Center for Independent Experts (CIE) Independent Peer Review Report on the 2016 South East Data, Assessment and Review (SEDAR 41) Workshop to Review the Assessments of South Atlantic Red Snapper and Gray Triggerfish

Michael J Armstrong[1]
Centre for Fisheries, Environment & Aquaculture Science (Cefas)
Pakefield Road
Lowestoft
Suffolk NR33 0HT
United Kingdom
mike.armstrong@cefas.co.uk

May 2016

[1] Representing the Center for Independent Experts

1. Executive Summary

The 2016 South East Data, Assessment and Review (SEDAR 41) Workshop to review the assessment of the status of South Atlantic Red Snapper and Gray triggerfish took place on 15-18th March 2016 in Charleston, South Carolina. The material presented for each stock was derived from a separate Data Workshop and an Assessment Workshop. The primary assessment model was the Beaufort Assessment Model (BAM), which is a length and age structured statistical model using fleet-specific landings, discards, length and age compositions and relative abundance indices. The data series for red snapper and gray triggerfish were based largely on the same data collection programmes and data processing methods, but differed in the exact mix of data inputs and the year ranges used. The BAM is an evolution of the one developed in SEDAR-15 and SEDAR-21 for red snapper.

To examine sensitivity of stock status evaluations and benchmarks to the type of model used, a surplus production model (ASPIC) was also run using the same fishery-dependent and fishery-independent data sources as used for the BAM, but excluding any length or age data. Some additional analyses using catch-curves were carried out for red snapper to examine mortality rates using only the landings-at-age data.

The main conclusion arising from the SEDAR-41 assessment of South Atlantic red snapper using BAM, and the sensitivity runs and uncertainty analysis of the assessment, is that there is a high probability that the stock is overfished and is experiencing overfishing. The perception of overfishing is driven by the addition of 2014 data, which revises all the 2010-2013 estimates upwards from values from runs of the same model terminating in those years. In other words, if the assessment had been done only to 2013, overfishing may not have been evident. The recent increase in fishing mortality estimates is evident primarily at ages 4 and over. It is uncertain how selectivity is changing since the moratorium in 2010 and the mini-seasons since 2012, and fishery data quality must be deteriorating due to the resultant increase in discarding, which is estimated with sampling error. There are also changes in recreational data collection involving intensive State surveys since 2012 to estimate private and charter recreational catches and discards, which is essentially a major change in design from just using the Marine Recreational Information Program (MRIP) in previous years. Additional sources of uncertainty for the model are the cessation of the fishery-dependent landings indices after 2009, concerns over altered fishing behaviour on catchability in the headboat discards index during the moratorium, and the short duration of the CVID index series. All these factors are likely to degrade the ability of the model to detect changes in fishing mortality, and the sudden change in perception with addition of 2014 data should be investigated further before concluding that overfishing is occurring and is increasing between 2010 and 2014. The conclusion that the stock is overfished and very slowly rebuilding in terms of egg production, however, appears robust. Some strong year classes formed in the 1980s, and since the mid 2000s have resulted in the stock being very abundant in numbers of young fish, which is corroborated by fishermen's observations, but the numbers of mature fish have increased only very slowly over the last decade and the total egg production remains very low. Testimony from recreational and commercial fishermen at the review meeting strongly disputed the model estimates that large red snappers remain comparatively rare, and considered that selectivity of commercial lines and the trap survey is not asymptotic as assumed for the model. Independent, empirical observations of fish behavior in relation to the gears are needed to resolve this dispute. The BAM conclusions on stock status of red snapper are not supported by the comparative ASPIC model that indicates that the stock is currently no longer overfished (using biomass rather than egg production) and that overfishing is no longer occurring. However, the assessment working group and the review panel agreed that the BAM was more appropriate as it explicitly models recruitment, age-based dynamics and fishery selectivity, which is important because the stock is strongly driven by widely-varying recruitment. An error in age compositions for the chevron trap survey was detected during the meeting, due to not correcting to true

age, requiring a re-run of the assessment to provide a new base model. Changes to the results were relatively small, and the sensitivity analyses, Monte Carlo bootstrap uncertainty analysis and retrospective analysis of the original base model are used for the present review.

The SEDAR-41 assessment of South Atlantic gray triggerfish using BAM had been fitted using both length and age compositions from the same trips, and the chevron trap series of relative abundance indices had been up-weighted by a factor of six to allow the model to fit the survey more closely than the age compositions, following some general guidelines by Chris Francis (New Zealand) who had commented on the SEDAR-24 assessment of red snapper. The review panel considered that this resulted in fitting too closely to the point estimates from the survey, including an unusually low value in the first year of the series (1990). A sensitivity run using only age data where both length and age were available resulted in poorer fits to the age composition data. The review panel requested that the assessment be put back to the assessment working group for a full review of how these different data sets should be weighted. Hence there is no final assessment run for gray trigger fish from SEDAR-41. All BAM sensitivity runs and uncertainty analysis presented at the review workshop indicated that the stock is not overfished, and that overfishing is not occurring, with declining fishing mortality and increasing spawning stock biomass in recent years. The comparative ASPIC model run fitted poorly due to lack of contrast in the data, but the conclusions on stock status were the same as from BAM. There does not therefore appear to be any current risk to stock status but further work on the BAM by the assessment working group is needed to confirm this finding.

2. Background

South East Data, Assessment, and Review (SEDAR) is a joint process for conducting stock assessments, and peer-reviewing their outcomes, for stocks of interest to the South Atlantic, Gulf of Mexico and Caribbean Fishery Management Councils, NOAA Fisheries, SEFC, SERO and the Atlantic and Gulf States Marine Fisheries Commissions. SEDAR is organized around separate data, assessment and review workshops.

The previous assessment of South Atlantic red snapper was conducted in 2010 (SEDAR-21), using a similar form of model to the one implemented at SEDAR-41. South Atlantic Gray triggerfish has not previously been benchmarked by SEDAR. Input data for the SEDAR-41 assessment were compiled during the Data Workshop (DW), and population models were developed during the subsequent Assessment Workshop (AW), taking into account recommendations from the SEDAR-21 independent peer review of the data and assessment models for red snapper and other subsequent developments.

3. Description of review activities

The SEDAR 41 Review Workshop (RW) took place at the Crowne Plaza Convention Centre at Charleston from 8:30 am Tuesday 15 March 2016 to around midday Friday 18 March. The assessment results and background were clearly presented by the experts at the meeting. Time was set aside each day to allow observers and Council representatives to provide testimony, and for the review panel to seek clarification from the assessment team and to request additional model runs, which were done very quickly and led to fruitful discussion that helped to clarify a number of important issues. All proceedings were on record, and the coordinator maintained notes each day that were circulated. The provisional agenda for the meeting is given in Annex 3 of Appendix 2.

The Review Panel itself comprised the Chair and two reviewers from SAFMC SSC, and three reviewers appointed by the CIE (Appendix 3). The assessment results were presented by two US technical experts from SEFSC Beaufort who were involved in the AW, with support from five assessment workshop participants from the same laboratory. The RW was also attended by the SEDAR coordinator, Council and Agency staff, Council Representatives and some appointed observers. All documentation, including background documentation provided to earlier DW and AW meetings, was provided to the Review Panel in advance of the review workshop, and was comprehensive for the job in hand.

The Review Panel developed a Summary Report during and after the meeting. The following report presents my personal evaluation of the review process together with more extended observations on the data and assessment models that are not necessarily shared with the other panel members. I accept all responsibility for any errors in my report due to misinterpretations of the data or analyses.

Due to the need to revise the red snapper assessment at the end of the RW, the Review Panel report could not be completed during or immediately after the RW. Due to other commitments, the present CIE personal review had to be submitted before completion of the Panel Report in order to meet the CIE deadline. It is therefore possible that there are different views expressed here compared with the Panel report for which the reasons are not explained.

4. Summary of findings by Term of Reference

4.1 South Atlantic Red Snapper

ToR 1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:

a) Are data decisions made by the DW and AW sound and robust?
b) Are data uncertainties acknowledged, reported, and within the normal or expected levels?
c) Are data properly applied within the assessment model?
d) Are data input series reliable and sufficient to support the assessment approach and findings?

A general statement on (a) – (d) is given below, followed by some supporting observations.

(a) Data decisions made by the DW and AW were sound and robust. The development of input data and parameters for the BAM and ASPIC models required an extremely thorough compilation and evaluation of all available data at the DW. Modifications made subsequently by the AW were fully explained.

(b) Data uncertainties were acknowledged, reported, and were within the normal or expected levels, as far as could be ascertained from information provided to the RW. The data varied widely in coverage and quality, and were often heavily manipulated and standardized to try and develop coherent time series from diverse data sources of differing designs, coverage and accuracy. The combined data will have biases that in some cases are poorly understood, especially in earlier years of the time series. All decisions made by the DW and AW in compiling data were explained and justified in detail. Data quality metrics were provided by the DW in terms of numbers of samples, CVs, or alternative plausible data series or biological parameter values. These were used by the AW to weight data series in the assessment model, estimate the uncertainty in the assessment results using the Monte Carlo /

bootstrap method, or to explore the sensitivity of the assessment to data decisions and uncertainty.

(c) The data were properly applied within the assessment model. Any issues with application of the data such as time periods for fitting, use of length and age data from the same sampling schemes, or weighting of data according to data quality metrics, were explored at the SEDAR-41 RW, if not previously evaluated by the DW and AW.

(d) Data input series were applied if considered reliable and sufficient to support the assessment approach and findings. Reliability and sufficiency was evaluated based on a-priori criteria where possible, supported by data quality metrics such as numbers of samples or CVs and by model fits. The assessment is supported primarily by a wide range of fishery-dependent data covering landings and discards, and therefore is heavily driven by these data and assumptions related to their reliability and use. An additional fishery-independent trap and video survey data set unfortunately covers only the period since 2010 due to very low incidence of red snapper catches prior to the recent increase in abundance due to strong year classes.

Additional supporting comments

Life history parameters
Overall, the DW and AW made sound and robust decisions on life history parameters, and the uncertainty in these, and the data were used appropriately in the assessment.
Natural mortality (M) and stock-recruit steepness are key parameters related to stock productivity and biological reference points, but there are no direct estimates for red snapper and they have to be inferred from a meta-analysis approach for M-at-age using growth parameters and maximum observed age, and from the ability of the BAM to estimate steepness (which always tended to 1.0 due to high recruitment at the lower spawning stocks). These are conventional approaches in statistical assessment modelling. Sensitivity analyses of the BAM were conducted for different time-invariant values for M, but not for any plausible trends over time in M as there is no information to guide this. The stock-recruit relationship uses total annual egg production estimated based on maturity, length, number of batches and batch fecundity, allowing the effect of age structure on reproductive output to be reflected in setting SSB reference points and stock status. Trends or more random variations in fecundity in red snapper, considered an indeterminate batch spawner, are possible sources of uncertainty that could not be fully explored as historical information was not available. The low estimate of age at first maturity in females (43% at age 1) was considered by the review panel to be unusual for snappers, and it was speculated if it has declined as a compensatory response to heavy exploitation. Annual maturity data from the SERFS chevron trap survey could not be used to test this because sample collections have been from different areas in different time periods.

Fishery removals
The DW and AW made sound and robust decisions on fishery removals data, and the uncertainty in these (which is largest for discards estimates), and the data were used appropriately in the assessment, using only those data considered reliable enough for this purpose.

Historical commercial and recreational fishery removals – landings and dead discards – were constructed back to 1950 to allow a sufficient burn-in period for the BAM with assumed stable age structure and low fishing mortality in the early period. The DW made a large number of decisions to infer historical values from more recent data or to calibrate data series where the survey design has changed, including adjusting NMFS Marine Recreational Fisheries Statistics Survey (MRFSS) estimates from 1981 to 2003 to be consistent with catches from the Marine Recreational Information Program (MRIP: 2004 to present), and developing combined recreational landings back to 1955 using effort data from the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation Survey (FHWAR: SEDAR41-DW17) combined with average MRFSS and SRHS CPUE data for 1981-83. All such calibrations and extrapolations may introduce additional biases in the data series.

Recreational landings of headboats are estimated from the Southeast Region Headboat Survey (SRHS) log book scheme which has improved in quality over time due to introduction of mandatory reporting in 1996 and improved logbook supply from 2008 onwards. A lengthy review of the headboat survey that was carried out to address concerns of fishermen about the quality of the data (SEDAR41_DW46) found little evidence of problems that would preclude use of the data. The DW proposed a declining CV in three blocks from 1981 onwards to reflect improved quality of the data, which is reasonable in relative terms, but will be biased to some extent in absolute terms as it is not based on actual sampling errors. For earlier years in the BAM Monte Carlo bootstrap (MCB) uncertainty analysis, deviations were applied equally to all years and not independently by year. In practice, errors in subsequent years, where only half of the logbooks were returned (Fig.4.1.1 from SEDAR41_DW46), may also be correlated between years if non-response is related to size of catch; but no information was available on this and the assumptions made about independence of errors in the MCB analysis were adequate although potentially underestimating uncertainty.

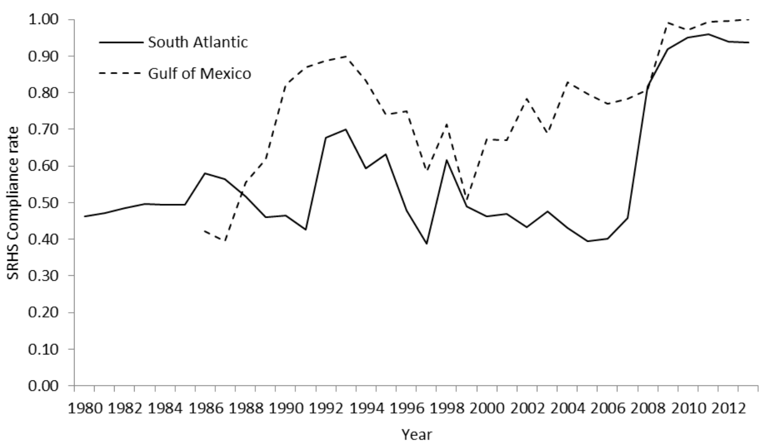

Fig. 4.1.1. Compliance rates in submission of headboat survey logbooks over time (from SEDAR41_DW46_NMFS-SEFC_HBDataEval_7.20.2015.pdf).

Private boat and charter boat landings since the early 1980s were estimated from MRFSS/MRIP, which has a robust and peer-reviewed statistical design that has substantially reduced bias and improved precision over time, and for which CVs are estimated directly based on efficient estimators.

Discards estimates from commercial and recreational fleets, where available, are subject to sampling error if based on observation, and potential for bias also exists where discards are recalled, as in the MRFSS/MRIP intercept surveys. Additional time-series error is introduced by extrapolating recent data to historical periods with no data. For example, commercial handline discard observations for 2002-2009 were extrapolated back to 1992, with zero discards assumed prior to that, due to low minimum landing size. Similarly, head boat discards estimates from log books and some at-sea observation since 2004 were extrapolated back in time based on changes in length frequencies recorded by dockside sampling before and after changes in minimum landing sizes, with zero discards assumed pre-1984. All these data manipulations can introduce additional error in the time series. Discards estimates from MRFSS/MRIP are self-reported by anglers intercepted at landing sites and are not verified.

The MRIP surveys provided too few estimates of red snapper landings or discards for the very brief mini-seasons since 2012, and additional State surveys using a variety of off-site and on-site methods were also used for these periods, based on collaboration between MRIP staff and State laboratories, which the Review Panel was advised is continuing in order to develop options for future sampling. Continued collaboration is encouraged. Whilst these surveys may be providing much more data than MRIP, it is not clear if this has resulted in a discontinuity due to different designs which may have different bias characteristics. The period covered by the mini-season could be treated as a separate post-stratum in MRIP for red snapper, albeit with low sampling frequency, and a statistically robust procedure then adopted to estimate discards within this stratum using additional samples collected by State laboratories, ensuring that sampling is representative and compatible with MRIP in terms of bias characteristics. This would avoid the more complex decision making processes on which data set to use. A draft manuscript by Sauls et al. provided to the Panel after the RW describes the relatively intensive on-site State survey in Florida.

Discarding of red snapper has increased over time due to changes in minimum landing size to 20 inches in 1992, increases in abundance of young fish from above-average year classes in some recent years, and the introduction of the moratorium in 2010 and 2011, and the small commercial catch limits and recreational bag limits in the mini seasons for 2012 onwards. Most of the catch is now discarded, and the inherent errors in discards estimation based on limited observer trips and fisher recalls means that the quality of total fishery removals estimates may therefore have deteriorated significantly, which will impact estimation of stock size and fishing mortality. Any initiatives to improve the quality of discards estimates would likely be beneficial, particularly as the BAM requires these and any landings estimates to be treated as precise.

Overall, the DW and AW made the best possible evaluation of fishery removals data and their uncertainties.

Length and age compositions
The DW and AW made sound and robust decisions on length and age composition data, and the uncertainty in these, and the data were used appropriately in the assessment, using only those data considered reliable enough for this purpose.

Length and age composition data for fisheries were collected by port-side and in some cases by at-sea sampling, and by sampling of fish in the trap surveys. The AW used age composition data in preference to length composition data in BAM where both data exist, except for discards where selectivity is fitted separately for discards and landings, but age data are available only for landings. Length composition data were fitted only for commercial handlines from 1984 – 1992, commercial discards in 2009 and 2013, and headboat discards from 2005 to 2014. Age compositions were fitted for commercial handlines landings from 1990 onwards, for headboat landings in two widely separated blocks in the 1980s and 2000s, for general recreational landings since 2001, and for the CVID survey from 2010. The CVID age data were found towards the end of the review meeting to have not been converted to calendar ages, and revised data were provided along with some preliminary assessment results which indicated some relatively small changes to the overall assessment results and stock status.

Samples below a specified number of trips and fish were excluded from the BAM, but there is still a widely varying number of trips sampled, which can be considered as the primary sampling units. The patchy nature of sampling of the fisheries means that the mix of data used in the model differs in the three time blocks where landings selectivity is fitted (1950-91; 1992-2009; 2010 onwards). Sampling for age was particularly intensive for headboats in the mid-1980s, commercial handlines and headboats in 2009 and for general recreational in 2014. The true effective sample sizes were not computed, but this would be a useful exercise to determine the relationship between numbers of trips, numbers of fish and the effective sample sizes, and hence evaluate the relative weighting factors for years.

Relative abundance indices
Overall, the DW and AW made sound and robust decisions on most of the relative abundance indices and the uncertainty in these, and the data were used appropriately in the assessment, using only those data considered reliable enough for this purpose. Three fishery-dependent indices were used in the assessment: commercial handline landings per unit effort (1993-2009), headboat landings per unit effort (1976-2009), headboat discards per unit effort from observer trips (2005-2014), and one fishery-independent survey, the chevron trap abundance index combined with video records from same survey (SERFS combined video and trap - CVID - 2010-2014). The general approaches to analysis of the data appear appropriate and are widely used for such data in assessments in the USA. Changes in red snapper management will affect targeting and catch-release behavior, and therefore any CPUE data from 2010 onwards are likely to be incompatible with earlier years, which is of concern for use of the headboat discards index.

The CVID index was used only from 2010 due to increased spatial coverage, integration with video data, and sufficient operations with red snapper catches to allow a reliable index calculation. A zero-inflated negative binomial model with trap soak minutes as an offset was used for trap and video, with stepwise removal of covariates (SEDAR41-DW54 for trap survey and DW45 for video survey). The covariates included in both the trap and video index were depth, latitude and bottom temperature, with some additional covariates in the video index. Subsequent to the DW, a combined analysis was conducted as the two data sets are not independent. The statistical methods adopted appear sound, but the series is unfortunately still very short and it is therefore difficult to evaluate if it provides abundance indices directly proportional to stock abundance in the longer term.

The DW adopted the widely used practices in the USA of filtering fishery CPUE data using the Stephens-MacCall (2004) method to remove fishing operations from areas and habitats with low probability of occurrence of red snapper based on species compositions observed, and then applying a delta-GLM with predictor variables to explain CPUE patterns accounting for spatial and temporal coverage, and trip type, among other factors. There was some discussion at the RW that, as these fisheries cover different depth ranges, a combined analysis might be beneficial external to the BAM, and this should be considered for future benchmark assessments. Possibly of more concern is the assumption of constant catchability in the fishery CPUE series, which if violated would bias the BAM (and ASPIC) model estimates. The short CVID and the longer handline index series show random residuals (AW report Figs 10 & 11) whilst the headboat fleet shows strongly serially correlated residuals with substantial under-fitting of the large 2008 and 2009 indices, with closest fit to the series of very low indices from 1992 – 2007. It is notable that the 2008 and 2009 indices correspond to a period of substantially increased compliance in logbook returns from less than 50% to almost 100% (Fig. 4.1.1 above). The estimation of total headboat landings assumes that vessels not supplying logbooks have the same CPUE on average as those supplying logbooks, and can therefore be applied to the headboat activity records (HAR) of fishing effort to generate catches where these are not supplied – this assumption should be investigated further by comparing the activities of vessels that now supply logbooks but persistently did not before 2008.

Further work is also warranted to better understand factors that could cause trends in catchability in commercial and recreational hook-and-line catch rates, such as changes in species targeting or effects of hook competition and handling time for other species on deck due to changes in their abundance relative to red snapper. It would be useful to consider trends in species compositions, particularly given the difficulty of BAM and ASPIC to fit the headboat indices for the early part of the series.
Other factors affecting CPUE trends include technology creep in catchability due to improvements in fishing gear, positioning (GPS) and communication systems, and also by rising fuel costs in recent years. SEDAR24 attempted account for improvements in technology (notably, GPS systems) by linearly increasing catchability of South Atlantic red snapper by 2% per year, beginning in 1976 for headboats, and 1993 for commercial lines, until 2003 and then holding it constant thereafter. This approach was not adopted in SEDAR41.

The Review Panel proposed that changes in management actions such as the moratorium, mini-season and reductions in bag limits are expected to alter fishing behavior and hence catchability in fishery-dependent indices, and should inform decisions on inclusion of data or periods of data in assessments. A member of the SAFMC stated on record at the SEDAR41 RW meeting that the behavior of anglers has changed substantially since the moratorium, to avoid catches and discarding of red snapper. The panel therefore considered that all the fishery CPUE series for red snapper should be applicable only to 2009, the year before the moratorium, and earlier years, and I agree with this conclusion.

ToR 2. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:

a) Are methods scientifically sound and robust?
b) Are assessment models configured properly and used consistent with standard practices?
c) Are the methods appropriate for the available data?

A general statement on (a) – (c) is given below, followed by some supporting observations.

a) The BAM stock assessment method used as the primary assessment tool for South Atlantic Red Snapper is scientifically sound and robust, and is based on well tried and tested integrated assessment models such as Stock Synthesis, and has been applied to a number of stocks in the USA. The ASPIC production model is also well tested worldwide, but the conditions under which it can give reliable results are more limited as it does not explicitly represent year class dynamics and fishery selectivity, and requires sufficient contrast in data to reliably estimate model parameters. Approaches to investigate uncertainty, including retrospective analysis, sensitivity analysis and Monte Carlo bootstrap analysis were all scientifically sound and robust.

b) The BAM and ASPIC assessment models are configured properly and their use is consistent with standard practices.

c) The methods used are appropriate given the nature of the available data. In particular, the patchy distribution of length and age sampling across time and fleets, and the variable sampling rates where sampling occurs, require a statistical modeling framework such as BAM to fit the underlying population model to the available observations in a statistically sound way. A downside is that patchy occurrence of short fishery composition series with very variable sampling rates can cause the model to over-weight and over-fit some series and years within series.

Additional comments are given below:

BAM vs ASPIC
The BAM is a superior model to the ASPIC production model as it makes direct use of fishery and survey length and age compositions to estimate annual recruitments (which have varied widely and are the most dynamic component of the intrinsic rate of increase (R) parameter in ASPIC for this stock), and can also account for changes in selectivity. In recent years, the total stock biomass has been dominated by young snappers, and therefore is heavily recruitment driven. The ASPIC model also proved difficult to fit.

The BAM has many assumptions and many estimated parameters, but the base model configuration appears to have reasonable assumptions and parameter estimates. As indicated by sensitivity analyses reported by the Assessment Workshop and during the Review Workshop, the most important data and modeling decisions were around choice and weighting of relative abundance indices, and modeling the form and time variation in selectivity.

Catch curves of age composition data were provided as exploratory information at the RW, but are not a valid basis for status determination as the apparent mortality rates do not reflect selectivity.

Configuration of the BAM base assessment and weighting of input data

The BAM fits to a wide range of fishery landings and discards data, multinomial length and age composition data, and relative abundance data. Where data are from a sampling design (rather than census), such as length compositions, recreational catches and trap surveys, the input data have associated estimates of sampling error. For catch estimates and surveys a CV is provided, whilst for length and age compositions an input sample size is included as a proxy for effective sample size – in this case the number of fishing trips sampled. The input CVs and sample sizes inform the model about the relative precision of input data. A penalized likelihood approach is used for fitting the model, using robust multinomial likelihoods for length and age compositions, and lognormal likelihoods for removals and index values. For parameters defining selectivities, CV of size at age, and recruitment variability (σ_R), normal priors were applied. An age error matrix is included to allow for this additional source of error when fitting age composition data.

The assessment follows the common practice of weighting compositional catch data and abundance indices in two stages. The input data are first assigned relative weights before the model is run using input CVs and sample sizes. Assuming the model converges correctly, the data weights are iteratively adjusted upwards or downwards during a series of model runs to improve model fit. The method of doing this for red snapper was to adjust the weights until the standard deviations of normalized residuals were near 1.0 (Francis 2011).

The relative input weights between years in a data set should reflect information about sample sizes, e.g. annual numbers of trips sampled for age as a proxy for effective sample size, or CVs calculated for annual abundance indices. For red snapper, this is done for length and age compositions, and for the headboat discards relative abundance index and the CVID index, but not for the commercial lines and headboat landings abundance indices, which were set to 0.2 for all years (for headboats this is at odds with the CVs used for the landings of this fleet in MCB runs of the BAM, where precision was assumed to improve substantially over time). This could lead to some bias in fitting the lines and headboat indices.

From experience in fitting such models, it has proved necessary to treat removals estimates as exact, even if obtained from surveys where sampling error may be high, otherwise convergence problems are experienced. For red snapper, this was achieved by inputting a small CV of 0.05.

The treating of key parameters such as selectivity and natural mortality as constant over time periods, treating uncertain estimates such as recreational catches as exact, and assuming errors in data are random when there are biases in data which may alter over time, can cause the model to provide fits to data that are much worse than suggested by input sample sizes and CVs, requiring iterative re-weighting. However, if the data sets are short, for example the CVID indices or certain blocks of length or age data, the model may not "see" the full extent of this additional process error and over-fit the data sets. It is notable that the BAM fits to headboat age compositions for many of the years are almost exact (e.g. 1981, 1982, 1984, 1985, 1986, 2008 and 2009 – years with comparatively large input sample sizes; Fig. 3 in AW report), which is nonetheless unexpected given the combination of sampling error, process error and ageing error. It is also notable that there are some years with large sample sizes but very poor fit, for example general recreational landings in the 2014 min-season and commercial handlines in 2007.

The base configuration of BAM from the Assessment Workshop was revised at the RW with corrected age compositions of the Chevron Trap survey. Results and diagnostics from the AW base model and the corrected base model ('newbase') were similar. The present CIE review of the BAM assessment is based on the Assessment Workshop report and the corrected base model.

Uncertainties due to imprecision in discards estimates
During the 2010-2014 moratorium, recreational discards estimates are one of the most important but also most uncertain sources of information for the assessment. Treating the estimates as exact by using an input CV of 0.05 appears especially problematic. The Review Panel requested an exploratory analysis that allowed error in the estimates of recreational discards, assuming the MRIP estimates of CV. Exploratory assessment models with more or less catch had similar estimates for the last 30 years (BAM runs S17–S20).

Choice of selectivity patterns and estimation of parameters
The AW made appropriate, pragmatic decisions on fishery selectivity, based on the evidence in the length and age composition data and other information available. However, treatment of landings and discards from a fishery as independent components with their own selectivity parameters estimated appeared unusual as I am more used to seeing selectivity fitted to the total catch, and then fitting a retention ogive as is done for the Gulf of Mexico red snapper assessment. The assessment team at the RW had specific arguments for their approach, but it might be useful to also try the alternative approach in future applications as the landings and discards sample data are not independent.

The relative length and age compositions in samples from each of the fishery-dependent and independent catches provides information on their relative selectivity patterns. In statistical models, it is advisable to fix the fishery selection at older ages in at least one significant fleet, often one considered to have asymptotic selectivity (least parameters), to avoid the model iterating towards steeply domed selectivity in all the fleets and a large unobserved ("cryptic") biomass of fish if there is no clear solution. The AW noted that an unconstrained parameter fit in BAM for fleets with domed selectivity iterated to a solution with almost zero selectivity at the older ages, and selectivity was therefore fixed above a given age.

In the BAM, commercial handlines and the chevron trap survey were assumed to have asymptotic selectivity, based mainly on consideration of depth range covered. A BAM sensitivity run with domed handline selectivity (BAM run S21) made little difference to the trends in fishing mortality and spawning output relative to reference points. However, as the CVID trap survey is also assumed to have asymptotic selectivity, it may not be sensible to allow only one of these two data sets to have domed selectivity if catches from both sources do have similar selectivity above a certain age. Another approach could be to fit domed for both, but explore different fixed parameters for fish above a given age.

Selectivity of the headboat fleet was assumed to be strongly dome-shaped, and the model estimated full selectivity at ages 3-4 and low selectivity of ages 10+. Selectivity of the general recreational fleet was also assumed to be dome shaped until 2010, with full selectivity at ages 3-4 and low selectivity of ages 10+. Results were not sensitive to how selectivity was estimated for ages 10+ (BAM run S31).

Since 2010 (during the moratorium, and the mini-seasons with 1-fish bag limit), selectivity of the general recreational fleet was assumed to be flat-topped, with full selection at ages 6+. It was argued that there may be targeting of large fish during the mini-seasons, shown by larger size compositions and older age compositions since 2010. The model was also unable to estimate lower selectivity at older ages. A problem is that recent strong year classes (age 1 in 2006-2008), for which estimates appear stable in retrospective analyses, are feeding progressively into the 5+ age groups from 2010 onwards, the period for which the BAM sees more adult fish and "wants" to estimate asymptotic selectivity for the general recreational fishery. The RW was not presented with any diagnostics (e.g. parameter correlation tables) that could show if the model has an issue estimating fully selected Fs in 2014 vs recruitment estimates for the strong year classes. A draft paper by Sauls *et al.* circulated by the Panel chair after the RW suggested that private boat anglers have been operating mainly in shallow water during the mini-season and not targeting large fish offshore as suggested to support the assumption of asymptotic selectivity for the general recreational fleet landings since 2010. Under the assumption of asymptotic selectivity, fishing mortality in the general recreational fleet is estimated to be increasing and is a major contribution to the conclusion that overfishing is occurring despite the conservation measures in place.

It is of some concern that the retrospective analysis (slide 102 on SEDAR41_RW_Day1corrected.pdf see Fig. 4.1.2 below) indicates a substantial upward adjustment of recent Fs with addition of 2014 data with corresponding downward adjustment of SSB, indicating that overfishing is occurring and is increasing since 2010. Assessments terminating in 2013 and earlier show recent Fs close to the F30% ref point (apical F values). The increasing F in the SEDAR 41 base model is mainly a feature of fish aged 4 and older. Therefore, there is a potential large uncertainty in the F estimates from the assessment including 2014 data. Some of the age composition data are very closely fitted in 2014 – the CVID comps (from the new base run with corrected age compositions) are fitted extremely closely in 2012 and 2014 and close in 2013, whilst the general recreational age comps are fitted very poorly in 2014 despite a very large sample size and may be an indication of problems with fitting the data for this fishery in 2014. It is noted that since 2012 the bulk of the data for the mini-seasons is from intensive State surveys using a variety of methods, and whilst these data may be more precise, the continuity with MRIP estimates is not clear. The Review Panel at the RW requested a sensitivity analysis in which selectivity of the recent general recreational fleet was assumed to be the same as the recent headboat fleet. The model fit to age composition data deteriorated, underestimating catch at older ages, but the perception of stock status was not changed. However, this may simply reflect the combined effect of other data and parameter estimates that push the model towards a solution with relatively large apical F in 2014.

The selectivity of the lines and trap catches remains a point of contention with red snapper fishermen who provided testimony at SEDAR41, also documented in SEDAR 41-RW6. They expressed concern that the BAM assessment underestimates the numbers of large, older red snappers. In their experience, these fish occur more frequently in midwater than is the case for smaller snappers, and therefore are less likely to enter traps, and also have behaviour and distribution that makes them less probable to be caught by commercial handlines, suggesting that all fisheries have domed selection. The scientific sampling of fishery catches shows that the incidence of large snappers is lowest in headboats operating inshore, highest in commercial lines operating in deeper water on average, and intermediate in recreational private and charter boats which typically operate in intermediate depths. The age composition of red snappers caught in the Chevron trap survey, which extends across a wide depth range, is considered to have an asymptotic selectivity. Broad spatial coverage of the commercial fishery and survey has been used by the DW and AW to justify asymptotic selectivity for these catches, but it is difficult to prove that the commercial fishery and Chevron trap survey have asymptotic selectivity based purely on model diagnostics or spatial fishery distribution. The Gulf of Mexico red snapper assessment includes an NMFS long line survey with assumed asymptotic selectivity and generates strongly dome-shaped selectivity in commercial line and longline fisheries. Studies are needed to provide independent data showing how red snapper behaviour affects the probability of encounter with a fishing operation or trap, and the probability of being caught when encountering the gear, to help define selectivity patterns and resolve the different perspectives on abundance of large snappers during the rebuilding period. Possible approaches could include the use of electronic data-storage tags to monitor daily depth distribution of fish of different sizes in different locations across the range of the stock.

The perception of current selectivity used to derive reference points and projections is conditional on recent fishing behavior, and projections of alternative management scenarios should consider alternative selectivity assumptions that are consistent with each scenario.

Fitting of relative abundance indices
Fishery CPUE indices when used on their own suggest a greater recent increase in stock biomass and lower mortality (BAM run S4) than when used with the CVID series. The Review Panel considered that the CVID fishery-independent index is informative and should be included in the assessment model despite the short time series from 2010. The decision to combine the Chevron Trap Survey and Video Survey is appropriate as the two series are not independent. An alternative model configuration that included 2005 – 2014 CVID data (BAM run S9) provided similar estimates as the base model historically but lower recent estimates of F, similar to the effect of up-weighting the 2010-2014 CVID data by a factor of 10 (BAM run S2).

ToR 3. Evaluate the assessment findings and consider the following:

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
b) Is the stock overfished? What information helps you reach this conclusion?
c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

a) The abundance, historical exploitation, and biomass estimates appear reliable and consistent with input data and population biological characteristics, although the ability of the BAM assessment to support status inferences in terms of probability of undergoing overfishing is likely to be degraded by data quality issues including increasing dependence on discard estimates and uncertainties around how selectivity has changed following the moratorium. Discarding has become more important since the inception of the moratorium and the mini-seasons with 1-fish per day limits on recreational fishing and small limits on commercial landings. The impact of this is explored in more detail below.

b) All sensitivity runs of the BAM assessment indicate that the stock remains severely overfished in terms of total annual egg production. In spite of very high abundance of young red snappers in recent years, the assessment indicates there are relatively low numbers of larger adult fish, which have relatively high fecundity and contribute disproportionately to total annual egg production. The ASPIC model reaches a different conclusion, suggesting that the stock has not been overfished since 2012. However, the consensus of the review panel and the AW was that the ASPIC model is not appropriate for this stock and proved difficult to fit, and that the BAM assessment more accurately represents the strong recruitment dynamics shown by the stock and the changing selectivity patterns in the fishery.

c) The baseline BAM, and all the sensitivity analyses and the Monte Carlo bootstrap uncertainty analysis indicate that overfishing is currently taking place and has been occurring more progressively since the moratorium and mini-seasons. The review panel could not find any evidence against the overfishing determination in the assessment but did have a number concerns that are discussed below. The panel also reflected on issues with using apical fishing mortality to monitor the impact of the fishery on the stock over time (see additional notes below). The perception of overfishing is driven by the addition of 2014 data which revises all the 2010-2013 estimates upwards from values from runs of the same model terminating in those years (Fig. 4.1.2). In other words, if the assessment had been done only to 2013, overfishing may not be evident. The recent increase in fishing mortality estimates is evident primarily at ages 4 and over. It is uncertain how selectivity is changing since the moratorium in 2010 and the mini-seasons since 2012, and fishery data quality must be deteriorating due to the resultant increase in discarding which is estimated with sampling error. There are also changes in recreational data collection involving intensive State surveys since 2012 to estimate private and charter recreational catches and discards, which is essentially a major change in design from just using MRIP in previous years. Additional sources of uncertainty for the model are the cessation of the fishery-dependent landings indices after 2009, concerns of the Panel over altered fishing behaviour on catchability in the headboat discards index during the moratorium, and the short duration of the CVID index series. All these factors are likely to degrade the ability of the model to detect changes in fishing mortality, and the sudden change in perception with the addition of 2014 data should be investigated further before concluding that overfishing is occurring and is increasing between 2010 and 2014. The main change between 2013 and 2014 was that landings and discards by the general recreational fleet were much higher in 2014 vs. 2013 by about 3.7 times for numbers landed and 3.4 times for discard numbers. Estimated increase in weight landed by the general recreational fleet was 3.4 times the 2013 landings. Fishing mortalities associated with general recreational landings and discards make up 78% of the 2014 apical F estimate (Table 14 of AW report). The mini-season in 2014 was longer than in previous years allowing additional landings.

d) There is not an informative stock-recruit curve for this stock. Recruitment deviations show a time-dependence, with strong year classes formed in the 1980s and since the mid 2000s, with weaker recruitment in the intervening years (Fig. 4.1.3). This leads to an apparent stock recruitment scatter with strong year classes near the lowest observed SSB, with no information on the shape of the true underlying relationship with other (e.g. environmental) factors removed. Any attempts to fit steepness resulted in a value tending towards 1.0 which is biologically impossible. BAM sensitivity run S11 with fixed steepness at 0.84 gave lower F estimates historically, especially from the mid-1960s to the mid-1980s, with correspondingly larger SSB, and increased F in 2014. The effect on perceived current overfished status of the spawning stock was unchanged. Better information on steepness might be obtained if environmental effects on recruitment can be disentangled from stock effects, given the large contrast in SSB over time.

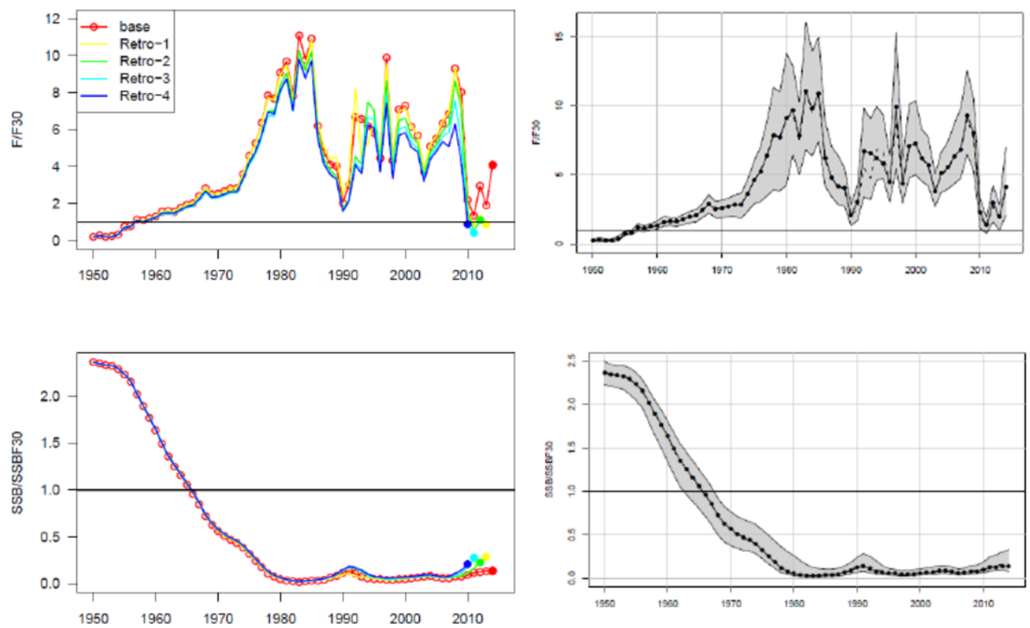

Fig. 4.1.2. Retrospective (left) and Monte Carlo bootstrap uncertainty analysis for SEDAR41 red snapper base model presented at the RW meeting (not the newbase model with corrected CVID age compositions).

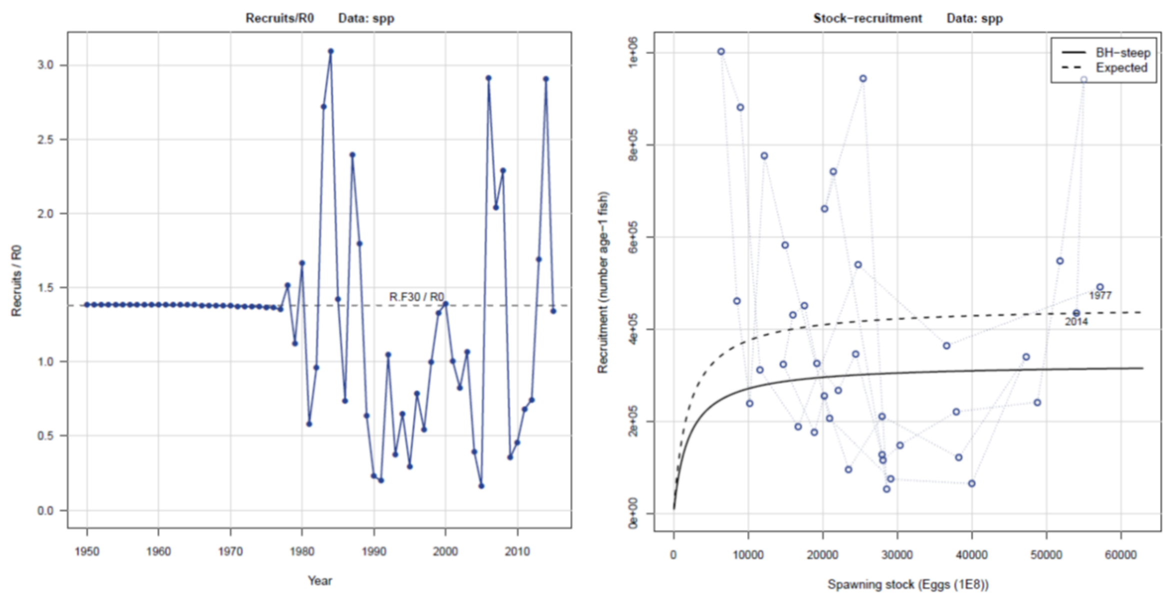

Fig.4.1. 3. Time series of relative recruitment estimates (left), and stock-recruit relationship (from new base run with corrected CVID age compositions).

e) The quantitative estimates of the status determination criteria for this stock appear fairly reliable in terms of biomass but possibly less so for recent fishing mortality rates. The strong adjustment of 2010-2013 fishing mortality from around F30% to well above F30%, when 2014 data are added (Fig. 4.1.2), due mainly to increased F at ages 4 and over, together with increased dependence on discards data, may indicate the model has difficulty in providing robust estimates of current fishing mortality and hence overfishing status. The possibility, of course, exists that the updated assessment is more correct than assessment runs terminating in 2013 or 2012, but I would recommend a more detailed investigation of what aspects of the data and parameter fitting are responsible for the large upward adjustment of F when 2014 data are added.

When severe management measures are introduced for any stock, leading to large changes in fishing patterns, fisher behaviour, catches and data quality, it is inevitable that the ability to accurately assess the stock will be degraded. This is especially the case in the absence of proven, long-term fishery independent survey data on relative abundance, and dependence on sampling surveys to estimate catches and catch composition. Any sudden changes in assessment results should be viewed critically. Currently, the only fishery independent data for red snapper is the CVID index which only extends back to 2010. The survey is only used from 2010, because the incidence of red snapper in earlier years was too low to allow meaningful calculation of an index using the modeling approach with covariates. In the current assessment, the annual red snapper discard CPUE the headboat fleet for 2005 to the present spans the fishery dependent indices in the earlier period up to 2009 and the CVID during the moratorium period. Although the assumption of constant catchability in the headboat discards index is questionable given the likely effects of the large changes in management controls since 2010, sensitivity runs removing the series (run S16) or only the data from 2010 onwards (run S12) had no discernable impact on the assessment.

The working document *SEDAR41_DW51_BallengerHistoricalChevronTrap* examines the utility of a longer series of trap data back to 1990, using a variety of model configurations to incorporate covariates. The raw data show increasing proportions of sets with red snapper and increasing numbers of fish.

Additional notes

The consistency of the stock status determinations for the combination of monitoring indices was evaluated through a series of sensitivity runs. These runs indicated that the determination of stock status was relatively insensitive to changes such as using the longer time series for the CVID (S9), removing the CVID (S4), up-weighting the fishery dependent indices (S3), dropping the headboat discard index for 2010 to the present (S12), dropping the headboat discard index altogether (S16), or only using the CVID (S23). All indices were well fit by the data, except for the headboat discard rate in the most recent years (Figure 13 of AW document).

All of these results suggest that the population trends in the model results probably have as much or more to do with the very close fit of the model to the landings, discard data, and associated age compositions as they do with the trends in abundance indices. CVs were set to 0.05 for the landings and discards, which seems unreasonably low for the MRIP estimates of the latter, but a higher CV of 0.20 for discards was investigated in MCB study and the results did not indicate a change in stock status from the base case.

The Review Panel looked at the robustness of the current metric for fishing mortality (apical F) for evaluating stock status. The following text is largely from the Panel report but is reproduced as it is informative. Evaluating trends in F over time requires a metric that is comparable among years and reflects exploitation across a range of ages. Apical F (maximum F at age, Figure 4.1.4a) is based on a different range of ages among years, because of changing fleet contributions and changes in fleet selectivity. Apical F also does not reflect F for partially selected ages.

Simple average F at age over a defined age range can reflect trends for similar ages (e.g., ages 2-3, ages 4+), and show different recent trends. During the moratorium, F remained low for ages 1-3, but more than tripled for ages 4+ (Fig. 4.1.4b).

Average F can be weighted by abundance at age or biomass at age to measure the average F exerted on the entire stock (Fig. 4.1.4c). With young ages typically having greater abundance, abundance weighted average F reflects patterns of F at young ages. Biomass peaks at different ages over the assessment time series (age-20 in 1950, age-2 in 2014), so biomass weighted average F reflects a varying age range.

Based on these metrics, the sharp increase in unweighted average F for 4+ year old fish in 2014 stands out as most unusual, whilst mean F for younger fish is at the low values estimated for the 1960s. Strong evidence is needed to explain why, under the current moratorium and mini-seasons, F on older snappers should have increased so dramatically in 2014, bearing in mind the retrospective upward adjustments in F in 2013 and previous recent years with addition of 2014 data.

Average F can also be weighted by exploitable abundance (the product of abundance at age and selectivity at age) or exploitable biomass (the product of biomass at age and selectivity at age) to measure the average F exerted on the exploitable stock (Fig. 4.1.4d). The two exploitable stock average F’s are similar, but the exploitable biomass weighted F reflects older ages (e.g., more than doubles during the moratorium), and the exploitable abundance weighted F reflects younger ages (e.g., remains low during the moratorium).

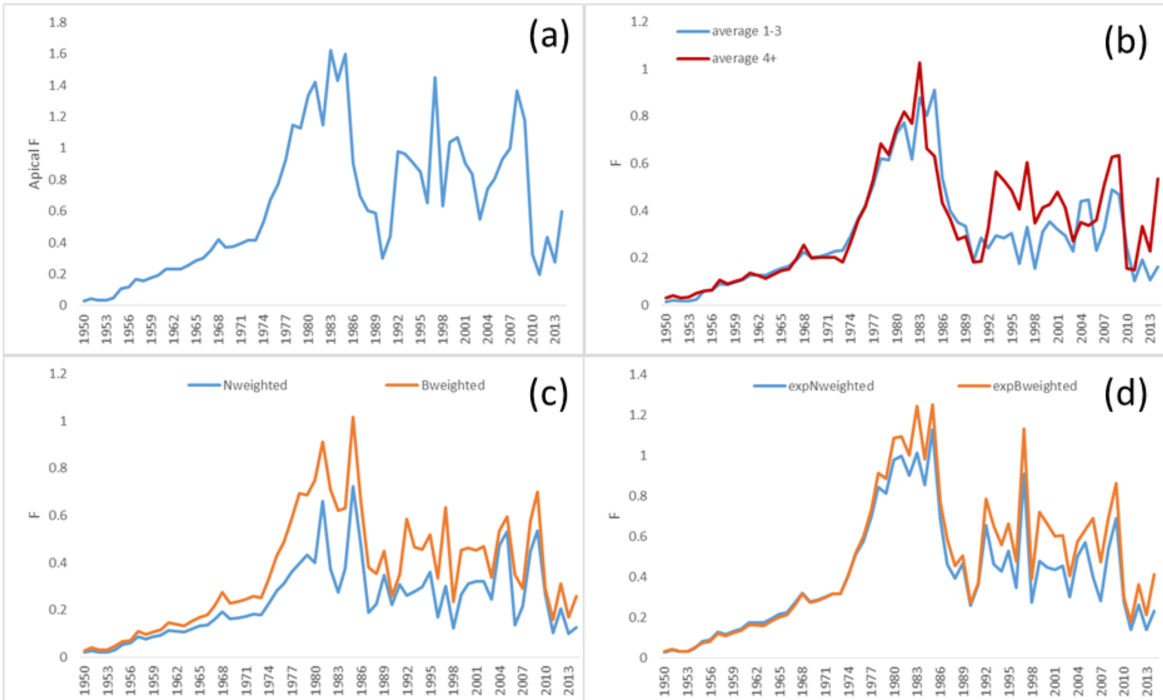

Fig. 4.1.4. Possible metrics of fishing mortality: (a) Apical fishing mortality from the BAM assessment; (b) unweighted average of F at age for the age ranges shown; (c) total F derived as mean of F-at-age weighted either by abundance (numbers) or biomass of fish; (d) average F weighted by exploitable abundance (the product of abundance at age and selectivity at age) or exploitable biomass (the product of biomass at age and selectivity at age).

The overfishing limit ($F_{30\%SPR}$) can be expressed in the same currency as the measure of F from the stock assessment. $F_{30\%}$ is currently expressed as Apical F, assuming the average selectivity for the last three years of the stock assessment, which peaks at age-5 (e.g., $F_{30\%}$ expressed as age-5 F is 0.15). The $F_{30\%SPR}$ is basically derived by scaling the combined fleet selectivity curve by a multiplier until the resultant F vector gives 30%SPR. Hence, if current F at age 5 (for example) is 2.8 times the value for F at age 5 at 30%SPR, the same will hold for any other ages or averages across ages (see table below). However, use of the weighted F metrics leads to different current status relative to $F_{30\%SPR}$ though for the current BAM base model, all metrics suggest overfishing. This conclusion is conditional on the base run of the BAM assessment presented at SEDAR41, and any biases in F estimates for recent years will translate into biases in stock status evaluation.

Metric	2012-2014 Geo.Mean	F30%	F/F30%
F(age-5)	0.43	0.15	2.8
F(ages 1-3)	0.15	0.06	2.7
F(age-4+)	0.35	0.12	2.8
F(Nwtd)	0.14	0.08	1.8
F(Bwtd)	0.24	0.11	2.1
F(expNwtd)	0.20	0.10	2.0
F(expBwtd)	0.31	0.12	2.5

ToR 4. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:

a) Are the methods consistent with accepted practices and available data?
b) Are the methods appropriate for the assessment model and outputs?
c) Are the results informative and robust, and are they useful to support inferences of probable future conditions?
d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

The projection method is consistent with those used widely in SEDAR assessments based on statistical models such as BAM and Stock Synthesis, and is consistent with the available data. A single selectivity curve was applied to calculate landings and one for discards, averaged across fleets using geometric mean F's from the last three years of the assessment period, similar to computation of $F_{30\%}$ benchmarks. Expected values of SSB (time of peak spawning), F, recruits, and removals were represented by deterministic projections using parameter estimates from the base run. These projections were built on the spawner-recruit relationship (with steepness $h = 0.99$) with bias correction, and were thus consistent with estimated benchmarks in the sense that long-term fishing at $F_{30\%}$ would result in $SSB_{30\%}$ on average.

The stock projections are appropriate for the BAM assessment model and outputs.

The results of the projections are informative and robust, and are useful to support inferences of probable future conditions. The projections provide the information needed to develop management advice, showing projections for F=0; F=Fcurrent (geometric mean of the last 3 years); F=F30%; F=Ftarget; F=Frebuild (max exploitation that rebuilds in greatest allowed time (2044)). An additional projection was carried out with F from discards only. Each projection shows the 10th and 90th percentiles of the replicate projections allowing an evaluation of the probability of overfishing occurring, or the stock being overfished, for each year in the rebuilding time frame up to 2044. The projections are robust in terms of propagating realistic levels of uncertainty from the accepted base model run.

Key uncertainties in the projections are acknowledged, discussed, and reflected in the projection results. The method used stochastic projections that extended the Monte Carlo/ Bootstrap (MCB) fits of the assessment model with added stochasticity in recruitment, and hence the propagation of uncertainty from the assessment into the projection period is internally consistent. The MCB runs included ranges of values of natural mortality, discard mortality and fecundity at age agreed by the assessment working group, together with bootstrap selection of data using well-justified error distributions and additional random process error in recruitment conditional on the fitted stock recruit pattern with steepness fixed at 0.99. Initial age structure at the start of 2015 was computed by the assessment model, and fishing rates for the projection started in 2017 following an initialisation period in 2015-2016 where fishing mortality rates were derived to represent the management measures in place. Projections assume the sharp increase in F up to 2014 and stock abundance at age consistent with this, and may be biased if the sharp retrospective upward revision of 2010-2013 F with addition of 2014 data is indicative of a biased model fit, which requires further investigation.

ToR 5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.

a) Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Uncertainties in the assessment were thoroughly explored through (1) a Monte Carlo and bootstrap (MCB) analysis; (2) sensitivity analyses around the base BAM run; (3) retrospective analysis; and (4) comparison with an alternative assessment model (ASPIC). In practice, the ASPIC model was not considered by the Review Panel and the Assessment Working group to be appropriate for red snapper, so it is difficult to use this to evaluate uncertainties in the BAM results. Overall, the DW and AW specified the most important sources of error and uncertainty in the sensitivity analysis and MCB for the BAM assessment. The implications of uncertainty in technical conclusions were clearly stated.

The MCB procedure involves parametric bootstrap resampling of abundance indices and fishery catches from error distributions specified by CVs. For multinomial age and length compositions, new distributions were drawn with replacement using the cell probabilities of the original data. For parameters that are constant in each assessment model realization, such as parameters of natural mortality, discard mortality, and fecundity, values were drawn using a Monte Carlo randomization process. Stock-recruit steepness was not included as an uncertain parameter and was held at 0.99 for all runs. Uncertainty in steepness was explored instead using sensitivity analyses. The MCB procedure allows exploration of the potential effects of errors in catch estimates that are treated as exact in the assessment, but are subject to sampling error according to CVs given by the Data Workshop, ex., discards estimates and recreational catch estimates. It is not the same as allowing for the errors when fitting the model, but provides useful guidance on uncertainties introduced by these sampling errors in terms of variance around management metrics such as biomass, fishing mortality and stock status (e.g. F/F30%).

In some cases, the error distributions for the bootstrap were inferences rather than direct estimates and will be biased to some extent. An example is the assumption of a declining CV in three blocks from 1981 onwards to reflect improved quality of the headboat data. The way the bootstrap is applied also was varied between catch data sets - deviations for the headboat landings were applied equally to all years in the early part of the series but independently by year thereafter. In practice, errors in these subsequent years, where only around half of the logbooks were returned (Fig.4.1.1, reproduced from SEDAR41_DW46), may also be correlated between years if non-response is related to size of catch, but no information was available on this and the assumptions made about independence of errors in the MCB analysis were adequate although potentially underestimating uncertainty.

The MCB estimates of uncertainty for the base BAM presented initially at the RW are shown above in Fig. 4.1.2 and indicate relatively large uncertainties in historical fishing mortality, but relatively low uncertainty in biomass. It indicates a very high probability that the stock is overfished and is undergoing overfishing.

The sensitivity runs provide additional support to the MCB that there is a very high probability that the stock is overfished and is undergoing overfishing (Fig. 4.1.5). The majority of the runs indicate higher values of SSB/SSBF30% and lower values of F/F30% than the base model, with the only runs suggesting significantly worse status being those with steepness = 0.84, low M, low catch values and the continuity BAM run using the SEDAR24 settings. The most extreme improvement in stock status was in using the fishery independent indices on their own (i.e. excluding the CVID index) where F/F30% approaches 1.0. In this run, the only relative abundance indices from 2010 onwards are the headboat discards index, which may have some biases due to changes in fishing patterns and fisher behavior since the moratorium.

The Review Panel noted that the sensitivity testing by alternating one factor at a time may not fully reflect the uncertainty in model outputs from a complex model such as BAM, where there are a large number of parameters where many are likely to be correlated. It was suggested that global sensitivity analysis (Saltelli *et al.* 2008) may be used to untangle the contribution of single factors/parameters and interactions between parameters to the overall variability in model output.

Model uncertainty was mainly explored by running an alternative Stock Production Model Incorporating Covariates (ASPIC software of Prager, Version 7.03, 2005) that relies on length-age aggregated catch and CPUE indices, with no compositional catch being included. The ASPIC runs resulted in biomass estimates above B_{MSY} and estimates of F below F_{MSY}, and hence do not place the stock in the "overfished-overfishing" category. The difference between the ASPIC and the BAM results can however be explained by the fact that ASPIC does not take into account the age-structure of the catches and the stock. Thus, a biomass made up largely by recruits can result in a stock status of not overfished and not overfishing. The BAM base configuration is therefore considered to provide the most appropriate basis for status determination, despite many sources of uncertainty.

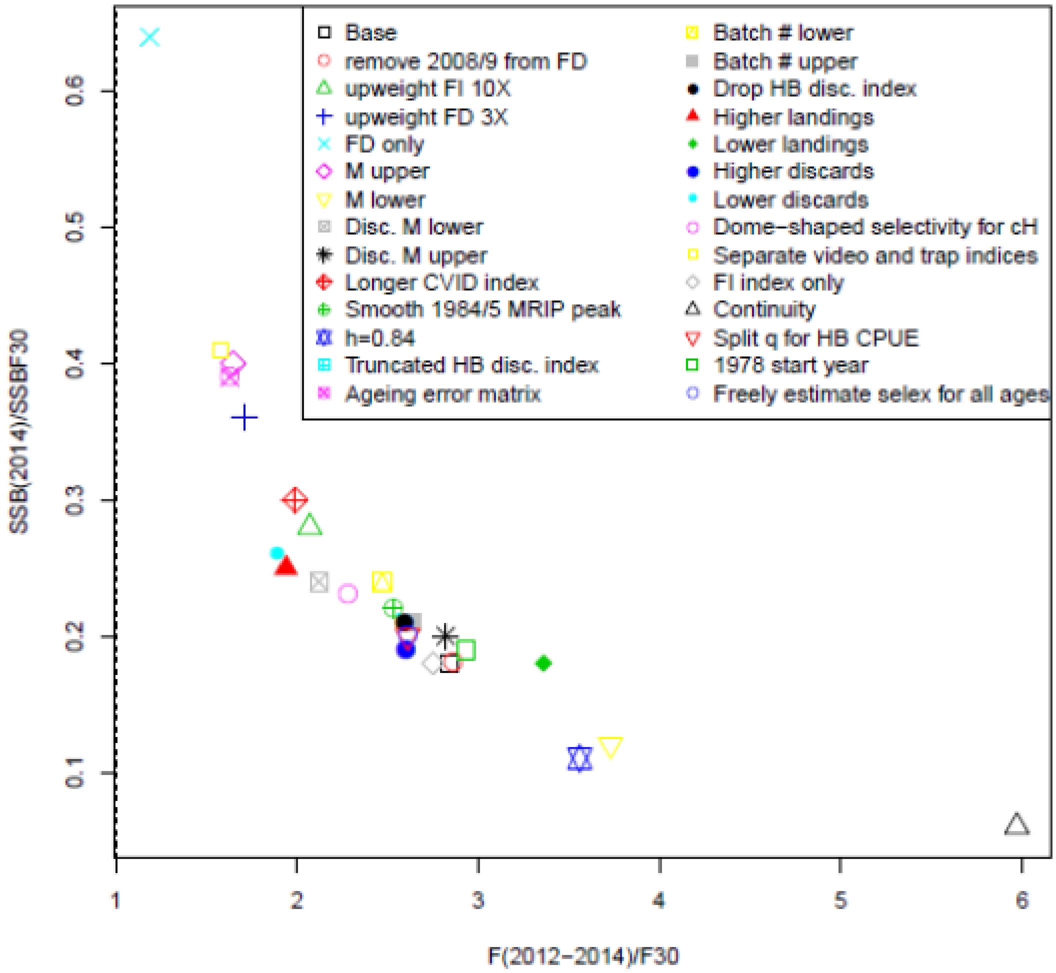

Fig. 4.1.5. Red Snapper: Phase plot of current stock status for sensitivity runs presented at the RW. Does not include additional sensitivity runs requested by the Review Panel.

ToR 6 Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

a) Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.

b) Provide recommendations on possible ways to improve the SEDAR process.

Research and monitoring

The Data Workshop and Assessment Workshops provided extensive lists of recommendations for research and monitoring that could improve the reliability of, and information provided by, future assessments (see Appendix 4). These are largely sensible, but there is no evaluation of which suggestions would have the greatest impact on improving the quality of the assessment. It is also important to consider the costs of the research or additional / new data collection in relation to the benefits in terms of improved accuracy of the assessments. As an independent reviewer, I have no information on this which makes it difficult to suggest any prioritization of the existing recommendations and can only indicate what I feel would be useful to resolve important uncertainties in the assessment. Some ICES expert groups that I am involved in (e.g. Planning Group on Data Needs for Assessment and Advice - http://www.ices.dk/community/groups/Pages/PGDATA.aspx) are promoting a cost-benefit framework running alongside the existing quality assurance frameworks to encourage the use of objective approaches to evaluating and prioritizing data needs according to impact on assessments and advice relative to costs of data collection. The handling of research and data collection recommendations within the SEDAR process could be improved by imposing a requirement to clearly describe how the research or data would improve the assessments and advice, and to evaluate the relative cost (even if in broad cost categories or equivalent metrics)

Some potential areas of research or enhanced data collection that could be useful to better understand and reduce uncertainties in the assessment include:

- Studies are needed to help resolve the size and age selectivity of the fisheries and fishery-independent survey independently of the BAM assessment, and how selectivity in fisheries is affected by management measures, particularly the commercial lines and general recreational fleet. Methods such as electronic data storage tagging in different areas could potentially provide useful information on size-related differences in fish behavior

- Further collaboration between MRIP and State agencies, building on the good work done in recent years, to develop a coherent, design-based and cost-efficient approach for the areas covered by the stock and that ensures continuity of total catch estimates over time.

- Further evaluation of the ability of the CVID survey to provide indices proportional to stock size and consider if the series can be permanently extended back in time as suggested in SEDAR41_DW51.

- Studies to better understand selectivity, catchability, and changes in these induced by fisher behavior in response to management measures as well as other factors such as hook competition with competing species.

- Continued age-calibration workshops and use of a reference collection of age material that is inserted at random and anonymously into annual age collections to monitor consistency of age readings.

Improving the SEDAR process
In my experience of attending several SEDAR review workshops as a CIE reviewer, the process is very effective in thoroughly reviewing the assessments and the basis for management advice. Presentations given by the lead stock assessors and colleagues are usually extremely thorough, easy to understand and well-focused on key issues, which was the case at SEDAR 41. The entire process involves a substantial investment of staff time and funds to operate the data and assessment workshops and associated preparation and reporting, but the benefit is a very thorough compilation and evaluation of available data and modelling approaches. The involvement of people from the commercial and recreational fishing community at the RW and earlier stages is an important part of the process to hear their views based on their own observations on the water, as well as to ensure transparency in the assessment and review process. Inevitably there are differences of opinion, which

in some cases can be a result of the different scales at which fishermen and scientists observe the ocean, but it is a learning process on both sides that must be maintained. It is hard to see how the overall process can be significantly improved.

One area of improvement already discussed above is that the handling of research and data collection recommendations within the SEDAR process could be improved by imposing a requirement to clearly describe how the research or data would improve the assessments and advice, and to evaluate the relative cost (even if in broad cost categories or equivalent metrics).

It is clearly important that SEDAR continues to bring in reviewers with the necessary experience and technical competence in data collection and assessment methods. Many assessments are now being conducted with complex statistical models such as BAM and Stock Synthesis, and it can be difficult to fully understand how the parameter estimation is driving a particular model result, particularly where there are multiple uses of the same data sets (e.g. catches and catch rates, or lengths and ages derived from the same data sets), correlation structures within data sets (e.g. multinomial distributions), and correlations between parameters. Uncertainty in assessment results is usually investigated using sensitivity analysis and MCB or MCMC analysis, but to varying degrees and the number of sensitivity runs can vary widely depending on the assessment working group and the perceived issues with the assessment. Considerable time can be spent producing additional sensitivity runs requested by the review panel. The SEDAR 41 review panel noted that such runs done one at a time do not fully reflect the true sensitivity of the model, which is built from multiple choices of data and model settings. As proposed by the Review Panel, the review process could be greatly helped by use of more advanced "global sensitivity analysis" methods for statistical models, which more comprehensively explore the sensitivity of particular outputs to the combination of different data and model settings. An investment to develop such methods for the different types of models used in SEDAR and other review processes would have considerable benefits for understanding the quality of assessments and advice and targeting efforts at improving model inputs and settings that have greatest impact on the results.

ToR 7. Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.

I consider that, with some caveats about the large change in perception of fishing mortality with addition of 2014 data, the BAM assessment for red snapper constitutes the best scientific information available, and fulfils the following criteria:

Relevance: The SEDAR 41 assessment is highly relevant as the red snapper stock is depleted and undergoing rebuilding under a moratorium with limited landings permitted and most catches being discarded. The data and assessment provide the best means of establishing the rate of recovery of the stock, determining if measures are preventing overfishing, and providing information that can be used to adjust management actions where appropriate.

Inclusiveness: The SEDAR 41 assessment includes all data that have been quality assured and proved adequate for use in the assessment. This includes data from State, as well as Federal, sampling schemes where needed, for example to estimate discards during the mini-season where MRIP sampling is too limited for such a short season length.

Objectivity: The SEDAR 41 BAM is a highly objective procedure based on well-tested statistical modeling principles, and using data sets and assumptions that have been rigorously documented and reviewed through the SEDAR data, assessment and peer-review process. Where fully objective decisions are difficult to make, such as some decisions on scenarios for historic catches where evidence is lacking, the uncertainties around the decisions made have been explored and included in sensitivity analyses, and the Monte Carlo Bootstrap evaluation of assessment uncertainty.

Transparency: All outputs of the data, assessment and review workshops in SEDAR 41 are fully documented and publicly available. The discussions at the review workshop are also recorded for record. All data sets are thoroughly explored and the quality of data on which the assessment is based is documented and transparent, as are all decisions related to the choice of assessment model, how it is implemented, and the results of the base run and sensitivity and uncertainty analyses.

Timeliness: The SEDAR process in general is arranged to provide timely fishery management advice where it is needed, and to ensure that assessments are benchmarked and reviewed at appropriate intervals.

Verification: The SEDAR 41 assessment process and deliverables comply with legal requirements under the Magnuson Stevens Act (2007) for developing and monitoring of fishery management plans and providing information on stock status.

Validation. The SEDAR 41 process is designed to meet the needs of fishery managers for peer-reviewed stock assessments and associated advice on stock status and future catches, and the process is open and fully transparent to the fishery managers and to stakeholders from commercial and recreational fisheries, conservation groups or others with a stake in the outcomes and who have opportunity to give their views on record.

Peer review: the SEDAR 41 process includes full peer-review by experts appointed by the Center for Independent Experts (CIE) and by reviewers from the SAFMC SSC. The review panel report and the independent CIE reviews are publicly available.

ToR 8. Compare and contrast assessment uncertainties between the Gulf of Mexico and South Atlantic stocks.

There appears to be little information to distinguish Red Snapper in the South Atlantic and Gulf of Mexico as being separate biological stocks. Tagging studies have indicated strong site fidelity so there may be strong metapopulation dynamics within and between the areas.

The SEDAR41-RD61 (Stock Assessment of Red Snapper in the Gulf of Mexico 1872-2013, with provisional 2014 Landings: SEDAR Update Assessment) provides an update of the SEDAR 31 benchmark assessment of red snapper in the U.S. Gulf of Mexico using updated data inputs through 2013, and provisional 2014 landings. Except as otherwise noted, the specifications of the model and data streams are identical to those of the base models identified in the SEDAR 31 final report. An important change incorporated in the update assessment is an adjustment to the recreational catch and discard estimates produced by the Marine Recreational Information Program (MRIP). The stock assessment was accepted (with 2014 provisional landings) by the GMFMC SSC at a special meeting on Feb 19, 2015.

The GOM assessment differs from the South Atlantic assessment in having a greater quantity of fishery-independent survey indices of abundance. These include the SeaMap reef video survey, SeaMap larval survey, SeaMap groundfish survey, the NMFS bottom longline survey and the Artificial Reef Survey using ROVs and stationary cameras. The report notes that "there continues to be little evidence from the NMFS fishery-independent survey of a large biomass in Gulf of Mexico waters beyond 150 m, because peak distributions of red snapper occurred between 50-100 m. However, the survey tends to capture a greater proportion of older fish than any of the fisheries (including commercial longline). This result suggests that older biomass is relatively less vulnerable to the fisheries (i.e., a dome-shaped selection/availability pattern)".

The GOM assessment uses Stock Synthesis which is directly analogous to the BAM, and the model is structured as follows:

- Age structured model: ages 0 to 20+, 1872-2013 (provisional landings in 2014).
- 2-stock/region model : East and West of the Mississippi River
- Time-varying recruitment in each region allowing a change in average productivity in recent years (1984-2013)
- Time-varying selectivity to account for implementation of IFQ program and circle hooks
- Time-varying retention to account for changes in size limits and IFQ
- Time-varying discard mortality to account for venting

There are 14 directed and by-catch fleet definitions, 8 fishery-dependent relative abundance indices (4 per area), and 11 fishery independent series across the two areas. For all of the fleets and surveys except for the NMFS bottom longline survey, age-specific selectivity parameters were specified for each age using a random walk approach. Discards are modelled using a retention ogive rather than using separate discards and landings selectivity as in the SA snapper assessment. The assessment is much more complex than for South Atlantic red snapper, uses more data sets and has longer continuous time series of abundance indices (e.g. from 1982 for the trawl survey and from 1986 for the longline survey). As with SA red snapper, all fishery removals data are treated as exact by specifying a small CV of 0.05. Likelihood profile in stock recruit steepness tended towards 1.0 as in South Atlantic red snapper, and steepness was fixed at 0.99.

The two red snapper populations show qualitatively similar trends over time but with different timing in growth of fishing mortality (Fig. 4.1.6). A progressive increase in fishing mortality occurred in the GOM from around 1950 to 1970, causing a sharp decline in SSB over the same period reaching low values by 1980. In the South Atlantic, fishing mortality also increased from around 1950, but peaked later than in GOM, in the 1980s. The South Atlantic snapper SSB also declined sharply from 1950, bottoming out around 1980, but there is no prior data to indicate trends prior to 1950. Recruitment trends differ in the two stocks, with a period of poor recruitment in GOM in the 1980s at a time when recruitment was relatively high in the SA stock. The period of poor recruitment in the SA stock in the 1990s is not observed in the GOM stock. The current stock status is different from the two assessments, noting that the F reference point differs slightly (FSPR26% in GOM and FSPR30% in SA). Note also that the basis of F is not the same – the SA stock assessment reports apical F and the GOM stock gives an exploitation rate calculated as total numbers killed for all fleets divided by total stock numbers. Both stocks are estimated to be overfished although the GOM stock is recovering faster. The large reduction in F in recent years in the GOM stock brings exploitation status close to the reference point and overfishing is not occurring. In the SA stock, F appears to have been increasing and the SEDAR 41 assessment indicates that overfishing is taking place.

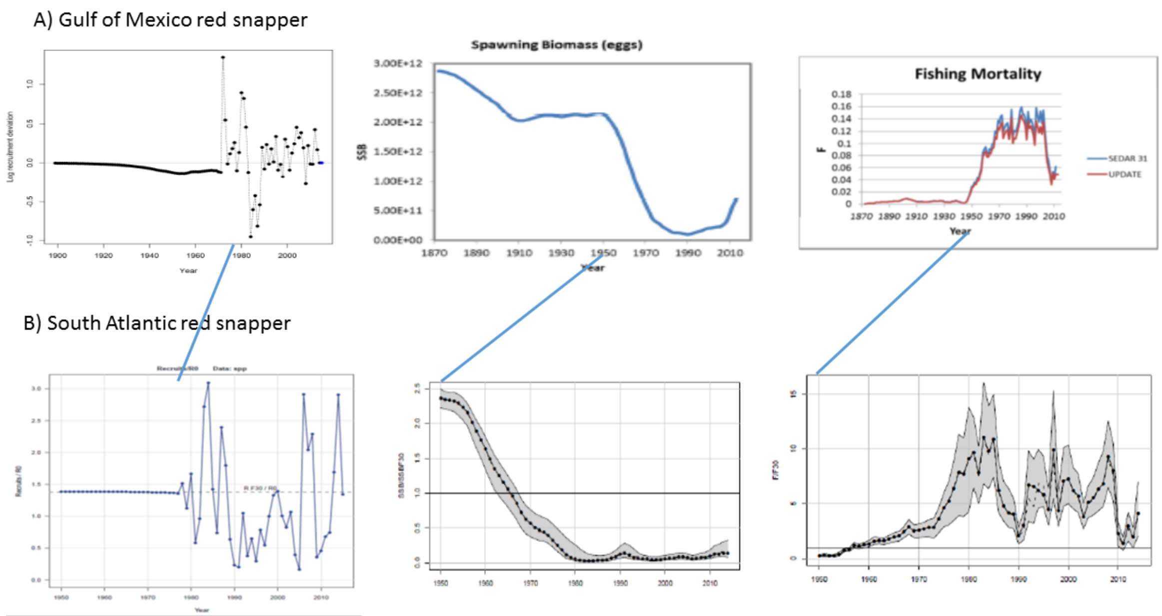

Fig. 4.1.6. Comparison between trends in recruitment deviations, spawning stock biomass (as eggs) and fishing mortality in the most recent Gulf of Mexico red snapper assessment and the SEDAR 41 South Atlantic red snapper assessment. Lines connect the same year in both assessments.

The SEDAR 41 assessment of SA red snapper conducted a much more comprehensive analysis of uncertainty using MCB analysis and a wide range of sensitivity runs. Uncertainty in the GOM assessment was evaluated mainly from asymptotic standard errors for each parameter calculated by inverting the Hessian matrix (i.e., the matrix of second derivatives) after the model fitting process. Asymptotic standard errors provide a minimum estimate of uncertainty in parameter values. Stability of convergence was examined using jitter analysis. Several sensitivity runs were conducted on the GOM assessment to examine high and low M values and upweighting of indices, and extensive comparisons were made to examine the effect of the MRFSS-MRIP calibrations (documented in Appendix 1 of the update assessment report). The calibration had little effect on time series of SSB relative to unfished condition and fishing mortality, and modest effect on recruitment. Estimates of OFL and ABC were more sensitive to the changes in recreational catches due to the MRIP calibration.

It can be concluded that comparison of uncertainties in the two stock assessments is difficult to make due to the more limited exploration of uncertainty in the GOM assessment. However, there are more extensive data sets in the GOM assessment including more and longer fishery-independent abundance indices that in most (though not all) cases appear to be reasonably well fitted by the model, but an additional complexity of a 2-area assessment. The general stock trends are qualitatively similar in the two stocks with the exception that the GOM model indicates that overfishing is not occurring. The result in terms of recent F relative to the F reference point is similar to the SA assessment terminating in 2013, before addition of the 2014 data.

ToR 9. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

As discussed in previous sections, a major issue with this assessment going forward into the future is deterioration in the quality of fishery-dependent data caused by the moratorium and other subsequent measures, which will have caused discarding of a larger part of the catch, changes in fisher behavior during and outside the mini-seasons, and changes in catchability and selectivity. It is essential that efforts are focused on obtaining the most reliable data on fishery removals and compositions that is compatible with available resources, and to obtain information needed to understand the causes of observed changes in fishery data.

At the same time, the only fishery-independent abundance index is the CVID trap/video survey and further work is needed to establish the ability of this survey to provide indices proportional to abundance and to confirm the selection pattern.

Given the increasing uncertainties in data and fishery selection patterns, and the patchiness of historical data, estimation of large numbers of parameters in the BAM may not be sustainable and more parsimonious models should also be explored that can take advantage of the age structure information so that year classes can be tracked and their future impact forecasted. Further investigations are needed to understand why addition of 2014 data caused a retrospective upward adjustment of Fs from 2010 – 2013 to values that indicate overfishing is taking place – this may be related to aspects of the model fitting process that are trying to interpret the recent strong year classes, as well as the related appearance of fish of these year classes in recent year's samples at ages 4+ that influence the choice and fitting of selectivity models for the fisheries since the moratorium started in 2010.

The Review Panel recommended the use of an assessment model (XSAM - Sondre Aanes, Norwegian Computing Center) recently developed in Norway that can explicitly account for the correlation structures in multinomial sample data that reduce effective sample sizes, and recently used in an ICES benchmark assessment for Norwegian Spring Spawning Herring, and approved as the standard assessment model.

ToR 10. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. Develop a list of tasks to be completed following the workshop. Complete and submit the Peer Review Summary Report in accordance with the project guidelines.

At the time of submission of the present report, the Peer Review Summary requested by the ToR was not complete.

References

Francis, RIC. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68: 1124-1138.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. Global Sensitivity Analysis - The Primer. John Wiley & Sons. Chichester.

Sauls, B.J., Cody, R.P., Strelcheck, A.J. (draft report provided by RW chair after RW). Creel survey methods to estimate red snapper (*Lutjanus campechanus*) landings during short season openings for a high-effort recreational fishery managed with a small annual catch limit.

Stephens, A., and MacCall, A. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research, 70: 299-310.

4.2 South Atlantic Gray Trigger Fish

ToR 1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:

a) Are data decisions made by the DW and AW sound and robust?
b) Are data uncertainties acknowledged, reported, and within the normal or expected levels?
c) Are data properly applied within the assessment model?
d) Are data input series reliable and sufficient to support the assessment approach and findings?

A general statement on (a) – (d) is given below, followed by some supporting observations.

(a) Data decisions made by the DW and AW were sound and robust. The development of input data and parameters for the BAM and ASPIC models required an extremely thorough compilation and evaluation of all available data at the DW. Modifications made subsequently by the AW were fully explained.

(b) Data uncertainties were acknowledged, reported, and were within the normal or expected levels, where this could be ascertained from information provided to the RW. The data varied widely in coverage and quality, and were often heavily manipulated and standardized to try and develop coherent time series from diverse data sources of differing designs, coverage and accuracy. The combined data will have biases that in some cases are poorly understood especially in earlier years of the time series. All decisions made by the DW and AW in compiling data were explained and justified in detail. Data quality metrics were provided by the DW in terms of numbers of samples, CVs, or alternative plausible data series or biological parameter values. These were used by the AW to weight data series in the assessment model, estimate the uncertainty in the assessment results using the Monte Carlo / bootstrap method, or to explore the sensitivity of the assessment to data decisions and uncertainty.

(c) The data were properly applied within the assessment model, though with some caveats about use of age and length data from the same sources in the same year. Any issues with application of the data such as time periods for fitting, use of length and age data from the same sampling schemes, or weighting of data according to data quality metrics, were explored at the SEDAR-41 RW if not previously evaluated by the DW and AW.

(d) Data input series were applied if considered reliable and sufficient to support the assessment approach and findings. Reliability and sufficiency were evaluated based on a-priori criteria where possible, supported by data quality metrics such as numbers of samples or CVs and by model fits. The assessment is supported primarily by a wide range of fishery-dependent data covering landings and discards, and therefore is heavily driven by these data and assumptions related to their reliability and use. An additional fishery-independent trap and (from 2010) video survey provides abundance indices from 1995 to present. The Review Panel was mostly concerned about how the fishery-dependent and fishery-independent data were treated in the BAM assessment than with the inherent reliability of the data, leading to the decision to put the assessment on hold pending a response from the AW as a whole to suggestions made by the Review Panel. This is explained under Term of Reference 2.

Additional supporting comments

Life history parameters

Life history data and assumptions used in the Gray Triggerfish assessment include stock structure, reproductive biology and natural mortality. An age-dependent, year-invariant estimate of M was determined by a meta-analysis approach using growth parameters and maximum observed age. Reproductive biology was included in the model by computing total annual egg production at age based on maturity, length, number of batches and batch fecundity, allowing the effect of age structure on reproductive output to be reflected in setting SSB reference points and stock status. Inter-annual variation in fecundity, a possible source of uncertainty, was not able to be included as historical information was not available.

Age in Gray Triggerfish is based on increments in dorsal spines (converted to calendar years), and those results are used in the BAM. Gray Triggerfish age readings show a broad distribution of length at age relative to the annual growth increment in length (Fig.4.2.1 – from presentation to SEDAR-41 RW), which in combination with selectivity assumptions will affect the ability for BAM to estimate annual age compositions through fit to sample length compositions.

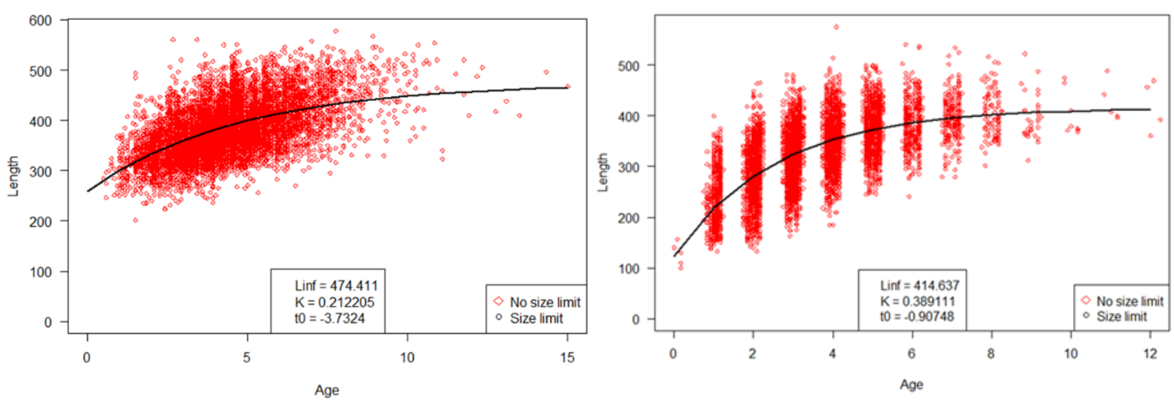

Fig.4.2.1. Left: fishery dependent length-age data for Gray Triggerfish; right: fishery-independent (CVID trap) data. From Gray Triggerfish presentation to SEDAR-41 RW

Fishery removals

The DW and AW made sound and robust decisions on fishery removals data, and the uncertainty in these (which is largest for discards estimates), and the data were used appropriately in the assessment, using only those data considered reliable enough for this purpose.

Reconstruction of a historical series of commercial and recreational fishery removals – landings and dead discards – was made back to 1988 for the assessment. This required a large number of decisions to impute missing values or to calibrate data series where design has changed, particularly for the change from the NMFS Marine Recreational Fisheries Statistics Survey (MRFSS) surveys (1981 to 2003) to the Marine Recreational Information Program (MRIP: 2004 to present), and for developing discards time series.

Landings of commercial handline fleets have improved in accuracy over time, and the DW proposed CVs that could be used for MCB uncertainty analysis in the assessment. Recreational landings of headboats are estimated from the Southeast Region Headboat Survey (SRHS) log book scheme, which has improved in quality over time due to introduction of mandatory reporting in 1996 and improved logbook supply from 2008 onwards. As shown for the red snapper assessment review in Section 4.1 above, headboat logbook returns had a roughly 50% non-response rate up to the mid-2000s which then increased to almost 100% in recent years, which may impact data, such as relative abundance indices, if the previous non-respondents had different catches or CPUE compared with respondents.

Private boat and charter boat landings since early 1980s were estimated from MRFSS/MRIP, which has a robust and peer-reviewed statistical design with improved design and precision over time, and for which CVs are estimated directly based on efficient estimators.

Discards estimates are inherently less reliable than landings for both the commercial and recreational fleets, and for commercial handlines involved extrapolating observations for 2001-2011 to other years back to 1988. Separate discards estimates for the open and closed seasons since 2012 were made for this fishery based on effort. Recreational head boat discards estimates are available from SRHS log books since 2004, but for previous years back to 1988 are inferred using MRIP charter boat data adjusted using ratios of SRHS to MRIP estimates for 2004-2013. Gray triggerfish discards estimates from SRHS and MRIP are self-reported and are not verified. All these uncertainties and data manipulations may introduce error in the time series.

Overall, the DW and AW made the best possible evaluation of fishery removals data and their uncertainties.

Length and age compositions
In comparison with red snapper, gray triggerfish has longer continuous series of length sampling from commercial handlines, headboat fishery landings and CVID survey, with relatively large sample sizes especially for headboat sampling. Discard lengths are only available from headboats since 2005, and age data are only available for commercial handlines and headboat catches from the early 2000s, but from 1991 for the CVID survey.

The CVID age data for Gray Triggerfish were found towards the end of the review meeting to have not been converted to calendar ages. Revised data were provided, and the assessment was rerun (see ToR 2 &3 for elaboration on this). The changes affected years 1991-2007 for which ages were shifted up by one year, whilst for 2008 – 2014, the age compositions were largely unchanged. The difference in magnitude of the change over time results from a change in sampling protocol. From 1991 to 2007, there were target numbers of age samples per length bin, and the sample requirements were typically collected early in the sampling season, requiring conversion to calendar year by adding one to the annuli count an age. Sampling after 2007 was based on a random sampling design. Thus, post 2007, fish retained for age determination were distributed throughout the entire sampling season, requiring less of a correction. The change in protocol suggests that age compositions up to 2007 may have variable biases related to non-random sampling and concentration of age sampling early in the sampling season.

For Red Snapper, the SEDAR-41 AW used age composition data in preference to length composition data in BAM where both data exist, but for Gray Triggerfish the AW fitted both length and age compositions for headboats and the CVID survey, which will result in some over-weighting of composition data from the early 2000s onwards.

There is a broad length at age distribution relative to annual growth increment for many of the age classes making up a large portion of fishery-dependent and fishery-independent data (see Fig. 4.2.1 above), which will affect the ability of BAM to estimate annual age compositions through fit to length composition sample data. The Review Panel requested a sensitivity run of BAM with length data omitted where age data were available. This resulted in a deterioration in fit to some age composition data. This could relate to adequacy of sampling for age, errors in ageing, correlations between length and age data collected from the same samples, the use of length data to weight the age compositions in each length class, and changes in the relative weighting of different fishery dependent and independent data sets (noting that the trap survey is upweighted by a factor of 6 in the base model). These issues need to be examined by the AW.

Relative abundance indices

Overall, the DW and AW mostly made sound and robust decisions on most of the relative abundance indices and the uncertainty in these, and the data were used appropriately in the assessment, using only those data considered reliable enough for this purpose (however see comments on data weighting later in review). Only one abundance index series was used in the BAM, the chevron trap abundance index combined with video records from same survey (SERFS combined video and trap - CVID - 2010-2014), which covers the depth ranges where gray triggerfish occur. A range of fishery-dependent indices were included in the ASPIC model, derived from the commercial handline landings, headboat landings, and MRIP landings data. The recreational abundance indices were converted to weight using smoothed mean weights, and filtered and standardized using approaches similar to those used for red snapper although a different approach was adopted to select trips where gray triggerfish could be caught (see DW report).

The general approaches to analysis of the data appear appropriate and are widely used for such data in assessments in the USA. The standardized index for the fisheries-independent survey based on a zero-inflated model accounts for yearly shifts in sampling distributions relative to covariates that affect catch rates for gray triggerfish. The *SEDAR41_DW52* by Ballinger et al. explains the use of the trap survey for gray trigger fish and how the data were analysed. The proportion of trap operations with gray triggerfish catches was very low in 1990 (0.11), increased to 0.45 in 1991, then in most years varied from 0.2 – 0.4. The first year (1990) of the survey, which had a very low abundance index, was after hurricane Hugo, and may have experienced drastic lower catching efficiency due to strong habitat disturbances. The Review Panel recommended that the inclusion of this first year be reconsidered. The final CVID model used soak time as an offset and fitted covariates latitude, depth, bottom temperature and day of year. As with red snapper, catch rates increased with increasing bottom temperature over most of the temperature range.

A paper by Bachelor et al. (2013; provided as SEDAR41_RD80) based on analysis of the SERFS trap survey data shows that catches of gray triggerfish plateaued once a moderate number of total individuals of all species were already caught in a trap, indicating interactions with other species in the trap and a saturation effect where fish are much less likely to enter a trap once it begins filling up. Discussion with an expert on this survey outside of the RW indicated that gray triggerfish are relatively aggressive towards other individuals around the trap. Saturation effects were not included in the analysis of the data to provide the abundance indices for SEDAR41.

ToR 2. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:

a) Are methods scientifically sound and robust?

b) Are assessment models configured properly and used consistent with standard practices?
c) Are the methods appropriate for the available data?

A general statement on (a) – (c) is given below, followed by some supporting observations.

a) The BAM stock assessment method used as the primary assessment tool for South Atlantic gray triggerfish is scientifically sound and robust, and is based on well tried and tested integrated assessment models such as Stock Synthesis, and has been applied to a number of stocks in the USA. The ASPIC production model is also well tested worldwide, but the conditions under which it can give reliable results are more limited as it does not explicitly represent year class dynamics and fishery selectivity, and requires sufficient contrast in data to reliably estimate model parameters, which is not the case for gray triggerfish. Approaches to investigate uncertainty, including retrospective analysis, sensitivity analysis, and Monte Carlo bootstrap analysis were all scientifically sound and robust.

b) The BAM and ASPIC assessment models are configured properly and their use is consistent with standard practices.

c) The methods used are appropriate for the available data. In particular, the patchy distribution of length and age sampling across time and fleets, and the variable sampling rates where sampling occurs, require a statistical modeling framework such as BAM to fit the underlying population model to the available observations in a statistically sound way. A downside is that patchy occurrence of short fishery composition series with very variable sampling rates can cause the model to over-weight and over-fit some series and years within series.

Additional comments are given below:

ASPIC assessment

The BAM is inherently a superior model to the ASPIC production model as it makes direct use of fishery and survey length and age compositions to estimate annual, and can account for changes in selectivity (although for gray triggerfish selectivity was assumed constant over the full time series, which appears a reasonable assumption given the evidence presented at the RW). The ASPIC model for gray triggerfish could only be fitted from 2000 onwards and proved difficult to fit as it requires good contrast in data over time to estimate the R and K parameters, and this is not the case for gray triggerfish. The AW attempted many model runs, many of which did not converge or provided unreasonable solutions. The run presented at the RW was the one most closely matching the BAM, and therefore cannot be considered as an independent assessment to corroborate the BAM results. It included upweighting the CVID indices by a factor of 3.

The fitted model presented at the RW showed reasonable fits to fishery-dependent abundance indices and a poorer fit to the up-weighted CVID series (Fig. 4.2.2). The model suggests that the stock has increased in biomass and is not overfished, and that the stock is very lightly exploited at well below the Fmsy. However, the parameter estimates for this model (particularly catchability of each series and the MSY estimate were close to the lower bounds of the priors (Fig. 53 in AW report for gray triggerfish) and the estimates of key management variables have large variances (AW report Fig. 54).

For reasons given above, the ASPIC model is not a reliable basis for providing management advice for the SA gray triggerfish.

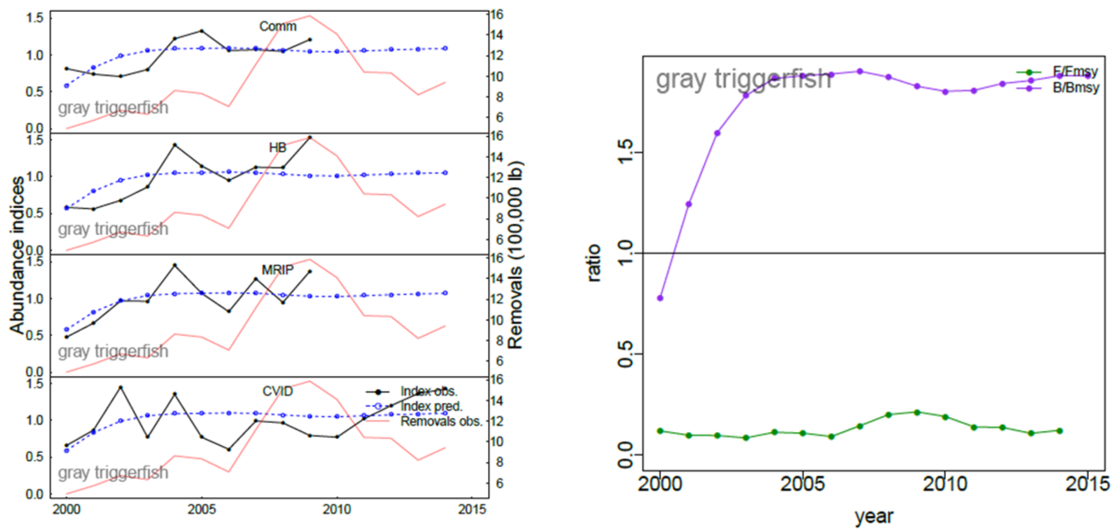

Fig. 4.2.2. Fits of ASPIC model to South Atlantic gray triggerfish abundance indices, and (right) time series of B/Bmsy and F/Fmsy,

BAM assessment

The BAM has many assumptions and many estimated parameters. In comparison with the red snapper BAM, the base model has more continuous time series of fishery composition data and relies on only one relative abundance series, which is a standardized fishery-independent series from the SERFS chevron trap survey combined also with video data since 2010. The most important data and modeling decisions were around weighting of the CVID relative abundance indices and modeling the form of selectivity.

The base configuration of the Beaufort Assessment Model (BAM) from the Assessment Workshop was revised with corrected age compositions of the Chevron Trap survey during the Review Workshop, as described above under ToR 1. The aggregate time-series fit to CVID age compositions for the original base run and the revised run with corrected age compositions shows the shift in modal age from 2 to 3, but the fit is not particularly close for either model (Fig. 4.2.3).

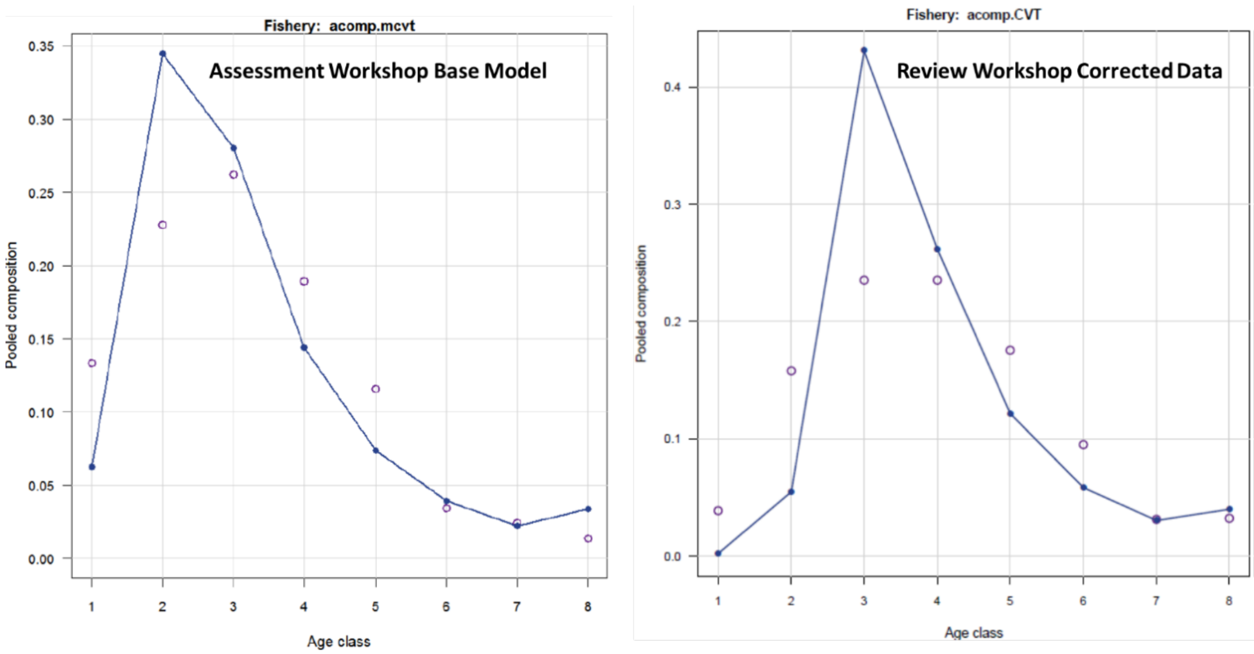

Figure 4.2.3. Observed and predicted age composition of the Chevron Video Trap Survey from the Assessment Workshop based model (left) and the corrected data and revised model from the Review Workshop (right).

The Review Panel had two concerns with the assessment, which they considered could have interacted in the model fit: firstly the upweighting of the CVID index by a factor of 6, and the combined use of length and age compositions from the same sources in fitting the model. The rationale for upweighting the index was the advice from Francis (2011) that relative abundance indices should be given priority in weighting an assessment over composition data. However, the double-use of length data (which is used for weighting age composition by length class) was thought by the Panel to have possibly led to over-weighting of the composition data, which in turn could have influenced the extent of CVID upweighting carried out by the AW.

The AW noted that with no additional weighting, the CVID series was relatively poorly fitted. The weighting was increased progressively until an acceptably good fit was obtained. This required upweighting by a factor of 6. Using the corrected base model with revised CVID age compositions, the fit to the CVID index can be seen to be very close (Fig. 4.2.4). In 15 out of 25 years, the model fits the index almost exactly, which appears highly unlikely given the variance in the indices and the additional structural uncertainty in the BAM. In this configuration, all the fishery removals estimates are fitted exactly and the only abundance index series is also fitted extremely closely. The model fits very closely to the 1990 survey index, which was very low and coincided with the recent passage of a hurricane. The resultant model estimates of biomass indicated a stock with low biomass in 1990, but increasing rapidly in biomass over the next few years. An alternative with no upweighting of the survey (sensitivity run 6) produced estimates of abundance in the first year of the assessment that were similar to the rest of the time series.

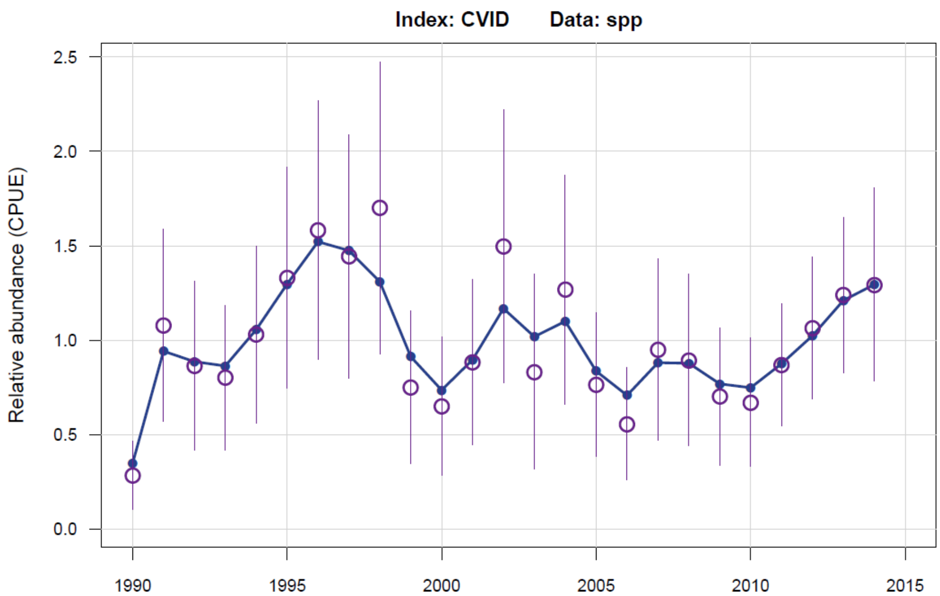

Figure 4.2. 4. Observed and predicted CVID trap/video Survey indices from the base model with corrected age compositions.

The Review Panel requested additional sensitivity analyses to help understand the results. An alternative BAM configuration with a starting year of 1974 estimated a series of low recruitment to explain the low survey index in 1990. An analysis that removed length compositions for fleets with age compositions, with no up-weighting of the survey, still did not fit the survey well, and the fit to age compositions deteriorated. A sensitivity analysis that removed the 1990 survey observation produced estimates of abundance in the first year of the assessment that were similar to the rest of the time series. The Chevron trap survey began in 1988, but the protocol was being refined in 1988 and 1990. There have been no changes to the design of the survey since 1990. However, Hurricane Hugo was 7-8 months prior to the 1990 survey. A study of Jamaican reef fish found changes in abundance, behavior, and distribution a year after Hurricane Allen (Kaufman 1983).

The Review Panel did not accept the BAM for South Atlantic gray triggerfish, and recommended that further modeling is needed to fit the corrected data and to resolve the fit to the survey, perhaps investigating a multispecies year effect in 1990, to consider possible effects from Hurricane Hugo and a justification for removing the 1990 survey observation.

ToR 3. Evaluate the assessment findings and consider the following:

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
b) Is the stock overfished? What information helps you reach this conclusion?
c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

As there was no agreed final assessment model, it is not possible to complete ToR 3 for SA gray triggerfish. The RW however noted that the current BAM explored at the meeting provides no evidence that the stock is overfished or that overfishing is taking place. The ASPIC model, which is also not considered suitable for providing management advice, also suggested the stock is lightly fished. However, these cannot be taken as providing unambiguous management advice and further work is needed on the BAM before advice can be given.

ToR 4. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:

a) Are the methods consistent with accepted practices and available data?
b) Are the methods appropriate for the assessment model and outputs?
c) Are the results informative and robust, and are they useful to support inferences of probable future conditions?
d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

In the absence of an accepted assessment model, stock projections cannot be evaluated.

ToR 5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.

a) Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Uncertainties in the assessment provided for the RW were thoroughly explored through (1) a Monte Carlo and bootstrap (MCB) analysis; (2) sensitivity analyses around the base BAM run; (3) retrospective analysis; and (4) comparison with an alternative assessment model (ASPIC). In practice, the ASPIC model was not considered by the Review Panel and the Assessment Working group to be appropriate for gray triggerfish, so it is difficult to use this to evaluate uncertainties in the BAM results. Overall, the DW and AW specified the most important sources of error and uncertainty in the sensitivity analysis and MCB for the BAM assessment. The implications of uncertainty in technical conclusions were clearly stated. The uncertainties will have to be re-evaluated when the BAM has been explored further by the AW to resolve the issues raised at the RW. However, the methods used appear appropriate. In general, the current model exhibits large historical uncertainty in biomass and exploitation based on the MCB, and some overestimation of F and underestimation of recruitment and biomass in 2009 and 2010 in retrospective analysis.

ToR 6 Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

a) Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.
b) Provide recommendations on possible ways to improve the SEDAR process.

Research and monitoring

The Data Workshop and Assessment Workshops provided extensive lists of recommendations for research and monitoring that could improve the reliability of, and information provided by, future assessments (see Appendix 5). These are largely sensible but there is no evaluation of which suggestions would have the greatest impact on improving the quality of the assessment. It is also important to consider the costs of the research or additional / new data collection in relation to the benefits in terms of improved accuracy of the assessments. As an independent reviewer, I have no information on this, which makes it difficult to suggest any prioritization of the existing recommendations and can only indicate what I feel would be useful to resolve important uncertainties in the assessment. Some ICES expert groups that I am involved in (e.g. Planning Group on Data Needs for Assessment and Advice - http://www.ices.dk/community/groups/Pages/PGDATA.aspx) are promoting a cost-benefit framework running alongside the existing quality assurance frameworks to encourage the use of objective approaches to evaluating and prioritizing data needs according to impact on assessments and advice relative to costs of data collection. The handling of research and data collection recommendations within the SEDAR process could be improved by imposing a requirement to clearly describe how the research or data would improve the assessments and advice, and to evaluate the relative cost (even if in broad cost categories or equivalent metrics). It is appreciated that there are many assessed stocks and limited funding for research and enhanced data collection, and if gray triggerfish are shown to be very lightly exploited, managers may wish to direct funding towards stocks with more urgent management and assessment quality issues, emphasizing the need to target funds to the most critical issues for gray triggerfish.

Some potential areas of research or enhanced data collection that could be useful to better understand and reduce uncertainties in the assessment include:

- Further investigation into the utility of the CVID survey for delivering abundance indices that are proportional to gray triggerfish abundance over a wide range of stock size. As the only abundance index in the assessment, it is critical to understand any potential biases associated with design or analysis of the data, or with other factors not currently accounted for such as saturation effects. Given the long time series of age compositions, some simple diagnostics should be developed such as overlays of indices by age (standardized to time-series average) plotted against year (to examine possible year-effects) or year class (to examine year class tracking). Fitting a simple model to the matrix of indices by age with separable (age and year) effects on total mortality is also possible to examine residuals and provide additional information on residual variance.

- Age validation studies, age-calibration workshops and use of a reference collection of age material should be carried out. Even experienced age readers can drift over time in their interpretation of age material, and ideally there should be a reference collection of known-age fish from across the age range (which is difficult) or at least a reference collection that is inserted at random and anonymously into annual age collections to monitor consistency of age readings.

Improving the SEDAR process

This is a repeat of the text for red snapper. In my experience of attending several SEDAR review workshops as a CIE reviewer, the process is very effective in thoroughly reviewing the assessments and the basis for management advice. Presentations given by the lead stock assessors and colleagues are usually extremely thorough, easy to understand and well-focused on key issues, which was the case at SEDAR 41. The entire process involves a substantial investment of staff time and funds to operate the data and assessment workshops and associated preparation and reporting, but the benefit is a very thorough compilation and evaluation of available data and modelling approaches. The involvement of people from the commercial and recreational fishing community at the RW and earlier stages is an important part of the process. It is important to hear their views based on their own observations on the water, as well as to ensure transparency in the assessment and review process. Inevitably there are differences of opinion, which in some cases can be a result of the different scales at which fishermen and scientists observe the ocean, but it is a learning process on both sides that must be maintained. It is hard to see how the overall process can be significantly improved.

One area of improvement already discussed above is that the handling of research and data collection recommendations within the SEDAR process could be improved by imposing a requirement to clearly describe how the research or data would improve the assessments and advice, and to evaluate the relative cost (even if in broad cost categories or equivalent metrics).

It is clearly important that SEDAR continues to bring in reviewers with the necessary experience and technical competence in data collection and assessment methods. Many assessments are now being conducted with complex statistical models such as BAM and Stock Synthesis, and it can be difficult to fully understand how the parameter estimation is driving a particular model result, particularly where there are multiple uses of the same data sets (e.g. catches and catch rates, or lengths and ages derived from the same data sets), correlation structures within data sets (e.g. multinomial distributions), and correlations between parameters. Uncertainty in assessment results is usually investigated using sensitivity analysis and MCB or MCMC analysis, but to varying degrees and the number of sensitivity runs can vary widely depending on the assessment working group and the perceived issues with the assessment. Considerable time can be spent producing additional sensitivity runs requested by the review panel. The SEDAR 41 review panel noted that such runs done one at a time do not fully reflect the true sensitivity of the model, which is built from multiple choices of data and model settings. As proposed by the Review Panel, the review process could be greatly helped by use of more advanced “global sensitivity analysis” methods for statistical models, which more comprehensively explore the sensitivity of particular outputs to the combination of different data and model settings. An investment to develop such methods for the different types of models used in SEDAR and other review processes would have considerable benefits for understanding the quality of assessments and advice, and targeting efforts at improving model inputs and settings that have greatest impact on the results.

ToR 7. Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.

I consider that the gray triggerfish BAM assessment in its present form may not constitute the best scientific information available, but it may do so if the issues identified at the RW are successfully addressed by the AW to provide a more robust assessment in future.

In relation to the following criteria:

Relevance: The SEDAR 41 assessment is relevant as gray triggerfish provide other fishing opportunities to fleets subjected to the red snapper moratorium or any other limits to fishing opportunities for other species managed in the same region. An assessment is needed at intervals to ensure that gray triggerfish do not become targeted to an extent that overfishing could occur or the stock is declining to an extent that it becomes overfished. The neighbouring Gulf of Mexico stock appears to have declined substantially over time.

Inclusiveness: The SEDAR 41 assessment includes all data that have been quality assured and proved adequate for use in the assessment.

Objectivity: The SEDAR 41 BAM is a highly objective procedure based on well-tested statistical modeling principles, and using data sets and assumptions that have been rigorously documented and reviewed through the SEDAR data, assessment and peer-review process. Where fully objective decisions are difficult to make, such as some decisions on scenarios for historic catches where evidence is lacking, the uncertainties around the decisions made have been explored and included in sensitivity analyses and the Monte Carlo Bootstrap evaluation of assessment uncertainty.

Transparency: All outputs of the data, assessment and review workshops in SEDAR 41 are fully documented and publicly available. The discussions at the review workshop are also recorded for the record. All data sets are thoroughly explored, and the quality of data on which the assessment is based is documented and transparent, as are all decisions related to the choice of assessment model, how it is implemented, and the results of the base run and sensitivity and uncertainty analyses.

Timeliness: The SEDAR process in general is arranged to provide timely fishery management advice where it is needed, and to ensure that assessments are benchmarked and reviewed at appropriate intervals.

Verification: The SEDAR 41 assessment process and deliverables comply with legal requirements under the Magnuson Stevens Act (2007) for developing and monitoring of fishery management plans and providing information on stock status.

Validation. The SEDAR 41 process is designed to meet the needs of fishery managers for peer-reviewed stock assessments and associated advice on stock status and future catches, and the process is open and fully transparent to the fishery managers and to stakeholders from commercial and recreational fisheries, conservation groups or others with a stake in the outcomes and who have opportunity to give their views on record.

Peer review: The SEDAR 41 process includes full peer-review by experts appointed by the Center for Independent Experts (CIE) and by reviewers from the SAFMC SSC. The review panel report and the independent CIE reviews are publicly available.

ToR 8. Compare and contrast assessment uncertainties between the Gulf of Mexico and South Atlantic stocks.

Until a final agreed assessment is produced, it is difficult to address this Term of Reference. Comments that apply to red snapper under ToR 8 in relation to differences in modelling and data apply also to a large extent to gray triggerfish, which is assessed in the GOM using Stock Synthesis, and also uses a wider range of fishery dependent and independent data sets. The latest assessment (SEDAR 43 Gulf of Mexico Gray Triggerfish stock assessment report: SEDAR41-RD77, 4 February 2016) indicates a dramatically different stock trend than given either by the SA BAM or the ASPIC model. Biomass in the GOM is in steep decline due to increasing fishing mortality up to the 1990s, as well as a progressive decline in recruitment (Fig. 4.2.5).

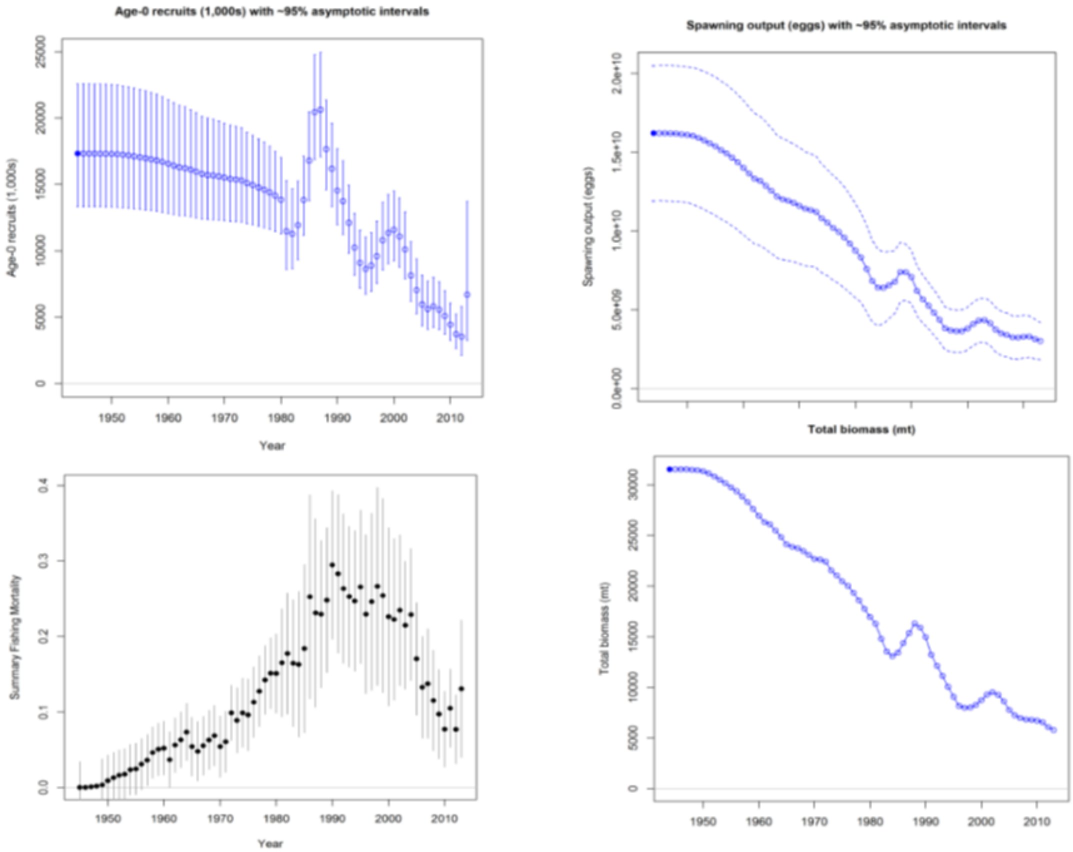

Fig. 4.2.5. Stock trends from the SEDAR 43 Gulf of Mexico Gray Triggerfish stock assessment report (SEDAR41-RD77, 4 February 2016)

ToR 9. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

The SEDAR 41 assessment of SA gray triggerfish was not accepted as the basis for providing advice and has gone back to the assessment working group to address issues related to weighting of composition data and the CVID trap survey index. This will be the focus of improvements needed to allow a more robust assessment with associated uncertainty analysis.

As only abundance index is the CVID trap/video survey, further work is needed to establish the ability of this survey to provide indices proportional to abundance and to confirm the selection pattern, taking account of other factors such as trap saturation which is known to occur. As noted earlier, given the long time series of age compositions, some simple diagnostics should be developed for the survey, such as overlays of indices by age (standardized to time-series average) plotted against year (to examine possible year-effects), or year class (to examine year class tracking). Fitting a simple separable model to the matrix of indices by age is also possible to examine residuals and provide additional information on residual variance.

ToR 10. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. Develop a list of tasks to be completed following the workshop. Complete and submit the Peer Review Summary Report in accordance with the project guidelines.

At the time of submission of the present report, the Peer Review Summary requested by the ToR was not complete.

References

Bachelor, N.M., Bartolino, V. and Reichert, M.J.M. 2013. Influence of soak time and fish accumulation on catches of reef fish in a multispecies trap survey. Fish. Bull. 111: 218-232.

Francis, RIC. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68: 1124-1138.

Kaufman, L.S. 1983. Effects of Hurricane Allen on reef fish assemblages near Discovery Bay, Jamaica. Coral Reefs 2: 43-47.

Appendix 1 Material provided for review

All material was available on the FTP site or through the SEDAR website.

DW and AW reports

S41_SA_RS_DWReport_FINAL_9.15.2015	South Atlantic Red Snapper Data Workshop Report
S41_SA_RS_AWReport_2.29.2016	South Atlantic Red Snapper Assessment Report
S41_SA_GTF_DWReport_FINAL_9.15.2015	South Atlantic Gray Triggerfish Data Workshop Report
S41_SA_GTF_AWReport_2.29.2016	South Atlantic Gray Triggerfish Assessment Report

A draft updated version of the red snapper assessment report was provided after the RW to include revised age composition data.

Documents outlining the fishery management plan and amendments, and the assessment history and review, were provided for each species.

A document S41RW06_S41_Public Comments_3.7.2016.pdf was provided containing the views of fishermen on the state of the stock and the quality of the assessment.

A large number of other background documents submitted to the DW and AW meetings were available.

Appendix 2: Statement of Work for Dr. Michael Armstrong (CEFAS)

External Independent Peer Review by the Center for Independent Experts

SEDAR 41 South Atlantic Red Snapper and Gray Triggerfish Assessment Review Workshop

Scope of Work and CIE Process: 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 Technical Representative (COTR), 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 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.

Project Description: SEDAR 41 will be a compilation of data, an assessment of the stocks, and CIE assessment review conducted for South Atlantic red snapper and gray triggerfish.
The review workshop provides an independent peer review of SEDAR stock assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and sensitivity runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. The stocks assessed through SEDAR 41 are within the jurisdiction of the South Atlantic Fisheries Management Council and the states of Florida, Georgia, South Carolina, and North Carolina. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop Terms of Reference. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in Charleston**, SC during March 15-18, 2016**.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also 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 COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) 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/
http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein.

Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator. 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 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.
CIE reviewers shall conduct an impartial and independent peer review of the assessment in accordance with the SoW and ToRs herein.

A description of the SEDAR Review process can be found in the SEDAR Policies and Procedures document:
http://sedarweb.org/docs/page/SEDARPoliciesandProcedures_Oct14_FINAL.pdf
The CIE reviewers may contribute to a Summary Report of the Review Workshop produced by the Workshop Panel.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

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 **Charleston, SC during March 15-18, 2016**.
3) Participate at the review meeting in **Charleston, SC during March 15-18, 2016** as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than April 11 2016, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to mshivlani@ntvifederal.net, and 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 ToR in **Annex 2**.

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

February 9, 2016	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
March 1, 2016	NMFS Project Contact sends the CIE Reviewers the pre-review documents
March 15-18, 2016	Each reviewer participates and conducts an independent peer review during the panel review meeting
April 11, 2016	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
April 25, 2016	CIE submits CIE independent peer review reports to the COTR
May 2, 2016	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR 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 CIE 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: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (Allen Shimada at allen.shimada@noaa.gov.

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

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

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:
Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen Shimada@noaa.gov Phone: 301-427-8174
Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136

Key Personnel:
NMFS Project Contact:
Julia Byrd
SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
(843)571-4366
julia.byrd@safmc.net

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, 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 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 CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, 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 the CIE Statement of Work
Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

SEDAR 41 South Atlantic Red Snapper and Gray Triggerfish Assessment Review Workshop: Terms of Reference

1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:
   a) Are data decisions made by the DW and AW sound and robust?
   b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
   c) Are data applied properly within the assessment model?
   d) Are input data series reliable and sufficient to support the assessment approach and findings?

2. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:
   a) Are methods scientifically sound and robust?
   b) Are assessment models configured properly and used consistent with standard practices?
   c) Are the methods appropriate for the available data?

3. Evaluate the assessment findings and consider the following:
   a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
   b) Is the stock overfished? What information helps you reach this conclusion?
   c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
   d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
   e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

4. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:
   a) Are the methods consistent with accepted practices and available data?
   b) Are the methods appropriate for the assessment model and outputs?
   c) Are the results informative and robust, and are they useful to support inferences of probable future conditions?
   d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   a) Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
   b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.
   a) Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.

b) Provide recommendations on possible ways to improve the SEDAR process.

7. Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.

8. Compare and contrast assessment uncertainties between the Gulf of Mexico and South Atlantic stocks.

9. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.

10. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. Develop a list of tasks to be completed following the workshop. Complete and submit the Peer Review Summary Report in accordance with the project guidelines.

Annex 3: Agenda

SEDAR 41 South Atlantic Red Snapper and Gray Triggerfish Review Workshop North Charleston, SC March 15-18, 2016

Tuesday
8:30 a.m. Convene
8:30a.m. – 9:00a.m. Introductions and Opening Remarks Coordinator
- *Agenda Review, TOR, Task Assignments &* **Chair**
9:00a.m. – 12:00p.m. Assessment Presentation and Discussion (RS*) Kate Siegfried
12:00p.m. – 1:30p.m. Lunch Break
1:30 p.m. - 3:30 p.m. Panel Discussion Chair
- *Assessment Data & Methods*
- *Identify additional analyses, sensitivities, corrections*
3:30p.m. – 3:45 p.m. Break
3:30 p.m. - 5:30 p.m. Panel Discussion and/or Work Session Chair
-Continue deliberations
5:30p.m. – 6:00p.m. Public Comment Chair
***Tuesday Goals*:** Initial RS* presentation completed, sensitivities and modifications identified.

Wednesday
8:30 a.m. – 12:00 p.m. Assessment Presentation and Discussion (GTF) Kevin Craig**
12:00 a.m. – 1:30 p.m. Lunch Break
1:30 p.m. - 3:30 p.m. Panel Discussion Chair
- *Assessment Data & Methods*
- *Identify additional analyses, sensitivities, corrections*
3:30p.m. – 3:45 p.m. Break
3:30 p.m. - 5:30 p.m. Panel Discussion and/or Work Session Chair
-Continue deliberations
5:30p.m. – 6:00p.m. Public Comment Chair
***Wednesday Goals*:** Initial GTF** presentation completed, sensitivities and modifications identified.

Thursday
8:30 a.m. – 12:00 p.m. Panel Discussion Chair
- *Review additional analyses, sensitivities*
12:00 a.m. – 1:30 p.m. Lunch Break
1:30 p.m. – 3:30 p.m. Panel Discussion Chair
-Continue deliberations
3:30 p.m. – 3:45 p.m. Break
3:45 p.m. – 5:30 p.m. Panel Discussion and/or Work Session Chair
-Consensus recommendations and comments
5:30 p.m. - 6:00 p.m. Public Comment Chair
Thursday Goals: Final sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun.

Friday
8:00 a.m. – 10:30 a.m. Panel Discussion Chair
- *Review additional analyses, final sensitivities*
- *Projections reviewed.*
10:30 a.m. – 10:45 p.m. Break
10:45 a.m. – 12:30 p.m. Panel Discussion or Work Session Chair
- *Review Consensus Reports*

12:30 p.m. – 1:00 p.m. Public Comment Chair
1:00 p.m. ADJOURN
Friday Goals: Complete assessment work and discussions. Final results available. Draft Summary Report reviewed.
* RS = South Atlantic red snapper **GTF = South Atlantic gray triggerfish

Appendix 3: Panel membership

REVIEW PANEL

Luiz Barbieri	Review Panel Chair (SAFMC SSC)
Steve Cadrin	Reviewer (SAFMC SSC)
Churchill Grimes	Reviewer (SAFMC SSC)
Mike Armstrong	CIE Reviewer (CIE)
Stephen Smith	CIE Reviewer (CIE)
Jon Helge Volstad	CIE Reviewer (CIE)

Appendix 4: Research Recommendations from SEDAR41 Data Workshop and Assessment Workshop: Red Snapper.

*1. **Data Workshop***

1.2 Life History

Red Snapper Mini Season

If this program, along with continued closure of the fishery, is to extend into future seasons, an exploration of methods to further incentivize angler participation would be useful. After brief interviews with participants from the recreational fishers group at SEDAR 41, the following suggestions were provided to increase angler participation:

- Free fish cleaning at donation site.
- As people may be tired after being out on the water all day and with busy boat ramps, short questionnaire from a biologist on-site could be used instead of the anglers filling the forms out or requiring fishermen to fill out a survey online after they return home.
- Advertise data collection at local bait & tackle shops.
- Use NOAA's announcement system on weather radio channel where they also announce season closures, etc. Since fishermen are frequently monitoring this channel for weather updates, it could be an effective communication route to announce the collection information (drop locations, reward information, etc.).
- Dry storage areas are a good place to sample; many people store boats there instead of trailering home.

Life History Research

- More research on red snapper movements and migrations in Atlantic waters is needed. Available data and the results of studies in the Gulf of Mexico indicate high site fidelity, but that tropical storms may cause greater than normal movement that might help dispersal to depleted areas. This needs to be confirmed in the South Atlantic. Additional acoustic and traditional tagging is needed on known spawning locations to document spawning migrations or aggregations, and return of fish to non-spawning areas.
- Evaluate more thoroughly the data/sample collection during the mini-season to improve utility for assessments. This should include what samples should be collected (e.g. reproductive information).
- Possible changes in life history parameters, in particular relative to reproduction, need to be further investigated.
- Much is unknown about the early life history of Red Snapper, in particular relative to spawning areas, larval and juvenile stages, including habitat and dispersal.
- Alternative methods of reproductive output. The methods described in Klibansky's SEDAR41-DW49 may provide a more accurate estimate of reproductive output than previously used. Further investigation into this modeling effort and use for future assessments should be investigated.

- Duration of spawning indicators. The definition of spawning indicators has received significant discussion recently. As this has significant implications for the estimates of reproductive output, further research is needed to define consistent criteria for spawning indicators in finfish.
- Continuing the age reading comparisons and calibrations between labs on a reference collection of known age fish would be beneficial for determining a more accurate aging error matrix and would provide accuracy to the age composition data.

1.2 Commercial Statistics

Landings

- Improve gear and effort data for each trip.
- Standardize methodology for developing average proportions to parse out unclassified landings.

Discards

- Investigate the validity and magnitude of "no discard" trips. This may include fisher interviews throughout the region.
- Examine potential impacts of "no discard" trips on estimated discards.
- Improve discard logbook data collections via program expansion or more detailed reporting (i.e. electronic logbooks, etc.)
- Establish an observer program that is representative of the fisheries in the South Atlantic

Biosampling

- Establish an observer program that is representative of the fisheries in the South Atlantic.
- Angler education with regards to recording depths on paper logbooks (i.e. standardized units); validation of additions to the logbook form still needed.
- Standardize TIP sampling protocol to get representative samples at the species level.
- Standardize TIP data extraction.

1.3 Recreational Statistics

- Complete analysis of available historic photos for trends in CPUE and mean size of landed Red Snapper and Gray Triggerfish for pre-1981 time period. (Ultimately all species).
- Formally archive data and photos for all other SEDAR target species.
- For Hire Survey (FHS) should collect additional variables (e.g. depth fished).
- Increasing sample sizes for at-sea headboat observers (i.e. number of trips sampled).
- Compute variance estimate for headboat landings.
- Mandatory logbooks for all federally permitted for-hire vessels.

1.4 Indices

- Compare existing methods and/or develop new methods to define effective effort in fishery dependent data.

- Estimate selectivity of video gear in the SERFS.
    - Tagging, stereo cameras
- For video reading, evaluate methods to score water clarity and habitat.
- Evaluate effect of (non) independence between chevron traps and videos, including methods to combine the indices.
- Continue exploring the use of continuous predictor variables (e.g., splines or polynomials) for ZIP and ZINB standardization models.
- Headboat at-sea observer program needs depth data from all states (not just FL) and increased coverage overall.
- SCDNR charterboat logbook program should be replicated by other states.
- Develop fishery independent hook-gear index (S41-DW08).

2. Assessment Workshop

- Increased fishery independent information, in particular reliable indices of abundance and age compositions.
- Red Snapper were modeled in this assessment as a unit stock off the southeastern U.S. For any stock, variation in exploitation and life-history characteristics might be expected at finer geographic scales. Modeling such sub-stock structure would require more data, such as information on the movements and migrations of adults and juveniles, as well as spatial patterns of larval dispersal and recruitment. In addition, it is unknown whether a spatial model would improve the assessment.
- More research to describe the life history of Red Snapper is needed, including more work to identify the location of juveniles before they recruit to the fishery.
- The effects on environmental variation on the changes in recruitment or survivorship.
- The Florida sampling program, during the mini-season in particular, provided invaluable data to this assessment. Programs such as these would be useful in all South Atlantic states, particularly if the management regulations continue to make established methods of index development or composition sampling from fleets less regular or possible.

Appendix 5: Research Recommendations from SEDAR41 Data Workshop and Assessment Workshop: Gray Triggerfish.

*1. **Data Workshop***

1.2 Life History

- More research on gray triggerfish movements and migrations in Atlantic waters is needed. Available data and the results of studies in the Gulf of Mexico indicate high site fidelity, but that tropical storms may cause greater than normal movement that might help dispersal to depleted areas. This needs to be confirmed in the South Atlantic. Additional acoustic and traditional tagging is needed on known spawning locations to document spawning migrations or aggregations, and return of fish to non-spawning areas.
- Age validation study that should include edge type and the potential for using various age structures for use in assessment. This should include the logistical feasibility of using these alternative structures for routine sampling and processing.
- Early life history is largely unknown. E.g. size and age at settlement and length of the pelagic stage.
- Estimates of delayed bycatch mortality are needed. This should include the effect of cloacal protrusion as a result of barotrauma.
- Tagging studies are needed to define spawning locations (only shelf edge or not) and, movement, the results of which could be used to help inform fishing mortality and natural mortality.
- Impact of climate change on mortality and recruitment.
- Research on spawning behavior/nesting and how it impacts survivorship and stock productivity.
- Determine fecundity type and estimate annual fecundity in Atlantic waters.
- Alternative methods of reproductive output. The methods described in Klibansky's SEDAR41-DW49 may provide a more accurate estimate of reproductive output than previously used. Further investigation into this modeling effort and use for future assessments should be investigated.
- Duration of spawning indicators. The definition of spawning indicators has received significant discussion recently. As this has significant implications for the estimates of reproductive output, further research is needed to define consistent criteria for spawning indicators in finfish.
- Investigate gray triggerfish competition for nests. The presence of competition for nest space may affect, among other things, the spawning success (reproductive output) and the choice of a spawner recruit relationship. Further investigation into the nesting behavior of gray triggerfish is needed to provide information to address these issues.

1.2 Commercial Statistics

Landings

- Require species level reporting in state trip ticket programs. Some states in process of instituting species level reporting for all species.
- Improve gear and effort data collections.

Discards

- Investigate the validity and magnitude of "no discard" trips. This may include fisher interviews throughout the region.
- Examine potential impacts of "no discard" trips on estimated discards.
- Improve discard logbook data collections via program expansion or more detailed reporting (e.g. more detailed logbook, electronic reporting)
- Establish an observer program that is representative of the fisheries in the South Atlantic.

Biosampling

- Standardize TIP sampling protocol to get representative samples at the species level.
- Standardize TIP data extraction.
- Establish an observer program that is representative of the fisheries in the South Atlantic.
- Increase untargeted sampling in NE and Mid-Atlantic observer programs.
- Increase untargeted dockside sampling in NE and Mid-Atlantic.

1.3 Recreational Statistics

- Complete analysis of available historic photos for trends in CPUE and mean size of landed Red Snapper and Gray Triggerfish for pre-1981 time period. (Ultimately all species).
- Formally archive data and photos for all other SEDAR target species.
- For Hire Survey (FHS) should collect additional variables (e.g. depth fished).
- Increasing sample sizes for at-sea headboat observers (i.e. number of trips sampled).
- Compute variance estimate for headboat landings.
- Mandatory logbooks for all federally permitted for-hire vessels.

1.4 Indices

- Compare existing methods and/or develop new methods to define effective effort in fishery dependent data.
- Estimate selectivity of video gear in the SERFS.
    - Tagging, stereo cameras
- For video reading, evaluate methods to score water clarity and habitat.
- Evaluate effect of (non) independence between chevron traps and videos, including methods to combine the indices.
- Continue exploring the use of continuous predictor variables (e.g., splines or polynomials) for ZIP and ZINB standardization models.
- Headboat at-sea observer program needs depth data from all states (not just FL) and increased coverage overall.
- SCDNR charterboat logbook program should be replicated by other states.
- Develop fishery independent hook-gear index (S41-DW08).

2. Assessment Workshop

- Increased fishery independent information, in particular reliable indices of abundance and age compositions.
- Increased age sampling and evaluation of ageing error over the stock area and from all fleets, particularly the general recreational fleet.
- In this assessment Gray Triggerfish were modeled as a unit stock off the southeastern U.S. For any stock, variation in exploitation and life-history characteristics might be expected at finer geographic scales. Modeling such sub-stock structure would require more data, such as information on the movements and migrations of adults and juveniles, as well as spatial patterns of larval dispersal and recruitment. In addition, it is unknown whether a spatial model would improve the assessment.
- More research to better understand the life history of Gray Triggerfish is needed, including natural mortality, maturity, and reproductive potential, particularly for the youngest ages.
- The effects on environmental variation on the changes in recruitment or survivorship.