Executive Summary

The Southern Demersal Working Group met from 27-28 April, 2009 at the Northeast Fisheries Science Center, Woods Hole, Massachusetts to address the terms of reference agreed by the NRCC for tilefish. The following members were in attendance:

Dan Farnhan	F/V Kimberly
Chris Legault	NEFSC
Richard McBride	NEFSC
Jose´ Montañez	MAFMC
Josh Moser	NEFSC
Paul Nitschke	NEFSC (Assessment Lead)
John Nolan	F/V Seacapture
Laurie Nolan	F/V Seacapture
Michael Palmer	NEFSC
Barbara Rountree	NEFSC
Gary Shepherd	NEFSC
Martin Smith	Duke University (SSC lead, phone)
Katherine Sosebee	NEFSC
Mark Terceiro	NEFSC (Chair)
Tiffany Vidal	NEFSC
Susan Wigley	NEFSC

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

Terms of Reference

1. Characterize the commercial catch including landings, effort and discards. Characterize recreational landings. Evaluate utility of study fleet results as improved measures of CPUE.

Total commercial landings (live weight) increased from less than 125 metric tons (mt) during 1967-1972 to more than 3,900 mt in 1979 and 1980. Annual landings have ranged between 666 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. Landing from 2005 to 2008 have been at or below the quota. Landings in 2007 and 2008 were 751 mt and 736 mt respectively. During 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 of the commercial landings are taken by the directed longline fishery. Discards in the trawl and longine fishery are a minor component of the catch. Recreational catches have also been low for the last 25 years (i.e., less than 1 mt caught annually).

A fishery independent index of abundance does not exist for 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 fished obtained from logbooks of tilefish fishermen. Two additional CPUE series were calculated from the NEFSC weighout (1979-1993) and the VTR (1995-2008) systems. The number of vessels targeting tilefish has declined over the time series; during 1994-2003, five vessels accounted for more than 70 percent of the total tilefish landings. The length of a targeted tilefish trip had been generally increasing until the mid 1990s. At the time of the last assessment (2005) trip lengths have shorten to about 5 days. Since then trip length has been increasing.

Six market categories exist in the database. From smallest to largest they are: small, kitten, medium, large and extra large as well as an unclassified category. The proportion of landings in the kittens and small market categories increased in 1995 and 1996. Evidence of two strong recruitment events can be seen tracking through these market categories. At the time of the last tilefish assessment (2005) the proportion of large market category has declined since the early 1980s. However more recently most of the landings come from the large market category as the last strong year class (1999) has grown. Commercial length sampling has been inadequate over most of the time series. However some commercial length sampling occurred in the mid to late 1990s. More recently there has been a substantial increase in the commercial length sampling from 2003 to 2008.

Study fleet analysis is addressed in Appendix A1.

2. Estimate fishing mortality and total stock biomass for the current year, and for previous years if possible, and characterize the uncertainty of those estimates. Incorporate results of new age and growth studies.

As in SARC 41 the 2009 Working Group accepted the formulation that began the analysis in 1973, separated the Turner, weighout and VTR CPUE into three series and fixed the $B1/B_{MSY}$ ratio at 1 as the final run (base run). The working group expressed some concern over whether the CPUE in this fishery is more a reflection of changes in fishing practices and changes in spatial distribution of the fish rather than fluctuations in population size. Commercial length data indicate that increases in total biomass are predominantly due to a strong 1999 year class. It appears that most of the commercial catch over the 2002-2007 period were derived from this year class. Process error in the ASPIC model associated with the recent large year class has increased at the end of the time series due to an assumed constant recruitment/growth parameter.

The Working Group examined results obtained from an alternative forward projecting age/size structured model (SCALE) due to the difficulties with ASPIC model fitting the CPUE index at the end of the time series. An earlier version of this model was call catch-length model in SARC 41. The SCALE model incorporates population growth and length information into the model framework. This allows for the estimation of strong recruitment events which can be seen in the commercial length frequency distributions over time. However the overall lack of data and issues with independence of the data sources is a source of concern with the SCALE model results. The lack of a recruitment index, inability to estimate uncertainty using mcmc, and questions with the estimated flat top selectivity curve are also sources of uncertainty. However SCALE model results suggests that the surplus production model may have overestimated the productivity of the stock.

New age and growth study is addressed in Appendix A2.

3. Update or redefine biological reference points (BRPs; estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, and F_{MSY}). Comment on the scientific adequacy of existing and redefined BRPs.

Biological reference points estimated by the 2009 ASPIC BASE run are moderately different from the 2005 SAW 41 assessment. B_{MSY} is estimated to be 11,400 mt (a 22% increase), F_{MSY} is estimated to be 0.16 (a 24% decrease), and MSY is estimated to be 1,868 mt (a 6% decrease), compared to B_{MSY} = 9,384 mt, F_{MSY} = 0.21, and MSY = 1,988 mt from the 2005 SAW 41 assessment.

SCALE yield per recruit biological reference points suggest that SSB_{MSY} is between 9,878 mt and 15,108 mt for the combine sex run using F_{40} or F_{MAX} as the F_{MSY} proxy. The separate sex run suggests female SSB_{MSY} is between 5,335 mt and 7,100 mt. For both the single sex and separate sex run the F_{MSY} is between 0.079 and 0.128 and MSY ranging from 1,072 mt to 1,200 mt using either F_{40} or F_{MAX} as the F_{MSY} proxy.

4. Evaluate stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 3).

The biomass-based surplus production model (ASPIC) indicates that the tilefish stock biomass in 2008 has improved since the last assessment in 2005. Total biomass in 2008 is estimated to be 104% of B_{MSY} and fishing mortality in 2008 is estimated to be 38% of F_{MSY}. The tilefish stock is not overfished and overfishing is not occurring. The SARC 48 review panel accepted the ASPIC model but concluded that the ASPIC model is likely over optimistic and that the stock has not rebuilt above B_{MSY}.

SCALE model result suggests a different status determination. The 2009 BASE SCALE model run (separate sex run) and the combined sex run results indicate that the 2009 Golden tilefish stock is at a low biomass (29% to 47% of SSB_{MSY}) and is overfished with respect to the update SSB reference points. Both SCALE runs also suggest recruitment and growth overfishing (147% to 260% of F_{MSY}) is occurring with respect to the F_{40} or F_{MAX} updated biological reference points. However fishing mortality has been decreasing and biomass has been increasing since the beginning of the FMP in 2001.

5. Develop and apply analytical approaches and data that can be used for conducting single and multi-year stock projections and for computing candidate ABCs (Acceptable Biological Catch).

a. Provide numerical short-term projections (2-3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. In carrying out

projections, consider a range of assumptions about the most important uncertainties in the assessment (alternate states of nature).

b. If possible, comment on the relative probability of the alternate states of nature and on which projections seem most realistic.

c. For a range of candidate ABCs, compute the probabilities of rebuilding the stock by November 1, 2011.

d. Describe this stock's vulnerability to becoming overfished, and how this could affect the choice of ABC.

The Working Group examined several ASPIC projections, including the current TAC of 905 mt. The ASPIC model indicates the stock is rebuilt and F in 2008 is low. Therefore the projections suggest the stock will continue to build if catches remain below MSY (1,854 mt). Projection scenarios that incorporated a possible future CPUE index illustrate the concern with the model stability due to the year class effects in the CPUE index. The scenario projections suggest that uncertainty with the stock status determination is much higher than what is suggested from the bootstrap uncertainty distributions and the standard projections.

Several options (age-based AGRPRO, deterministic SCALE projection) are available for 63SCALE model projections depending on whether growth is model as a single sex or with the sexes separated. Continued stock rebuilding is projected in the SCALE model with status quo conditions. Uncertainty estimates were not possible likely due to the overall lack of data in the model. Results of the SCALE model should be considered as a possible alternative state of nature for judging the extent of the overall uncertainty in the assessment when setting an ABC.

6. Review, evaluate and report on the status of the research recommendations offered in recent SARC reviewed assessments. Identify new research recommendations, including recruitment estimation.

Most of the research recommendations were addressed through the new study fleet project and updated growth study. Several new research recommendations were also suggested at the working group meeting, including continuation of the tilefish study fleet program or possibly modifying the study fleet program into an industry based survey that could obtain a recruitment index as part of the sampling design. Research recommendations TOR 6 are summarized on pages 32-33.

Introduction

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

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

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

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

Term of Reference 1: Commercial Fishery

TOR 1: Characterize the commercial catch including landings, effort and discards. Characterize recreational landings. Evaluate utility of study fleet results as improved measures of CPUE.

See Appendix A1 for details on the utility of study fleet results as an improved measures of CPUE.

Data Sources

Commercial catch data

Total commercial landings (live weight) increased from less than 125 mt during 1967-1972 to more than 3,900 mt in 1979 and 1980 (Table A1, Figure A1). Landings stabilized at about 2,000 mt during 1982-1986. An increase in landings occurred in 1987 to 3,200 mt but subsequently declined to 450 mt in 1989. Annual landings have 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 above the quota at 1,130 mt and 1,215 mt respectively. Landing from 2005 to 2008 have been at or below the quota. Landings in 2007 and 2008 were 751 mt ant 736 mt respectively. Over 75% of the landings came from Statistical Areas 537 and 616 since 1991 (Table A2). Since the 1980s, over 85% of the commercial landings of tilefish in the MA-SNE region have been taken in the longline fishery (Table

A3, Figure A2). During the late 1970s and early 1980s Barnegat, NJ was the principal tilefish port; more recently Montauk, NY has accounted for most of the landings. The shift in landings can be seen in the proportion of the landings by state in Table A4 and Figure A3. In the late 1970s and earlier 1980s a greater proportion of the landings were taken in quarters 1 and 2 (Table A5, Figure A4). Recent landings have been relatively constant over the year.

Commercial discard data

Very little discarding (< 1%) of tilefish was reported in the vessel trip report (VTR) from longline vessels that target tilefish and there is little reported discarding of tilefish in the trawl fishery in the VTR data (SARC 41). Recent observer directed tilefish longline trips also suggest that discards of tilefish is minimal. Observer trawl data produce more variable discard estimates across years for tilefish. Discard to kept ratios for trawl trips that either kept or discarded tilefish in the observer data varied from 0 in 1993 to 1.4 in 2001 (Table A6). Twelve of the sixteen years had less than 15 trips sampled that caught tilefish from 1989 to 2003. The number of observer trips that caught tilefish has increase from 2004 to 2008 (average 47). Trawl discards were not expanded to derive total discards due to the relativity minor component of the trawl landings to the total and due to the high uncertainty associated with the hindcast estimates.

Commercial CPUE data

Analyses of catch (landings) and effort data were confined to the longline fishery since directed tilefish effort occurs in this fishery (e.g. the remainder of tilefish landings are taken as bycatch in the trawl fishery). Most longline trips that catch tilefish fall into two categories: (a) trips in which tilefish comprise greater than 90% of the trip catch by weight and (b) trips in which tilefish accounted for less than 10% of the catch. Effort was considered directed for tilefish when at least 75% of the catch from a trip consisted of tilefish (NEFSC 1993).

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 CPUE series were calculated from the NEFSC weighout (1979-1993) and the VTR (1995-2008) systems as well as a combined 1979-2008 series. Effort from the weighout data was derived by port agents' interviews with vessel captains whereas effort from the VTR systems comes directly from mandatory logbook data. In this assessment and in the 1998 and 2005 tilefish assessments we used Days absent as the best available effort metric. In the 1998 assessment an effort metric based on Days fished (average hours fished per set / 24 * number of sets in trip) was not used because effort data were missing in many of the logbooks and the effort data were collected on a trip basis as opposed to a haul by haul basis. For this assessment effort was calculated as:

Effort = days absent (time & date landed - time & date sailed) - number of trips.

For some trips, the reported days absent were calculated to be a single day. This was considered unlikely, as a directed tilefish trip requires time for a vessel to steam to near the edge of the continental shelf, time for fishing, and return trip time (Grimes et al. 1980). Thus, to produce a realistic effort metric based on days absent, a one day steam time for each trip (or the number of trips) was subtracted from days absents and therefore only trips with days absent greater than one day were used.

The NEFSC Weighout and VTR CPUE series were standardized using a general linear model (GLM) incorporating year and individual vessel effects (Mayo et al. 1994). The CPUE was standardized to an individual longline vessel and the year 1984; the same year used in the last assessment. For the VTR series the year 2000 was used as the standard. Model coefficients were back-transformed to a linear scale after correcting for transformation bias (Granger and Newbold 1977). The full GLM output for the Weighout and the VTR CPUE series is included as Appendix A3.

The number of vessels targeting tilefish has declined over the time series (Table A7, Figure A5); during 1994-2003, five vessels accounted for more than 70 percent of the total tilefish landings (Table A8, Figure A6). The number of vessels targeting tilefish has remained fairly constant since the last assessment in 2005. The length of a targeted tilefish trip had been generally increasing until the mid 1990s. At the time of the last assessment (2005) trip lengths have shorten to about 5 days. Since then trip length has been increasing (Figure A5). In the weighout data the small number of interview is a source of concern; very little interview data exists at the beginning of the time series (Table A7, Figure A7). The 5 dominant tilefish vessels make up almost all of the VTR data with the exception of 2004 when there appears to be more vessels targeting tilefish (Figure A6). In some years there were higher total landings reported in the VTR data than the Dealer data for the 5 dominant tilefish vessels. After the FMP was implemented the IVR (Interactive Voice recorder) database was developed to monitor the quota. In 2005 the IVR database had the highest landings level despite that this system only applies to the limited access tilefish fishery. The IVR 2005 total was assumed to be a better estimate of the total landings in that year then the other data sources. The IVR total landing in 2005 was used as the total removals in all tilefish modeling.

The number of targeted tilefish trips declined in the early 1980s while trip length increased at the time the FMP was being developed in 2000 (Figures A5 and A8). During the last assessment in 2005 the number of trips became relatively stable as trip length decreased. Since the last assessment trip length has increased. The interaction between the number of vessels, the length of a trip and the number of trips can be seen in the total days absent trend in Figure A8. Total days absent remained relatively stable in the early 1980s, but then declined at the end of the weighout series (1979-1994). In the beginning of the VTR series (1994-2004) days absent increased through 1998 but declined to 2005. Since 2005 total days absent has increase somewhat. Figure A8 also shows that a smaller fraction of the total landings were included in the calculation of CPUE compared to the VTR series.

Figure A9 illustrates difference between the nominal CPUE and vessel standardized (GLM) CPUE with the weighout and VTR data combined. CPUE trends are very similar for most vessels that targeted tilefish (Figure A10). A sensitivity test of the GLM using different vessel combinations was done in SARC 41. The SARC 41 GLM was found not to be sensitivity to different vessels entering the CPUE series.

Very little CPUE data exist for New York vessels in the 1979-1994 weighout series despite the shift in landing from New Jersey to New York before the start of the VTR series in 1994. The small amount of overlap between the weighout and VTR series is illustrated in Figures A11 and A12. Splitting the weighout and VTR CPUE series can be justified by the differences in the way effort was measured and difference in the tilefish fleet between the series. In breaking up the series we omitted 1994 because there were very little CPUE data. The sparse 1994 data that existed came mostly from the weighout system in the first quarter of the year. Very similar trends exist in the four years of overlap between Turner (1986) CPUE and the weighout series (Figure A13).

Since 1979, the tilefish industry has changed from using cotton twine to steel cables for the backbone and from J hooks to circle hooks. The gear change to steel cable and snaps started on New

York vessels in 1983. In light of possible changes in catchability associated with these changes in fishing gear, the working group considered that it would be best to use the three available indices separately rather than combined into one or two series. The earliest series (Turner 1986) covered 1973-1982 when gear construction and configuration was thought to be relatively consistent. The Weightout series (1979-1993) overlapped the earlier series for four years and showed similar patterns (Figure A13) and is based primarily on catch rates from New Jersey vessels. The VTR (1995-2004) series is based primarily on information from New York vessels using steel cable and snaps.

In SARC 41 a month vessel interaction was significant but explained only a small amount of the total sum of squares (6%*).* Adding a month - vessel interaction term to the GLM model had very little influence on the results at SARC 41 and was not updated for this assessment. The GLM output for the Weighout and VTR CPUE series standardized for individual vessel effects can be seen in Appendix A3.

In this assessment the sensitivity of the assumed error structure used in VTR GLM CPUE index was explored. The nominal VTR CPUE data distribution does appear over-dispersed relative to normal or lognormal distribution, suggesting that a model with poisson or negative binomial distribution may be more appropriate (Figure A14). However the GLM CPUE indices using different error assumptions showed very little differences in the CPUE trends (Figure A15). Therefore the lognormal error distribution was retained.

Commercial market category and size composition data

Six market categories exist in the database. From smallest to largest they are: small, kitten, medium, large and extra large as well as an unclassified category. In 1996 and 1997, the reporting of tilefish by market categories increased, with the proportion of unclassified catch declining to less than 20% (Table A9, Figure A16). The proportion of landings in the small and kitten market categories increased in 1995 and 1996. Small and kitten market categories had similar length distributions and samples from 1995 to 1999 were combined. Evidence of several strong recruitment events can be seen tracking through the market category proportions (Figures A16 and A17). At SARC 41 the proportion of the large market category has declined since the early 1980s (Figure A16). Landings data obtained directly from the New York tilefish industry shows a similar decline in the proportion of the large market category between 1980 and 1990 (Figure A18). Landings by market category has shifted from smalls and kittens in 2004 to larges in 2007 and 2008 which is likely the result of a strong year class effect (Figure A17).

Extensive size sampling was conducted in 1976-1982 (Grimes *et al.* 1980, Turner 1986) however that data are not available by market category (Figure A19). Since then commercial length sampling has been inadequate in most years (Table A4). However some commercial length sampling occurred in the mid to late 1990s. More recently there has been a substantial increase in the commercial length sampling in 2003 and 2004. Commercial length sampling in New York has also increased since the last assessment in 2005 (Table A4). Expanded length frequency distributions from 1995 to 1999 from SARC 41 are shown in Figure A20. In this assessment expanded length frequency distributions were estimated form 2002 to 2008 (Figure A21 and A22). The stratification used in the expansion can be seen in table A10. The large market category length frequencies appear to have been relatively stable for years when more than 100 fish were measured. However the small market category exhibits shifts in the size distribution in certain years as strong year classes move through the fishery (Figure A23). The tracking of a year class can be seen as the cohort grows over the year in 2003 and 2004 (Figure A23). The strong 1998/1999 year class seen in the kept length frequency distributions from tilefish longline observer trips matches well with the expanded commercial length frequency distributions (Figures A24). In addition, the 2008 study fleet length distribution looks similar to the 2008 commercial landings distribution (Figure A25).

Smaller fish sizes are seen in the trawl gear length distributions for the small and kitten market category compared to longline gear (Figure A26). Therefore trawl length frequency distribution where not used to characterize the catch (Table A10). Longline tilefish fishermen often receive forecasts from the draggers of when a strong year class will be entering the fishery. There is some anecdotal information from draggers for the existence of a stronger year class in 2009.

Commercial length frequencies were expanded for years where sufficient length data exist (1995-1999 and 2002-2008) (Table AC10). The large length frequency samples from 1996 to 1998 were used to calculate the 1995 to 1999 expanded numbers at length while the large length samples from 2001 and 2003 were used to calculate the 2002 expanded numbers at length. Evidence of strong 1992/1993 and 1998/1999 year classes can be seen in the expanded numbers at length in the years when length data existed (1995-1999 and 2002-2008) (Figure A20). The matching of modes in the length frequency with ages was done using Turner's (1986) and Vidal's (2009) aging studies. In 2004 and 2005 the 1998/1999 year class can be seen growing into the medium market category and in 2006 and 2007 the year class has entered the large market category (Figure A20). From 2002 to 2007 it appears that most of the landings were comprised of this year class. The catch appears to be comprised of multiple year classes in 2008 after catch rates have declined in the VTR series. An increase in the landings and CPUE can be seen when the 1992/1993 and 1998/1999 year classes recruit to the longline fishery. As the year classes gets older the catch rates decline (Figure A13 and A21).

Recreational data

A small recreational fishery occurred briefly in the mid 1970s (< 100 mt annually, Turner 1986) but subsequent recreational catches have been quite low for the last 30 years (i.e., less than 1 mt caught annually) (Table A11). Party and charter boat vessel trip reports also show low numbers of tilefish being caught since 1994 (Table A12).

NEFSC Trawl survey data

Only a few fish per survey are caught during NEFSC bottom trawl surveys. This survey time series is not useful as an index of abundance for tilefish.

Term of Reference 2: Mortality and stock size estimates

TOR2: Estimate fishing mortality and total stock biomass for the current year and for previous years if possible, and characterize the uncertainty of those estimates. Incorporate results of new age and growth studies.

See Appendix A2 for details on the new age and growth study.

ASPIC Surplus production model

The ASPIC surplus production model (Prager 1994; 1995) was used to determine fishing mortality, stock biomass and biological reference points (F_{MSY}, and B_{MSY}) for the development of the tilefish FMP in 2001. SARC 41 in 2005 accepted the ASPIC model as a basis for determining whether the stock was on schedule for rebuilding by 2011.

As a first step in the surplus production modeling, the landings and index data from the 2005

SAW41 assessment were used as input in the latest version (5.33) of the ASPIC software and compared with the results from the 2005 SAW 41 assessment, which was run in ASPIC version 3.93. There were no significant differences in the results due to the ASPIC version update (Table A13). The three commercial fishery CPUE index series (Turner 1973-1982; NEFSC Weighout 1982-1993; and VTR 1995-2004) as configured in the 2005 SAW 41 assessment were retained in constructing the 2009 ASPIC model configurations. The VTR CPUE index of abundance and commercial fishery landings were updated through 2008 to create the 2009 BASE run. A bootstrap with 1000 iterations was used to estimate confidence intervals for annual F and stock biomass estimates and biological reference points. Several sensitivity runs were made to further evaluate the impact on results of the assumption for the B1/K ratio starting condition (equivalent to the $B1/B_{MSY}$ ratio in the 2005 SAW 41 assessment ASPIC v3.93). A retrospective analysis of the BASE run was made to evaluate model performance.

The trends in fishing mortality (F; in the ASPIC model, this is the ratio of annual catch to average annual stock biomass) were very similar in the 2005 SAW 41 and in the 2009 BASE results through 2004. The 2005 SAW 41 F estimates generally followed the 75%ile of the 2009 BASE estimates of F (i.e., were generally somewhat higher), while the 2005 SAW 41 biomass estimates followed the 25%ile of the 2009 BASE estimates of biomass (i.e., were generally somewhat lower; Figures A27 and A28). The early period (Turner 1973-1982) indices fit better (higher r2 value) in the 2009 BASE run than in the 2005 SAW 41 assessment; conversely, the two later series (NEFSC Weighout 1982-1993 and VTR 1995-2008) fit worse (lower r2 values) (Figure A29). Catchability coefficients (q) decreased for all three index series (Turner by 34%; NEFSC Weighout by 22%; VTR by 34%). The biomass reference points (B_{MSY} and K) increased by 22% from the 2005 SAW 41 run to the 2009 BASE run, while FMSY decreased by 22% and MSY decreased by 6%. The 2009 BASE run estimates provide a more optimistic evaluation of stock status in 2004 than did the 2005 SAW 41 model estimates (e.g., the B_{2004}/B_{MSY} ratio; Table A13).

As in the last assessment, sensitivity runs were made to explore the effect of the value of the B1/K ratio on results (B1 is the stock biomass in the first year of the analysis time series; K is the carrying capacity of the stock, equivalent to the biomass when fishing mortality is zero over the long-term). In the 2009 BASE run configuration the B1/K ratio was fixed at 0.50 (equivalent to the $B1/B_{MSY}$ ratio = 1.00 in the 2005 SAW 41 ASPIC v3.93). The BASE results were compared with runs fixing B1/K at 0.10, 1.00, and a run in which B1/K was estimated at 1.19. The run with B1/K fixed at 0.10 provides a value for the Root Mean Squared Error (RMSE) value over 50% higher than the BASE run and negative r2 values for all 3 CPUE index series. The estimates of K (carrying capacity), MSY (Maximum Sustainable Yield), and FMSY (fishing mortality rate providing MSY) for this run are infeasible given the historical pattern and magnitude of fishery landings and the life history characteristics of tilefish (Table A13, dashed lines in Figures A30 and A31).

The runs fixing B1/K = 1.00 and estimating B1/K = 1.19 provided results and diagnostics comparable to the BASE run with B1/K = 0.50. Estimates of F and biomass for 1979 and later years are nearly identical to the BASE run. The major differences are for 1973-1978, when the B1/K = 1.00 and B1/K = 1.19 runs obviously indicate that the stock declined from a high biomass level near K. Estimates of MSY and K for these sensitivity runs are about 10% (B_{MSY}) and 16% (K) lower than the BASE run, while estimates of F_{MSY} are 10-15% higher (Table A13, Figures A30 and A31). The runs fixing/estimating B1/K ratio near 1.00 in 1973 imply that the stock was near carrying capacity in the early 1970s, which is unlikely given the historical pattern and magnitude of fishery landings. The 2005 SAW 41 review concluded that the most likely assumption for the B1/K ratio was 0.50 (equivalent to $B1/B_{MSY}$ = 1.00). That assumption is again supported by the current sensitivity analysis results, and so has been retained for the 2009 BASE run configuration.

A retrospective analysis (sequential removal of the last year of data) was conducted for the 2009 BASE run configuration with ten "peels" (ten years sequentially removed from the end of the analysis). The BASE run results are fairly stable for the 1999, 2002-2008 terminal years, both in terms of time series trends (Figures A32 and A33) and in the estimated catchability coefficients and reference points (left side of Table A14). For the 1998, 2000-2001 terminal years, however, the 2009 BASE run converged at a different solution but with a comparable value of the RMSE. For the 1998, 2000-2001 runs, the estimated catchability coefficients were about 25-50% of the 1999, 2002-2008 runs, and the estimated reference points were infeasible given the historic trend and magnitude of the fishery landings (right side of Table A14). These results indicate that the current 2009 BASE model solution is stable for the last several terminal years, but also indicates that future runs should continue to be examined in a similar manner (multiple retrospectives and sensitivity analyses) to evaluate performance.

The 2009 BASE run indicates that the tilefish stock biomass has continued to increase since the 2005 SAW 41 assessment (Figures A28 and A29). Fishing mortality (F = 0.06) is estimated to be 38% of F_{MSY} and stock biomass in 2008 (B = 11,910 mt) is estimated to be 4% above B_{MSY} (Table A13). Bootstrap (1000 iterations) estimates of the 2008 F were 0.05 (25%ile) to 0.07 (75%ile), with a median of 0.06 (50%ile; Figure A34). Bootstrap estimates of the 2008 stock biomass were 9,550 mt (25%ile) to 13,538 mt (75%ile), with a median of 11,767 mt (50%ile; Figure A35). The complete ASPIC model output with bootstrap results is included as Appendix A3.

Expanded landing length frequency distributions and trends in the VTR CPUE suggest recent strong year class effects in the fishery. The recent strong 1998/1999 year class results in increase process error with the fit to the VTR series in the ASPIC model since the surplus production model assumes constant growth/recruitment (Figure A30). The increase in error is reflected in the comparison of the r^2 from the SARC 41 ASPIC assessment (0.54) with the updated assessment (0.20).

SCALE Model

The working group investigated the use of an age and size structured forward projection model (SCALE) for assessing the tilefish stock due to the inability of the ASPIC surplus production model in fitting the observed year class effects. Incomplete or lack of age-specific catch and survey indices often limits the application of a full age-structured assessment (e.g. Virtual Population Analysis and many forward projecting age-structured models). Stock assessments will often rely on the simpler size/age aggregated models (e.g. surplus production models) when age-specific information is lacking. However the simpler size/age aggregated models may not utilize all of the available information for a stock assessment. Knowledge of a species growth and lifespan, along with total catch data, size composition of the removals, recruitment indices and indices on numbers and size composition of the large fish in a survey can provide insights on population status using a simple model framework.

The Statistical Catch At LEngth (SCALE) model, is a forward projecting age-structured model tuned with total catch (mt), catch at length or proportional catch at length, recruitment at a specified age (usually estimated from first length mode in the survey), survey indices of abundance of the larger/older fish (usually adult fish) and the survey length frequency distributions (NOAA Fisheries Toolbox 2008a). The SCALE model was developed in the AD model builder framework. The model parameter estimates are fishing mortality and recruitment in each year, fishing mortality to produce the initial population (Fstart), logistic selectivity parameters for each year or blocks of years and Qs for each survey index.

The SCALE model was developed as an age-structured model that does NOT rely on age-specific information on a yearly basis. The model is designed to fit length information, abundance indices, and recruitment at age which can be estimated by using survey length slicing. However the model does require an accurate representation of the average overall growth of the population which is input to the model as mean lengths at age. Growth can be modeled as sex-specific growth and natural mortality or growth and natural mortality can be model with the sexes combined. The SCALE model will allow for missing data.

Model Configuration

The SCALE model assumes growth follows the mean input length at age with predetermined input error in length at age. Therefore a growth model or estimates of mean length at age are essential for reliable results. The model assumes static growth and therefore population mean length/weight at age are assumed constant over time. A depiction of model assumed population growth at age using the input mean lengths at age and variation can be seen in Table A15).

The SCALE model estimates logistic parameters for a flattop selectivity curve at length in each time block specified by the user for the calculation of population and catch age-length matrices or the user can input fixed logistic selectivity parameters. Presently the SCALE model can not account for the dome shaped selectivity pattern

The SCALE model computes an initial age-length population matrix in year one of the model as follows. First the estimated populations numbers at age starting with age-1 recruitment get normally distributed at one cm length intervals using the mean length at age with the assumed standard deviation. Next the initial population numbers at age are calculated from the previous age at length abundance using the survival equation. An estimated fishing mortality (Fstart) is also used to produce the initial population. This F can be thought of as the average fishing mortality that occurred before the first year in the model. Now the process repeats itself with the total of the estimated abundance at age getting redistributed according to the mean length at age and standard deviation in the next age (age+1).

This two step process is used to incorporate the effects of length specific selectivities and fishing mortality. The initial population length and age distribution is constructed by assuming population equilibrium with an initial value of F, called F_{start}. Length specific mortality is estimated as a two step process in which the population is first decremented for the length specific effects of mortality as follows:

$$N^{*}_{a,len,y_1} = N_{a-1,len,y_1} e^{-\left(PR_{len} F_{start} + M\right)}$$

In the second step, the total population of survivors is then redistributed over the lengths at age *a* by assuming that the proportions of numbers at length at age *a* follow a normal distribution with a mean length derived from the input growth curve (mean lengths at age).

$$N_{a,len,y_1} = \pi_{len,a} \sum_{len=0}^{L_\infty} N^{*}_{a,len,y_1}$$

where

$$\pi_{len,a} = \Phi\left(len+1 \mid \mu_a, \sigma_a^2\right) - \Phi\left(len \mid \mu_a, \sigma_a^2\right)$$

where

$$\mu_a = L_\infty\left(1 - e^{-K(a-t_0)}\right)$$

Mean lengths at age can be calculated from a von Bertalanffy model from a prior study as shown in the equation above or mean lengths at age can be calculated directly from an age-length key. Variation in length at age a = σ_s^2 can often be approximated empirically from the growth study used for the estimation of mean lengths at age. If large differences in growth exist between the sexes then growth can be input as sex-specific growth with sex-specific natural mortality. However catch and survey data are still fitted with sexes combined.

This SCALE model formulation does not explicitly track the dynamics of length groups across age because the consequences of differential survival at length at age a do not alter the mean length of fish at age a+1. However, it does more realistically account for the variations in age-specific partial recruitment patterns by incorporating the expected distribution of lengths at age.

In the next step the population numbers at age and length for years after the calculation of the initial population use the previous age and year for the estimate of abundance. Here the calculations are done on a cohort basis. Like in the previous initial population survival equation the partial recruitment is estimated on a length vector.

$$N^*_{a,len,y} = N_{a-1,len,y-1} e^{-\left(PR_{len}F_{y-1}+M\right)}$$

second stage

$$N_{a,len,y} = \pi_{len,a} \sum_{len=0}^{L_\infty} N^*_{a,len,y}$$

Constant M is assumed along with an estimated length-weight relationship to convert estimated catch in numbers to catch in weight. The standard Baranov's catch equation is used to remove the catch from the population in estimating fishing mortality.

$$C_{y,a,len} = \frac{N_{y,a,len}\, F_y\, PR_{len}\left(1 - e^{-\left(F_y PR_{len}+M\right)}\right)}{(F_y \quad PR_{len}) + M}$$

Catch is converted to yield by assuming a time invariant average weight at length.

$$Y_{y,a,len} = C_{y,a,len} W_{len}$$

The SCALE model results in the calculation of population and catch age-length matrices for the starting population and then for each year thereafter. The model is programmed to estimate recruitment in year 1 and estimate variation in recruitment relative to recruitment in year 1 for each year thereafter. Estimated recruitment in year one can be thought of as the estimated average long term recruitment in the population since it produces the initial population. The residual sum of squares of the variation in recruitment $\sum(\mathrm{Vrec})^2$ is than used as a component of the total objective function. The weight on the recruitment variation component of the objective function (Vrec) can be used to penalize the model for estimating large changes in recruitment relative to estimated recruitment in year one.

The model requires an age-1 recruitment index for tuning or the user can assume relatively constant recruitment over time by using a high weight on Vrec. Usually there is little overlap in ages at length for fish that are one and/or two years of age in a survey of abundance. The first mode in a survey can generally index age-1 recruitment using length slicing. In addition numbers and the length frequency of the larger fish (adult fish) in a survey where overlap in ages at a particular length occurs can be used for tuning population abundance. The model tunes to the catch and survey length frequency data using a multinomial distribution. The user specifies the minimum size (cm) for the model to fit. Different minimum sizes can be fit for the catch and survey data length frequencies.

The number of parameters estimated is equal to the number of years in estimating F and recruitment plus one for the F to produce the initial population (Fstart), logistic selectivity parameters for each year or blocks of years, and for each survey Q. The total likelihood function to be minimized is made up of likelihood components comprised of fits to the catch, catch length frequencies, the recruitment variation penalty, each recruitment index, each adult index, and adult survey length frequencies:

$$\mathrm{L}_{\mathrm{catch}} = \sum_{years}\left(l\,\mathrm{n}(Y_{\mathrm{obs,y}}+1) - \ln\left(\sum_{a}\sum_{len} \mathrm{Y}_{\mathrm{pred,len,a,y}} + 1\right)\right)^2$$

$$L_{catch\_lf} = -N_{eff} \sum_{y}\left(\sum_{inlen}^{L_\infty}\left(\left(C_{y,len}+1\right)\ln\left(1+\sum_{a} C_{pred,y,a,len}\right) - \ln\left(C_{y,len}+1\right)\right)\right)$$

$$\mathrm{L}_{\mathrm{vrec}} = \sum_{y=2}^{Nyears}\left(Vrec_y\right)^2 = \sum_{y=2}^{Nyears}\left(R_1 - R_y\right)^2$$

$$\sum L_{rec} = \sum_{i=1}^{Nrec} \left[\sum_{y}^{Nyears} \left(\ln\left(I_{rec_i, inage_i, y}\right) - \ln\left(\sum_{len}^{L_\infty} N_{y, inage_i, len} * q_{rec\,i} \right) \right)^2 \right]$$

$$\sum L_{adult} = \sum_{i=1}^{Nadult} \left[\sum_{y}^{Nyears} \left(\ln\left(I_{adult_i, inlen+_i, y}\right) - \left(\sum_{a} \sum_{inlen_i}^{L_\infty} \ln\left(N_{pred, y, a, len} * q_{adult\,i}\right) \right) \right)^2 \right]$$

$$\sum L_{lf} = \sum_{i=1}^{Nlf} \left[-N_{eff} \sum_{y} \left(\sum_{inlen_i}^{L_\infty} \left(\left(I_{lf_i, y, len} + 1\right) \ln\left(1 + \sum_{a} N_{pred, y, a, len}\right) - \ln\left(I_{lf_i, y, len} + 1\right) \right) \right) \right]$$

In equation L_{catch_lf} calculations of the sum of length are made from the user input specified catch length to the maximum length for fitting the catch. Input user specified fits are indicated with the prefix "in" in the equations. LF indicates fits to length frequencies. In equation L_{rec} the input specified recruitment age and in L_{adult} and L_{lf} the input survey specified lengths up to the maximum length are used in the calculation.

$$Obj\ fcn = \sum_{i=1}^{N} \lambda_i L_i$$

Lambdas represent the weights to be set by the user for each likelihood component in the total objective function.

Tilefish SCALE Model Configuration and results

Two growth studies are available for Golden tilefish (Figure A36 and A37). Turner's aging study was done during the development of the longline fishery (1978-1982). Vidal updated growth from fish collected recently after three decades of fishing in 2008 (Appendix A2). Inferences on the assumed natural mortality were made using Turner's aging work since landings were relativity low before this period. Tilefish have sexual dimorphic growth with the males growing larger than the females. There is some indication from the study fleet length distributions by sex that a greater proportion of the larger fish are males (Figures A38 and A39). Natural mortality may be higher on male than females judging from the number of older fish seen by sex in Turner's sample (Table A16 and A17). In general Turner saw fewer older males than females during his study. Vidal's study was done after a long period of fishing in which the directed longline fishery was active. Large fish were present in Vidal's sampling but very few older fish (>20) were aged. The lack of older fish in Vidal study made the estimation of L infinity more difficult. The sensitivity of the SCALE model results to the assumed growth model (Turner's and Vidal's) was examined (Table A18). The modeling of growth as a combined sex model or with sex specific growth was also investigated. A natural mortality rate of 0.15 on males and 0.1 on females was assumed in runs when sex specific growth was used. In the combined sex model a natural mortality rate of 0.1 was used. The assumed variation around the mean lengths at age can be seen in Table A15 and Figure A40. The sensitivity of the assumed variation (run 5) around the mean lengths at age was also examined with a run were the variation in the mean lengths at age was increased (Table A18). The length weight relationship was updated using the data collected from the study fleet and growth study (Figure A41). The update relationship was used in the SCALE model. However the update relationship did not differ greatly from Turner's estimate.

A model which used Vidal's growth by sex and estimated selectivity in two time blocks (1971-1981, 1982-2008) was used as the base run (Table A18 and Figure A42 through A46). The SCALE model was dimensioned from ages 1-35, lengths 1-120 cm from years 1971-2008 as either a combined sex or separate sex model. A recruitment index does not exist for tilefish so a straight line index (constant recruitment index) was used as a proxy for an index with the model allowed to loosely fit the recruitment index (Figure A42). A low penalty weight (0.05) on recruitment variation was use in fitting the recruitment. The SCALE model appears to be able to pick up a recruitment signal from the commercial expanded length frequency distributions. The same general recruitment trend is estimated by the model even when yearly selectivity blocks were used. However this model run was not used since large changes in selectivity on a yearly basis seem unrealistic. A proxy for a recruitment index was developed as a sensitivity run (Table A18; run 6). This was done by through the redistribution of the VTR CPUE index according to the proportion of the expanded landing length frequency distribution and then slicing out the 40-50 cm fish as an age 5 index of recruitment (Figure A47). The CPUE indices were fit to fish sizes that were approximate according to the landing length frequency distributions. Turner's CPUE series was fit to 47+ cm fish and the Weighout and VTR series were fit to 37+ cm fish.

The catch length frequency distributions are an important component of the SCALE model. Turner collected landing length frequency information in 1974 and from 1976 to 1982. Note that Turner's length frequency data is only available in 5 cm blocks. NEFSC expanded landing size information exist from 1995 to 1999 and from 2002 to 2008. There appears to be a shift to smaller fish sizes between 1981 and 1982 in Turner's size distributions. Two selectivity blocks were assumed in the SCALE model (1971-1981, 1982-2008). The sensitivity of assuming a single selectivity block (run 3) over the time series was also tested. However in some years this run has trouble fitting the left side of the catch length frequency distribution due to the apparent change in selectivity over the time series.

The SCALE model time series starts in 1971 at the beginning of the tilefish directed longline fishery. However the SCALE model estimates an Fstart close to 0.2. This estimated equilibrium F that is assumed to occur before the beginning the time series appears to be on the high end since there was only a small limited fishery before 1971. A strong retrospective pattern did not exist in the base run (Figure A48). Little differences in the results are seen among the different model configurations (Table A18). There is a general concern with the lack of data and with the data independence used in the SCALE model. The lack of tuning information may result in little difference between the sensitivity runs. The lack of data, in particular the lack of recruitment index, could be preventing the mcmc from producing realistic results so uncertainty estimates around a particular model run could not be estimated. The estimated selectivity curve is also a source of concern given the tilefish longline fleet has some ability to target certain fish sizes by fishing different areas and depths. The SCALE model estimates of F during the late 1990s appear to be unrealistically high (over ten times F_{MSY}), while estimates of biomass in that period were correspondingly unrealistically low.

Term of Reference 3: Biological Reference Points

TOR3: Update or redefine biological reference points (BRPs; estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, and F_{MSY}). Comment on the scientific adequacy of existing and redefined BRPs.

ASPIC Surplus Production Model

Biological reference points estimated by the 2009 BASE run are moderately different from the 2005 SAW 41 assessment (Table A19). B_{MSY} is estimated to be 11,400 mt (a 22% increase), F_{MSY} is estimated to be 0.16 (a 24% decrease), and MSY is estimated to be 1,868 mt (a 6% decrease), compared to B_{MSY} = 9,384 mt, F_{MSY} = 0.21, and MSY = 1,988 mt from the 2005 SAW 41 assessment. The bootstrap (1000 iterations) median estimate (50%ile) of B_{MSY} was 10,135 mt; quartiles were 8,974 mt (25%ile) and 11,436 mt (75%ile). The bootstrap mean estimate of B_{MSY} was 10,336 mt, with a standard deviation (sd) of 2,089 mt and coefficient of variation (cv; sd/mean) of 20%. The bootstrap median (50%ile) estimate of F_{MSY} was 0.19; quartiles were 0.16 (25%ile) and 0.23 (75%ile). The bootstrap mean estimate of F_{MSY} was 0.20, with a standard deviation (sd) of 0.06 and coefficient of variation (cv; sd/mean) of 30%. The bootstrap results indicated that deterministic point estimates of the reference points are likely to be more precise than those accepted for the 2005 SAW 41 assessment, and are negatively biased by about 9% for B_{MSY} and positively biased by about 21% for F_{MSY} (Table A19).

SCALE model

Non-parametric yield per recruit (F_{MAX}) and spawners per recruit (F_{40}) biological reference points (BRP) were developed for SCALE base run 1 (separate sex model, two selectivity blocks) and run 2 (combined sex model, two selectivity blocks) (Table A20). BRPs were estimated both within the SCALE model and by converting the YPR inputs (selectivity, maturity schedule, stock and catch weights) to age based equivalents for use in an age based yield per recruit model (Table A21). The update maturity schedule from Vidal was used in the SPR analysis (Figure A49). MSY and SSB_{MSY} BRPS were estimated from the product of the model estimated initial recruitment (long term average recruitment) and the YPR or SSB per recruit estimates. The conversion to an age based YPR recruit model and an age based projection using AGEPRO is only possible in SCALE runs which modeled growth with the sexes combined (Figure A50). Similar BRPs are seen between the two methods (age based and SCALE). Uncertainty in recruitment can be incorporated into the AGEPRO projection by resampling from the CDF of the recruitment estimates. Reference points can also be estimated from long term projections with the CDF of recruitment and a F_{MSY} proxy. An example for run 2 using the CDF for the entire time series of recruitment and F_{MAX} produced a higher estimate of SSB_{MSY} at 14,000 mt relative to the simple product calculation of around 10,000 mt in Table A20 (Figure A51). The SSB_{MSY} estimate for the separate sex run is based on female fish (run 1). Note that a female estimate of SSB_{MSY} is not possible using the age based YPR model. In addition the age based projections in AGRPRO can not account for the sex specific effects that exist in the separate sex model. However for the separate sex model a simple deterministic projection can be done within the SCALE model.

The estimates of F_{MAX} and F_{40} were similar to the estimates from SARC 41 (F_{MAX} = 0.138 and F_{40}=0.08). F_{MAX} is estimated from a well defined yield curve (Figure A52). The predicted terminal year age and length distributions were slightly truncated in comparison to the equilibrium distribution at F_{MAX} for both run1 and run 2 (Figure A53). Run 2 has a greater proportion of larger fish in the F_{MAX} equilibrium distribution relative to run 1 because run 1 assumes a higher natural mortality rate on males (Figure A52). SCALE YPR BRPs suggest that SSB_{MSY} is between 9,878 mt and 15,108 mt for the combine sex run using F_{40} or F_{MAX} as the F_{MSY} proxy (Table A20). The separate sex run suggests female SSB_{MSY} is between 5,335 mt and 7,100 mt. For both the single sex and separate sex run the F_{MSY} is between 0.079 and 0.128 and MSY ranging from 1,072 mt to 1,200 mt using either F_{40} or F_{MAX} as the F_{MSY} proxy.

Term of Reference 4: Stock Status

TOR4: Evaluate stock status with respect to the existing BRPs, as well as with respect to updated or redefined BRPs (from TOR 3).

ASPIC Surplus Production Model

The 2009 BASE model run results indicate that the Golden tilefish stock is not overfished and that overfishing is not occurring. With respect to the reference points from the 2005 SAW 41 assessment, fishing mortality in 2008 was estimated to be 0.06, 29% of F_{MSY} = 0.21, and total biomass in 2008 was estimated to be 11,910 mt, 127% of B_{MSY} = 9,384 mt. For this TOR note that for the ASPIC surplus production model it may not be appropriate to compare stock status relative to biological reference points from a different model run.

With respect to the updated reference points from the 2009 BASE run, fishing mortality in 2008 was estimated to be 0.06, 38% of F_{MSY} = 0.16. Total biomass in 2008 was estimated to be 11,910 mt, 104% of B_{MSY} = 11,400 mt (Table A13, Figure A54 and A55). The 50% confidence interval (range between the 25%ile and 75%ile) for the 2008 F/F_{MSY} ratio was between 0.25 and 0.42 and for the 2008 B/B_{MSY} ratio was between 0.87 and 1.46. The SARC 48 review panel accepted the ASPIC model but concluded that the ASPIC model is likely over optimistic and that the stock has not rebuilt above B_{MSY}.

SCALE Model

With respect to the existing reference points from the 2005 SAW 41 assessment, SCALE base run 1 fishing mortality in 2008 was estimated to be 0.188, 90% of F_{MSY} = 0.21, and total biomass in 2008 was estimated to be 4,950 mt, 53% of B_{MSY} = 9,384 mt. For this TOR note that this is a comparison of terminal year F (fully selected) and biomass from an age/size structured model relative to biological reference points from the SARC 41 surplus production model. This comparison results in a different status determination (no overfishing and not overfished) than if the update biological reference points were used.

With respect to the updated reference points from the SCALE BASE run (separate sex run), fishing mortality in 2008 was estimated to be 0.188, 147% of F_{MSY} = 0.128 using F_{MAX} as the proxy for F_{MSY}. Total female SSB in 2009 was estimated to be 2,520 mt, 47% of SSB_{MSY} = 5,335 mt using F_{MAX} as the proxy for F_{MSY}. With respect to the updated reference points from the SCALE (run2) combined sex run, fishing mortality in 2008 was estimated to be 0.205, 169% of F_{MSY} = 0.121 using F_{MAX} as the proxy for F_{MSY}. Total SSB in 2009 was estimated to be 4,399 mt, 41% of SSB_{MSY} = 10,794 mt using F_{MAX} as the proxy for F_{MSY}.

The 2009 BASE SCALE model run (separate sex run) and the combined sex run results indicate that the 2009 Golden tilefish stock is at a low biomass (29% to 47% of SSB_{MSY}) and is overfished with respect to the update SSB reference points. Both SCALE runs also suggest recruitment and growth overfishing (147% to 260% of F_{MSY}) is occurring with respect to the F_{40} or

F_{MAX} updated biological reference points. However fishing mortality has been decreasing and biomass has been increasing since the beginning of the FMP in 2001. Comparison of F to F_{MSY} and Biomass to B_{MSY} ratios over time between the ASPIC and SCALE model can be seen in figures A56 and A57.

Term of Reference 5: Projections

TOR 5: Develop and apply analytical approaches and data that can be used for conducting single and multi-year stock projections and for computing candidate ABCs (Acceptable Biological Catch).

- **a. Provide numerical short-term projections (2-3 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. In carrying out projections, consider a range of assumptions about the most important uncertainties in the assessment (alternate states of nature).**
- **b. If possible, comment on the relative probability of the alternate states of nature and on which projections seem most realistic.**
- **c. For a range of candidate ABCs, compute the probabilities of rebuilding the stock by November 1, 2011.**
- **d. Describe this stock's vulnerability to becoming overfished, and how this could affect the choice of ABC.**

ASPIC Surplus Production Model

Standard ASPIC model projections can either project fishery yield (i.e., total catch) for a given trajectory of F or project F for a given trajectory of yield. In neither case are any assumptions made about the future trajectory of the calibration indices - for tilefish, the commercial fishery VTR CPUE index series. For this assessment, two types of projections have been made. The first type is the standard ASPIC projection just described. The second type of projection makes assumptions about the future trajectory and magnitude of the VTR CPUE series in addition to projected F, catch, and biomass, and is intended to further respond to TOR5. The projections with the CPUE assumptions, however, result in changes in the overall model fit, re-scaling of the historical development of the stock, and different reference points. These results are therefore not directly comparable to the 2009 BASE run results, but should be useful in demonstrating how stock status might change in the future given some possible trends in fishery CPUE.

The standard projections were made for 2009-2011 assuming A) constant *status quo* catch = 905 mt, B) constant MSY catch = 1,868 mt, and C) constant F_{MSY} = 0.16. The *status quo* catch = 905 mt (1.995 million lb) has been the TAC since the FMP was implemented in 2001. Status determination was evaluated with respect to the updated reference points from the 2009 BASE run (threshold F_{MSY} = 0.16, target B_{MSY} = 11,400 mt, threshold B_{MSY} = 5,700 mt). Projection results for these three scenarios indicate 15%, 39%, and 45% chances that the stock will decline below the biomass target of B_{MSY} by 2011, and <1% chance that the stock will decline below the biomass threshold of ½ B_{MSY} by 2011. The projections indicate 0%, 40%, and 50% chances that F will exceed the fishing mortality threshold of F_{MSY} by 2011 (Table A22, Figures A58 and A59).

For the projections incorporating the CPUE index, runs were made with constant *status quo* catch = 905 mt, and 2009-2011 index assumptions of A) constant at the 1995-2008 average VTR CPUE = 2.095 (mt/da), B) constant at the 2001-2008 average VTR CPUE = 2.6475 C) increasing an average rate of +25% per year, D) decreasing at an average rate of 25% per year, constant at the 2008 value of 1.434 (mt/da), and F) constant at the 2008 value rounded up to 1.4 (mt/da). Options C and D were specified to loosely mimic the ~25% average annual rate of increase in VTR CPUE during 2000-2005 that was followed by a ~33% decrease during 2005-2008. Status determination was evaluated with respect to the different reference points calculated in each run. For runs A, B and E (different mean levels of CPUE), the estimates of F_{MSY} increase and B_{MSY} and MSY decrease, relative to the 2009 BASE run estimates. These scenarios indicate about a 10% or less chance that biomass will decline below the target biomass B_{MSY} by 2011, and <1% chance that biomass will decline below the biomass threshold ½ B_{MSY} by 2011. For scenario C (increasing CPUE), F_{MSY}, B_{MSY}, and MSY all decrease, but like scenarios A and B, the projection indicates about a 10% chance that biomass will decline below B_{MSY} by 2011, and <1% chance that biomass will decline below ½ B_{MSY} by 2011 (Table A23, Figures A60 and A61). CPUE projection scenario E is *status quo* for both the fishery TAC and CPUE index, and so is considered the most likely in the short-term. Scenario E provides estimates of fishing mortality, stock biomass, and reference points in line with those from scenarios A, B and C. Scenario F is similar to the status quo CPUE of scenario E with the exception that the CPUE was rounded up to1 decimal place (CPUE was 1.4 instead of 1.434). This minor difference resulted in a large change in the results of the ASPIC model (Figure A62).

Projection scenario D (decreasing CPUE) re-scales the stock size and changes the reference points by a larger amount than the other four CPUE projection scenarios, and is particularly relevant to TOR5d. F_{MSY} decreases by about 60%, while B_{MSY} increases by 32% and MSY decreases by about 50%. These changes indicate a stock with lower resilience and productivity when compared to the other scenarios, in that the recent *status quo* TAC = 905 mt is above the estimated MSY. For scenario D, the time series estimates of F and B indicate that the stock has been below B_{MSY} since the late 1980s and F has consistently been above F_{MSY} since about 2000. The scenario D projection indicates a greater than 75% chance that fishing mortality will be above F_{MSY} and biomass will be below the target B_{MSY} by 2011, and a greater then 50% chance that biomass will be below the threshold ½ B_{MSY} by 2011 (Table A23, Figures A58 and A59). This projection scenario illustrates that the stock is vulnerable to being classified as "overfished" (below the threshold ½ B_{MSY}) if the VTR CPUE continues to decrease during 2009-2011 even as the catch remains near the recent *status quo*.

SCALE Model

As noted under TOR 3 age based projections can not be done in AGEPRO for SCALE separate sex model runs (base run 1). However, a deterministic projection can be done within the SCALE model by fixing the parameters in the model at the model solution and projecting into future years. Figure A63 and Figure A64 are examples of deterministic projections from run 1 at F_{MSY} = F_{MAX} = 0.13 and F_{2008} = 0.19, respectively. Combined sex model runs can be converted to an age based equivalent and projected using the AGEPRO projection program. Some uncertainty in recruitment can be accounted for in AGEPRO through resampling of the CDF of recruitment estimated from the SCALE model. Constant catch projections for run 2 (combined sex run) using agepro are shown in Figure A65. Note that using constant catches over 500 mt allows overfishing (F_{MSY} = F_{MAX}) in the first year of the projection.

Conclusions

The possibility of unknown refuge effects due to conflicts with lobster and trawl gear, effects of targeting incoming year classes, and the unknown effects on tilefish CPUE due to competition/interference from increased dogfish abundance introduce uncertainty in interpreting CPUE from this fishery as a measure of stock abundance. CPUE index of abundance and catch length frequency distributions are likely a reflection of both the population abundance and the unaccounted changes in fishing practice.

The Working Group accepted the ASPIC model solution but noted that there is very high uncertainty regarding whether the stock is rebuilt. The SARC 48 review panel concluded that the ASPIC model is likely over optimistic and that the stock has not rebuilt above B_{MSY}. The surplus production model inability to fit the decline in CPUE due to at year class effect at the end of the time series is a source of concern. The bootstrap uncertainty estimates from the ASPIC model likely do not capture the true uncertainty in this assessment. Results from the SCALE model which incorporates the species lifespan, growth, and recruitment dynamics evident in the commercial length distributions provide reason to be concerned that the stock is not rebuilt. However the overall lack of data within the scale model and questions on the estimated selectivity may result in a pessimistic stock status determination. The uncertainty in this assessment is encompassed by the results from two very different models which resulted in different status determinations. However increases in biomass and lower fishing mortality rates since the beginning of the FMP are evident in the results from both models. Consideration should be given to the possibility that the SCALE model results may be a reflection of the true state of nature when setting ABCs rather then using the results of the ASPIC surplus production model which states that the stock is rebuilt.

Term of Reference 6: Research Recommendations

TOR 6: Review, evaluate and report on the status of the research recommendations offered in recent SARC reviewed assessments. Identify new research recommendations, including recruitment estimation.

New research recommendations from 2009 SARC 48

1) Continue the development of an improved haul based fishery dependent cpue index (i.e., continue the current study fleet project) or design a tilefish longline survey as a semi fishery independent index of abundance that could be conducted by an existing longline vessel and the study fleet platform. If a tilefish longline survey is developed then size information should be incorporated into the survey design for the estimation of a recruitment and size specific index of abundance which could improve the tilefish assessment.

2). For the study fleet project and any potential semi fishery independent survey, include additional information on conflicts with lobster and trawl gear, the possibility of unknown effects on tilefish CPUE due to competition/interference from an increased abundance of dogfish, the unknown effects of bait type on tilefish CPUE (e.g., substitutes for the preferred squid).

3). Develop protocols to ensure consistency between dealer, VTR, and IVR reports of the tilefish landings.

4). Develop protocols to ensure consistency in market category designation among fishing ports.

5). Explore the influence of water temperature and other environmental factors on trend in the commercial fishery CPUE index of stock abundance.

Research recommendations from the 2005 SARC 41 review

1) Conduct a hook selectivity study to determine partial recruitment changes with hook size.

Determine catch rates by hook size. Update data on growth, maturity, size structure, and sex ratios at length.
Hook selectivity study was not done. Funding was initially available, but subsequently rescinded. Updated growth, maturity, and size structure studies were completed.
2) Collect data on spatial distribution and population size structure. This can help answer the question of the existence of a possible dome shaped partial recruitment pattern where larger fish are less vulnerable to the fishery due to spatial segregation by size.
This research recommendation was examined in the study fleet data.
3) Continue to develop the forward projecting catch-length model as additional length data becomes available. Investigate the influence of adding a tuning index of abundance and model estimated partial recruitment (logistic) to the catch-length model.
This research recommendation was completed. The improved catch-length model was renamed as the SCALE model.
4) Collect appropriate effort metrics (number and size of hooks, length of main line, soak time, time of day, area fished) on a haul basis to estimate commercial CPUE.
This research recommendation was examined with the study fleet analysis.
5) Initiate a study to examine the effects of density dependence on life history parameters between the 1978-82 period and present.
This research recommendation was examined with the update growth and maturity study.
6) Increased observer coverage in the tilefish fishery to obtain additional length data.
Observer coverage has improved in the tilefish fishery.
7) Develop a bioeconomic model to calculate maximum economic yield per recruit.
This research recommendation has not been initiated.

Research recommendations from 1999 Science and Statistical Committee review
1) Ensure that market category distributions accurately reflect the landings. Sampling of the commercial lengths has improved over the last six years. Small, kitten, and medium market category distributions can shift from one year to the next due to the growth of a strong yearclass. Intensive length sampling of the landings by market categories is needed to account for possible shifts in the distribution within a market category over time. Similar landings distributions were seen among the observer, study fleet, and commercial port sampling data sources.
2) Ensure that length frequency sampling is proportional to landings by market category.
Commercial length sampling has been sporadic during the beginning of the time series. In particular length samples from the large market category have been lacking. However commercial length sampling has greatly improved over the last six years with a higher proportion of the sampling coming from Montauk where most of the fish are landed.
3) Increase and ensure adequate length sampling coverage of the fishery.
See comments for research recommendations 1 and 2.
4) Update age- and length- weight relationships.
This TOR has been addressed.
5) Update the maturity-at-age, weight-at-age, and partial recruitment patterns.
This TOR has been addressed.
6) Develop fork length to total length conversion factors for the estimation of total length to weight relationships.
This work was addressed in SARC 41.
7) Incorporate auxiliary data to estimate r independent of the ASPIC model.
This TOR has not been addressed. SARC 41 questioned if this can be done or should be done.

However SARC 48 SCALE results suggest that r is overestimated in the ASPIC model.

References

Freeman BL, Turner SC. 1977. Biological and fisheries data on tilefish, *Lopholatilus chamaeleonticeps* Goode and Bean. NEFSC Tech Rep. 5.

Granger CW, Newbold P. 1977. Forecasting Economic Time Series. New York: Academic Press.

Grimes CB, Able KW, Turner SC. 1980. A preliminary analysis of the tilefish, *Lopholatilus chamaeleonticeps*, fishery in the Mid-Atlantic Bight. Mar Fish Rev. 42(11): 13-18.

Grimes CB, Idelberger CF, Able KW, Turner SC. 1988. The reproductive biology of tilefish, *Lopholatilus chamaeleonticeps*, Goode and Bean, from the United States Mid-Atlantic Bight, and the effects of fishing on the breeding system. Fish Bull. 86(4): 745-762.

Mayo RK, Helser TE, O'Brien L, Sosebee KA, Figuerido BF, Hayes D. 1994. Estimation of standardized otter trawl effort, landings per unit effort, and landings at age for Gulf of Maine and Georges Bank cod. NEFSC Ref Doc. 94-12.

Mid-Atlantic Fishery Management Council. 2000. Tilefish fishery management plan. NOAA award No. NA57FC0002.

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

Nitschke P, Shepherd G, Terceiro M. 1998. Assessment of tilefish in the middle Atlantic – southern New England region. NEFSC. 1-12.

Prager MH. 1994. A suite of extensions to a nonequilibrium surplus-production model. Fish Bull. 92: 374-389

Prager MH. 1995. Users manual for ASPIC: a stock-production model incorporating covariates. SEFSC Miami Lab Doc. MIA-92/93-55.

Turner SC, Grimes CB, Able KW. 1983. Growth, mortality, and age/size structure of the fisheries for tilefish, *Lopholatilus chamaeleonticeps*, in the Middle Atlantic-Southern New England region. Fish Bull. 81(4): 751-763.

Turner SC. 1986. Population dynamics of and, impact of fishing on tilefish, *Lopholatilus chamaeleonticeps*, in the Middle Atlantic-Southern New England region during the 1970's and early 1980's. New Brunswick, N.J.: Rutgers University. Ph.D. dissertation.