple correlated predictors (e.g., environmental variables like bottom temperature, salinity, microplankton abundance) may lead to collinearity issues. Collinearity can inflate standard errors, making it difficult to assess the significance of individual predictors and potentially obscuring true relationships. For nonlinear relationships in GAMs, concurvity can also be an issue for model identifiability.

2. Non-stationarity: The assumption of stationarity (relationships between predictors and response remain constant over time) may not hold, especially in ecological systems where environmental conditions and species distributions may change over time.

3. Consider several distributions (negative binomial, Tweedie, and Quasi Poisson) which can deal with overdispersion. However, both the negative binomial and the quasi-poisson are for count data. The Tweedie would be the most appropriate since it is defined for non-negative reals; it can deal with zeros, and some overdispersion.

Developing alternative models for exploring the link between environmental variables and the population dynamics:

The progress made in defining the full life cycle of the golden tilefish and in identifying significant environmental factors that may affect its population dynamics may be harnessed to develop parsimonious models of how the ecosystem affects golden tilefish population dynamics. These models can help understand and potentially predict the effect of a changing marine environment on the golden tilefish population dynamics, both on temporal and spatial scales.

1. **Identifying spawning sites using Larval drift models:**

The availability of CTD, presence-absence, and oceanographic data (currents, winds) can aid in identifying potential spawning sites, using particle tracking with back propagation models. This approach may be feasible, given the amount of ichthyoplankton data available. A potential challenge will be how to capture the empirically observed depth-stratified larvae distribution.

2. **Modeling space-space species behavior using Agent-Based Models (ABMs)**

An Agent-Based Model (ABM) can be a suitable approach for simulating the full life cycle of a fish species like golden tilefish. ABMs are computational models that simulate the actions and interactions of individual agents (e.g., fish) within a defined environment (e.g., ocean habitat). Each agent operates according to a set of rules and behaviors, and the collective behavior of the agents gives rise to emergent patterns at the population level.

Using an ABM can enhance the understanding of, among other things, habitat preferences, feeding (selectivity) patterns, and how these may be linked to the population dynamics in a changing environment. For the golden tilefish, knowledge of the full life cycle (see Figure 1) and empirical data can help develop a basis ABM that can be improved as more information becomes available.

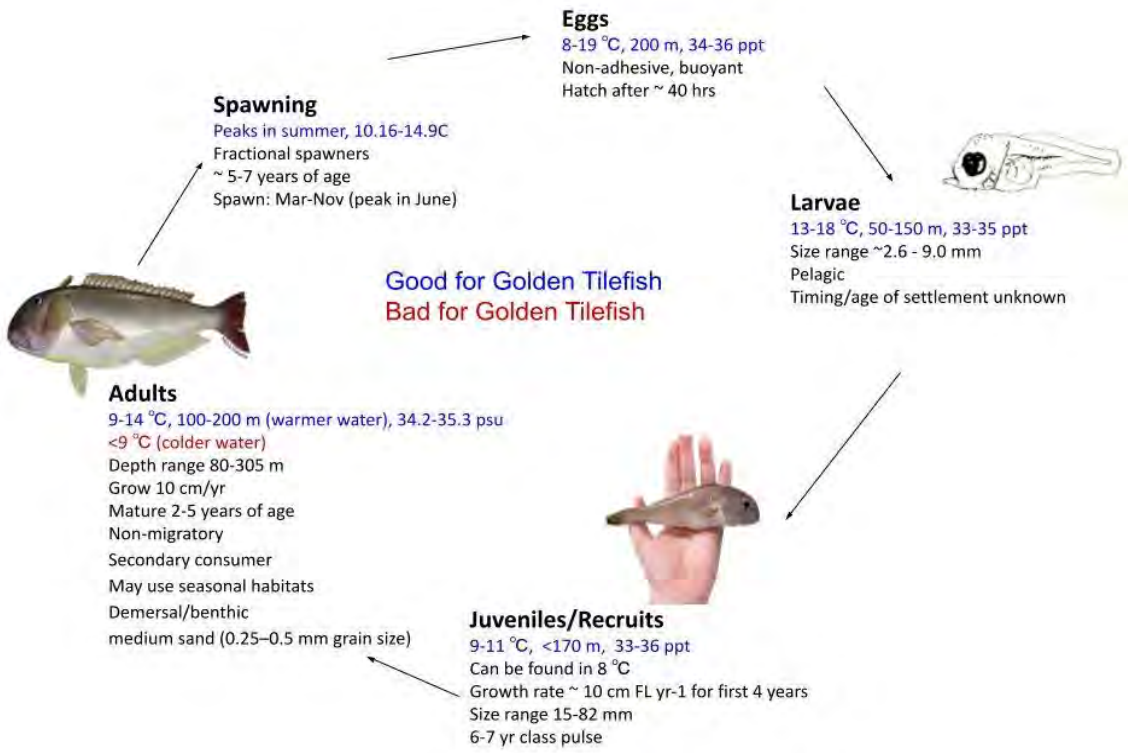

Figure 1. Life cycle of the golden tilefish (reproduced from WP01).

The suggested models can potentially provide input to TOR 3 (Survey) and TOR 4.

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

The Panel agreed that this ToR **was fully met**.

The catch data presented during the Review Workshop included commercial data from the directed longline as well as incidental removals from non-target fleets. Recreational catch data from both the Marine Recreational Information Program (MRIP) and Large Pelagics Survey (LPS) surveys were also presented.

Commercial data sources include the Dealer Database, the Vessel Trip Reports (VTR) as well as the more recent combination of all removal data sources under CAMS (Catch Accounting Monitoring System). Despite the potential climatic/environmental changes in this region, landings have been relatively stable in both amount and location since 2000 (Figures 2 and 3).

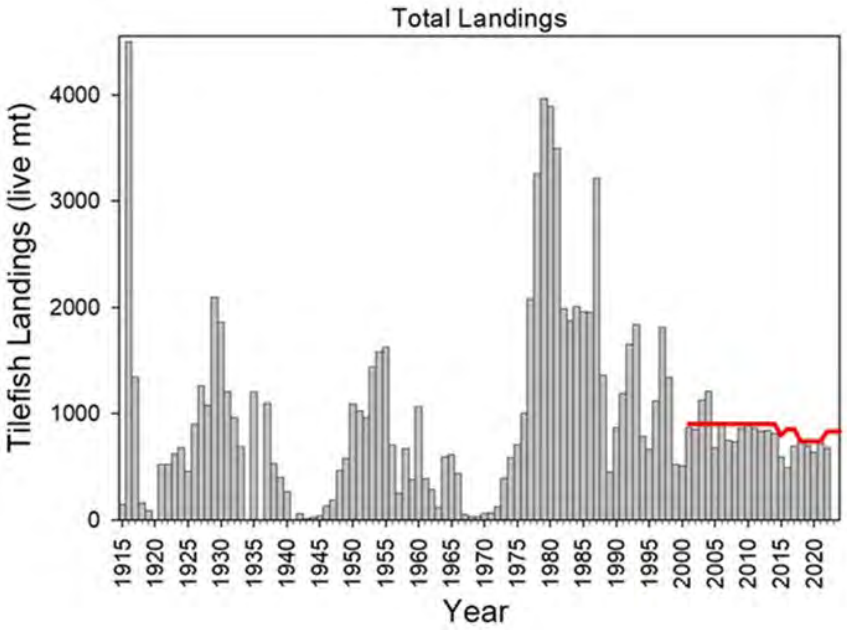

Figure 2. Landings of tilefish in live metric tons from 2001-2022. Red line is the TAL

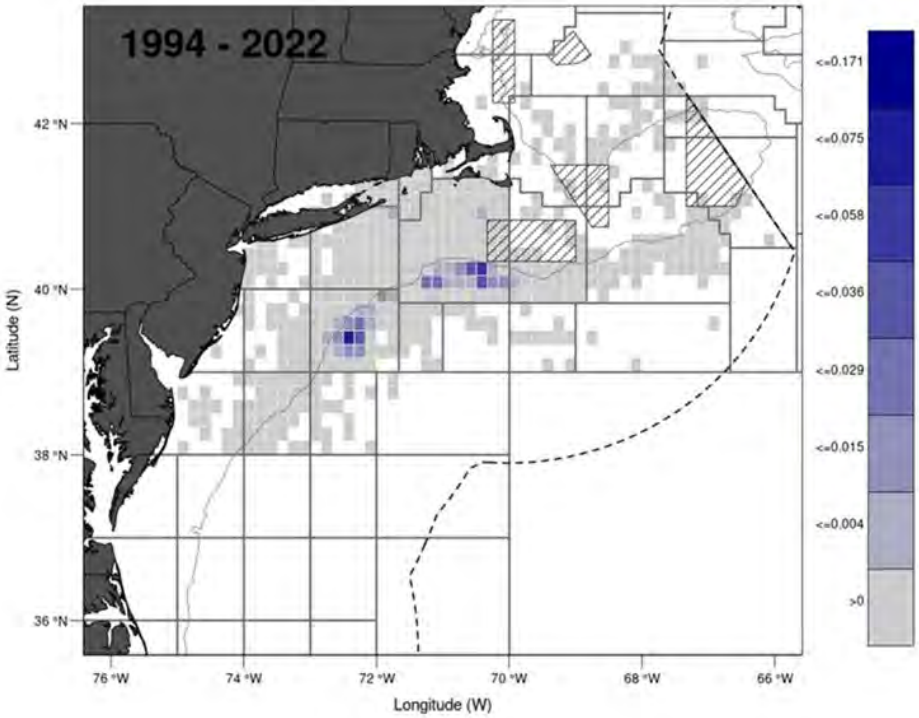

Figure 3. Map of 1994- 2022 VTR reported landings of tilefish by ten minute square.

Given the low number of participants and the completeness of both the dealer and VTR datasets, there appears to be little uncertainty concerning the commercial removals. Likewise, discards are very low, though the assessment team made a substantial effort to examine the discards in the commercial fleet for this Research Track assessment. These two factors combined reduced the level of uncertainty surrounding the overall commercial removals for Golden Tilefish

Recreational data sources include both the MRIP and the LPS as well as the recent party/charter Vessel Trip Reports (VTRs). Interestingly the catch by the recreational fleet was similar to the commercial fishery in locations, despite the recreational catches occurring at a much lower level.

Given the low removals by the recreational fleet (approximately 3% of the removals), the Panel agreed with the Assessment team that this fleet is not as important to overall fishing mortality when compared to the commercial longline fleet. However, there are indications that recreational removals are increasing relative to commercial removals, so it may be important in the future to reexamine this source of mortality. The Panel noted that the LPS data had substantially lower proportional standard errors (PSEs) than the MRIP. After consideration of the presentations and the working papers supplied to the Panel, this TOR was considered to have been met. The Panel was impressed with the amount of work conducted by the Assessment Group on estimating discards and recreational catches. This work constitutes potential new data streams, even if the overall removals are small. As such it is recommended that all recreational sources of removals be used.

It was noted, however, that not a lot of information was supplied concerning the fishery-dependent sampling. While some comparisons in length sampling between commercial at-sea observers and portside samples were conducted, little information on the aging process and methods was provided. In the future, the Panel suggests that this ToR include not only fishery removals but also fishery-dependent sampling for both length and ages, as appropriate, and include descriptions of how age-length keys were generated.

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

The Panel agreed that this TOR **was met.**

The Panel recognizes, and appreciates work done by the WG in addressing this TOR. Results from the analysis show that the data (LPUE, Trawl CPUE, Longline) are consistent in trend. However, the analyses lacked a characterization of uncertainty in the different data sources, which could be different, despite the fact that the trends were similar. The panel suggests that uncertainty be estimated, as this will provide information about data precision for use in the golden tilefish assessment and a rationale for the choice of data.

The Panel also notes that data from trawl (potential index) and longline surveys did not provide any new indices that were included in the assessment model. It is also unclear how the choice of year to begin using the CAMS data was made, where a jump of approximately 10% in CPUE is apparent. Directly linked to this consideration is the question about the start year for beginning to use CAMS LPUE time series (Figure 4). The panel acknowledges that the transition from the VTR to the CAMS does not indicate a change in the data source per se. Rather, it signifies a shift in the foundational database infrastructure, requiring the merging of the existing VTR series with the extensive new database. A suitable start date for the CAMS LPUE will be the first year of minimum deviation from the VTR LUE. The panel recommends, therefore, 2010 as a start date.

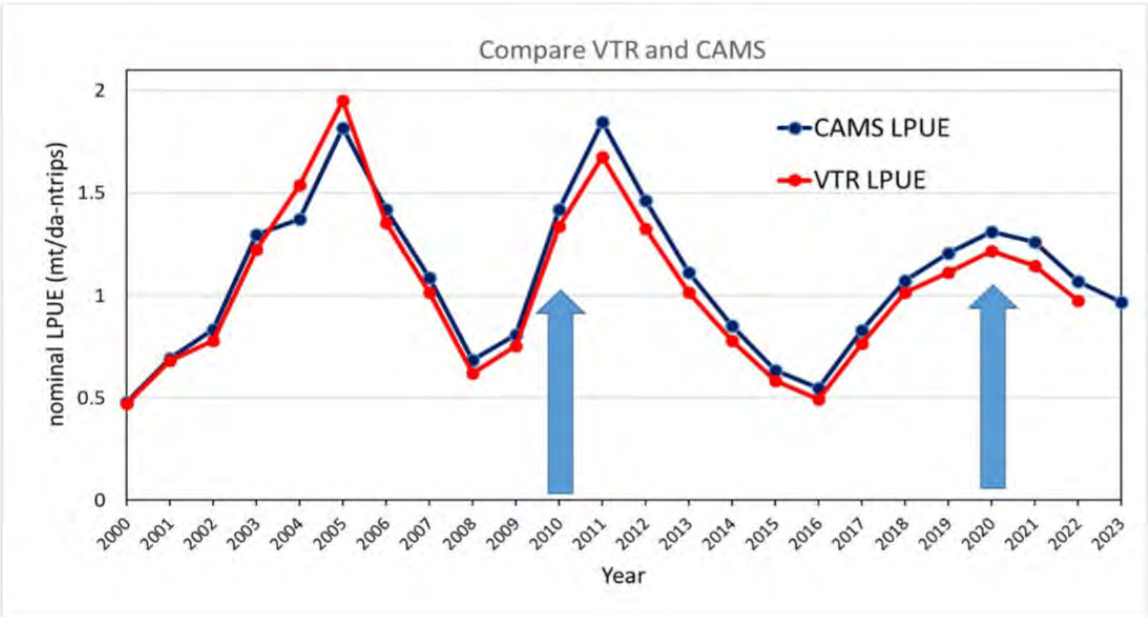

Figure 4. Comparison of CAMS and VTR LPUE time series. Arrows indicate potential years to switch from the VTR to the CAMS LPUE series.

It is the opinion of the Panel that additional analysis may be required to disentangle the hook effect and the bait size effect in selectivity of the tilefish longline survey. Furthermore, although the fishery and the survey covered the same area, the survey covered some shallow areas where the fishermen did not go. Comparing the catch rates of the survey and the fishery may potentially provide new insights, which could also be further corroborated by comparing the study fleet data to the survey data.

There is a need for further analysis to understand the observed dome-shaped selectivity pattern in the proposed base assessment model and whether it is justified. From the report (presentations), dome-shaped selectivity patterns in golden tilefish fisheries may be influenced by several factors (including, but not limited to):

1. Depth Distribution: Golden tilefish inhabit deep waters, typically ranging from 250 to 1,500 feet or more. Fishing gear used to target them, such as longlines, may be more effective at certain depths where golden tilefish are abundant. Intermediate depths within their range might be particularly productive, leading to a dome-shaped selectivity pattern.

2. Bait Preferences: Golden tilefish are opportunistic feeders, consuming a variety of bottom-dwelling organisms like crustaceans, mollusks, and small fish. Fishing baits commonly used to target golden tilefish may be more attractive to fish of intermediate sizes, leading to higher catch rates for this size range.

3. Gear Selectivity: The type of gear used to catch golden tilefish can influence selectivity patterns. Hooks used in longline setups may be sized and baited to target fish within a specific size range. Intermediate-sized golden tilefish may be more susceptible to being caught by these gears compared to smaller or larger individuals.

4. Behavioral Patterns: Golden tilefish exhibit certain behavioral patterns that can make them more vulnerable to fishing gear at particular sizes. Intermediate-sized fish may be more active feeders or more likely to aggregate in specific areas, increasing their encounter rates with fishing gear.

Understanding which of the above factors may be able to explain the shape of the selectivity curve may require a modeling approach, similar to a GAM. Such a formal statistical modeling approach could incorporate the specification of underlying assumptions so that the resulting estimates of gear selectivity have a clear interpretation. The Panel, therefore, recommends following the approach (SELECT) in Millar and Fryer (1999) for evaluating the hypothesis that hook size may be responsible for the dome-shaped selectivity pattern, which can be conducted with existing longline survey data.

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 Panel agreed that this ToR **was partially met**.

An underlying goal of the Golden Tilefish WG was to move from the Age Structured Assessment Program (ASAP) modeling platform used in previous Golden Tilefish assessments to the WHAM platform. Although the WHAM platform was designed to utilize state-space model formulations, there are many options where the model can be simplified such that it can be very similar to ASAP. Additionally, a Research Track Review of the WHAM approach (Summary Report of the Applied State Space Models Research Track Stock Assessment Peer Review, February 12-15, 2024, Northeast Fisheries Science Center, Woods Hole, MA) was recently conducted and the recommendations of that Panel provide some guidance for the development of the WHAM assessment applications. However, the Applied State-Space review focused on data-rich applications in their case study evaluations. Whereas, the Golden Tilefish assessment is considered data-poor because survey data are extremely limited and have not historically been used in the assessment. Thus, the tradeoffs of model construction of state-space models when data are limited have not been fully explored. The efforts of the Golden Tilefish WG initiated that sort of evaluation, but the results are incomplete and not definitive.

Nevertheless, the Applied State-Space Panel made several cogent recommendations that are applicable here for the Golden Tile assessment and which the Golden Tile WG adhered to:

> The Applied State-Space Panel suggests "*that an important consideration for selecting preferred state-space model formulations is convergence. Models that do not converge frequently in simulations or retrospective analyses are not preferable for the specific stock being investigated. Alternative and usually simpler model formulations should be investigated with a good convergence rate (i.e., > 90% in simulation).*"
>
> *"Treat recruitment as random effects so that variance and correlation parameters can be estimated."*

> *"Consider as many sources of process error as might be plausible and practical, but be aware of unintended implications for management reference points and catch advice."*

Given this background, the Golden Tilefish WG addressed the items in TOR 4

The basic data used in the assessment are: 1) a catch history; 2) a single Landings per Unit Effort standardized index from the commercial directed fishery (a small number of participants who exhibit degrees of cooperation), age length keys (although pooled over some years) and length sampling. Several alternative index sources were explored but have not yet been included in the assessment.

A bridging analysis was constructed to define a WHAM model that mimics the most recent ASAP-based assessment (Figure 4).

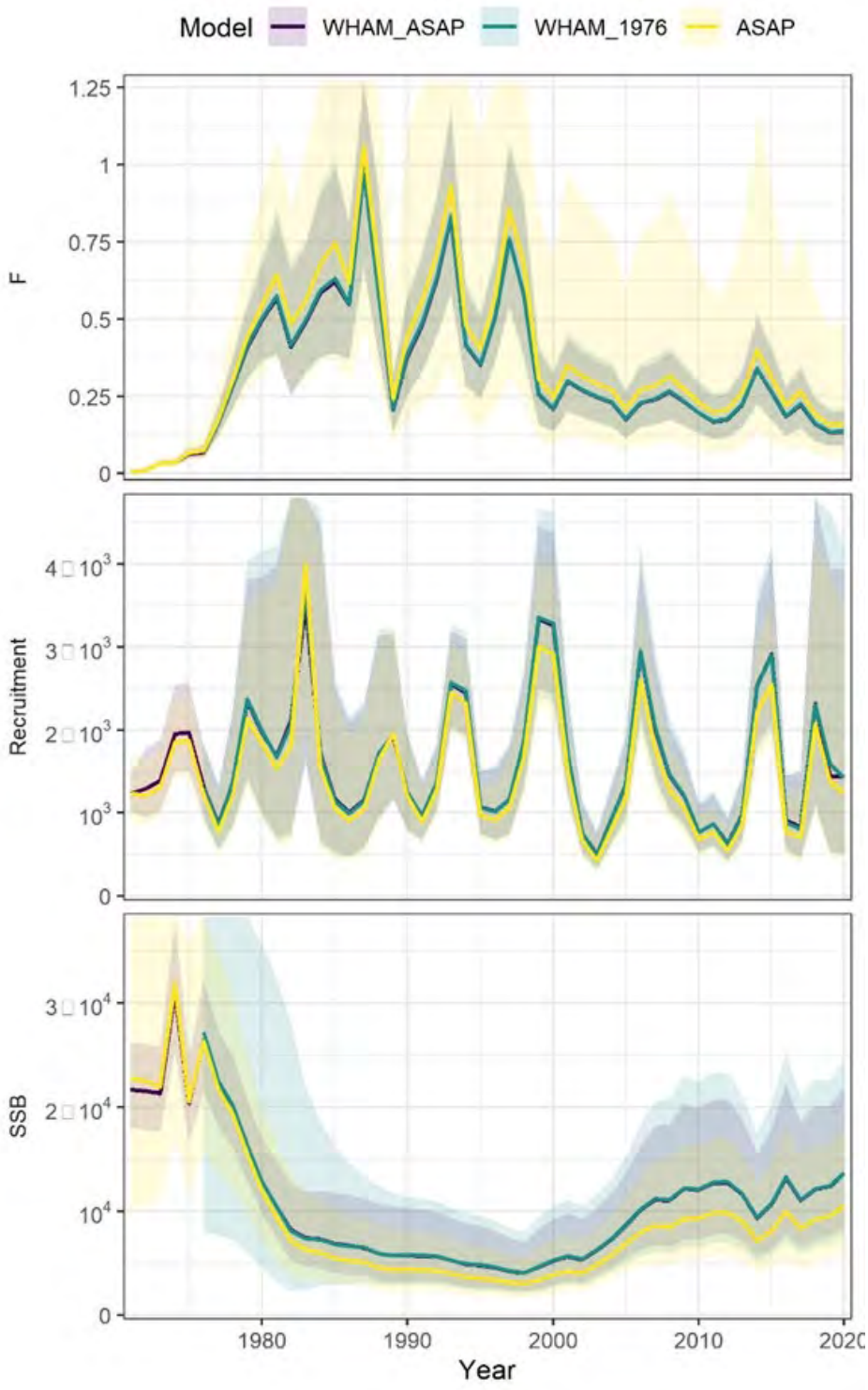

Figure 5. Bridge comparison of ASAP and WHAM

The main differences between the two were that ASAP estimates recruitment deviations around an estimated mean over the time series (with the standard deviation fixed at a prespecified

value). Whereas, WHAM recruitment was modeled with random effects on recruitment (with the standard deviation estimated). Additionally, the ASAP and WHAM constructions used two selectivity time-blocks.

A number of diagnostic features were utilized as candidate criteria for model selection (see Hennen Table 2 of WP 12):

dAIC is the change in AIC value.
rho_SSB and rho_Fbar are the Mohn's rho values in SSB and F.
OSAsd is the standard deviation of the OSA residuals.
OSApVal is the p value from the Shapiro- Wilks test for normality.
NAA cor is the correlation structure of the random effects on survival.
Selx cor is the correlation structure of the random effects on selectivity.
Age comp likelihood is the likelihood distribution used to fit the age composition data

Thus, the WG has provided a suite of statistical diagnostics and metrics to guide model selection that should be considered in final model development for the management track assessment. However, biological plausibility should also be considered as an important factor in that decision process.

A number of renditions of the WHAM model were constructed with varying approaches for random effects on numbers at age (NAA) and on selectivity. Additionally, a "best possible diagnostics" model was developed through adding additional random effects on NAA and selectivity (Full_RE). The results of that exercise lead to the following conclusions:

Models without random effects (NAA & selectivity) estimate large domes at the end of the time series which produces results that are very similar to the 2021 MT ASAP assessment with reasonable diagnostics. Forcing a flattening of the selectivity in the second block with this configuration produces retrospective issues.

Adding additional random effects to the model with a single selectivity pattern results in relative improvements in the model diagnostics while also estimating a flattening of the selectivity curve. One notable difference between the models with two selectivity blocks and the models with one selectivity pattern with random effects is that the cryptic biomass in the 10+ group was much lower in the model with a single selectivity pattern. This occurs because the model estimates a larger reduction in biomass since the development in the directed longline fishery in the 1970s with relatively lower rebuilding with the inception of management in 2001 than the model with two selectivity time blocks.

The estimated yield (msy) at $SSB_{40\%}$ increases in runs when additional random effects are added to the model which results in poor stock status (biomass and overfishing) assuming an $F_{40\%}$ proxy.

The Panel believes these results are incomplete in that these explorations have indicated directions of outcomes without a clear understanding of causes and how these interact with the limited data availability in the Golden Tilefish assessment. For example, during the review, the Panel requested an additional run with the younger age classes (age-1 and age-2) removed because there were few observations of the younger animals. Removing these younger ages from the model appeared to improve the model diagnostics, but it was only examined for a single model and with both age classes removed. Similarly, the large block of years with no catch-at-age data may be causing issues with estimability. Since additional years of data will be added before the management track assessment, the interaction of data and model structure will need to be examined further at that time. Therefore, the Panel suggests that the management track WG be allowed some flexibility to do that at that time. Specifically, the Panel suggests rerunning many of the models that were presented in the assessment report, but with age-1 and potentially age-2 removed and considering models that begin in 1995 (the first year for which age-composition data are available after the block of missing data). Because the data are being changed in the suggested runs, it will be inappropriate to use AIC for model selection choices of start year or ages to include. Furthermore, AIC does not penalize the complexity introduced by random effects in the same manner as it does for extra parameters in nested models. Alternative model selection criteria such as the Bayesian Information Criterion (BIC), or the Deviance Information Criterion (DIC) might be more suitable options.

Additionally, the WG should explore the "biological plausibility" of the individual model runs. For example, one interpretation of the age specific random effects is that they are a measure of additional natural mortality. Are the estimated natural mortalities (M+eps(t)) at age consistent with expectations of M at age? Also, are the estimated historical patterns of exploitation consistent with the historical status estimates (from the same model)? These activities might provide guidance when the data are insufficient to do so on their own.

5. Update or redefine status determination criteria (SDC; point estimates or proxies for BMSY, $B_{THRESHOLD}$, FMSY 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 Panel agreed that this TOR **was met.**

Stock-recruitment is modeled in the assessment as random effects associated with the mean recruitment over the time period. As such, no analytic model-based estimates of B_{MSY}, F_{MSY}, and

MSY are available. A proxy for F_{MSY} has previously been established as $F_{40\%}$ SPR and has been in place for many years. This specification is consistent with the life history characteristics of golden tilefish and consistent with established practices for other species of similar characteristics in both the Northeast region and elsewhere. No new information was presented to suggest any deviation from this determination. Similarly, the standard practice in the Mid-Atlantic region is to specify $B_{THRESHOLD}$ equal to 50% of $B_{SPR40\%}$.

The WG provided estimates of these SDC quantities relative to "current" conditions for the base model. However, these values will, of course, change with the completion of the Management Track assessment.

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 Panel agreed that this TOR **was met.**

The WG performed projections in line with this TOR, which were similar to previous assessments and are calculated within the WHAM approach. These were shown for illustrative purposes as projections for setting catch advice are normally done as part of the Management Track process.

Projections were conducted for 10 years, were based on the time series recruitment minus the most recent two years, using the most recent selectivity block, and used the ten-year average weight at age. Projections were conducted for both base and full random effects (Full RE) model configurations to illustrate the difference between them. Only one projection scenario was examined: fishing at $F_{40\%}$.

Unsurprisingly, the higher stock status implied by the base run resulted in fishing down the stock until it reached its $SSB_{40\%}$. Conversely, the Full RE showed an increase in stock status above the threshold reference point to target, effectively rebuilding the stock at $F_{40\%}$ within 10 years. While both projections were uncertain, the Full RE model has much higher uncertainty associated with its projection, compared to the base run.

The Panel agreed with the assessment team that the uncertainty within the projections was likely underestimated. Such projections do not fully account for potential model misspecification, the uncertainty in the year class strength near the end of the time series, nor does it capture the effects of the model's sensitivity to random effects. Additionally, the panel noted that the results were similar to previous ASAP projections conducted during the last Management Track, though those were only conducted for three years as opposed to 10 here. While there was some discussion that the 10-year average might be an alternative compared to the 5-year average, the

lack of difference between average weight using the 10 or the 5-year average presented by the team suggested that either approach was acceptable.

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 2 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 Panel agreed that this TOR **was partially met**.

The WG conducted a substantial amount of work to respond to previous research recommendations, particularly for estimating catches that had not been included in previous assessment (recreational and discards), consideration of new survey results, and consideration of environmental drivers of recruitment. However, due to time constraints, the WG was unable to agree upon a full list of research recommendations. Sub groups of the WG provided research recommendations, but they had not yet been reviewed by the full WG for inclusion or prioritization.

The Panel's list of research recommendations is below.

Necessary for management track

Incorporation of more recent (to 2023) data should be a priority for the next Management Track. The WG should focus on an assessment model that produces stable results (i.e., models with similar assumptions produce similar results) with acceptable diagnostics with good rates of convergence. Potential options for improving model stability include removing the ages that are rarely observed and only using the later years that do not have holes in the catch-at-age data. Additional sensitivity analyses should also be conducted to compare among models with different assumptions about selectivity and random effects structures. The WG should avoid using similarity with past results and management consequences as criteria for model plausibility. Rather, the most acceptable base run should be determined based on model diagnostics and biological plausibility. Estimates of selectivity, particularly differences between the early and latter part of the fishery, should be explainable. The Panel was concerned about the substantial effect of the low selectivity of older ages in the second time block of the base model. If such a model is to be used for management, there should be a plausible reason why the dome

in selectivity exists. The last benchmark assessment justified the severe doming of selectivity by arguing that the fishery only occurred in a small portion of the golden tilefish range. However, subsequent fishery independent longline surveys have indicated that there are likely not large amounts of older fish outside the area where the fishery occurs.

The Panel acknowledges that not fully endorsing a base model creates challenges. However, the Panel did not want to recommend a model without seeing the diagnostics of that model, and it was not possible to run the full set of models the Panel suggests considering at the peer review meeting. One potential criterion for choosing a base model is a comparison of the trend in estimated biomass with the "stitched together" LPUE index. The "stitched together" LPUE index would involve calculating an estimate of the amount of change in biomass by calculating adjustments for each section of the index (e.g., Turner LPUE, weighout LPUE, etc.) The stitched together index could then then be used to provide an estimate of the percentage change in biomass over the time series. An important assumption to recognize is that comparing the percentage change in the index with the percentage change in estimated biomass assumes that selectivity and catchability are comparable between the beginning and ending periods. Given that catchability of commercial fisheries is generally expected to increase over time, models that estimate less decline in biomass than the index would likely not be plausible.

The data on all the removals including recreational catch and commercial and recreational discards should be included in the management track assessment.

Given the extra work that will likely be necessary to address these recommendations and that will most likely not be externally reviewed, the Panel recommends that golden tilefish be given extra time at the management track peer review meeting for a comprehensive review of model performance, model selection, and diagnostics.

Priority research recommendations

The recommended base assessment model included a substantial dome in the selectivity function which did not seem to be explained by either the location of the fishery or fishing gear. The Panel recommends using the results from the tilefish longline surveys (2017, 2020, 2023) to estimate length-based selectivity of different hook sizes using the SELECT approach (Millar and Fryer 1999). (**Short term**)

The stock assessment uses a pooled age-length key for multiple years that had no or inadequate aging to develop year-specific age-length keys. Pooled age-length keys make assumptions about relative year-class strength being constant over years, which do not seem to be appropriate for golden tilefish. As an extreme example, consider if there is only one age class in the population. An age-length key will assign all the fish to that age. If that age-length key is used in another year, all the fish will be assigned to the wrong age. A similar effect happens with pooling age-length keys. The panel recommends using inverse age-length keys (Ailloud and Hoening 2019) to estimate the age composition of the catch for years with inadequate aging to avoid assumptions about relative yearclass strength from pooled age-length keys. (**Medium term**)

Given sexual dimorphism in growth, collect sex additional data on fishery landings. This could be potentially done with observer data from the directed longline fleet or from the tilefish

longline survey. The potential consequences of any sex specific differences on age-length keys should be considered. An exploration with existing data would be a first step. (**Medium term**)

Collect and examine length frequency samples from observers to estimate age-compositions of the trawl bycatch index to potentially include it in the assessment model. (**Long term**)

Develop an approach to include the tilefish longline survey as a fishery-independent index in the assessment. The Panel suggests an approach that considers catches of all hook sizes together instead of as separate indices because separate indices would have some level of correlation that would need to be accounted for. (**Medium term**)

Collect appropriate metrics and perform further analysis for LPUE standardization and estimation. The current stock assessment model relies heavily on LPUE as an index of biomass for the stock. However, the current method of standardization does not account for changes in set length, soak time, or number of hooks. (**Long term**)

Increase observer coverage. If port sampling of golden tilefish continues to decrease, alternative sources of size and age composition data for the fishery will be needed. These could potentially be collected by the observer program. (**Long term**)

Develop a lifecycle model of golden tilefish to better understand potential environmental drivers on population and fishery dynamics. (**Medium term**)

Consider characterization and inclusion of aging error in the assessment (**Medium term**)

Examine the assumed value(s) for natural mortality using both literature review and simulation/sensitivity analysis if possible. (**Medium term**)

Develop an age- and length-structured assessment model that avoids needing to borrow age data for years with only length composition data. (**Long term**)

The Panel also agrees with the research priorities presented by the WG:

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?

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 Pane agreed that this ToR was partially met.

The WG recommended using a simpler WHAM model as the plan B approach, and recent constant catch as a plan C approach. The panel suggests reverting back to ASAP as a plan B approach if the WHAM model cannot be implemented. The panel suggests extreme caution be used with the proposed plan C of recent constant catch because some preliminary assessment results suggested that recent catch levels may have caused overfishing to occur in recent times.

9. Identify and consider any additional stock specific analyses or investigations that are critical for this assessment and warrant peer review, and develop additional TOR(s)* to address as needed.

No additional TORs were developed by the WG.

References

Ailloud, L.E., and J.M. Hoenig. 2019. A general theory of age-length keys: combining the forward and
inverse keys to estimate age composition from incomplete data. ICES Journal of Marine Science 76: 1515–1523. doi:10.1093/icesjms/fsz072

Russell B. Millar and Robert J. Fryer, “Estimating the size-selection curves of towed gears, traps, nets and hooks”, Reviews in Fish Biology and Fisheries 9: 89–116, 1999.

Appendix 1 - Terms of Reference for Golden Tilefish Research Track Stock Assessment

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.

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.

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.

4. Use an 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.

5. Update or redefine status determination criteria (SDC; point estimates or proxies for BMSY, BTHRESHOLD, FMSY 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.

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.

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 2 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.

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.

Appendix 2 – Initial agenda for Golden Tilefish Research Track Assessment Peer Review meeting, March 11-14, 2024.

Golden Tilefish Research Track Assessment Peer Review Meeting
March 11-14, 2024

Google Meet joining info: https://meet.google.com/rgd-unsq-quh
Or dial: (US) +1 929-324-9258 PIN: 225 517 513#

DRAFT AGENDA* (v. 2/20/24)
**All times are approximate, and may be changed at the discretion of the Peer Review Panel chair. The meeting is open to the public; however, during the Report Writing sessions we ask that the public refrain from engaging in discussion with the Peer Review Panel.*

Monday, March 11, 2024

Time	Topic	Presenter(s)	Notes
9:30 a.m. - 9:45 a.m.	Welcome/Logistics/ Agenda	Michele Traver, Assessment Process Lead, Kristan Blackhart, PopDy Branch Chief, Mike Wilberg, Panel Chair	
9:45 a.m. - 10:15 a.m.	Introduction/ Executive Summary	José Montañez, WG Chair	
10:15 a.m. - 11 a.m.	Term of Reference (TOR) #1	Sarah Salois, Kimberly Hyde, Stephanie Owen, and Adelle Molina	Ecosystem
11 a.m. - 11:15 a.m.	Break		
11:15 a.m. - 12:15 p.m.	TOR #2	Paul Nitschke	Removals (commercial)
12:15 p.m. - 12:30 p.m.	Discussion	Panel	

Time	Topic	Presenter(s)	Notes
12:30 p.m. - 12:45 p.m.	Public Comment	Public	
12:45 p.m. - 1:45 p.m.	Lunch		
1:45 p.m. - 2:45 p.m.	TOR #2 cont.	José Montañez	Removals (recreational)
2:45 p.m. - 3:45 p.m.	TOR #3	Paul Nitschke	Indices
3:45 p.m. - 4 p.m.	Break		
4 p.m. - 4:45 p.m.	TOR #3 cont.	Paul Nitschke	Indices
4:45 p.m. - 5 p.m.	Discussion	Panel	
5 p.m. - 5:15 p.m.	Public Comment	Public	
5:15 p.m.	Adjourn		

Tuesday, March 12, 2024

Time	Topic	Presenter(s)	Notes
9:30 a.m. - 9:35 a.m.	Welcome/Logistics/ Agenda	Michele Traver, Assessment Process Lead, Kristan Blackhart, PopDy Branch Chief, Mike Wilberg, Panel Chair	
9:35 a.m. - 10 a.m.	TOR #3 cont.	Andy Jones	Indices
10 a.m. - 11 a.m.	TOR #3 cont.	Paul Nitschke and Jason Boucher	Indices
11 a.m. - 11:15 a.m.	Break		
11:15 a.m. - 11:45 a.m.	TOR #8	José Montañez and Paul Nitschke	Alternative approach
11:45 a.m. - 12 p.m.	Discussion	Panel	
12 p.m. - 12:15 p.m.	Public Comment	Public	
12:15 p.m. - 1:15 p.m.	Lunch		
1:15 p.m. - 3:15 p.m.	TOR #4	Paul Nitschke	Models
3:15 p.m. - 3:30 p.m.	Break		

3:30 p.m. - 4:30 p.m.	TOR #4 cont	Dan Hennen	Models
4:30 p.m. - 4:45 p.m.	Discussion	Panel	
4:45 p.m. - 5 p.m.	Public Comment	Public	
5 p.m.	Adjourn		

Wednesday, March 13, 2024

Time	Topic	Presenter(s)	Notes
9:30 a.m. - 9:35 a.m.	Welcome/Logistics/ Agenda	Michele Traver, Assessment Process Lead, Kristan Blackhart, PopDy Branch Chief, Mike Wilberg, Panel Chair	
9:35 a.m. - 10 a.m.	TOR #5	Paul Nitschke	BRPs
10 a.m. - 10:45 a.m.	TOR #6	Paul Nitschke	Projections
10:45 a.m. - 11 a.m.	Break		
11 a.m. - 11:30 a.m.	TOR #7	José Montañez and Paul Nitschke	Research Recommendations
11:30 a.m. - 11:45 a.m.	Discussion	Panel	
11:45 a.m. - 12 p.m.	Public Comment	Public	
12 p.m. - 1 p.m.	Lunch		
1 p.m. - 3 p.m.	Summary/Meeting Wrap Up	Panel	
3 p.m. - 5 p.m.	Report Writing	Panel	
5 p.m.	Adjourn		

Thursday, March 14, 2024

Time	Topic	Presenter(s)	Notes
9:30 a.m. - 12 p.m.	Report Writing	Panel	
12 p.m.	Adjourn		

Appendix 3 - Performance Work Statement (PWS) - Center for Independent Experts (CIE) Program – Golden Tilefish Research Track Peer Review

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1].

---

[1] https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2005/m05-03.pdf

Scope

The Research Track Peer Review meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The research track peer review is the cornerstone of the Northeast Region Coordinating Council stock assessment process, which includes assessment development, and report preparation (which is done by Working Groups or Atlantic States Marine Fisheries Commission (ASMFC) technical committees), assessment peer review (by the peer review panel), public presentations, and document publication. The results of this peer review will be incorporated into future management track assessments, which serve as the basis for developing fishery management recommendations.

The purpose of this meeting will be to provide an external peer review of the golden tilefish stock. The requirements for the peer review follow. This Performance Work Statement (PWS) also includes: **Annex 1**: TORs for the research track, which are the responsibility of the analysts; **Annex 2:** a draft meeting agenda; **Annex 3:** Individual Independent Review Report Requirements; and **Annex 4:** Peer Reviewer Summary Report Requirements.

Requirements

NMFS requires three reviewers under this contract (i.e. subject to CIE standards for reviewers) to participate in the panel review. The chair, who is in addition to the three reviewers, will be provided by either the New England or Mid-Atlantic Fishery Management Council's Science and Statistical Committee; although the chair will be participating in this review, the chair's participation (i.e. labor and travel) is not covered by this contract.

Each reviewer will write an individual review report in accordance with the PWS, OMB Guidelines, and the TORs below. Modifications to the PWS and TORs cannot be made during the peer review, and any PWS or TORs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor. All TORs must be addressed in each reviewer's report. The reviewers shall have working knowledge and recent experience in the use and application of index-based, age-based, and state-space stock assessment models, including familiarity with retrospective patterns, model diagnostics from various population models, and how catch advice is provided from stock assessment models. In addition, knowledge and experience with simulation analyses is helpful.

Tasks for Reviewers

- Review the background materials and reports prior to the review meeting

  - Two weeks before the peer review, the Assessment Process Lead will electronically disseminate all necessary background information and reports to the CIE reviewers for the peer review.

- Attend and participate in the panel review meeting

  - The meeting will consist of presentations by NMFS and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers

- Conduct an independent peer review in accordance with the requirements specified in this PWS and TORs, in adherence with the required formatting and content guidelines.

- Reviewers are not required to reach a consensus. Individual reviewer perspectives should be provided in their individual reports, and any lack of consensus should be clearly described in the panel's summary report.

- Each reviewer shall assist the Peer Review Panel Chair with contributions to the Peer Review Panel's Summary Report.

- Deliver individual Independent Reviewer Reports to NMFS according to the specified milestone dates.

- This report should explain whether each research track Term of Reference was or was not completed successfully during the peer review meeting, using the criteria specified below in the "Tasks for Peer Review Panel."

- If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.

- During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments and research topics may be raised. Comments on these questions should be included in a separate section at the end of the Independent Report produced by each reviewer.

- The Independent Report can also be used to provide greater detail than the Peer Reviewer Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

Tasks for Review panel

- During the peer review meeting, the panel is to determine whether each research track Term of Reference (TOR) was or was not completed successfully. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted. Where possible, the Peer Review Panel chair shall identify or facilitate agreement among the reviewers for each research track TOR.

- If the panel rejects any of the current BRP or BRP proxies (for B_{MSY} and F_{MSY} and MSY), the panel should explain why those particular BRPs or proxies are not suitable, <u>and</u> the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs or BRP proxies are the best available at this time.

- Each reviewer shall complete the tasks in accordance with the PWS and Schedule of Milestones and Deliverables below.

<u>Tasks for Peer Review Panel chair and reviewers combined</u>:

Review the Report of Golden Tilefish Research Track Working Group.

The Peer Review Panel Chair, with the assistance from the reviewers, will write the Peer Reviewer Summary Report. Each reviewer and the chair will discuss whether they hold similar views on each research track Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the peer review meeting. For terms where a similar view can be reached, the Peer Reviewer Summary Report will contain a summary of such opinions.

The chair's objective during this Peer Reviewer Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. Again, the CIE reviewers are not required to reach a consensus. The chair will take the lead in editing and completing this report. The chair may express their opinion on each research track Term of Reference, either as part of the group opinion, or as a separate minority opinion. The Peer Reviewer Summary Report will not be submitted, reviewed, or approved by the Contractor.

Place of Performance

The place of performance shall be hybrid, via in person at the Northeast Fisheries Science Center (NEFSC, 166 Water Street, Woods Hole, MA 02532) and Google Meet video conferencing.

Period of Performance

The period of performance shall be from the time of award through April 2024. Each reviewer's duties shall not exceed **14** days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Milestone	**Deliverable**
Within 2 weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
March 11-14, 2024	Panel review meeting
Approximately 2 weeks later	Contractor receives draft reports
Within 2 weeks of receiving draft reports	Contractor submits final reports to the Government

* The Peer Reviewer Summary Report will not be submitted to, reviewed, or approved by the Contractor.

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Confidentiality and Data Privacy

This contract may require that services contractors have access to Privacy Information. Services contractors are responsible for maintaining the confidentiality of all subjects and materials and may be required to sign and adhere to a Non-disclosure Agreement (NDA).

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790), and all contractor travel must be approved by the COR prior to the actual travel. Any travel conducted prior to the receipt of proper written authorization from the COR will be done at the Contractor's own risk and expense. International travel is authorized for this contract. Travel is not to exceed $12,500.00.

NMFS Project Contact

Michele Traver, NEFSC Assessment Process Lead
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
Michele.Traver@noaa.gov

Annex 1. Generic Research Track Terms of Reference

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.

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.

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.

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.

5. Update or redefine status determination criteria (SDC; point estimates or proxies for BMSY, BTHRESHOLD, FMSY 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.

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.

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 2 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.

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.

9. Identify and consider any additional stock specific analyses or investigations that are critical for this assessment and warrant peer review, and develop additional TOR(s)* to address as needed.

Research Track TORs:

General Clarification of Terms that may be Used in the Research Track Terms of Reference

Guidance to Peer Review Panels about "Number of Models to include in the Peer Reviewer Report":

In general, for any TOR in which one or more models are explored by the Working Group, give a detailed presentation of the "best" model, including inputs, outputs, diagnostics of model adequacy, and sensitivity analyses that evaluate robustness of model results to the assumptions. In less detail, describe other models that were evaluated by the Working Group and explain their strengths, weaknesses and results in relation to the "best" model. If selection of a "best" model is not possible, present alternative models in detail, and summarize the relative utility each

model, including a comparison of results. It should be highlighted whether any models represent a minority opinion.

On "Acceptable Biological Catch" (DOC Nat. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):

Acceptable biological catch (ABC) is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of Overfishing Limit (OFL) and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*

ABC for overfished stocks. For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect the annual catch that is consistent with the schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*

NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)

ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, Optimal Yield (OY) does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

On "Vulnerability" (DOC Natl. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):

"*Vulnerability.* A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce Maximum Sustainable Yield (MSY) and to recover if the population is depleted, and susceptibility is the potential for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Participation among members of a Research Track Working Group:

Anyone participating in peer review meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

Annex 2. Draft Review Meeting Agenda

{Final Meeting agenda to be provided at time of award}

Golden Tilefish Track Assessment Peer Review Meeting

March 11-15, 2024

Video link: meet.google.com/rgd-unsq-quh

Join by phone: (US) +1 929-324-9258 PIN: 225 517 513#

For Details, Please see the following link: https://www.fisheries.noaa.gov/event/golden-tilefish-2024-research-track-peer-review

Annex 3. Individual Independent Peer Reviewer Report Requirements

1. The independent Peer Reviewer report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs. The independent report shall be an independent peer review, and shall not simply repeat the contents of the Peer Reviewer Summary Report.

a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.

b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

c. Reviewers should elaborate on any points raised in the Peer Reviewer Summary Report that they believe might require further clarification.

d. The report may include recommendations on how to improve future assessments.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of this Performance Work Statement
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 4. Peer Reviewer Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the Research Track Peer Review Panel chair that will include the background and a review of activities and comments on the appropriateness of the process in reaching the goals of the peer review meeting. Following the introduction, for each assessment /research topic reviewed, the report should address whether or not each Term of Reference of the Research Track Working Group was completed successfully. For each Term of Reference, the Peer Reviewer Summary Report should state why that Term of Reference was or was not completed successfully. It should also include whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.)

   To make this determination, the peer review panel chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing fishery management advice. If the reviewers and peer review panel chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

   The report may include recommendations on how to improve future assessments.

2. If any existing Biological Reference Points (BRPs) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

3. The report shall also include the bibliography of all materials provided during the peer review meeting, and relevant papers cited in the Peer Reviewer Summary Report, along with a copy of the CIE Performance Work Statement.

   The report shall also include as a separate appendix the assessment Terms of Reference used for the peer review meeting, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 4 - Materials provided or referenced during the Golden Tilefish Research Track Stock Assessment Peer Review meeting

Working papers and presentations were available on a NEFSC website (https://apps-nefsc.fisheries.noaa.gov/saw/sasi.php) by selecting the species and year of assessment.

Working Papers and Background Documentation:

Report
Assessment Report: 2024_GoldenTilefish_RT_assessment.pdf

Models
base_diagnostics.pdf
base_iid_diagnostics.pdf
base_iid_input_data.pdf
base_iid_misc.pdf
base_iid_ref_points.pdf
base_iid_results.pdf
base_iid_retro.pdfF
base_input_data.pdf
base_misc.pdf
base_ref_points.pdf
base_results.pdf
base_retro.pdf
ful_RE_retro.pdf
full_RE_diagnostics.pdf
full_RE_input_data.pdf
full_RE_misc.pdf
full_RE_noage1&2_diagnostics.pdf
full_RE_noage1&2_input_data.pdf
full_RE_noage1&2_misc.pdf
full_RE_noage1&2_ref_points.pdf
full_RE_noage1&2_results.pdf
full_RE_noage1&2_retro.pdf
full_RE_ref_points.pdf
full_RE_results.pdf

Background
Readme First File.pdf
Tilefish_WP B1_SAW58_Assessment_Report_AOB_011314_PN_final.pdf
WP 01 Salois_et al 2024 - GTF ESP.pdf
WP 02 Nitschke 2024 - Commercial Data.pdf
WP 03 Montanez et al 2023 - Recreational Data-1.pdf
WP 04 Nitschke 2024 - Commercial CPUE.pdf
WP 05 Nitschke 2024 - Longline Study Fleet CPUE.pdf

WP 06 Jones et al 2024 - Trawl CPUE Exploration.pdf
WP 07 Boucher et al 2023- Tilefish Survey Stratified Indices at Length.pdf
WP 08 Frisk et al 2018 - 2017 Tilefish Pilot Survey.pdf
WP 09 Olin et al 2020 - 2020 Golden Tilefish Survey Report.pdf
WP 10 2018 - Report on the 2017 Pilot Tilefish Survey Review.pdf
WP 11 Nitschke 2024 - Stock Assessment Modeling - ASAPtoWHAM.pdf
WP 12 Hennen 2024 - Full Random Effects Model.pdf
WP 13 Montanez et al 2024 - History of Assessment Work.pdf
WP 14 Nitschke 2024 - Tilefish NEFSC Bottom Trawl Survey Plots.pdf
stock_miller_21.pdf

Presentations
GTF WG TOR0 Introduction & Overview.pdf
GTF WG TOR1 Ecosystem Influences.pdf
GTF WG TOR2a_commercial Removals.pdf
GTF WG TOR2b Recreational Catch.pdf
GTF WG TOR3a_LPUE.pdf
GTF WG TOR3b Trawl CPUE.pdf
GTF WG TOR3c_longline_survey.pdf
GTF WG TOR4a_models_ASAPtoWHAM.pdf
GTF WG TOR4b WHAMtileExtensions.pdf
GTF WG TOR5-6_BRPs_projections.pdf
GTF WG TOR7 Research Recommendations.pdf
GTF WG TOR8 Plan B.pdf
miller_wham_intro.pdf

Appendix 5 - Meeting attendees at the Golden Tilefish Research Track Stock Assessment Peer Review meeting

Golden Tilefish Research Track Peer Review Attendance
March 11-14, 2024

GARFO - Greater Atlantic Regional Fisheries Office
MAFMC - Mid Atlantic Fisheries Management Council
NEFSC - Northeast Fisheries Science Center
SSC - Science and Statistical Committee
~~~
~~~

Larry Alade - NEFSC
Lindsey Nelson - NEFSC
Liz Brooks - NEFSC
Paul Nitschke - NEFSC
Paul Rago - MAFMC SSC
Sarah Salois - NEFSC
Sefatia Romeo Theken - Deputy Commissioner for MA Fisheries and Game
Stephanie Owen - NEFSC
Susan Wigley - NEFSC
Tim Miller - NEFSC
Toni Chute - NEFSC
Tony Wood - NEFSC