Report on the 58th Northeast Regional Stock Assessment Review Committee (SARC 58)

Independent reviewer's report on the 2014 Butterfish, Tilefish, and Northern shrimp Benchmark Stock Assessments

27-31 January 2014

Northeast Fisheries Science Center, National Marine Fisheries Service, Woods Hole, MA

Ian Jonsen

31 Belvedere Dr

Dartmouth, Nova Scotia, Canada

B2X 2N3

Prepared for:

The Center for Independent Experts

February 2014

1 Executive Summary

During 27-31 January 2014, the 58th Stock Assessment Review Committee (SARC) meeting was held at the Northeast Fisheries Science Center (NEFSC) in Woods Hole, MA, to review the butterfish (*Peprilus triacanthus*), tilefish (*Lopholatilus chamaeleonticeps*), and northern shrimp (*Pandalus borealis*) assessments. I participated as one of three CIE reviewers.

Butterfish

The working group carried out a thorough assessment of butterfish and all the terms of reference were completed, except one, regarding consumptive removals of butterfish by its predators, which was mostly completed. There is currently no directed fishery for butterfish, with discards comprising the bulk of removals. A number of spring/fall, inshore/offshore (NEFSC and NEAMAP) and local, state-level surveys are used to monitor butterfish but due to conflicts between spring and fall offshore surveys, the SARC recommended that only the fall surveys be used in the assessment. Innovative research was conducted to determine the affect temperature has on butterfish availability to the surveys. Results from this analysis were combined with efficiency analysis to calculate a habitat-informed annual catchability parameter as an input to the assessment model. Unfortunately, the trend in the habitat index was surprisingly flat over the time series considered and therefore contributed relatively little to the assessment. The SARC panel requested modifications to the ASAP model presented to deal with large residual patterns in fits to the fall and spring surveys. The accepted configuration was fit to the fall surveys only and excluded the habitat index, which was deemed superfluous. The BRP's were accepted as presented and the probabilities of overfishing occurring and the stock being overfished were less than 1%.

Tilefish

The working group carried out a thorough assessment of tilefish and all the terms of reference were completed, except one, regarding the relationship between temperature, tilefish distribution and thermal tolerance, which was partly completed. Nearly all catch currently comes from the directed longline fishery and is well characterized, however there is inconsistency among dealers/fishers in market size designations that has potential to affect the size composition information feeding into the assessment. There is currently no reliable fishery-independent data for tilefish and commercial LPUE is the only index used in the assessment. In addition to adopting a longline survey in collaboration with the fishing industry, a more comprehensive standardization, including assessment of potential structuring environmental variables, of the LPUE data and better characterization of fishing effort should be explored. The ASAP model, one of three presented by the working group, was accepted, including the assumption of a strongly domed selectivity function that was supported by the catch at age data. The BRP's presented were reasonable, the stock

is below the overfishing reference and appears to be rebuilt. Because the model has no information to reliably estimate recruits in the terminal 3 years, the stock projections were empirically adjusted based on retrospective analysis results for ages 1-4 in these years. This approach seems valid given that it yielded projections more consistent with stock abundance trends.

Northern shrimp

The working group carried out a thorough assessment of northern shrimp, but an extremely large and rapid fluctuation in stock abundance/biomass from 2005 to 2007 created a number of technical issues that rendered the three models to be considered inappropriate for management purposes. As a consequence, several of the terms of reference were not completed. The SARC recommended that annual updates on stock status and management advice be based on survey indices, catches and CPUE. The working group should continue to refine the UME and CSA models and the next benchmark assessment should occur as soon as possible.

2 Background

This report is an independent review of benchmark assessments of butterfish, tilefish, and northern shrimp conducted at the Stock Assessment Workshop (SAW 58). Results of these assessments were presented at the Northeast Regional Stock Assessment Review Committee (SARC 58) meeting, held on 27-31 January 2014 at the Northeast Fisheries Science Center, Woods Hole, MA. Prior to the meeting, the Review Panel (RP; Appendix 3) was provided with a Statement of Work (SoW; Appendix 2), including the Terms of Reference (ToR) for each assessment and the ToR for the RP. Assessment working papers and background material (Appendix 1) were provided via a secure server approximately 2 weeks before the meeting. During the meeting there was general consensus among the RP on all of the main discussion points and findings of the RP as described in the Summary Report. This report presents a summary of those findings and focuses on my own views about the assessments.

3 Review Activities

The review consisted of three sequential tasks: (1) a review of the assessment and background documents; (2) a panel review meeting; (3) completion of an individual report. The CIE panel review meeting was held between 27-31 January 2014, at the NEFSC, NOAA Fisheries Service, Woods Hole, MA. The review took a format of

presentations by various contributors to the butterfish, tilefish, and northern shrimp assessments, followed by questions and discussion.

Prior to the meeting I reviewed the assessment and background documents provided. During the meeting all reviewers were equally responsible for each assessment. Stock assessment presentations for butterfish were made on 27 January, with follow-up evaluations of additional model runs and discussion on 29 January. Presentations for tilefish and northern shrimp were made on 28 January, with follow-up evaluations of additional model runs and discussion on 29 January. Additional follow-up evaluations and discussion of northern shrimp was conducted on 30 January. Assessment summary reports for each stock were edited in-session on 30 January. The SARC panel met on 31 January to draft the SARC summary report for each stock. Additional drafts of the reports were reviewed by the panel after the meeting.

I participated as one of three CIE reviewers. This report presents my review findings and recommendations, adhering to the review meeting Terms of Reference (ToRs - see Appendix 2, annex 2).

4 Findings

4.1 Butterfish

4.1.1 Characterize the commercial catch including landings, effort and discards by gear type. Describe the magnitude of uncertainty in these sources of data.

The ToR was completed. There has been no directed fishery for butterfish between 2002 and 2012, but in 2013 a fishery was re-established. Discards comprised the majority of butterfish catch at 58% from 1989-2001 and 67% from 2002-2012. Recreational catch appears to be negligible. The catch time series was started in 1989 to mitigate the lack of CVs on foreign fishing fleet and lack of observed small mesh trips prior to 1989. The decision to start the time series in 1989 seems well-justified and was additionally explored in follow-up analyses requested by the RP (see ToR 5). Fishing effort information was, appropriately, not used in the assessment due to the large proportion of bycatch. Due to the high bycatch levels, observer coverage is very important to adequately characterize catch in this fishery and coverage should be maintained or even increased.

4.1.2 Characterize the survey data that are being used in the assessment. Describe the magnitude of uncertainty in these sources of data.

The ToR was completed. Spring and fall inshore (1989-2008) and offshore (1989-2012) NEFSC surveys, along with the inshore Northeast Area Monitoring and Assessment Program (NEAMAP) spring (2008-2012) and fall (2007-2012) surveys were used in the assessment. Several additional state surveys were also available but these were not used primarily due to their individually small spatial coverage (yet survey coverage maps were not presented to demonstrate this clearly) and therefore are unlikely to be representative of the entire stock. Presentation of correlations between these local surveys and the broader NEFSC and NEAMAP surveys suggested that the state surveys generally track the broader stock dynamics. In general, the fall offshore NEFSC survey is believed to provide the most reliable index of abundance, but based on correlation results presented in the Wednesday follow-up session, it appears less able to track cohorts than the NEFSC spring offshore survey. Unfortunately, the index correlations also provided reasonable evidence of conflict between the spring and fall offshore surveys, as was apparent from Fig. A2.7 (Butterfish Working Paper 1). This conflict was not addressed in the working paper, but was a source of concern among the RP especially in light of the model fits to the main survey indices presented under ToR 5 (4.1.5).

It was pointed out by the SARC chair, and I concur, that efforts should be taken to assimilate these local surveys with the broader surveys using model-based approaches such as spatio-temporal GAMs or hierarchical models. The obvious benefit to such an approach would be the single estimated index that could feed into the assessment, rather than a more complex amalgam of somewhat conflicting indices that was presented at the meeting. Clearly, taking the approach of combining survey indices would require substantial effort but at a minimum it should provide more insight into the reason(s) for the conflict between the offshore surveys. At best, the conflict may be resolvable and the assessment would likely benefit from having more of the disparate survey data informing the model than is currently the case (see 4.1.5).

4.1.3 Characterize oceanographic and habitat data as it pertains to butterfish distribution and availability. If possible, integrate the results into the stock assessment (TOR-5).

The ToR was completed. The working group presented a very innovative and rigorous approach to understanding butterfish habitat preferences, using temperature estimates from an oceanographic model, and how these preferences affected availability of the stock to the surveys. Estimates of availability from this analysis and estimates of efficiency of the surveys were used to parameterize catchability in the assessment model. The result, however, was somewhat disappointing as it did not reveal any trend in availability over time, despite evidence of climate change affecting other stock dynamics (eg., Richards et al. 2012).

It was surprising that the initial GAM to estimate the relationship between temperature and the survey catches only accounted for 30% of the deviance in the catches. This seems quite low if temperature is indeed an important environmental driver of habitat preference and hence availability to the survey. The approach of removing the larger catches from this analysis seems somewhat suspect. It would be useful to explore fitting the parametric model to the full catch dataset, rather than removing the putative outliers as these data may have important information to contribute in estimation of the relationship.

The complexity and volume of the work presented at the meeting and in the working paper (which could have been presented in a clearer, more succinct fashion, especially given the strong oceanographic component) made it difficult to fully comment on the suitability of the approach and analyses. A concern is that this analysis ultimately contributed to a fixed catchability parameter in the final assessment model configuration, which is a very strong assumption to make (although it was mitigated by estimating M in the model). I urge the assessment working group to revisit the habitat analyses to ensure that the true temperature - catch relationship is being estimated as well as possible under their methodology. The group should consider applying the analyses to catches from the other butterfish surveys to determine if there is generally conformity in the results. Once these issues are resolved, the group should consider applying this approach to other stocks and/or exploring other potential environmental drivers, such as productivity.

4.1.4 Evaluate consumptive removals of butterfish by its predators. If possible, integrate results into the stock assessment (TOR-5).

The ToR was mostly completed. Results presented implied that butterfish consumption by 6 main predators (smooth dogfish, spiny dogfish, silver hake, summer flounder, bluefish and goosefish) were relatively constant and similar to one another over the past 3.5 decades. This could be consistent with the assessment assumption of constant M.

I have no problem with the analyses presented, but the description in the working paper was somewhat confusing, with equations - although simple - poorly explained. The RP asked for plots of predator abundance trends for some context and these really should have been included in the working paper. Given the long time span of the stomach contents data collected by the NEFSC and their strong ecosystem modelling expertise, it was disappointing that more had not been accomplished under this ToR, especially given the evidence of major ecosystem changes in the broader northwest Atlantic. This is an area where expertise outside the assessment group should be sought to explore potential ecosystem drivers of butterfish abundance further.

4.1.5 Use assessment models to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a comparison with previous assessment results and previous projections.

The ToR was completed. The working group used the ASAP model within the NOAA Fisheries Toolbox for the butterfish assessment. This model has been peer-reviewed and is widely viewed as a useful, generally applicable age-structured assessment model. The base configuration of the model presented used catchability, q, as an input rather than estimated in the model; where the NEFSC fall offshore survey annual q was calculated as the product of availability (estimated from the habitat analyses in ToR 3, 4.1.3) and efficiency (estimated from fishing comparisons between the *FSV Albatross IV* and the *FSV Henry B. Bigelow*, conservatively assuming the surveys using the latter vessel were 100% efficient with respect to butterfish). Natural mortality was estimated in the model. The model was fit to the catch time-series and all 6 of the major NEFSC and NEAMAP survey indices and age composition data. The approach of fixing q and estimating M is unconventional, but given the comprehensive work undertaken to estimate habitat-based availability to the fall offshore survey and relative survey efficiency, with the important conservative assumption of 100% efficiency of the *Bigelow* surveys, the approach seems reasonable.

The major issue with the base configuration was the strong and contrasting residual patterns in the fits to the NEFSC spring and fall offshore surveys. The contrasting residual patterns between spring and fall offshore surveys highlighted their conflicting trends. Fits to the NEAMAP surveys were generally difficult to evaluate as these time-series are relatively short (5 and 6 years). Fits to the age composition data were generally good, although there were runs of negative or positive residuals in the ages 1 and 2 for the spring offshore survey data.

To explore the conflicting signals from the spring and fall offshore surveys, the working group was asked to run additional configurations of the ASAP model: (1) fitting only to the spring indices, which appeared to track cohorts best; (2) fitting only to the fall indices, which appeared to reflect abundance best; and (3) fitting only to the fall indices without the habitat index, which appeared to be superfluous. There was also a question about whether the results were sensitive to the choice of starting the time series in 1989. To address this, the working group was asked to construct ``prospective'' plots by refitting the model to time series starting in 1990-1993.

Configuration (1) could not be accepted because M was fixed using the estimate from the rejected base configuration. This approach was taken because the spring survey q's had to be estimated in the model; availability and efficiency were not estimated outside of the model for these surveys. Configurations (2) and (3) yielded acceptable and nearly identical fits, with less severe residual patterns in fits to the indices and age compositions. Given the lack of sensitivity to exclusion of the

habitat-index, the simpler model should be preferred as the new base model for advice.

The requested prospective plots suggested little sensitivity to the choice of the 1989 start to the time-series. A fuller exploration would, of course, include fits to time series starting prior to 1989 but far more work would be involved in doing that and I suspect the results would simply confirm the current conclusion.

4.1.6 State the existing stock status definitions for "overfished" and "overfishing". Given that the stock status is currently unknown, update or redefine biological reference points (BRPs; point estimates for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and B_{MSY}, or their proxies) and provide estimates of their uncertainty. Consider effects of environmental factors on stability of reference points and implications for stock status.

The ToR was completed. The working group proposed F_{MSY} and B_{MSY} proxies as new BRP's. The F_{MSY} proxy was 2/3M and the B_{MSY} proxy was the median of the 50-yr SSB projection with removals at the F_{MSY} proxy. With these BRPs, overfishing occurs when F > 2/3 M and the stock is overfished when SSB < ½ B_{MSY} proxy. These proxies seem reasonable given that the stock appears to be healthy and the current lower abundance relative to that implied by the large catches prior to 1989 suggests the stock may be driven by factors, possibly environmental, other than F.

4.1.7 Evaluate stock status with respect to a newly proposed model and with respect to "new" BRPs and their estimates (from TOR-6). Evaluate whether the stock is rebuilt.

The ToR was completed. Fishing mortality (F_{2012} = 0.02) appears low and well below the overfishing reference point (2/3 M = 0.81). The current stock biomass is rebuilt (SSB_{2012} = 79,451 mt), nearly double the overfished reference point (1/2 B_{MSY} proxy = 44,730 mt). Given these estimates, the working group's conclusions that overfishing is not occurring and the stock is not overfished are warranted.

4.1.8 Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs). a. Provide numerical annual projections (2 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g.,

terminal year abundance, variability in recruitment). Comment on which projections seem most realistic. b. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

The ToR was completed. The projections from the new base model fit to the fall-only surveys and without the habitat index were generated using a standard and acceptable methodology. Recruitment was sampled from the full time-series of model-estimated recruitment and removals were set at the known 2013 butterfish catch, the 2014 butterfish ABC and the F_{MSY} proxy of 2/3 M thereafter. These projections implied extremely low probabilities (< 1%) of overfishing and of being overfished in 2017.

4.1.9 Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

The ToR was completed. The working group made substantial progress on the research recommendations from the previous SARC (SARC 49), including the innovative work on exploring temperature-driven habitat preference of butterfish. The new research proposed by the working group is sensible, including the suggestion to not conduct additional assessments until a fishery develops that can influence stock biomass. The working group should, however, continue to monitor annually the stock status via survey indices and the catch. I encourage the group to explore potential reasons for a lack of trend in the habitat-availability index, including updating their approach to fitting the parametric model to the full survey catch dataset, rather than removing the larger catches. Use of a negative binomial or a gamma error distribution may prove effective here. Once these details are resolved, the group should apply their approach to the spring offshore survey and perhaps also consider other potential structuring variables such as productivity. More work could be accomplished on predator removals of butterfish and general ecosystem interactions that have the potential to influence butterfish stock dynamics. The NEFSC is blessed with impressive stomach contents data and have a team of talented ecosystem researchers that should be drawn in to further explore how ecosystem considerations can be incorporated into the butterfish assessment. Finally, considerable work could be done on using a statistical approach to combine survey indices and thereby assimilate the state survey data into the assessment.

4.2 Tilefish

4.2.1 Estimate catch from all sources including landings and discards. Describe the spatial and temporal distribution of landings, discards, and fishing effort.

Characterize the magnitude of uncertainty in these sources of data.

The ToR was completed. The working group presented estimates of total catch from 1915-2012, but concluded that the assessment should be based on the period 1971-onward as that coincides with a directed longline fishery for tilefish. Currently, nearly all landings appear to come from the longline fishery. There appears to be relatively little contribution from discards or the recreational fishery, but it is unclear if recreational landings reports are reliable. The commercial fishery was initially comprised of vessels from Barnegat, New Jersey but effort shifted roughly in the mid- to late-1980's to vessels originating from Montauk, New York. The working group did a comprehensive job in assembling a useful catch series from disparate data sources.

Perhaps the biggest concern regarding the catch data, aside from the lack of CV estimates, was the inconsistent and potentially confusing market size categories used by different dealers. This source of uncertainty is potentially affecting the size composition information used by the assessment. However, since the bulk of the fishery is conducted by a small number of fishers and dealers there appears to be opportunity to work with them to standardize these categories and improve the quality of this key data input to the assessment. Similarly, aging of commercial landings was initiated only recently; age data available from 2007-2012. The age data are essential for the ASAP model to be used for the assessment and so further aging of the catch samples must continue.

4.2.2 Characterize commercial LPUE as a measure of relative abundance. Consider the utility of recreational data for this purpose. Characterize the uncertainty and any bias in these sources of data.

The ToR was completed. Effort for the LPUE index is calculated as the number of days absent from port (minus 1 d for steaming) per trip. Fishing gear is consistent across the 4 main fishing vessels that comprise about 80% of the landings. Three LPUE series are available for the period 1973-82 (Turner series); 1979-93 (Weighout series); 1995-2013 (VTR series). These data provide the only information on abundance trends and cohort dynamics, which is in itself a concern but the working group has no other viable fisheries-independent information at the present time. The commercial data are able to track large cohorts. As with the catch data, annual estimates of uncertainty in the unstandardized LPUE indices were not reported.

The working group conducted a simple GLM standardization of the LPUE data with year and vessel main effects. Additional covariates such as month/season, at a minimum, should be explored in the standardization. Additionally, basing effort estimates on trip length is particularly crude and more detailed information such as number of hooks per set, line length, and soak time should be sought out from the fishers. Again mentioned above, the bulk of the fishery is now comprised of 4 vessels

so seeking out this information or adopting practices to formally record the information going forward seem achievable. Finally, the apparent spatial limitation of the fishery and targeting of particular size classes suggest some bias, currently unquantifiable, in the LPUE index. Adopting a longline survey in collaboration with the fishery would help overcome this potential bias in future years.

4.2.3 For the depth zone occupied by tilefish, examine the relationship between bottom temperature, tilefish distribution and thermal tolerance.

The ToR was partly completed. Most tilefish caught in the NEFSC surveys were primarily in a narrow temperature range (11-13° C), implying that temperature at some scale could be a strong driver of tilefish abundance and/or recruitment. Surprisingly little information was presented under this ToR. Given the paucity of data to address this ToR, the working group somewhat understandably chose to focus on more immediate aspects of the assessment. Additional work however should focus on further analyses of the LPUE index to ascertain the potential influence of environmental drivers, such as the North Atlantic Oscillation index (Marsh et al. 1999), on tilefish abundance and/or recruitment. For example, cross-correlations between the NAO index and the recruitment estimates from the ASAP model could be calculated to ascertain whether there is any relationship between the two. If correlations exist at some lag (12-18 months was suggested in the working paper with respect to the NAO's impact on the region's hydrography) then the index could be a useful predictor of recruitment.

4.2.4 Use assessment models to estimate annual fishing mortality and stock size for the time series, and estimate their uncertainty. Include a historical retrospective to allow a comparison with previous assessment results.

The ToR was completed. The working group completed an impressive amount of work in applying three different types of model: ASPIC, SCALE, and ASAP. The working group showed that the ASPIC model had strong retrospective patterns and failed to capture the periodic strong year classes that were evident in the age composition data. This model should no longer be considered in future assessments.

The SCALE model was used in the previous (SARC 48) assessment, and was better able to capture the periodic strong year classes that caused biomass to fluctuate. However, the model does not fit the LPUE or size composition data very well and it appears to strongly under-estimate numbers at age of older fish.

The ASAP model was used for the first time in this assessment. The working group tested a number of configurations, arriving at a base configuration of: ages 1-10+;

pooled age data; M=0.15; a strong dome-shaped selectivity with sharp peak at age 7 for 1971-1982 and at age 5 for 1982-2012. This configuration fit the LPUE indices and age compositions well. The only concern about the configuration was regarding evidence for the domed selectivity function. This is not a conservative assumption and the working group did not initially present much information to support it. Discussion on this issue was sustained during the follow up session with the working group. The key line of evidence supporting the domed-selectivity was the lack of older fish in the catch-at-age data. This was confirmed at the meeting with anecdotal information volunteered by industry representatives that they do not target large/old fish due to market pressures and because they are generally located on the shelf break in areas where the potential for gear loss and conflict with lobster and other fisheries is high.

4.2.5 State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY or for their proxies) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

The ToR was completed. The working group used the fact that the stock was managed with a constant quota since 2002 to develop BRP's. An F_{MSY} proxy was set at $F_{25\%}$ based on results of a yield-per-recruit analysis using the average F = 0.37 during the constant quota period. A B_{MSY} proxy = 5153 mt was obtained from stock projections, sampling from estimated recruitment values over the full time-series and F held at a constant 0.37. Estimates of uncertainty of the BRP's were not provided. This approach seems appropriate given the assessment suggests no sign of stock decline in recent years.

4.2.6 Evaluate stock status with respect to the existing ASPIC model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt. a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates. b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-4).

The ToR was completed. The BRP's arising from the ASPIC model are not comparable to the proxies adopted from the accepted ASAP model. Because the ASPIC model did not fit the data well it should not be used for stock status determination.

The current F (F_{2012} = 0.27) is below the overfishing reference point, F_{MSY} proxy ($F_{25\%}$ = 0.37). The current stock biomass (SSB_{2012} = 5229 mt) appears to be rebuilt, B_{MSY} proxy = 5153 mt.

4.2.7 Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs). a. Provide numerical annual 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. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment). b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions. c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

The ToR was completed. Because the model has no information to reliably estimate recruits in the terminal 3 years (ages 1-3 are not seen in the fishery), the working group employed an empirical correction to these estimates for ages 1-4 based on the results of a retrospective analysis of the estimated recruit series. Using these adjusted recruit values for 2010-2012, the working group employed a standard forward projection, sampling from the full time-series of recruit estimates. This approach yielded biomass projections that were more consistent with trends in stock abundance estimated by the model, when compared to projections sampling from unadjusted recruit estimates. The ``recruit-adjusted'' approach therefore seems valid.

4.2.8 Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

The ToR was completed. The working group has made some progress on research recommendations from the previous SARC (SARC 48). But, relatively little progress has been made on: (1) developing protocols to ensure consistency in market category designations among dealers and fishers; (2) developing an industry based survey that could serve as a proxy for a fully fishery-independent survey; and (3) understanding causes in the pattern and variability in recruitment. These are issues that were discussed by the RP in the current SARC and it was a little disappointing that the working group didn't have more to offer on these points. The industry

appears to be dominated by a small number of key participants and it therefore seems tractable to work with them to develop a survey and a consistent market size designation protocols.

Additionally, more refined measures of fishing effort beyond trip length should be developed to better characterized LPUE. With the adoption of the ASAP model, sustained or enhanced levels of aging of commercial landings and age validation work must be carried out. Finally, given the apparent narrow temperature range tolerated by tilefish, research on potential environmental drivers, such as the NAO, of recruitment and abundance should be conducted. There is an apparent cyclical pattern in the LPUE index and model-estimated recruitment that could imply periodic external forcing. Given the fishery doesn't see age 1-3 fish, a better understanding of recruitment dynamics would clearly provide useful information to the assessment.

4.3 Northern shrimp

4.3.1 Present the Gulf of Maine northern shrimp landings, discards, effort, and fishery-independent data used in the assessment. Characterize the precision and accuracy of the data and justify inclusion or elimination of data sources.

The ToR was completed. The working group did an excellent job of summarizing the data inputs to the assessment. Commercial catches were presented from 1960-2013. Catches in recent years (2010-2013) have surpassed the levels recommended by the working group in past assessments. Annual fishing effort is highly variable, in part due to the large changes in fleet size over relatively short time periods. Catches may not adequately reflect true abundance due to these rapid changes in fleet size and due to the constant quota management in recent years.

The working group deemed commercial CPUE an unreliable indicator of abundance and did not include it as an input to the assessment. This decision should be revisited as even if the CPUE data are of relatively poor quality (CV's were not presented) they can still inform the model(s) and may help overcome the possible poor indexing of abundance by the commercial catch.

Of the five survey datasets presented, the working group chose to include only the ASMFC (state-federal) summer shrimp survey and the NEFSC spring and fall trawl surveys. Other smaller state surveys were excluded primarily because of their small spatial coverage. This choice seems appropriate but more work could be done to determine whether combining surveys in a modelling framework (spatio-temporal GAM's, hierarchical models, etc...) could allow more of the available information to flow into the assessment.

4.3.2 Estimate population parameters (fishing mortality, biomass, and abundance) using assessment models. Evaluate model performance and stability through sensitivity analyses and retrospective analysis, including alternative natural mortality (M) scenarios. Include consideration of environmental effects where possible. Discuss the effects of data strengths and weaknesses on model results and performance.

The ToR was not completed. The working group conducted a tremendous amount of work on implementing and evaluating 3 very different assessment models (ASPIC, CSA, and UME). The work was rigourous and of a high standard given the time available, but the massive 2006 spike in the survey biomass and abundance indices followed by an immediate decline to near average levels in 2007, proved too difficult for any of the models to capture. I list here the main issues with each model.

ASPIC. The working group acknowledged that this model was not intended for assessment purposes but rather as a comparison for the CSA and UME model results. The high variability in northern shrimp year class strength was, as expected, not captured by the ASPIC model. Given the stock dynamics, including highly variable recruitment, this model should no longer be evaluated.

CSA. This model was accepted in the previous peer-reviewed assessment (SARC 45) and should generally be an appropriate option for northern shrimp as, unlike ASPIC, the model explicitly accounts for recruitment dynamics. The model however failed to fit the data adequately, with a repeating trend of positive and negative residuals of the fit to the survey indices. Clearly, the model could not handle the rapid increase in survey biomass and abundance from 2005 to 2006, followed by the large and immediate decline in 2007. Additionally, too much emphasis was placed on the catch series owing to its low CV and equal likelihood weightings. It is highly unlikely that the catch series adequately represents the temporal pattern in abundance due to the constant catch quota imposed in recent years and the highly annual variability in the fishing fleet size. The working group was asked to explore scenarios with the catch series down-weighted by varying degrees and with M set as either a constant 0.5 or 0.5 * the predation pressure index. The changes in M from the U-shaped function and down-weighting of the catch series did improve the residuals patterns somewhat. Unfortunately, changing the weights on the catch series produced large changes in F and consequently large differences in management advice. This implementation of the CSA model can not be used for management advice.

UME. The UME model is a size-structured model that in principle should be the right assessment approach for northern shrimp. However, the model failed to adequately fit the survey size composition data or the survey indices, even when the catch series was down-weighted (as was done for the CSA model) and the 2006 survey data point was excluded. The implementation of a U-shaped function of M at size was complex and apparently not data-driven. It was also a concern that, although the model has been used in assessments of other species, there appears to be no peer-review of

the general model to determine it's robustness and general applicability. Ultimately, the poor fit to the survey data rendered the model inadequate for management advice purposes.

4.3.3 Update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, SSB_{MSY}, F_{MSY}, MSY). Evaluate stock status based on BRPs.

The ToR was not completed. The proposed BRP's could not be accepted as they were based on failed assessments. Similarly, stock status based on BRP's could not be evaluated. Until a benchmark assessment is accepted, stock status and management advice should be based on insight from the survey index and, when a fishery occurs, observed catches and CPUE information.

4.3.4 Characterize uncertainty of model estimates of fishing mortality, biomass and recruitment, and biological reference points.

The ToR was not completed. Evaluation of uncertainty of model estimates and BRP's was not possible given the failure of the proposed assessment models.

4.3.5 Review the methods used to calculate the annual target catch and characterize uncertainty of target catch estimates.

The ToR was not completed. The general approach used was appropriate but could not be implemented or evaluated given the failure of the proposed assessment models.

4.3.6 Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made before the next benchmark assessment.

The ToR was completed. The working group put forth a detailed and appropriate list of research recommendations. In addition to further refinement of the UME size-structured model, the working group should consider developing BRP's appropriate for a changing environment and/or predator community. Similarly, they should consider exploring environmental drivers of recruitment; the use of physical ocean model output, specifically bottom temperature perhaps as a spatio-temporal index, could be informative here.

In the meeting, the working group provided a compelling rationale for assuming a terminal molt may exist in northern shrimp, but could not provide data or citations specific to this stock to support this assumption. A tagging study, lab-based at least initially, is needed to assess whether a terminal molt indeed exists for this northern shrimp stock. Only with supporting data should a U-shaped natural mortality function be considered in the assessment, as it is far too strong an assumption to consider otherwise.

The working group should consider incorporating fishing effort information or a CPUE index as an additional refinement to the CSA model. Although the commercial catch series is highly precise, it is unlikely to track abundance well, at least in the most years given the constant quota management of the fishery. Incorporating effort or and CPUE series would help alleviate this issue.

Finally, given that no fishery will occur in 2014, it is vital that the survey continues to provide information on the stock's status.

4.3.7 Based on the biology of species, and potential scientific advances, comment on the appropriate timing of the next benchmark assessment and intermediate updates.

The ToR was completed. The working group should continue with annual updates on the stock's status, providing management advice based on survey indices (and fishery catches and CPUE if the fishery is re-established in the near future). This simplified update approach should allow the group time to further develop the UME size-structured and CSA models. Given adequate time to ensure at least one of the UME or CSA models are viable, the next benchmark assessment should occur as soon as possible.

5 Comments on the NMFS review process

In general, the review process was rigorous, well organized, collegial, and from this reviewer's perspective worked very well. Clear strengths of the process include:

- the effective chairing of the review meeting, including effective guidance from the SAW chair to ensure that the review panel did not stray from the required outcomes of the process;
- the thoroughness of the assessment work, including comparisons among multiple assessment models and numerous sensitivity tests;

- the willingness of the assessment leads to conduct additional analyses upon request;

- the availability of presentations, rapporteur notes, and ``homework assignments'' prior to final adjournment of the meeting.

The only weakness to the process is the number of assessments that were reviewed at the meeting. Having three full assessment reviews squeezed into five days does not really provide adequate time to do justice to the large amount of work accomplished, and hence substantial volume of material presented, by each of the assessment teams. I would suggest that steps be taken to either: (1) reduce the volume of information presented, e.g., through reduction of duplication of results in tables and/or figures; (2) extend meetings beyond the 5 days normally allotted; or (3) strive to review a maximum of two assessments per meeting. Or some combination of these. In reality, none of these small suggestions are likely to resolve the issue as I suspect the NEFSC has become a ``victim of it's own success'' in being able to conduct increasingly sophisticated assessments of a, probably, relentlessly increasing list of species.

6 Acknowledgements

I thank Jim Weinberg and Paul Rago for their hospitality and advice throughout the meeting. Thanks to Rob Latour for an excellent job as the SARC chair, guiding the review panel throughout the meeting. I thank the other panel members, Cathy Dichmont and Stewart Frusher for enjoyable discussions during the meeting and at dinner each evening. Manoj Shivlani and Roberto Koeneke for coordinating the review and providing travel arrangements. Finally, I thank the assessment teams led by: Chuck Adams/Tim Miller (butterfish); Paul Nitschke (tilefish); Kelly Whitmore/Anne Richards/Katie Drew (northern shrimp). Their hard work in preparing thorough assessment documents and presentations; willingly submitting to requests for additional model runs; and being open to the panel's recommendations during the meeting was greatly appreciated.

Appendix 1: Bibliography of Material Provided for Review

Cadrin SX, Boutillier JA, & Idoine JS. 2004. A hierarchical approach to determining reference points for Pandalid shrimp. Canadian Journal of Fisheries and Aquatic Sciences 61(8): 1373-1391. 19 p.

Chen Y, Hunter M, Vadas R, Beal B. 2003. Developing a growth-transition matrix for the stock assessment of the green sea urchin (*Strongylocentrotus droebachiensis*) off Maine. Fishery Bulletin 101:737-744. 8 p.

Chen Y, Kanaiwa M, & Wilson C. 2005. Developing and evaluating a size-structured stock assessment model for the American lobster, *Homarus americanus*, fishery. New Zealand Journal of Marine and Freshwater Research (Special issue on Lobster Biology and Management). 39:645-660. 16 p.

Clark SH, Cadrin SX, Schick DF, Diodati PJ, Armstrong MP, and McCarron D. 2000. The Gulf of Maine Northern Shrimp (*Pandalus borealis*) Fishery: a Review of the Record. J. Northw. Atl. Fish. Sci. 27: 193-226. 34 p.

Cotter J. 2009. Northeast Stock Assessment Review Committee (SARC 49) for Atlantic Surfclams in the US EEZ and Butterfish. 30 Nov – 3 Dec 2009. Woods Hole, MA. 49 p.

Link JS & Idoine JS. 2009. Estimates of Predator Consumption of the Northern Shrimp Pandalus borealis with Implications for Estimates of Population Biomass in the Gulf of Maine. North American Journal of Fisheries Management, 29(6): 1567-1583. 18 p.

McBride et al. 2013. Changes in size and age at maturity of the northern stock of Tilefish *(Lopholatilus chamaeleonticeps*) after a period of overfishing. Fish Bull. 111: 161-174. 14 p.

Marsh et al. 1999. The 1882 tilefish kill - a cold event in shelf waters off the north-eastern United States? Fish. Oceanogr. 8:1, 39-49, 1999. 11 p.

Northeast Fisheries Science Center. 2007. 45th Northeast Regional Stock Assessment Workshop (45th SAW): 45th SAW assessment report. US Dep Commer, Northeast Fish Sci Cent Ref Doc 07-16; 370 p.

Northeast Fisheries Science Center. 2009. 48th Northeast Regional Stock Assessment Workshop (48th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-15; 834 p.

Northeast Fisheries Science Center. 2009. 48th Northeast Regional Stock Assessment Workshop (48th SAW) assessment summary report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 09-10; 50 p.

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Summary Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-01; 41 p

Northeast Fisheries Science Center. 2010. 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-03; 383 p.

Northeast Fisheries Science Center. 2009. Summary Report of the 49th Northeast Regional Stock Assessment Review Committee (SARC 49). 30 Nov – 3 Dec 2009. Woods Hole, MA. 37 p.

Northeast Fisheries Science Center. 2013. Butterfish model diagnostics for the ASAP3 base model which served as the basis for developing the final model for SAW58 in 2014. 73p.

Richards RA, Fogarty MJ, Mountain DG, Taylor MH. 2012. Climate change and northern shrimp recruitment variability in the Gulf of Maine. Marine Ecology Progress Series 464:167-178. 12 p.

Smith MT. Northeast Regional Stock Assessment Review Committee (SARC 49). 2009. Independent reviewer's report on the 2009 Atlantic Surfclam and Butterfish Benchmark Stock Assessments. December 2009. 30 Nov – 3 Dec 2009. Woods Hole, MA. 51 p.

Sparholt H. Northeast Regional Stock Assessment Review Committee (SARC 49). 2009. SARC 49: Butterfish and Atlantic surfclam Benchmark Stock Assessments. 30 Nov – 3 Dec 2009. Woods Hole, MA. 33 p.

Working Papers

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Report of Butterfish. Working Paper #1. SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Report of Tilefish. Working Paper #1.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Report of Northern shrimp. Working Paper #1.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Summary Report of Butterfish. Working Paper #2. SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Summary Report of Tilefish. Working Paper #2.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Working Group, Stock Assessment Workshop (SAW 58) 2014. Stock Assessment Summary Report of N. shrimp. Working Paper #2.SAW/SARC 58. January 27-31, 2014, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA.

Appendix 2: CIE Statement of Work for Dr. Ian Jonsen

58th Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for butterfish, tilefish, and northern shrimp

Statement of Work (SOW) for CIE Panelists
(including a description of SARC Chairman's duties)

BACKGROUND

The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Representative (COR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are independently selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

SCOPE

Project Description: The Northeast Regional Stock Assessment Review Committee (SARC) 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 SARC is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development (SAW Working Groups or ASMFC technical committees), assessment peer review, public presentations, and document publication. This review determines whether the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fishery management in the northeast region.

Brief description of the science to be peer reviewed, and its relevant importance: The purpose of this meeting will be to provide an external peer review of benchmark stock assessments for **butterfish**, **tilefish**, and **northern shrimp**. This review determines whether the scientific assessments are adequate to serve as a basis for

developing fishery management advice. Results form the scientific basis for fishery management in the northeast region.

OBJECTIVES

The SARC review panel will be composed of three appointed reviewers from the Center of Independent Experts (CIE), and an independent chair from the SSC of the New England or Mid-Atlantic Fishery Management Council. The SARC panel will write the SARC Summary Report and each CIE reviewer will write an individual independent review report.

Duties of reviewers are explained below in the "**Requirements for CIE Reviewers**", in the "**Charge to the SARC Panel**" and in the "**Statement of Tasks**". The stock assessment Terms of Reference (ToRs) are attached in **Annex 2**. The draft agenda of the panel review meeting is attached in **Annex 3**. The SARC Summary Report format is described in **Annex 4.**

Requirements for the reviewers: Three reviewers shall conduct an impartial and independent peer review of the striped bass and summer flounder stock assessments, and this review should be in accordance with this SoW and stock assessment ToRs herein. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include statistical catch-at-age, state-space and index methods. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points that includes an appreciation for the varying quality and quantity of data available to support estimation of Biological Reference Points. SARC 58 will address fishery stock assessments of **butterfish, tilefish,** and **northern shrimp.** For shrimp and butterfish, experience in the following is desirable: assessment of short-lived species, stocks where the environment and environmental change can impact recruitment and availability in research surveys. Specifically for tilefish: experience with assessments based on commercial catch per unit of effort.

PERIOD OF PERFORMANCE

The contractor shall complete the tasks and deliverables as specified in the schedule of milestones within this statement of work. Each reviewer's duties shall not exceed a maximum of 16 days to complete all work tasks of the peer review described herein.

Not covered by the CIE, the SARC chair's duties should not exceed a maximum of 16 days (i.e., several days prior to the meeting for document review; the SARC meeting in Woods Hole; several days following the open meeting for SARC Summary Report preparation).

PLACE OF PERFORMANCE AND TRAVEL

Each reviewer shall conduct an independent peer review during the panel review meeting scheduled in Woods Hole, Massachusetts during dates of January 27-31, 2014.

STATEMENT OF TASKS

Charge to SARC panel: During the SARC meeting, the panel is to determine and write down whether each stock assessment Term of Reference (ToR) of the SAW (see **Annex 2**) 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 SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment Term of Reference of the SAW.

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 following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Tasks prior to the meeting: The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, phone number, FAX number, and a CV suitable for the public) to the COR, who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COR prior to the commencement of the peer review.

<u>Foreign National Security Clearance</u>: The reviewers shall participate during a panel review meeting at a government facility, and the NMFS Project Contact will be

responsible for obtaining the Foreign National Security Clearance approval for the reviewers who are non-US citizens. For this reason, the reviewers shall provide by FAX (or by email if necessary) the requested information (e.g., 1.name [first middle and last], 2.contact information [address, telephone number], 3.gender, 4.country of birth, 5.country of citizenship, 6.country of permanent residence, 7.whether there is dual citizenship, 8.country of current residence, 9.birth date [mo, day, year], 10.passport number, 11.country of passport) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/.

Pre-review Background Documents and Working Papers: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the SARC chair and CIE reviewers the necessary background information and reports (i.e., working papers prepared by the SAW Working Group) for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

Tasks during the panel review meeting: Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the stock assessment ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

(SARC chair)
Act as chairperson, where duties include control of the meeting, coordination of presentations and discussions, making sure all stock assessment Terms of Reference of the SAW are reviewed, control of document flow, and facilitation of discussion. For each assessment, review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to discuss the stock assessment and to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

(SARC CIE reviewers)
For each stock assessment, participate as a peer reviewer in panel discussions on assessment validity, results, recommendations, and conclusions. From a reviewer's point of view, determine whether each stock assessment Term of Reference of the SAW was completed successfully. Terms of Reference that are completed successfully are likely to serve as a basis for providing scientific advice to management. If a reviewer considers any existing Biological Reference Point or BRP proxy to be inappropriate, the reviewer should try to recommend an alternative, should one exist. Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

During the question and answer periods, provide appropriate feedback to the assessment scientists on the sufficiency of their analyses. It is permissible to request additional information if it is needed to clarify or correct an existing analysis and if the information can be produced rather quickly.

Tasks after the panel review meeting:

SARC CIE reviewers:
Each CIE reviewer shall prepare an Independent CIE Report (see **Annex 1**). This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified above in the "Charge to SARC panel" statement.

If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent CIE 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 may be raised. Comments on these questions should be included in a separate section at the end of the Independent CIE Report produced by each reviewer.

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

SARC chair:
The SARC chair shall prepare a document summarizing the background of the work to be conducted as part of the SARC process and summarizing whether the process was adequate to complete the stock assessment Terms of Reference of the SAW. If appropriate, the chair will include suggestions on how to improve the process. This document will constitute the introduction to the SARC Summary Report (see **Annex 4**).

SARC chair and CIE reviewers:
The SARC Chair, with the assistance from the CIE reviewers, will prepare the SARC Summary Report. Each CIE reviewer and the chair will discuss whether they hold similar views on each stock assessment 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 SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.

The chair's objective during this SARC Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion.

The SARC Summary Report (please see **Annex 4** for information on contents) should address whether each stock assessment Term of Reference of the SAW was completed successfully. For each Term of Reference, this report should state why that Term of Reference was or was not completed successfully. The Report should also include recommendations that might improve future assessments.

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

The contents of the draft SARC Summary Report will be approved by the CIE reviewers by the end of the SARC Summary Report development process. The SARC chair will complete all final editorial and formatting changes prior to approval of the contents of the draft SARC Summary Report by the CIE reviewers. The SARC chair will then submit the approved SARC Summary Report to the NEFSC contact (i.e., SAW Chairman).

DELIVERY

Each reviewer shall complete an independent peer review report in accordance with the SoW. Each reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each reviewer shall complete the independent peer review addressing each stock assessment ToR listed in **Annex 2**.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting at the Woods Hole, Massachusetts scheduled during January 27-31, 2014.
3) Conduct an independent peer review in accordance with this SoW and the assessment ToRs (listed in **Annex 2**).
4) No later than February 14, 2014, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and to Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in **Annex 1**, and address each assessment ToR in **Annex 2**.

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

December 16, 2013	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
January 13, 2014	NMFS Project Contact will attempt to provide reviewers the pre-review documents
January 27-31, 2014	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
January 31, 2014	SARC Chair and CIE reviewers work at drafting reports during meeting at Woods Hole, MA, USA
February 14, 2014	Reviewers submit draft independent peer review reports to the contractor's technical team for independent review
February 17, 2014	Draft of SARC Summary Report, reviewed by all CIE reviewers, due to the SARC Chair *

February 21, 2014	SARC Chair sends Final SARC Summary Report, approved by CIE reviewers, to NEFSC contact (i.e., SAW Chairman)
February 28, 2014	Contractor submits independent peer review reports to the COR who reviews for compliance with the contract requirements
March 7, 2014	The COR distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC Summary Report will not be submitted, reviewed, or approved by the CIE.

The SAW Chairman will assist the SARC chair prior to, during, and after the meeting in ensuring that documents are distributed in a timely fashion.

NEFSC staff and the SAW Chairman will make the final SARC Summary Report available to the public. Staff and the SAW Chairman will also be responsible for production and publication of the collective Working Group papers, which will serve as a SAW Assessment Report.

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COR within 10 working days after receipt of all required information of the decision on substitutions. The COR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: The deliverables shall be the final peer review report from each reviewer that satisfies the requirements and terms of reference of this SoW. The contract shall be successfully completed upon the acceptance of the contract deliverables by the COR based on three performance standards:

(1) each report shall be completed with the format and content in accordance with **Annex 1**,
(2) each report shall address each stock assessment ToR listed in **Annex 2**,
(3) each report shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Upon the acceptance of each independent peer review report by the COR, the reports will be distributed to the NMFS Project Contact and pertinent NMFS science director, at which time the reports will be made publicly available through the government's website.

The contractor shall send the final reports in PDF format to the COR, designated to be William Michaels, via email William.Michaels@noaa.gov

Support Personnel:

William Michaels, Program Manager, COR
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

Key Personnel:

Dr. James Weinberg, NEFSC SAW Chairman, NMFS Project Contact
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov (Phone: 508-495-2352) (FAX: 508-495-2230)

Dr. William Karp, NEFSC Science Director
Northeast Fisheries Science Center
166 Water St., Woods Hole, MA 02543
william.karp@noaa.gov Phone: 508-495-2233

Annex 1: Format and Contents of Independent Peer Review Report

1. The independent peer review 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 main body of the report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Findings of whether they accept or reject the work that they reviewed, and an explanation of their decisions (strengths, weaknesses of the analyses, etc.) for each ToR, and Conclusions and Recommendations in accordance with the ToRs. For each assessment reviewed, the report should address whether each ToR of the SAW was completed successfully. For each ToR, the Independent Review Report should state why that ToR was or was not completed successfully. To make this determination, the SARC chair and reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice.

   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, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the SARC Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The independent report shall be a stand-alone document for others to understand the proceedings and findings of the meeting, regardless of whether or not others read the SARC Summary Report. The independent report shall be an independent peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

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

Annex 2: 58th SAW/SARC Stock Assessment Terms of Reference

(file vers.: 8/2/2013)

A. Butterfish

1. Characterize the commercial catch including landings, effort and discards by gear type. Describe the magnitude of uncertainty in these sources of data.

2. Characterize the survey data that are being used in the assessment. Describe the magnitude of uncertainty in these sources of data.

3. Characterize oceanographic and habitat data as it pertains to butterfish distribution and availability. If possible, integrate the results into the stock assessment (TOR-5).

4. Evaluate consumptive removals of butterfish by its predators. If possible, integrate results into the stock assessment (TOR-5).

5. Use assessment models to estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series, and estimate their uncertainty. Include a comparison with previous assessment results and previous projections.

6. State the existing stock status definitions for "overfished" and "overfishing". Given that the stock status is currently unknown, update or redefine biological reference points (BRPs; point estimates for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY, or their proxies) and provide estimates of their uncertainty. Consider effects of environmental factors on stability of reference points and implications for stock status.

7. Evaluate stock status with respect to a newly proposed model and with respect to "new" BRPs and their estimates (from TOR-6). Evaluate whether the stock is rebuilt.

8. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
    a. Provide numerical annual projections (2 years). Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment). Comment on which projections seem most realistic.
    b. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

9. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

B. Tilefish

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

2. Characterize commercial LPUE as a measure of relative abundance. Consider the utility of recreational data for this purpose. Characterize the uncertainty and any bias in these sources of data.

3. For the depth zone occupied by tilefish, examine the relationship between bottom temperature, tilefish distribution and thermal tolerance.

4. Use assessment models to estimate annual fishing mortality and stock size for the time series, and estimate their uncertainty. Include a historical retrospective to allow a comparison with previous assessment results.

5. State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY or for their proxies) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs.

6. Evaluate stock status with respect to the existing ASPIC model (from previous peer reviewed accepted assessment) and with respect to a new model developed for this peer review. In both cases, evaluate whether the stock is rebuilt.
    a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
    b. Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-4).

7. Develop approaches and apply them to conduct stock projections and to compute the statistical distribution (e.g., probability density function) of the OFL (overfishing level) and candidate ABCs (Acceptable Biological Catch; see Appendix to the SAW TORs).
    a. Provide numerical annual 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. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
    b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
    c. Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.

8. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

C. Northern shrimp

1. Present the Gulf of Maine northern shrimp landings, discards, effort, and fishery-independent data used in the assessment. Characterize the precision and accuracy of the data and justify inclusion or elimination of data sources.

2. Estimate population parameters (fishing mortality, biomass, and abundance) using assessment models. Evaluate model performance and stability through sensitivity analyses and retrospective analysis, including alternative natural mortality (M) scenarios. Include consideration of environmental effects where possible. Discuss the effects of data strengths and weaknesses on model results and performance.

3. Update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, SSB_{MSY}, F_{MSY}, MSY). Evaluate stock status based on BRPs.

4. Characterize uncertainty of model estimates of fishing mortality, biomass and recruitment, and biological reference points.

5. Review the methods used to calculate the annual target catch and characterize uncertainty of target catch estimates.

6. Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made before the next benchmark assessment.

7. Based on the biology of species, and potential scientific advances, comment on the appropriate timing of the next benchmark assessment and intermediate updates.

Annex 2 (cont.):

Appendix to the SAW Assessment TORs:

Clarification of Terms used in the SAW/SARC Terms of Reference

Appendix to the Assessment TORs:

Explanation of "Acceptable Biological Catch" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 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)

Explanation of "Vulnerability" (DOC Natl. Standard Guidelines, Fed. Reg., vol. 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 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)

Rules of Engagement among members of a SAW Assessment Working Group:

> Anyone participating in SAW assessment working group 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 3: Draft Agenda

58th Northeast Regional Stock Assessment Workshop (SAW 58)
Stock Assessment Review Committee (SARC) Meeting

January 27-31, 2014

Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

DRAFT AGENDA* (version: 25 October 2014)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Monday, Jan. 27

10 – 10:30 AM

Welcome	**James Weinberg**, SAW Chair		
Introduction	**Robert Latour**, SARC Chair		
Agenda			
Conduct of Meeting			

10:30 – 12:30	Assessment Presentation (A. Butterfish)		
	TBD	TBD	**TBD**

12:30 – 1:30 PM	Lunch

1:30 – 3:00	Assesssment Presentation (A. Butterfish)		
	TBD	TBD	**TBD**

3:00 – 3:15	Break

3:15 – 5:15	**S**ARC Discussion w/ Presenters (A. Butterfish)		
	Robert Latour, SARC Chair		**TBD**

5:15 – 5:45	Public Comments (A. Butterfish)

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Tuesday, Jan. 28

8:45 – 11 AM	Assessment Presentation (B. Tilefish)		
	TBD	TBD	**TBD**
11:00 - 11:15	Break		
11:15 – 12:15	SARC Discussion w/presenters (B. Tilefish)		
	Robert Latour, SARC Chair		**TBD**
12:15 – 12 :30	Public Comments (B. Tilefish)		
12:30 – 1:45 PM	Lunch		
1:45 – 4:15	Assessment Presentation (C. Northern shrimp)		
	TBD		**TBD**
4:15 - 4:30	Break		
4:30 – 5:45	SARC Discussion w/presenters (C. Northern shrimp)		
	Robert Latour, SARC Chair		**TBD**
5:45 – 6:00	Public Comments (C. Northern shrimp)		
7:00	**(**Social Gathering **)**		

TOPIC	PRESENTER(S)	SARC LEADER	RAPPORTEUR

Wed. Jan. 29

9:00 – 11:15 AM Revisit with presenters (A. Butterfish)
Robert Latour, SARC Chair **TBD**

11:15 – 11:30 Break

11:30 – 12:30 Revisit with presenters (B. Tilefish)
Robert Latour, SARC Chair **TBD**

12:30 – 1:30 PM Lunch

1:30 -2:30 (cont) Revisit with presenters (B. Tilefish)
Robert Latour, SARC Chair **TBD**

2:30 – 2:45 Break

2:45 – 5:15 Revisit with presenters (C. Northern shrimp)
Robert Latour, SARC Chair **TBD**

Thur. Jan. 30

8:30 – 11:30 Review/edit Assessment Summary Report (A. Butterfish)
Robert Latour, SARC Chair **TBD**

11:30 – 12:30 PM Lunch

12:30 – 2:45 Review/edit Assessment Summary Report (B. Tilefish)
Robert Latour, SARC Chair **TBD**

2:45 – 3:00 Break

3:00 - 5:30 Review/edit Assessment Summary Report (C. Northern shrimp)

Robert Latour, SARC Chair **TBD**

Friday, Jan. 31

9:00 AM – 5:00 PM SARC Report writing. (closed meeting)

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public, except where noted.

The NMFS Project contact will provide the final agenda before the meeting.

Reviewers must attend the entire meeting.

Annex 4: Contents of SARC Summary Report

1. The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background, a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the SARC chair and CIE reviewers should consider whether the work provides a scientifically credible basis for developing fishery management advice. Scientific 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 the CIE reviewers and SARC 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 (BRP) 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 SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

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

Appendix 3: Review Panel Membership

Robert Latour	SARC Chair
Catherine Dichmont	CIE Reviewer
Stewart Frusher	CIE Reviewer
Ian Jonsen	CIE Reviewer