Report on the 2010 South East Data, Assessment and Review (SEDAR 24) Workshop to Review the Assessment of the Status of South Atlantic Red Snapper, 12-14th October 2010, Savannah, Georgia.

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

[1] Representing the Center for Independent Experts.

1. Executive Summary

The 2010 South East Data, Assessment and Review (SEDAR 24) Workshop to review the assessment of the status of South Atlantic Red Snapper took place on 12-14th October 2010 in Savannah, Georgia. The material presented was derived from a separate Data Workshop and an Assessment Workshop, and included a revision of the assessment to take into account recommendations from a CIE reviewer who commented on the first draft of the assessment workshop report. 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 three series of fishery-dependent relative abundance indices. The model is an evolution of the one applied in SEDAR-15 and includes a number of improvements: (1) more plausible dome-shaped selectivity models for recreational fisheries; (2) the addition of the headboat discard recruitment index; (3) avoidance of using length and age data from the same sources; and (4) iterative re-weighting of the contribution of data components to the statistical likelihood used for estimating model parameters.

The fishery-dependent data for each fleet were not continuous over the time series, and were in some cases subject to changes in collection methods. This required the use of methods to impute missing values or adjust time series to ensure consistency. Age composition data were used in preference to length data from the same years and fleets, if sample sizes were adequate. This resulted in some blocks of years with age data, and some with length data only. The BAM model was developed specifically to deal with data of this nature.

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 data sources as used for the BAM, but excluding any length or age data. Some additional analyses using catch-curves were carried out to examine mortality rates using only the landings-at-age data. The main conclusion arising from the SEDAR-24 assessment of South Atlantic red snapper is that there is a high probability that the stock is overfished and is experiencing overfishing. This conclusion is supported by the two different assessment models (BAM and ASPIC) and by sensitivity analysis around the base model configurations. However, the extent of overfishing, and the extent to which the stock is overfished, is relatively poorly determined because of uncertainties in data and model structure. For example, the stock-recruit function is poorly determined making it difficult to estimate MSY and B_{MSY} benchmarks. The BAM and ASPIC models also give different interpretations of the extent of biomass decline over time.

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 2008 (SEDAR-15), using a similar form of model to the one implemented at SEDAR-24. Input data for the SEDAR-24 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-15 independent peer review of the data and assessment models and other subsequent developments. In 2010, an additional review of the draft AW report had been carried out by a CIE reviewer, and the subsequent revisions to the assessment were included in the AW report provided to the October 2010 Review Workshop.

3. Description of review activities

The SEDAR 14 Review Workshop (RW) took place at the Hilton de Soto hotel in Savannah, from 2:00pm Tuesday 12 October to 4:00pm Thursday 14 October, 2010. The assessment results and background were clearly presented by the experts at the meeting. The Review Panel requested a number of additional model runs and extraction of other supporting data, and these were done very quickly and led to fruitful discussion that helped to clarify a number of important issues. Some further clarifications were requested from the lead assessor after the review meeting. The provisional agenda for the meeting is given in Annex 3 of Appendix 2.

The Review Panel itself comprised the Chair, three reviewers appointed by the CIE and a South Atlantic Science and Statistical Committee representative (Appendix 3). The assessment results were presented by three US technical experts who were involved in the AW. The RW was also attended by the SEDAR coordinator, a number of NMFS and fishery committee representatives, and some members of the public. 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 documents relevant to the Review Workshop were received approximately one week before the panel convened, rather than the two weeks stated in the Terms of Reference.

During the course of the Review Workshop members of the Review Panel received hard copies and e-mails from the fishing public that contained new data to consider during their deliberations. The panel considered it more appropriate that this type of information be submitted during the data review workshop, where it can be evaluated along with other data sets being considered for use in the stock assessment.

The Review Panel provided a Summary Report. 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.

4. Summary of findings by Term of Reference

ToR 1. Evaluate the adequacy, appropriateness, and application of data used in the assessment.

The data used in the Beaufort Assessment Model (BAM) and the ASPIC surplus production model were all derived from the recreational and commercial fisheries, and excluded any fishery-independent indices of abundance. Available fishery independent data had been reviewed at the DW and considered unsuitable for this stock. The primary data for the BAM are summarized in Tables 1 – 3 of the AW report, with input parameters summarized in Tables 4 -5 of the AW report.

Landings data

Annual fishery landings were derived from a mixture of census estimates and MRFSS survey estimates for recreational catches. A number of approaches were adopted to generate a complete and consistent series of recreational landings and discards for the assessment period, in order to deal with gaps in historical data and changes in data collection methods. This involved imputation of missing values (e.g. mean weights) and calibration methods (ratio estimators) to predict data series that are missing in earlier years or have been subject to improvements in collection methods in recent years (e.g. prediction of 1981-85 charter boat landings, pre-1981 charter and private boat landings and pre-1972 headboat landings). The resultant time series of landings and discards estimates therefore contain a number of different forms of error, including any biases in census estimates (e.g. where catches are self-reported); bias and imprecision in MRFSS survey estimates; imprecision of ratio estimators and errors in assumptions of constancy of ratios back in time. The adequacy and appropriateness of the landings and discards input data are time-dependent, with much greater uncertainties in the earlier years. The methods for adjusting or in-filling historical data required pragmatic assumptions that were generally not validated using independent data. However the effects of some of these assumptions were explored by the AW through sensitivity analyses using BAM. Consideration should be given to developing routines within the model fitting procedure to estimate missing values such as historical recreational catches.

An additional manipulation of the data involved smoothing techniques (cubic spline fits) to reduce the influence of “spikes” in the catch history data. The use of smoothing masks the extent of observation error, and it might be better to explore weighting procedures to control the influence of high-leverage data values known to have high estimation error (inverse-variance weighting, or tapered weighting if historical values are considered to have generally higher estimation error).

A significant source of bias in landings data could arise from the use of a mean weight of 9lb to convert commercial line landings in 1955-1980 to fish numbers for predicting historical recreational landings using more recent ratios of commercial to recreational landings numbers. The BAM estimates of mean weight in the commercial landings, calculated from fitted selection curves and the population structure, were much higher than the 9lb input figure (Fig. 1, provided by the analysts at the Review Workshop). If the model estimates of mean fish weight (14-18lb) in the commercial landings were correct, this would imply biases in the historical recreational catch numbers using the 9lb figure in the ratio method, as the numbers in the commercial catches would be overestimated. Furthermore, BAM estimates for historical recreational catches are well above the 4.2lb figure assumed for the ASPIC input data. The inconsistent treatment of weights in the BAM model appears responsible for large differences in landings biomass trends from the ASPIC data and BAM estimates. Model estimates of mean weight for the commercial line fleet are also influenced by the choice of asymptotic selectivity, and will also reflect the age profile in the population according to the estimated F and assumed M. The historical mean fish weights for the different fleets should be thoroughly reviewed using additional evidence that may be available, and the BAM model adapted to ensure consistency between model inputs and outputs.

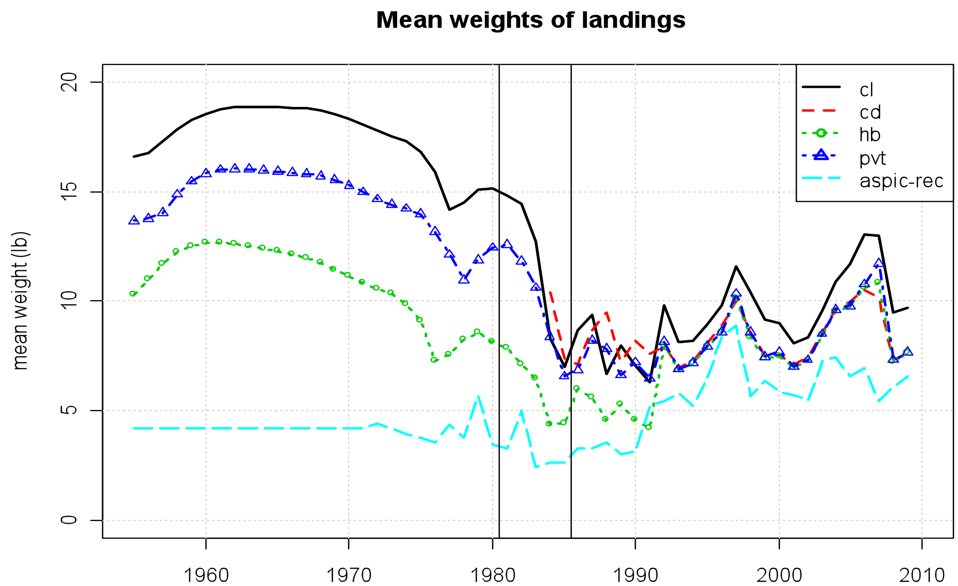

Fig. 1. BAM estimates of mean fish weight in the commercial line fishery (cl), commercial dive fishery (cd), recreational for-hire (headboat) fishery (hb) and private boats (pvt), calculated conditional on fitted selectivity patterns and model estimates of population structure. The assumed mean weights for converting recreational catch numbers to catch biomass for use in ASPIC are also shown (aspic-rec).

Abundance indices

Time series of relative abundance of were obtained from three age-aggregated data sources: headboat fleet CPUE (retained fish per angler-hour) in 1976-2009; commercial line fishery CPUE (retained weight (pounds) per hook-hour in 1993-2009), and observer based estimates of CPUE of discarded red snapper in the headboat fishery in 2005-2009 (used as a recruitment index). All CPUE data series were screened using the Stephens and MacCall (2004) method to include only those trips where red snapper could have been caught, based on significant associations with other species in the catches. This appears a suitably objective approach to try and remove trips where there is zero or very low probability of catching red snapper, particularly if the occurrence of such trips could have varied over time.

To account for improvements in technology (notably, GPS systems), catchability was linearly increased by 2% per year, beginning in 1976 for headboats and 1993 for commercial lines, until 2003 and holding it constant thereafter. Confirmation of long-term trends in catchability should be sought if there are other species taken with red snapper for which there are fishery independent and fishery dependent indices. The assumption of constant catchability since 2003 should also be reviewed in light of factors other than GPS that could affect catchability. Although the CPUE data were standardized using factors including year, area, season, trip type and number of anglers as categorical variables, catch rates in mixed-species hook and line fisheries could also be influenced by trends in abundance of other species, as the index depends on proportionality between density of red snapper and the numbers caught per unit effort. The CPUE will be affected by hook competition and additional handling time in catching other species on the same trip. 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.

The indices, as formulated, are adequate and appropriate for tuning the BAM, as the observed and predicted indices are of comparable structure (i.e. conditional on fleet selectivity). However, the ASPIC model will treat the CPUEs as indices of total stock biomass, not biomass weighted by selectivity. Furthermore, changes in CPUE caused by altered selectivity patterns, such as associated with the increase in MLS to 20 inches in 1992, will be interpreted by ASPIC as tracking changes in total biomass. The effect may be small as most of the change in selectivity is related to age 2. However, the headboat CPUE for ASPIC (Fig. 3.72 in AW report) shows a smooth gradual increase in CPUE in 1986 – 1991, followed by an equally smooth increase from 1992 onwards, but separated by a downward step change of around 50% which looks more than just a coincidence. Fig 3.22 in the AW report shows that the BAM also has difficulty explaining the steep drop in CPUE in numbers in 1992 despite the internal adjustment for a selectivity change, although the residuals are no worse than in some earlier years. It is possible that the increase in MLS could have altered targeting patterns causing at least a temporary change in catchability over and above the assumed 2% increase per year over this period. Further information should be sought on changes in behavior in the recreational fishery when the MLS changed in 1992, particularly if there was a persistent change over a period of years.

Length and age compositions

An important feature of the BAM is the variable use of length and age data. Age data were used for the for-hire recreational fishery landings in 1978 – 1987, 1989-1990, 1996, and 2002 -2009, whilst length data only were used for this fleet in other years since 1976. Age and length data for the commercial line fleet were more restricted: age data were used for 1996 – 2000, 2004, 2007 and 2009, whilst length data were used for other years since 1985 other than 2005, 2006 and 2008 when no length or age data were used. Some length and age data for the commercial dive fishery were also available since 1999, and length compositions of for-hire (headboat) discards were available and used from 2005 - 2009. The AW report shows observed and fitted length and age compositions, and bubble plots of residuals, that indicate very variable fits to individual years and fleets, and non-random residual patterns across lengths and ages. There are also some year-class effects in residuals (e.g. 2002-2009 for-hire landings: Fig. 3.12 in AW report). It would be useful to see some analyses of the internal consistency of the data series in terms of tracking cohorts, independent of the model fit, to evaluate the information content of the age composition data.

Some usable length data were discarded from the BAM inputs. The AW removed all length data for years / fleets where adequate age data were available, to avoid the use of non-independent data from the same sources. The length compositions could be retained for trips where no age data were collected, rather than discarding all the samples, provided there is sufficient independence of the length and age samples.

Fishery selectivity

The selectivity assumptions were well motivated in a working paper from the assessment workshop (AW-05). The recreational fishery is assumed to have a domed selectivity saturating at 0.3, which is effectively attained at age 10. The catch-at-age data for the recreational fleet show a very high proportion of zero catches in the age-year cells at ages 10 and above. Given the very low sample sizes for 10+ fish in the recreational fishery samples, it would probably be difficult to validate the use of a complex double logistic model with user-fixed saturation selectivity as opposed to a more continuous decline in selectivity with age beyond full selection (e.g. the more parsimonious gamma function used to describe selectivity in the hook fisheries for Alaska Sablefish). However, the BAM estimates of stock status relative to benchmarks appeared relatively insensitive to the fixed saturation value used. Tagging studies could be used to obtain direct estimates of fishery selectivity.

Growth parameters
Growth data are needed in the model to predict length-at-age distributions for age classes/years where there were length composition data only. The DW explored spatial patterns, and took selectivity into account when estimating growth parameters, but did not explore any temporal trends or year-class effects in growth rates that could influence model fits. Such analyses should be carried out.

Natural mortality
Choice of natural mortality values not only affects the scaling of population numbers but also affects estimates of parameters related to the productivity of the stock, such as MSY benchmarks. The DW recommended the Lorenzen (1996) method to predict the pattern of M at age from weight at age, rescaled to an estimate of M=0.07 for the oldest age classes based on an observed maximum age of 54 years (similar to the Gulf of Mexico stock). The Review Panel determined that the maximum age of 54 had been recorded in the last decade, and the probability of catching a fish this old must therefore have been influenced to some extent by the probability of being caught by a fishery (fishing mortality is estimated to be well above the assumed value of M for most age classes during the recent decade). Tagging studies show that red snapper have strong fidelity to a particular site (other than at spawning time) and it may be possible for fish in some areas to experience a very low probability of being caught or eaten during most of the year, and this may not be representative of the population as a whole. In general, in the absence of direct estimates of M at age, the values to use in the assessment must be considered undetermined.

Trends in M may be a factor in some of the problems in determining the F_{init}, for example the large plus-group residuals at low F_{init}, and in estimating long-term stock-recruit parameters under the assumption of stationarity of the parameters. An inability to account for trends in M is no different to most single-species assessments, and reconstructing historical M values is probably intractable. However, a more detailed review of the potential for temporal and spatial variability in predation mortality on red snapper could provide useful insights, particularly for developing scenarios for estimating stock depletion in the first half of the 20^{th} century and deriving starting age compositions for the model. Although sensitivity of the BAM to choice of M was explored, the relative fit of the model for different M values was not presented.

ToR 2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.

The assessment presentation included three methods: the Beaufort statistical catch-age model (BAM), surplus-production models (ASPIC), and catch curve analyses. The BAM was selected at the AW to be the primary assessment model, whilst ASPIC was run as a comparison using a totally different type of model that is not dependent on length or age data, although using the same fishery removals and CPUE data sources. Catch curve analyses were presented as a check of mortality estimates from BAM.

Beaufort statistical catch-age model (BAM)

BAM was the primary model in the assessment, and is a statistical catch-at-age model based on a similar approach to the SCA model used in SEDAR 15. Improvements (mainly in response to CIE reviews at SEDAR 15 and the assessment workshop of SEDAR 24) were: (1) more plausible dome-shaped selectivity models for recreational fisheries; (2) the addition of the headboat discard recruitment index; (3) avoidance of using length and age data from the same sources; and (4) iterative re-weighting of the contribution of data components to the statistical likelihood used for estimating model parameters. BAM has previously been applied to other SEDAR assessments of reef fishes in the U.S. South Atlantic, such as red porgy, black sea bass, tilefish, snowy grouper, gag grouper, greater amberjack, vermilion snapper, Spanish mackerel, and red grouper.

BAM was adequate and appropriate for this assessment, given the limitations of the available data such as discontinuous age composition data. The method was developed specifically to accommodate the type of assessment data available for this stock. The BAM appears to have been applied correctly.

A number of issues with the BAM application that should be explored in future assessments are discussed below.

Weighting of removals estimates
The input fishery removals data are a mixture of census data (log books; trip reports) and survey estimates that have associated precision estimates (MRFSS). A range of other uncertainties in the data also exist, related to imputation of missing data or ratio methods to estimate or adjust historical values based on more recent data. All the information on relative precision or potential bias in the data are ignored in the BAM fit, and the data on removals by fleet are given such high weighting that the data are all fitted exactly. The errors in the removals data are therefore translated into large variability in parameter estimates of importance for management. For example, the time-series of F estimates show low interannual variability up to 1980, due to the smoothing applied to earlier catch data, whereas post-1980 F estimates show extreme interannual fluctuations that probably reflect imprecision in non-smoothed landings, discards and length/age compositions. If total fishing effort has shown smooth trends over time, it would be expected that variability in F would be damped. It is suspected that the model may not converge if larger and more realistic CVs are applied by fleet. However, suitable penalties could be applied to constrain interannual variation in F or estimated catch (see B-ADAPT model applied to North Sea cod). Model formulations should be explored that allow the observation error in the catch estimates to be better reflected in the model fit. If alternative model formulations are developed along these lines, it would be valuable to test their relative robustness within the framework of management strategy evaluations.

Use of iterative re-weighting
Iterative re-weighting of the contribution of data components to the statistical likelihood was recommended by a CIE reviewer following a review of the first draft SEDAR-24 AW report, and this was implemented along with a number of other changes in preparing the AW report for the October Review Workshop. Although this is a recommended procedure for such models, it was noted by the Review Panel that some other undesirable features had emerged following implementation of iterative re-weighting. These included the appearance of a strong retrospective bias that was not apparent in the first AW draft report, and a relative down-weighting of the headboat landings CPUE series which the Data Workshop had considered a more robust indicator of stock trends due to the mixed species nature of the fishery. The relatively large CPUE values in the last two years of the headboat CPUE series and the last year of the commercial line CPUE became large positive residuals rather than being interpreted as a sharp increase in abundance. The recreational fishery landings-at-age data were substantially down-weighted following iterative re-weighting. The detailed results of some re-runs carried out by the analysts at the Review Workshop to explore the effect of iterative re-weighting, and manual weighting of the head-boat landings CPUE series, are given in the Review Panel Summary report and are not repeated here.

Iterative re-weighting has some well-known problems when the lengths of the data component series are quite different. For tuning indices, it is well known that iterative re-weighting can give too much weight to short time-series. The problem may be related to well-known biases in maximum likelihood estimates of variance parameters, in which variances are under-estimated when sample sizes are small and the number of model parameters is high. Some other assessment methods with iterative re-weighting, such as Extended Survivors Analysis used widely at ICES, include user-defined limits to the weighting applied to individual CPUE series (e.g. capping the weighting if the series is becoming fitted too precisely). In the case of red snapper, the CPUE series range from 5 years (headboat discards) to 34 years for the headboat landings CPUE. The discards index includes large values in 2007 and 2008 which are fitted almost exactly in the BAM. On balance, there are arguments for different approaches to re-weighting, and the review panel agreed to adopt the iterative re-weighted model as the base case, as proposed by the AW. However, the procedures for iterative re-weighting should be reviewed before the next assessment, to prevent over-weighting of short data series.

Retrospective bias

Retrospective analysis of the BAM is limited (and possibly misleading) because of the short headboat discards index, which has only five annual values and has relatively large variability. The base-case BAM presented at the Review Workshop has a pronounced retrospective bias (AW report Fig. 3.61) with some large downward adjustments of F and large upward adjustments of recruitment in 2006. The absolute level of retrospective under-estimation of SSB is very small, but in relative terms would be more pronounced. The first draft of the SEDAR-24 AW report (SEDAR24-AW13), prior to implementation of the changes requested by the CIE reviewer, showed no retrospective bias in F and SSB (no plot for recruitment was given). The headboat discards observer-based index was less precisely fitted than is the case with the current model configuration, which may have been a contributing factor. However, the current BAM implementation fits the short headboat discards index series very closely, even in additional runs requested by the Review Panel in which all CPUE series were equally weighted rather than iteratively re-weighted. The source of the discrepancy in the retrospective patterns in the initial and revised BAM assessments from SEDAR-24 should be explored.

The appropriateness of the assumed Beverton-Holt stock-recruit function fitted within the model

The current implementation of BAM for red snapper fits an implausibly large steepness value (close to 1.0) if this parameter is freely estimated. This is effectively equivalent to assuming a geometric mean for the expected recruitment at almost all stock sizes. Forcing smaller values of steepness (0.85 was used in the base case, based on the mode of a meta-analysis for a range of stocks) drives the R_0 and B_{msy} to higher values. In the base case model, the additional bias correction using an assumed residual error (σ^2) resulted in an R_0 within the range of the highest recruitment values estimated historically in the model. An additional run was requested by the Review Panel to examine the result of specifying a Ricker model. This resulted in R_0 hitting the upper threshold specified in the model, leading to an almost linear stock recruit curve over the range of estimates, which would imply implausibly large recruitment and SSB at low F (assuming the natural mortality profile used in the assessment). Although the run with the Ricker model did not converge, it illustrated that the model results could be quite sensitive to the assumed form of stock recruitment. The robustness of fitting stock-recruit models within the BAM model fitting procedure should be reviewed for future assessments.

Derivation of the numbers at age in the first year (1955)

The initial population numbers at age in 1955 (the first year used in the BAM model) were derived from the stable age structure computed from expected recruitment and the initial age-specific total mortality rate. This mortality rate was the sum of natural mortality and fishing mortality, where fishing mortality was the product of an initial fishing rate (F_{init}) and catch-weighted average selectivity. The initial fishing rate was chosen using an iterative approach. First, the assessment model was run using the nearly complete catch history (starting from the year 1901) provided by the DW, to indicate a plausible level of biomass depletion in 1955 (B1955/B0 ≈ 0.8). Then, F_{init} was adjusted to approximate that level of depletion. The value used in the base model run was F_{init} = 0.02. The model using the complete catch history to indicate the level of depletion in 1955 was not reviewed by the Review Panel. An indication that F_{init} of 0.02 may have been too low was a larger than expected number of fish in the plus group (i.e., age 20+) in the recreational catches in comparison with the observed age composition of the fishery catches prior to 1990 (AW report Fig. 3.12). To an extent this could have resulted from an inappropriately low natural mortality value assumed for earlier years in the series, but could also mean that the cumulative fishing mortality prior to the 1980s was higher than expected from F_{init} = 0.02. The Review Panel requested that the base model configuration be rerun while increasing the value of F_{init}, until the plus group residuals were removed (see Review Panel summary report for model outputs). Higher values of F_{init} resulted in a better fit to the recreational fishery 20+ age compostions, and a better fit to the data overall. However, the implied depletion of the stock in 1955 was very large for values of F_{init} greater than 0.1 (depletion to 21% of unfished stock at F_{init} = 0.15, and to 8% at F_{init} = 0.25). Because the poor fit in the base run may also be explained by a misspecification of the for-hire fishery selectivity, the Review Panel decided not to recommend a change to the F_{init} value used in the base run. The concept of relative biomass depletion in 1955 may however be a construct that should not be over-interpreted (or possibly even presented), as the dynamics of the stock (growth, natural mortality, maturity, etc.) may have been quite different between the 1900s and the 1950s, or if the fishery removals in this period are under-estimated.

Handling missing data, and consistent treatment of weight data
The historical removals estimates are incomplete, or require manipulation to ensure consistency over the time series. Much of this is done external to the model, for example the ratio method to estimate historical recreational landings. As noted under ToR1, there are large inconsistencies between mean weights used for converting historical commercial landings weights into numbers for predicting recreational fishery removals, and the BAM model estimates of mean weight for the different fleets. It would be an improvement to reconfigured BAM to estimate missing data internally where possible, rather than using external data manipulations that can lead to inconsistencies in variables such as input and output mean weights. The estimation could make use of prior distributions for values such as mean fish weight.

Surplus Production model (ASPIC)

In principle, ASPIC is an adequate and appropriate method to explore the robustness of the results from the BAM to structural assumptions. The ASPIC model appears to have been applied correctly. The F/F_{msy} values from ASPIC were at a lower scale compared to BAM, indicating a lower level of over-fishing. The values of B/B_{msy} from ASPIC were below 1.0 over the entire assessment time frame (1955-2009), whereas BAM indicated biomass above B_{msy} prior to 1970. BAM also indicated that current (2009) biomass is much less than B_{msy} (i.e. 10%), whereas ASPIC is somewhat more optimistic (B_{2009}/B_{msy} = 0.39; B_{2010}/B_{msy} =0.25). ASPIC is run from January 1, so the 2009 and 2010 biomass ratios bracket the BAM estimate, which is computed at the time of peak spawning (mid-year). It is noteworthy that both the BAM and ASPIC base runs resulted in positive residuals in the 1976 – 1982 headboat CPUE data and negative residuals in the 1992 - 2000 data. In general, ASPIC is a more limited stand-alone assessment model for red snapper because it does not use available age and length data, although this conclusion is conditional upon the available length and age data having sufficient coverage and quality to support the implementation of a more complex age-structured model.

The annual fishery removals (in weight) input to the ASPIC base-case model were not directly comparable with the values used in BAM. This is due to a ~ 3 times lower mean fish weight applied to the recreational fishery landings numbers in the 1950s-1980s than the mean fish weight for these fisheries estimated within BAM according to fishery selectivity patterns and population structure (see Fig. 1 above). This leads to a totally different interpretation of the trend in fishery removals weights input to ASPIC and estimated by BAM (Fig. 2: provided by assessment team at the RW). This leads to quite different interpretations of historical stock trends and initial stock depletion. ASPIC estimates of F/F_{msy} since the 1980s are around 50% of the BAM estimates, and the estimated rate of decline in biomass between the 1960s and the 1990s is an order of magnitude less than given by BAM. The base ASPIC run nonetheless indicates a very high probability that the stock is overfished and that overfishing is occurring, although the estimates of current stock status are relatively imprecise.

The RW asked for additional ASPIC runs using the BAM removals estimates as inputs. This resulted in very similar F/F_{MSY} values for recent years to the base case ASPIC run. The recreational CPUE series do not appear to have been adjusted to use the mean fish weights predicted by BAM, so it is not clear if the ASPIC comparisons using the two different landings streams are directly comparable. It is again suggested that historical observed mean weights, and the method of using these figures in the assessment, are reviewed to ensure internal consistency in BAM and consistent use of such data between BAM and other models such as ASPIC.

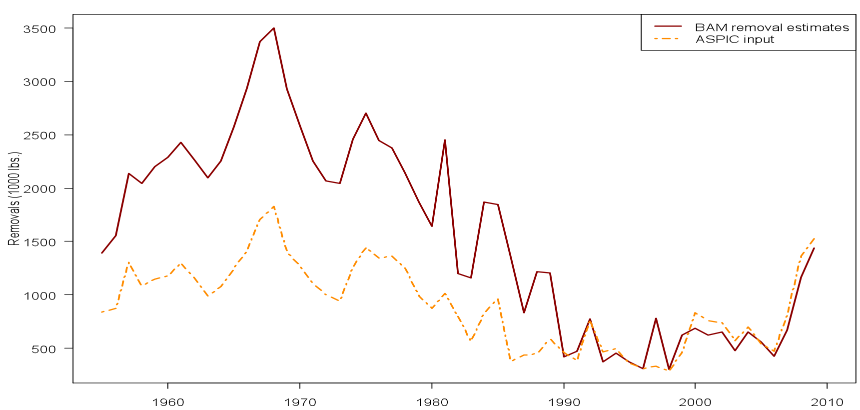

Figure 2. Landings (1000 lbs) as input to the ASPIC model and as estimated in BAM, 1955-2009.

As discussed above under ToR1, *Abundance indices*, the BAM accounts for the effects on CPUE indices of changes in selectivity, whereas the ASPIC model will treat the CPUEs as indices of total stock biomass, not biomass weighted by selectivity. The headboat CPUE for ASPIC (Fig. 3.72 in AW report) shows a ~ 50% downward step change in 1992 which coincides with the large change in MLS to 20 inches in that year. As suggested earlier, possible changes in fishing behavior associated with the MLS should be investigated to determine if there could have been a persistent change in catchability in addition to the change in selectivity.

Catch curve analyses

Catch curve analyses were carried out using a regression estimator (slope of log catch at age values, with various treatment of zero values) and the Chapman-Robson method based on mean age above a specified recruitment age (SEDAR24-AW7). Estimates were based on annual age compositions ("synthetic cohorts") rather than tracking cohorts through time. The age range chosen for the analysis was 4 – 12. The catch curve Z estimates for 1995-2005 in the commercial line fishery (which is assumed to have asymptotic selectivity from age 4 since 1992) were generally in the range 0.4 – 1.1. The equivalent mean Z from the BAM base run calculated as the arithmetic mean across ages 4-12, years 1995-2005, was 0.30 (AW report Table 3.7). A simple linear regression of log of mean population numbers at age against age for 1995-2005 from the BAM base run gives a slope of 0.43 (R^2 0.99). The catch curve analyses therefore tend to give larger Z estimates than the BAM. The differences probably reflect the fitting of selectivity curves in BAM. (The mean of the BAM apical values at age 3 over the same period was 0.66, however these are not comparable to the catch curve estimates due to domed overall selectivity.) The catch curve analyses provide weak support for the assumption that the selectivity of the headboat fisheries was more domed-shaped than the selectivity of commercial fisheries.

The catch-curve Z estimates for 1978 – 1989 headboat landings tend to be lower than those from 2002 onwards, illustrating a less steep age profile in the earlier period that will reflect mortality during this period as well as the cumulative mortality down cohorts in years prior to 1978. The BAM indicates a lower mean F over ages 4-12 during the 1970s compared to subsequent decades, which would explain the less steep age profile in 1978-89 compared with 2002 onwards. I would conclude that catch-curve analyses based on “synthetic cohorts” are informative only as a method of summarizing changes in catch age compositions over time, and cannot be used as a confirmation of the BAM results. Both methods are estimating F based on the same catch age composition data, the only real difference (apart from effects of CPUE tuning in recent years) being the imposition of selectivity patterns in BAM, including user-fixed parameters such as the saturation selectivity.

Other methods

A stochastic stock reduction analysis (SSRA) was briefly reviewed at the assessment workshop, but not included in the workshop report or Review Panel presentation.

ToR 3. Recommend appropriate estimates of stock abundance, biomass, and exploitation.

The AW provided estimates of stock abundance, biomass and exploitation based on the BAM base-case run. As the BAM application was considered by all the reviewers to be appropriate and adequate for the type of data available, the AW adequately addressed their equivalent ToRs. As discussed above, there are considerable uncertainties in historical fishery removals and composition, and a number of concerns regarding the application of the BAM and the absolute estimates of benchmarks such as B_{msy}. The BAM base-case run is one realization of a number of plausible runs and is conditioned on particular assumptions made about the data and population dynamics model that may change in future assessments. However, all sensitivity runs of the BAM model carried out by the AW, and additional ones requested by the Review Panel, show the same qualitative results indicating the stock is overfished and suffering from overfishing. A range of model configurations provided apparently plausible interpretations of the underlying data sets that could lead to qualitatively different projection results; however, it is difficult, on the basis of the material provided, to identify a unique ‘best estimate’ model run. For example, the iterative re-weighting procedure introduced following the AW meeting is an appropriate method for fitting this type of statistical model, but may need reconfiguring to avoid spurious over-fitting of short data series. Model runs with and without iterative re-weighting provide different interpretations of current abundance and fishing mortality that could affect projections, but there are equally valid arguments for either model formulation.

Estimates of stock abundance, biomass and exploitation were also provided from the ASPIC runs. There are a number of issues concerning the input data for the ASPIC which will have affected the results (see above). However, the base ASPIC run indicates a very high probability that the stock is overfished and that overfishing is occurring, although the estimates of current stock status are relatively imprecise. In this respect the ASPIC and BAM models provide the same conclusion about stock status, although the ASPIC model suggests a lower extent of overfishing in recent years, and an order-of-magnitude smaller relative decline in biomass between the 1960s and the 1990s, compared with the BAM model results.

ToR 4. Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies*); recommend appropriate management benchmarks, provide estimated values for management benchmarks, and provide declarations of stock status.

The AW adequately addressed their ToR on benchmarks and management parameters by providing values from on the BAM base case, and evaluating the sensitivity to different assumptions regarding input data and parameters. Alternative estimates were derived from the ASPIC model. The ability to compute accurate estimates of MSY and B_{msy} using the BAM is dependent on being able to define the shape of the stock-recruit function. Unfortunately, the shape and parameters of the stock-recruit function are poorly defined in the present assessment, and this has resulted in considerable uncertainty regarding absolute benchmarks. Within the domain of uncertainty explored in the sensitivity analyses, estimates of the status of the current stock relative to the benchmarks (i.e. the ratios $F_{current}/F_{msy}$ and $SSB_{current}/SSB_{msy}$) appeared relatively robust. In all the BAM sensitivity runs, the ratios estimated the stock to be overfished and experiencing overfishing, despite the absolute values of the individual quantities in the ratios varying substantially. The ASPIC model also indicates the stock is currently overfished and experiencing overfishing, although to a lesser extent than indicated using the BAM. The conclusion of the status of the stock therefore appears quite robust to a wide range of model configurations.

Recent F estimates could be compared to yield-per-recruit proxies for MSY benchmarks, such as $F_{40\%}$ or other similar values. The form of the yield curve obtained from BAM would be expected to be similar to the yield-per-recruit curve if stock-recruit steepness is close to 1.0 (i.e. recruitment effectively independent of SSB). This is confirmed by an F_{msy} value of ~0.20 from the sensitivity run with steepness set at 0.95 or estimated (giving a value>0.95), which is close to the F_{max} from yield-per-recruit. Values of F for 30%, 40% and 50% depletion of spawning potential ($F_{30\%}$, $F_{40\%}$ and $F_{50\%}$), from the BAM, are given in Table 3.14 of the AW report as 0.17, 0.125 and 0.092. These are close to the F for the same depletion of spawning potential given in the yield-per-recruit analysis (AW report Fig. 3.40). Hence, given the large steepness value for the Beverton-Holt S/R function imposed in the BAM fit, there is little difference between the benchmark F reference points from the fitted production curve and the equivalent values from yield-per-recruit.

During the Review Workshop, the Review Panel requested that the BAM model be run using a Ricker stock-recruit model in a base model configuration. Preliminary results from this analysis suggest a substantial change in the estimated stock-recruitment scatter plot with current stock status closer to the MSY benchmarks than the base-case run fitting a Beverton-Holt function with fixed steepness. The run fitting Ricker parameters resulted in R_0 hitting the upper limit set in the model (i.e. no convergence) and recruitment appeared almost linearly related to SSB over the observed recruitment range. This would suggest that the calculation of MSY benchmarks is sensitive to the choice of recruitment function in BAM. The form of stock-recruitment relationship for this stock, and the ability of BAM to fit a stock-recruitment relationship internally, should be investigated further.

The ASPIC runs indicated that the stock status was closer to F_{msy} than given by the BAM. This could have resulted (at least partly) from the different catch streams used in the respective stock assessment models.

ToR 5. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; recommend appropriate estimates of future stock condition (e.g., exploitation, abundance, biomass).

Projections carried out by the AW are conditioned on the base run of the BAM, which the Review Panel agreed was adequate and appropriate for characterizing the current stock abundance, age structure, and fishing mortality rates as one of a range of plausible runs. The method involves a deterministic projection assuming a 10% reduction in fishing effort in 2010 caused by the moratorium, and an assumption that all catches under a moratorium would be discarded and subject to the discard mortality rate used in the assessment. A stochastic model was also used to project the Monte Carlo and bootstrap runs of the base case model with additional uncertainty in the F reduction in 2010 (reduction to between 80% and 100% of current estimates) and process error in recruitment based on the assumed variance of log recruitment residuals (σ^2). The methods used in the projection are adequate and appropriate, but there are a number of concerns regarding the application:

- The anticipated reduction in effort under the moratorium was based on expert opinion, but the basis for that decision is not clear;
- Future stock growth is critically dependent on the values of predicted recruitment. The deterministic projection uses a bias-corrected stock recruit function according to the assumed $\sigma^{2,}$ rather than the non-bias corrected version that might be considered to provide the most probable values. The AW did not provide the criteria for this choice, although it is likely to be to ensure compatibility between the future abundance and catches from deterministic projection and the arithmetic means from the stochastic projections. The choice of σ^2 also affects the estimation of benchmarks.
- Although the stochastic projections include uncertainty obtained from the Monte Carlo bootstrap runs, these substantially underestimate the true uncertainty in the current stock status used to initiate the projections. This reduces the accuracy of the projections aimed at estimating the probability of achieving management target.

The use of deterministic projections to evaluate the relative rebuilding time under different management scenarios remain useful as a guideline. It is clear that current levels of exploitation are likely to lead to further stock depletion in the long term and, given the present level of depletion relative to the estimated B_{msy}, rebuilding times under the explored scenarios of reduced exploitation will be very long (on the order of decades).

The BAM model estimates of population numbers indicate the current stock is mainly fish of ages 1 to 12, and hence the estimated current population numbers will contribute substantially to the short-term projections. Therefore, the short-term projections are more reliable.

A moratorium or other measures restricting retained catches of red snapper without an equivalent reduction in effort will cause discarding over the full size range, and thus the accuracy of the projection outcomes become critically dependent on the accuracy of the discard mortality estimates. The projections indicate that under an assumed 10% reduction in effort during a continued moratorium, discard mortality will prevent recovery to B_{msy}. Any future measures to reduce discard mortality will benefit the stock, but it has not been possible to explore possible scenarios for this in the present projections.

ToR 6. Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Provide measures of uncertainty for estimated parameters. Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Uncertainty in the assessment was explored using a Monte Carlo bootstrap of the assessment, a range of sensitivity analyses around the base BAM run; and the use of ASPIC as an alternative assessment model. These are appropriate approaches, although the results will only reflect the form and range of uncertainty explored which for the red snapper assessment were probably somewhat restricted. For example, the Monte-Carlo parametric bootstrap of the base BAM assessment run used CVs set for some input parameters that appeared very small, especially on quantities such as landings and F_{init} that are not well known and which will likely underestimate the uncertainty in the MSY quantities. Also, the bootstrap procedure only included the measurement error CVs for CPUEs, and not the model residuals, which in some cases are quite large and probably reflect additional year-effects in catchability.

Sensitivity runs were comprehensive in investigating the likely areas of uncertainty in the BAM model, and all sensitivity runs resulted in the same stock status of overfished and suffering overfishing. However, the range of perturbation for each parameter was generally quite small. This means the analysis will provide estimates of the direction and rate of change near the nominal values, but will not necessarily explore the full range of plausible assessment runs. More analyses are required to explore sensitivity to the structural assumption about recruitment, F_{init}, and the effect of iterative re-reweighting on the model fit. A trial run of the BAM with a Ricker curve for recruitment suggested this effect could be large and merits further investigation. A problem with the sensitivity analysis is that the relative plausibility of sensitivity runs is not evaluated. For example, does a lower M result in worse model fit than a higher M?

The robustness of the benchmarks and management variables to choice of assessment model was explored only through the application of a surplus production model (ASPIC). The intention was mainly to demonstrate the effect of using an assessment that excludes length and age data, but otherwise utilizes the same data on fishery removals and abundance indices. Also, the implied stock-recruit function in the model differs from the Beverton-Holt model implemented in BAM. In practice, the review Panel discovered that the data were not equivalent, as the pre-1990 time series of total landings estimates were very different in the ASPIC inputs and the BAM model predictions. The Review Panel requested a re-run of the ASPIC using the BAM-predicted landings. However the latter are also subject to bias due to the estimation of mean fish weights. It is therefore difficult to interpret ASPIC results as a measure of sensitivity to model structure.

The main feature of the ASPIC-BAM comparison is that ASPIC runs also place the stock in the 'overfished-overfishing' category, although current F is estimated to be much closer to F_{msy} than given by the BAM model. The relative rate of depletion of biomass over time is also much less in ASPIC than in BAM, although biomass is estimated to be below Bmsy over the full series (if catchability is assumed to have a 2% annual increase up to 2003). Diagnostics of the ASPIC model fit show a poor fit to the early headboat CPUE data, although this is also a feature of the base-case BAM run.

The use of different approaches is important in exploring model uncertainty and is a valuable element of the assessment report, especially in getting some insight into the uncertainty in the catch and how this affects the level of stock depletion. However, it would be valuable to try other models that make different structural assumptions to get a wider view of the robustness of the assessment. One obvious candidate would be a state-space (e.g., Kalman filter) analysis.

ToR 7. Ensure that stock assessment results are clearly and accurately presented in the Stock Assessment Report and that reported results are consistent with Review Panel recommendations.*

The stock assessment results were clearly and accurately presented in the SEDAR Stock Assessment Report and the results are consistent with Review Panel recommendations.

ToR 8. Evaluate the SEDAR Process as applied to the reviewed assessment and identify any Terms of Reference which were inadequately addressed by the Data or Assessment Workshops.

The SEDAR process applied to the south Atlantic red snapper assessment was extremely effective in collating and evaluating all available sources of information, and ensuring that all decisions related to use of data or application of assessment models were made objectively. A potential shortcoming of separate data and assessment workshops is that the DW may devote a lot of time to adjusting or imputing historical data when some of this would be better done internally in the assessment model. Issues such as this would better be addressed collaboratively between the data and assessment experts at the time of the DW.

The documents relevant to the Review Workshop were received approximately one week before the panel convened, rather than the two weeks stated in the Terms of Reference. The delay reduced the time available to read background documents, although this was mitigated by the thorough presentations provided by the stock experts.

The Data and Assessment Workshops adequately addressed their Terms of Reference. The reports were generally clear and well laid out. Due to the complexity of the data and the various methods used for adjustment of historic data or imputation of missing data, the DW report was often difficult to follow and there was no clear summary of the source of annual data input to the BAM and any manipulations that were carried out. Tables 5.1.1 and 5.2.1 in the DW report provide useful summaries for CPUE data, but there were no equivalent summaries for the other data (although there is some overlap). An example of a useful summary table could include a column for each type of data input to the model (e.g. landings by fleet; CPUE by fleet; length / age by fleet; mean weights by fleet, and an entry by year to indicate that the data were used, with superscripts and footnotes to identify the source of the data, the units (e.g. numbers or weight) the imputation method (if used) or the adjustment/re-scaling method used (e.g. which ratio method, if used). This would greatly help the reviewers who are presented with a very large amount of written material.

While recognizing that resources within the government available to conduct stock assessment are limited, I agree with the overall Review Panel view expressed in the RW Summary Report that the assessment of red snapper would have benefited by having more than one assessment team deriving the benchmarks. This would broaden perspectives, and use of alternative models and data structures to cross-validate the information that is ultimately used to provide the scientific basis for management advice.

An extended discussion took place at the RW over the iterative re-weighting procedure that had been applied following advice from a CIE reviewer following the initial draft AW report. Whilst the reviewer's advice was statistically valid, the assessment team had relatively little time to consider how this could best be implemented. I agree with the overall Review Panel view in their summary report in encouraging re-thinking of the way in which CIE expertise is used during the Stock Assessment Workshop, so that the most appropriate solutions can be developed to address advice from external experts.

ToR 9. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment, and whether a benchmark or update assessment is warranted.

A large number of recommendations for further work were proposed by the DW and AW. These are mostly sensible proposals that should lead to improvements in future assessments. Some of these proposals are already being implemented (e.g. a fishery independent survey), and this is to be commended. The proposed development of spatial assessment models, whilst sensible in principle (especially given the strong site fidelity of red snapper), tend to be very hard to implement.

The Review Panel emphasized some DW/AW recommendations and added some additional recommendations, categorized as more important (Tier 1) and less important (Tier 2). I agree with all of these and list them as given in the RW Summary Report:

Tier 1

- Investigate alternate stock recruitment models, and in particular the robustness of stock status conclusions to reasonable alternative stock-recruit assumptions.
- Consider estimating missing catch (e.g., recreational) within the model to improve consistency. An example of such an approach is the B-ADAPT model applied to North Sea cod.
- Review historical records for determining historical average weights of fish. This is consistent with a DW recommendation.
- The Review Panel agreed with the DW and AW recommendations to improve age sampling. In particular, this should improve the estimation of fishing mortality in BAM.
- The Review Panel agreed with the DW and AW recommendations to continue developing fishery-independent abundance indices, especially because assumed changes in catchability of CPUE indices for red snapper are uncertain.
- Explore changes in catchability in light of other species involved in the mixed species fisheries that catch red snapper. The Review Panel anticipates that changes in catchability may be consistent among some of these species.

Tier 2

- Consistent with the AW recommendation regarding "plasticity in life-history traits", the Review Panel recommends investigating for temporal variation in growth and maturation rates, especially when such characteristics often show a density-dependent response.

- Tagging studies can provide relatively direct estimates of fishing mortality and selectivity, growth rates, and other stock assessment parameters. Where possible, information from tagging studies that are representative of the stock as a whole should be incorporated into the assessment.

A number of other suggestions for additional analyses are given throughout my report.

The next benchmark should not be done until sufficient new data are available to warrant a full assessment. For example, if a fishery-independent survey is initiated for red snapper, it will take several years before data collected in that survey are useful for assessment purposes.

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 no later than November 1, 2010.

This report constitutes my independent peer review of the SEDAR-24 South Atlantic red snapper assessment. The Review Panel Summary Report was completed and submitted in the week following the Review Workshop.

5. Conclusions and recommendations

The main conclusion arising from the SEDAR-24 assessment of South Atlantic red snapper is that there is a high probability that the stock is overfished and is experiencing overfishing. This conclusion is supported by two different assessment models (BAM and ASPIC) and by sensitivity analysis around the base model configurations. However, the extent of overfishing, and the extent to which the stock is overfished, is relatively poorly determined because of uncertainties in data and model structure. For example, the stock-recruit function is poorly determined making it difficult to estimate MSY and B_{MSY} benchmarks. The BAM and ASPIC models also give different interpretations of the extent of biomass decline over time.

A moratorium on red snapper landings will place greater emphasis on estimation of discards, and on discard survival rates. The CPUE series will be strongly affected by continuation of the moratorium or by any spatial management measures or catch limits that might be introduced. This further emphasizes the need for good fishery-independent data on trends in abundance and stock structure, and for sufficiently precise estimates of catches, discards and length/age compositions, so that recovery of the stock towards management targets can be accurately estimated.

Appendix 1 Material provided for review

DW and AW reports

SEDAR24-DWR	South Atlantic Red Snapper SECTION II: Data Workshop Report	DW participants: June 2010 + corrections July 30
SEDAR24-SAR	South Atlantic Red Snapper SECTION III: Assessment Report	AW participants: September 2010

Other papers prepared specifically for October Review Workshop

SEDAR24-RW01	The Beaufort Assessment Model (BAM) with application to red snapper: mathematical description, implementation details, and computer code	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-RW02	Paper not completed, withdrawn on 9-29-10	
SEDAR24-RW03	Red snapper: Iterative re-weighting of data components in the Beaufort Assessment Model	Sustainable Fisheries Branch, NMFS 2010

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

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

External Independent Peer Review by the Center for Independent Experts

SEDAR 24 Review on South Atlantic Red Snapper Assessment

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 24 will be a compilation of data, a benchmark stock assessment, and an assessment review for US South Atlantic red snapper conducted under the Southeast Data, Assessment and Review (SEDAR) process. The assessment will be conducted for the South Atlantic Fishery Management Council (SAFMC), which has responsibility for management of the South Atlantic snapper-grouper complex fishery, of which red snapper is a member. The lead assessment agency will be the Southeast Fisheries Science Center of the US National Marine Fisheries Service (NMFS). Other entities involved in the data evaluation and assessment development processes will be the four US South Atlantic States, the NMFS Southeast Regional Office, other NMFS data providers and analysts, and fisheries representatives.

Red snapper is an important commercial and recreational fishery resource and is a focal species in the management of the US South Atlantic Snapper-Grouper complex. The most recent assessment was a benchmark accomplished in 2008, via SEDAR 15. Additional discard mortality and historic recreational fishery data have been acquired since SEDAR 15. The SEDAR 24 peer review will involve a panel composed of a chair named by SAFMC from its Science and Statistics Committee (SSC), two reviewers from the SAFMC SSC, and three CIE reviewers. The duties of CIE panelist shall not exceed 12 workdays; several days prior to the meeting for document review; three workshop days; and several days following the workshop to complete the independent peer review in accordance with the Terms of Reference, and to ensure final review comments and document edits are provided to the Chair. 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 expertise, background, and recent experience in stock assessment, statistics, fisheries science, and marine biology sufficient to complete their primary tasks (1) to conduct an impartial and independent peer review in accordance with the Review Workshop Terms of Reference to determine if the best available science is utilized for fisheries management decisions, and (2) to present the review in writing. 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 Savannah, Georgia during 12-14 October 2010.

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

<u>Prior to the Peer Review</u>: 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.

<u>Pre-review Background Documents</u>: 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.

<u>Panel Review Meeting</u>: 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.

<u>Contract Deliverables - Independent CIE Peer Review Reports</u>: 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.

<u>Other Tasks – Contribution to Summary Report</u>: 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 Savannah, Georgia during 12-14 October 2010.
3) During 12-14 October 2010 in Savannah, Georgia as specified herein, and conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than 22 October 2010, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and CIE Regional Coordinator, via email to David Sampson 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.

7 September 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
29 September 2010	NMFS Project Contact sends the CIE Reviewers the pre-review documents
12-14 October 2010	Each reviewer participates and conducts an independent peer review during the panel review meeting
22 October 2010	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
29 October 2010	CIE submits CIE independent peer review reports to the COTR
1 November 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. 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 (William Michaels, via William.Michaels@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) each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) each 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:

William Michaels, Contracting Officer's Technical Representative (COTR)
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov   Phone: 301-713-2363 ext 136

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

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

Key Personnel:

NMFS Project Contact:

Kari Fenske, SEDAR Coordinator
4055 Faber Place Drive, Suite 201, North Charleston, SC 29405
Kari.fenske@safmc.net   Phone: 843-571-4366

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 24 Review Workshop

1. Evaluate the adequacy, appropriateness, and application of data used in the assessment.
2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.
3. Recommend appropriate estimates of stock abundance, biomass, and exploitation.
4. Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies*); recommend appropriate management benchmarks, provide estimated values for management benchmarks, and provide declarations of stock status.
5. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; recommend appropriate estimates of future stock condition (e.g., exploitation, abundance, biomass).
6. Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Provide measures of uncertainty for estimated parameters. Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty. Ensure that the implications of uncertainty in technical conclusions are clearly stated.
7. Ensure that stock assessment results are clearly and accurately presented in the Stock Assessment Report and that reported results are consistent with Review Panel recommendations.*
8. Evaluate the SEDAR Process as applied to the reviewed assessment and identify any Terms of Reference which were inadequately addressed by the Data or Assessment Workshops.
9. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment, and whether a benchmark or update assessment is warranted.
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 no later than November 1, 2010.

* The panel shall ensure that corrected estimates are provided by addenda to the assessment report in the event corrections are made in the assessment, alternative model configurations are recommended, or additional analyses are prepared as a result of review panel findings regarding the TORs above.

Annex 3: Agenda

Savannah, Georgia during 12-14 October 2010

Tuesday

1:00 p.m.	**Convene**	
1:00 – 1:30	**Introductions and Opening Remarks** **Coordinator** - *Agenda Review, TOR, Task Assignments*	
1:30 – 3:30	**Assessment Presentation**	**TBD**
3:30 – 4:00	**Break**	
4:00 – 6:00	**Continue Presentation/Discussion**	**Chair**

Tuesday Goals**:** Initial presentations completed, sensitivity and base model discussion begun

Wednesday

8:00 a.m. – 11:30 a.m.	**Panel Discussion** **-** *Assessment Data & Methods* *- Identify additional analyses, sensitivities, corrections*	**Chair**
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion** - *Continue deliberations* *- Review additional analyses*	**Chair**
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session** *- Recommendations and comments*	**Chair**

Wednesday Goals**:** sensitivities and modifications identified, preferred models selected, projection approaches approved, Report drafts begun

Thursday

8:00 a.m. – 11:30 a.m.	**Panel Discussion** *- Final sensitivities reviewed.* *- Projections reviewed.*	**Chair** **Chair**
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion or Work Session** *- Review Reports*	**Chair**
4:00 p.m.	**ADJOURN**	

Thursday Goals: Complete assessment work and discussions, final results available. Draft Reports reviewed.

Appendix 3: Panel membership

Chair:	Anne M. Lange (CIE)	
Panel members::	Mike Armstrong	(CIE)
	Robin Cook	(CIE)
	Noel Cadigan	(CIE)
	John Boreman	(South Atlantic Science and Statistical Committee)