SEDAR

SouthEast Data, Assessment, and Review

SEDAR 24

South Atlantic Red Snapper
SECTION V: Review Workshop Report

October 2010

SEDAR
4055 Faber Place Dr., Suite 201
North Charleston, SC 29405

Section V: Review Workshop Report

Contents

1 Introduction

1.1 Workshop Time and Place

The SEDAR 24 Review Workshop was held October 12-14, 2010, in Savannah, Georgia.

1.2 Terms of Reference

Review Workshop Terms of Reference

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.

1.3 List of Participants

Attendees	Role	Affiliation
Michael Armstrong	Review panelist	CIE
John Boreman	Review panelist	SA SSC
Noel Cadigan	Review panelist	CIE
Robin Cook	Review panelist	CIE
Anne Lange	Review chair	SA SSC
Rob Cheshire	Analyst	SEFSC - Beaufort
Kyle Shertzer	Analyst	SEFSC - Beaufort
Erik Williams	Analyst	SEFSC - Beaufort
Kenny Fex	Appointed observer	Snapper/Grouper AP
George Geiger	Council representative	SA Council
Charlie Phillips	Council representative	SA Council
Myra Brouwer	Council staff	SAFMC
John Carmichael	Council staff	SAFMC
David Cupka	Observer	SA Council
Nick Farmer	Observer	SERO
Kari Fenske	SEDAR 24 coordinator	SEDAR
Patrick Gilles	IT support	SEFSC - Miami
Rachael Lindsay	Administrative support	SEDAR
Julie Neer	SEDAR coordinator	SEDAR
Gregg Waugh	Council staff	SAFMC
Rusty Hudson	Observer	
Kathy Knowlton	Observer	

1.4 List of Review Workshop Working Papers and Documents

Documents Prepared for the 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

2. Red Snapper Review Panel Summary Report

The stock assessment presented by the SEDAR 24 Assessment Workshop (AW) provided the Review Panel with outputs and results from two statistical assessment models and a catch curve analysis. The primary model was the Beaufort Assessment Model (BAM), while a secondary, surplus-production model (ASPIC) provided a comparison of model results. Based on the assessment provided, the Review Panel concludes that the stock is overfished and overfishing is occurring. The current level of spawning stock biomass (SSB_{2009}) is estimated to be about 10% of MSST (SSB_{2009}/MSST= 0.09), and the current level of fishing is four times F_{MSY} ($F_{2007\text{-}2009}/F_{MSY}$= 4.12). Numerous sensitivity analyses were also presented in the assessment, all of which agreed with the base model run conclusions of stock status. However, there were significant areas of uncertainty identified in both the data and in components to the model. The most significant sources of this uncertainty include: landings, the stock-recruitment relationship, and CPUE catchability.

The terms of reference from the Data Workshop (DW) and AW, in general, were met.

2.1. Terms of Reference

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

Overall, the Review Panel concluded that the data used in the assessment are adequate and appropriate for that purpose. The Review Panel did note some caveats that should be considered when interpreting the results of the assessment. First, and foremost, there is no reliable set of fishery-independent indices of abundance for red snapper in the region, which prevents validation of the fishery-dependent indices used in the assessment. Use of CPUEs from the commercial and recreational fisheries lack the adequate statistical design and spatial coverage that one would expect from a fishery-independent survey.

The data sets used in the assessment had gaps in historical information on catch, discards, and key biological characteristics, requiring use of various methods to fill in the missing data points. Although the methods used (indexing against commercial landings, averaging adjacent years, etc.) were adequate, the Review Panel notes that the methods required pragmatic assumptions that cannot be verified.

Data-smoothing techniques (cubic spline fits) were used to reduce the influence of "spikes" in the catch history data. The Review Panel questions the use of smoothing, since the smoothing process masks uncertainty associated with variability in the landings data stream. Caution should be used in the interpretation of the smoothed data sets in that regard.

Although the Data Workshop addressed potential spatial differences in growth and maturation rates of red snapper throughout its range in the South Atlantic, changes in those rates over time were not examined. One might expect to see a change in the rates as the overall population abundance declined to its current low levels.

The Review Panel noted that a more detailed review of the catch-at-age data might have helped to understand why the age data were down-weighted in the BAM. For example, an examination would be useful of how well age sampling tracked year classes through the fishery.

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. The Review Panel questions the decision to hold catchability constant since 2003, feeling it is somewhat counter-intuitive since factors other than GPS proficiency (e.g., rising fuel costs, improved means of communications) may also have affected catchability in recent years. It also might be useful to explore catchability of other species in mixed fisheries to determine if trends are evident.

2.1.2. 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. 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 was the recommended approach in the last assessment of red snapper (SEDAR 15). It is a statistical catch-at-age model implemented in ADMB, and developed by staff at the Beaufort laboratory. The software was customized to deal with the specifics of the red snapper stock, which is an advantage of using "inhouse" software. 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.

The implementation of BAM for SEDAR 24 was improved in several aspects compared to the version used in SEDAR 15. Most improvements were in response to CIE reviews at SEDAR 15 and the assessment workshop of SEDAR 24. The improvements 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.

It is noteworthy that the selectivity assumptions were well motivated in a working paper from the assessment workshop (AW-05).

The Review Panel concluded that BAM was adequate and appropriate for this assessment. The method was developed specifically to accommodate the available assessment data for this stock. The Review Panel concluded that BAM was applied correctly.

Surplus Production model (ASPIC)

The Review Panel concluded that ASPIC was an adequate and appropriate method to explore the robustness of the results from the BAM to other structural assumptions. ASPIC was applied correctly. Note that BAM fits to the available fishery catch statistics in the form in which they were collected (biomass for commercial landings and numbers for recreational landings), whereas ASPIC requires conversion of catch numbers to catch weight.

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).

The differences between BAM and ASPIC results are partially related to differences in the catch biomass time-series used by ASPIC, and the catch biomass time series inferred by BAM (see additional analyses requested: Section 2.2). ASPIC is a more limited stand-alone assessment model for red snapper because it does not use available age and length data.

Catch curve analyses

The Review Panel concluded that the catch curve analyses were adequate and appropriate for checking mortality rates estimated by BAM. The methods were applied correctly.

The catch curve values of Z and values for natural mortality suggested that the fully-selected fishing mortality rate was on the scale of 0.32 to 0.92, which is generally consistent with estimates from BAM.

These analyses also support the conclusion that the selectivity of the headboat fisheries was more domed-shaped than the selectivity of commercial fisheries.

Other methods

A virtual population analysis (VPA) was not considered, primarily because catch age composition data are only available for years with adequate sampling for age, resulting in blocks of years with missing data for the dominant fleets. The review group agreed that any reconstruction of the catch at age over the assessment time series (1955-2009) would contain substantial uncertainty in catches such that the application of standard VPA packages (e.g., ADAPT) would be tenuous, at best. It may be possible to develop a shorter, contemporary time series of catch at age with sufficient precision for the application of VPA, but this would be less useful for evaluating current stock status relative to MSY benchmarks.

A stochastic stock reduction analysis (SSRA) was briefly reviewed at the assessment workshop, but not included in the workshop report or Review Panel presentation. The Review Panel could offer no conclusions on this application.

2.1.3. Recommend appropriate estimates of stock abundance, biomass, and exploitation.

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, the panel found it 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 over-fitting the very short headboat discards index series, which includes a year with apparently large recruitment. 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.

The panel suggests using the AW base case model to provide historical and current estimates of stock abundance, biomass, and exploitation (AW Table 3.4), but cautions that this 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.

The panel considered the ASPIC model runs could potentially provide useful supporting information, as it is a quite different type of model that excludes length and age data. However, information requested by the Review Panel showed that the removals weights up to 1990 in the ASPIC input data were about half what the BAM predicted, whilst the recent data were more comparable (also see Sections 2.1.1 and 2.2, below). 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.

2.1.4. Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY*, F_{msy}, B_{msy}, *MSST, MFMT, or their proxies*); recommend appropriate management benchmarks, provide estimated values for management benchmarks, and provide declarations of stock status.

The most important aspect of population benchmarks and management parameters is to be able to judge relative position of the current stock to the benchmarks. In this context, absolute values of F_{msy}, SSB_{msy} are less important than the ratios $F_{current}/F_{msy}$ and $SSB_{current}/SSB_{msy}$. In all the model sensitivity runs and the ASPIC model the ratios estimated the stock to be overfished and experiencing overfishing, despite the absolute values of the individual quantities varying substantially. The conclusion of the status of the stock therefore appears quite robust to a wide range of model configurations and the panel felt this was the appropriate classification given our current knowledge of the stock.

One of the principal difficulties with the BAM model estimate of the stock recruitment parameters is that the steepness estimate appears unrealistically high. To address this, the AW used the mode of steepness values from a meta-analysis (0.85, while the mean in that analysis was 0.75). In addition, there are no data in the assessment to adequately define the asymptote of the Beverton-Holt function, and hence estimates of MSY indicators cannot be considered reliable. During the RW 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 suggested a substantial change in the estimated stock-recruitment relationship, and a substantial change in the assessment of stock status (e.g., Fs much closer to F_{msy}). This suggests that the calculation of MSY benchmarks is sensitive to the choice of recruitment function and needs to be investigated further.

The ASPIC runs indicated that the stock status was closer to F_{msy} than given by the BAM. This could partially result from the different catch streams used in the respective stock assessment models (see section on uncertainty below, Section 2.1.6, and 2.2), although additional runs using BAM-predicted landings, requested by the Review Panel, indicated that post-1980 estimates of F/F_{msy} from ASPIC were relatively insensitive to the catch streams used.

A general difficulty with the BAM-estimated MSY benchmarks is that the implied stock sizes lie well beyond the range of the data. It should be noted that these quantities are theoretical values derived from estimated population dynamics observed since the mid-1970s, and the assumptions currently used to derive MSY (M, maturity, growth, selectivity, productivity, etc.) may not hold at substantially higher stock sizes.

The benchmark values in the assessment are point estimates that do not consider stochasticity in recruitment. Values derived from a stochastic analysis would differ.

2.1.5. 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 panel considers 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 mortality 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 panel considers that the methods used in the projection are adequate and appropriate, but had a number of concerns regarding the application:

- The anticipated reduction in F 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, the panel considers these to substantially underestimate the true uncertainty in the current stock status used to initiate the projections (see 2.1.6). 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 F 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.

2.1.6. 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 has been explored using three general approaches:

- a Monte Carlo bootstrap of the assessment;
- a sensitivity analysis around the base BAM run; and
- the use of alternative assessment models.

These approaches are appropriate given their limiting conditioning assumptions. Overall, the Review Panel felt that the analyses were probably somewhat restricted in the range of uncertainty explored.

The base BAM assessment run was bootstrapped using a Monte-Carlo parametric bootstrap procedure, drawing values from predefined distributions on some of the input values. These runs provide distributions for management values of interest such as MSY benchmarks. Some of the CVs set for the input parameters appear to be rather 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 larger source of variation related to the precision of CPUEs for measuring trends in stock size (i.e., model residual variations).

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. Areas where the Review Panel felt more analyses are required are 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.

Model uncertainty was explored mainly through the application of a surplus production model (ASPIC, see 2.1.2 and 2.1.3). Unlike BAM, ASPIC cannot use age-structured data and relies on aggregate catch and CPUE indices alone. Nevertheless, it provides a valuable comparison, especially as the implied stock-recruit function in the model differs from the Beverton-Holt model implemented in BAM. While the ASPIC runs also place the stock in the 'overfished-overfishing' category, it is noticeable that F is much closer to F_{msy} than given by the BAM model. The difference between the ASPIC analysis and the BAM is at least in part the result of the way the catch data enter the respective models (see Section 2.1.2 and 2.1.3).

In addition to ASPIC, a simple catch curve analysis was performed that tended to support the Z values estimated from the BAM (see Section 2.1.2 for a description of this comparison).

The use of three 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 makes sense 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.

2.1.7. 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 Review Panel ensured that the stock assessment results were clearly and accurately presented in the SEDAR Summary Report for Red Snapper and that the results were consistent with the Review Panel recommendations.

2.1.8. Evaluate the SEDAR Process as applied to the reviewed assessment and identify any Terms of Reference which were not adequately addressed by the Data or Assessment Workshops.

The Review Panel members noted that the documents relevant to the Review Workshop were received approximately one week before the panel convened, rather than the two weeks stipulated in the Terms of Reference. This delay hampered a more thorough review by some of the panel members, although this was mitigated by the thorough presentations provided by the stock experts.

During the course of the Assessment 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 Review Panel considers 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.

While recognizing that resources within the government available to conduct stock assessment are limited, the Review Panel felt the assessment of red snapper would have benefitted 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.

The Review Panel suggests that future Assessment Workshop reports contain only figures and tables that are most important to the assessment, and put the remaining ones in an appendix.

Finally, the Review Panel encourages re-thinking of the way in which CIE expertise is used during the Stock Assessment Workshop. Having only one CIE expert reviewing the draft assessment report runs the risk of the expert's comments being biased in the direction of personal preferences and philosophy. Also, the CIE expert is asked to review and provide a critique of the draft report emanating from the assessment workshop, leaving little time for the analytical team to respond to the reviewer's suggestions, especially if major changes are made to the assessment model formulation and input data, before the assessment report is due to the Review Panel (a "sequential" review). Having CIE and some other form of independent expertise at the assessment workshop, even perhaps functioning on the assessment panel where they can interact directly with the other panel members (an "integrated" review), might allow more time to improve the assessment before it is delivered to the Review Panel.

2.1.9. 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.

The next benchmark should not be done until sufficient new data/information 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.

Research Recommendations

The Review Panel agreed with the DW and AW recommendations. However, the Review Panel was unsure of the specific benefits of pursuing spatial assessment models, which tend to be very hard to implement.

The Review Panel added some additional recommendations, categorized as more important (Tier 1) and less important (Tier 2).

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.

2.1.10. 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 Summary Report no later than October 28, 2010.

This report constitutes the Review Panel’s summary evaluation of the stock assessment and discussion of the Terms of Reference. The Review Panel will complete edits to its report and submit to SEDAR by 10/28/10.

2.2. Summary Results of Analytical Requests *(Sensitivities, corrections, additional analyses, etc.)*

The Review Panel suggested using the AW base-case model to provide an assessment of the red snapper stock, but cautions that this was one realization of a number of plausible runs. During the Review Panel’s deliberations a number of analyses were requested to clarify model results and to explore a number of the areas of uncertainty that were identified by the assessment. The following summarizes the issues for which the Review Panel required additional information and the analyses requested to address them.

1) The iterative re-weighting of the contribution of data components to the statistical likelihood 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.

The iterative re-weighting may have given too much weight to the HB discard index, which was a very short time-series. Also, the HB recreational index was given less weight, although the DW felt that this was the best among the three indices they recommended. The Review Panel requested the following analyses:

a- Provide MSEs for all components of the base (iterative re-weighting) and the equal-internal weight model runs, and runs with increased weights given to the HB index relative to the iterative re-weighted run.

Table 2.2.1. Mean Square Errors for headboat CPUE, commercial line CPUE and headboat discards under varying input weights, compared to base model.

Runs	weights	HB CPUE MSE	CL CPUE MSE	HB discard MSE
1	Base (iterated) hb=0.11	0.247	0.122	0.087
2	all 1	0.103	0.125	0.053
3	hb=0.2	0.165	0.158	0.085
4	hb=0.25	0.133	0.203	0.084
5	hb=0.3	0.108	0.257	0.08

The results demonstrated that the mean squared error (MSE) for HB CPUE residuals from the iterative re-weighted base run were over double the MSEs derived from the equal-weighted run. The HB discard index was actually fitted worse with re-weighting. The down-weighting of the HB CPUE may be related to 1-2 large residuals early in the time series. The Review Panel could not determine if any of the weighting schemes in the above table were more appropriate, and concluded that the base run (with iterative re-weighting) should be used to estimate stock status.

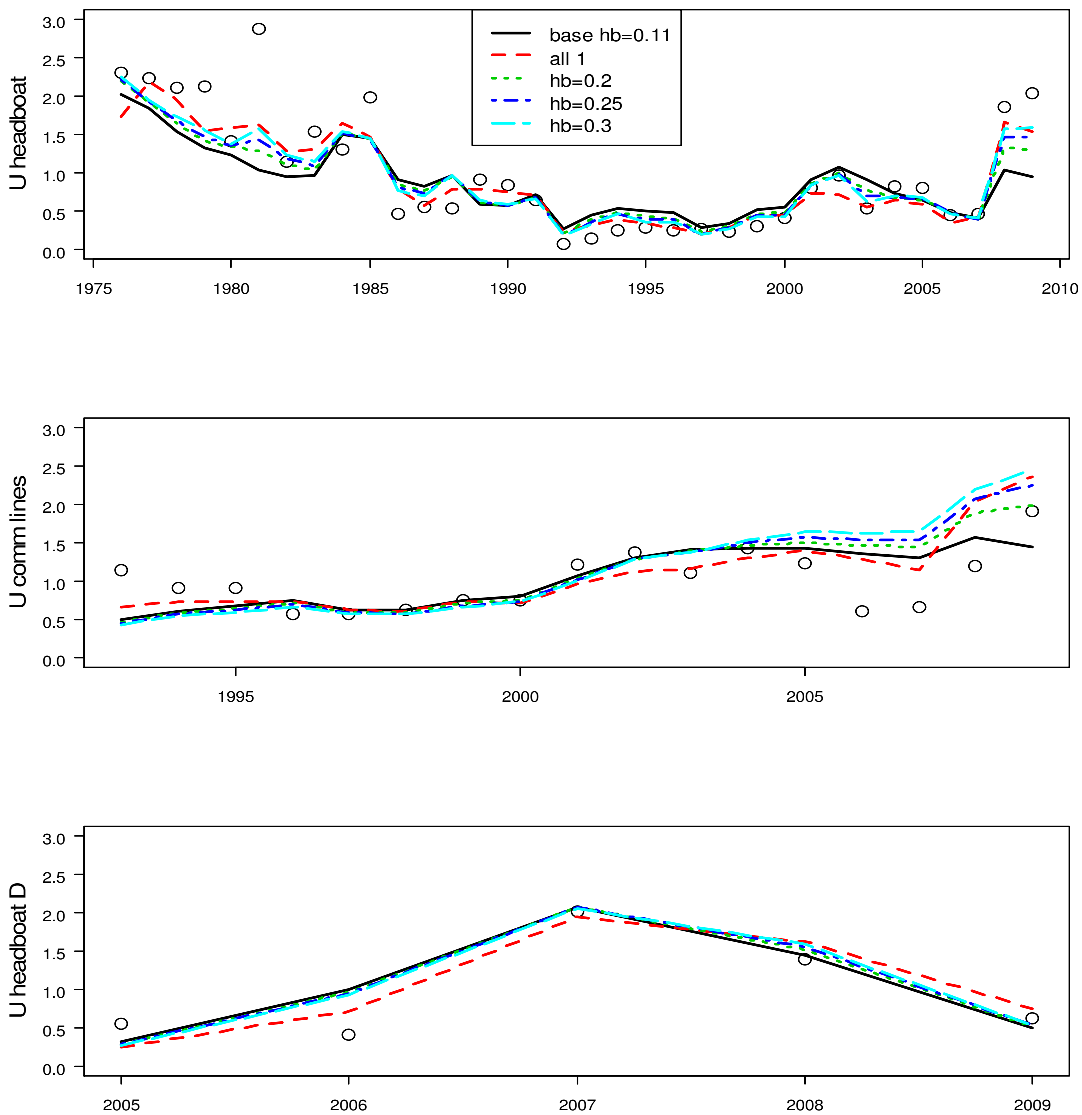
Perhaps one of the bigger consequences of re-weighting was the large reduction in weighting given to the recreational landings age-compositions. Age compositions are usually an important source of information for estimating F.

b- Rerun the base configuration model, while increasing the internal weight of the HB landings index, until the MSEs fall between the base run and the equal-internal weight runs resulting from task 1a.

Table 2.2.2. Standard deviation of normalized residuals (SDNRs) from data components, as indicated in row and column headings, using different configurations of data-component weighting. In the column labeled Iteration, "base" indicates base-run weights, including the headboat index weight of 0.11; "all 1" indicates all weights equal to one; "hb=0.X" indicates base-run weights, but with the headboat index weight increased to the value shown.

		For hire		Commercial	
Iteration	Data type	Landings	Discards	Lines	Dive
1 Base	CPUE	1.04	1.00	1.01	-
	Length comp	0.96	0.92	1.00	1.00
	Age Comp	1.05	-	1.01	1.00
2 All = 1	CPUE	5.60	1.67	5.20	-
	Length comp	2.93	2.96	2.84	1.04
	Age Comp	10.02	-	3.55	1.38
3 hb=0.2	CPUE	1.46	1.06	1.12	-
	Length comp	0.95	1.02	1.02	0.99
	Age Comp	1.02	-	.099	1.02
4 hb=0.25	CPUE	1.62	1.07	1.20	-
	Length comp	0.96	1.04	1.10	0.98
	Age Comp	1.01	-	0.96	1.00
5 hb=0.3	CPUE	1.73	1.07	1.29	-
	Length comp	0.97	1.05	1.22	0.96
	Age Comp	1.00	-	0.93	0.99

Figure 2.2.1 Fits to indices of abundance using different configurations of data-component weighting. In the legend, "base" indicates base-run weights, including the headboat index weight of 0.11; "all 1" indicates all weights equal to one; "hb=0.X" indicates base-run weights, but with the headboat index weight increased to the value shown. Top panel shows fits to the headboat index; middle panel shows fits to the commercial line index; and bottom panel shows fits to the headboat discard index.



2) Initial numbers at age in 1955 (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; 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. However, the low value of F_{init} resulted in a large plus group (i.e., age 20+) abundance, and was not consistent with the age composition information for the for-hire recreational landings during 1976-1990, which did not indicate a large plus group. This was the only source of age-composition data for this period.

To address concerns that the F_{init} appeared lower than would have been expected, the Review Panel requested that the base model configuration be rerun while increasing the value of F_{init}, until the plus group residuals are removed.

Table 2.2.3. Sensitivity analysis for BAM runs with increasing values of F_{init}

Sensitivity to F.init values

F.init	Fmsy	SSBmsy	MSY	F.Fmsy	SSB.MSST	steep	R0(1000)	B(1955)/B0	Avg 1978-83 HB ac resid 20+
0.02	0.178	156.01	1842	4.12	0.09	0.85	535	0.78	-0.095
0.05	0.175	168.35	1997	4.08	0.08	0.85	579	0.56	-0.076
0.1	0.173	186.32	2222	4.04	0.07	0.85	643	0.34	-0.055
0.15	0.172	203.74	2439	4.01	0.07	0.85	704	0.21	-0.04
0.2	0.17	223.91	2694	3.97	0.06	0.85	776	0.13	-0.029
0.25	0.168	255.52	3098	3.94	0.05	0.85	890	0.08	-0.017

F.init	nLL(data)	nLL(penalized)	nLL(SR)	U.cl	U.fh	U.fhd
0.02	858.838	889.041	15.902	8.259	17.758	2.013
0.05	856.753	885.547	14.816	8.259	17.456	1.961
0.1	854.43	881.513	13.629	8.273	17.04	1.904
0.15	852.803	878.459	12.759	8.288	16.699	1.863
0.2	851.451	875.627	11.958	8.292	16.371	1.829
0.25	850.107	872.358	11.03	8.269	15.963	1.791

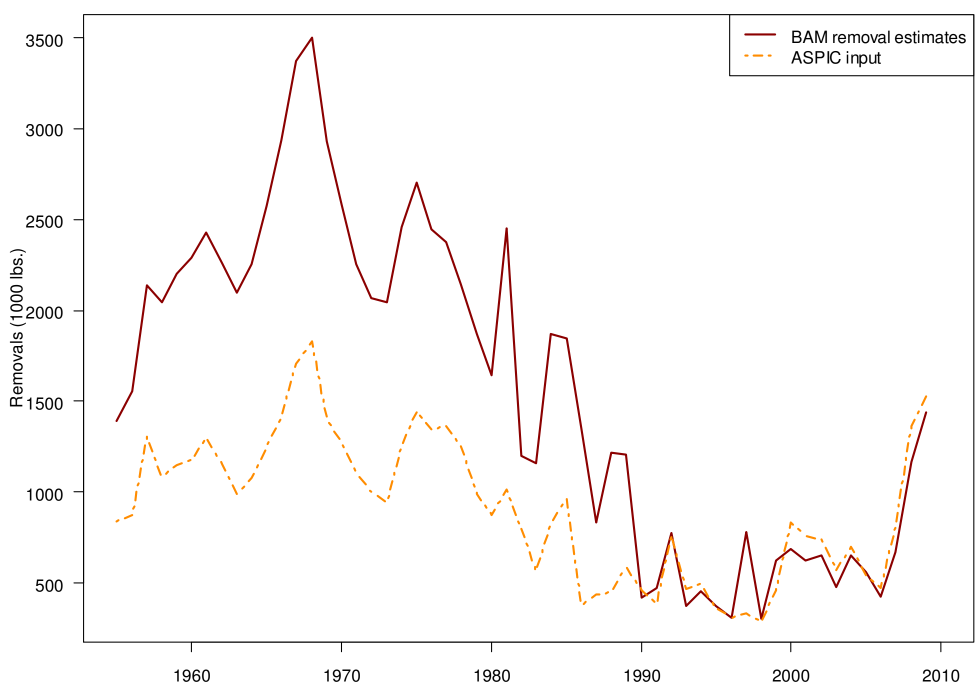
These results demonstrate that higher values of F_{init} resulted in a better fit to the HB 20+ age compostions, and a better fit to the data overall. However, the implied depletion of the stock in 1955 seemed implausible for values of F_{init} greater than 0.1. Because the poor fit in the base run may 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.

F.init	len.cl	len.cd	len.fh	len.pvt	len.cl.D	len.hb.D	age.cl	age.cd	age.fh	age.pvt
0.02	585.626	34.59	137.387	24.515	1.207	5.497	14.623	18.082	9.141	0.054
0.05	584.825	34.603	136.843	24.304	1.197	5.485	14.403	18.186	9.102	0.053
0.1	583.959	34.616	136.254	24.072	1.187	5.467	14.19	18.293	9.058	0.052
0.15	583.359	34.623	135.872	23.897	1.18	5.452	14.047	18.368	9.041	0.051
0.2	582.841	34.627	135.634	23.726	1.173	5.44	13.902	18.437	9.071	0.051
0.25	582.262	34.625	135.585	23.508	1.165	5.43	13.69	18.516	9.201	0.05

3) In order to better understand the differences in the results from the surplus production (ASPIC) and BAM models, the analysts were asked to provide total annual weights of landings that went into ASPIC, and annual total landings in weight estimated in the BAM base model run.

The analysts presented a plot of these catch streams and also reran ASPIC using each catch stream.

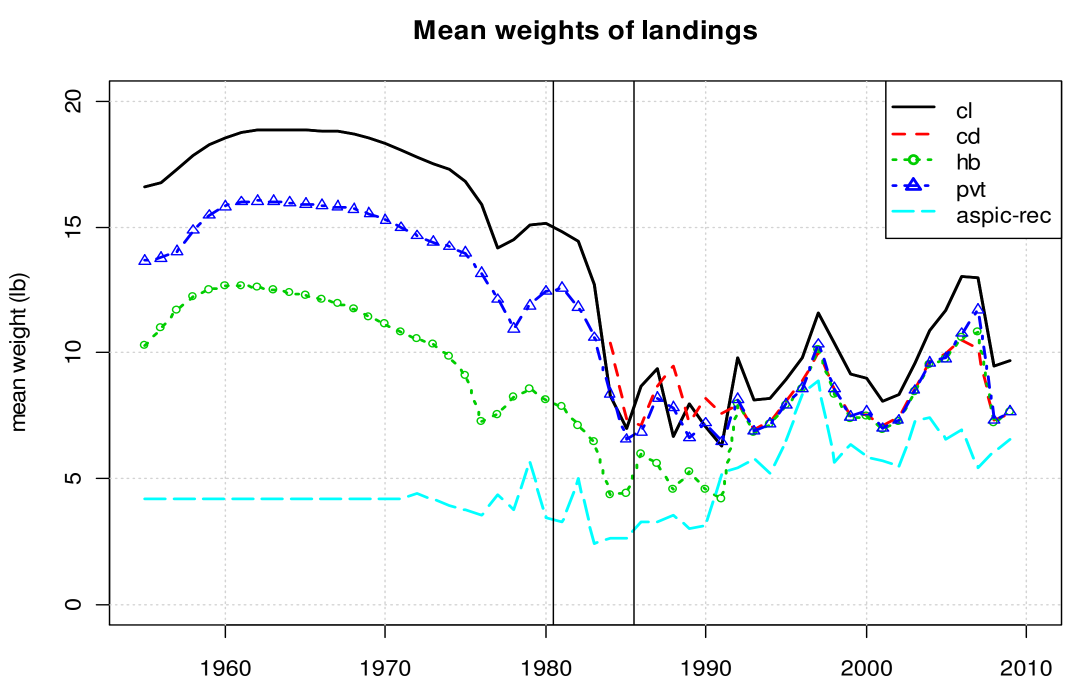
Figure 2.2.2 Landings (1000 lbs) as input to the ASPIC model and as estimated in BAM, 1955-2009.



The two landings series are very similar from 1990 onwards, but the BAM estimates for previous years are around double the figures used for the ASPIC run.

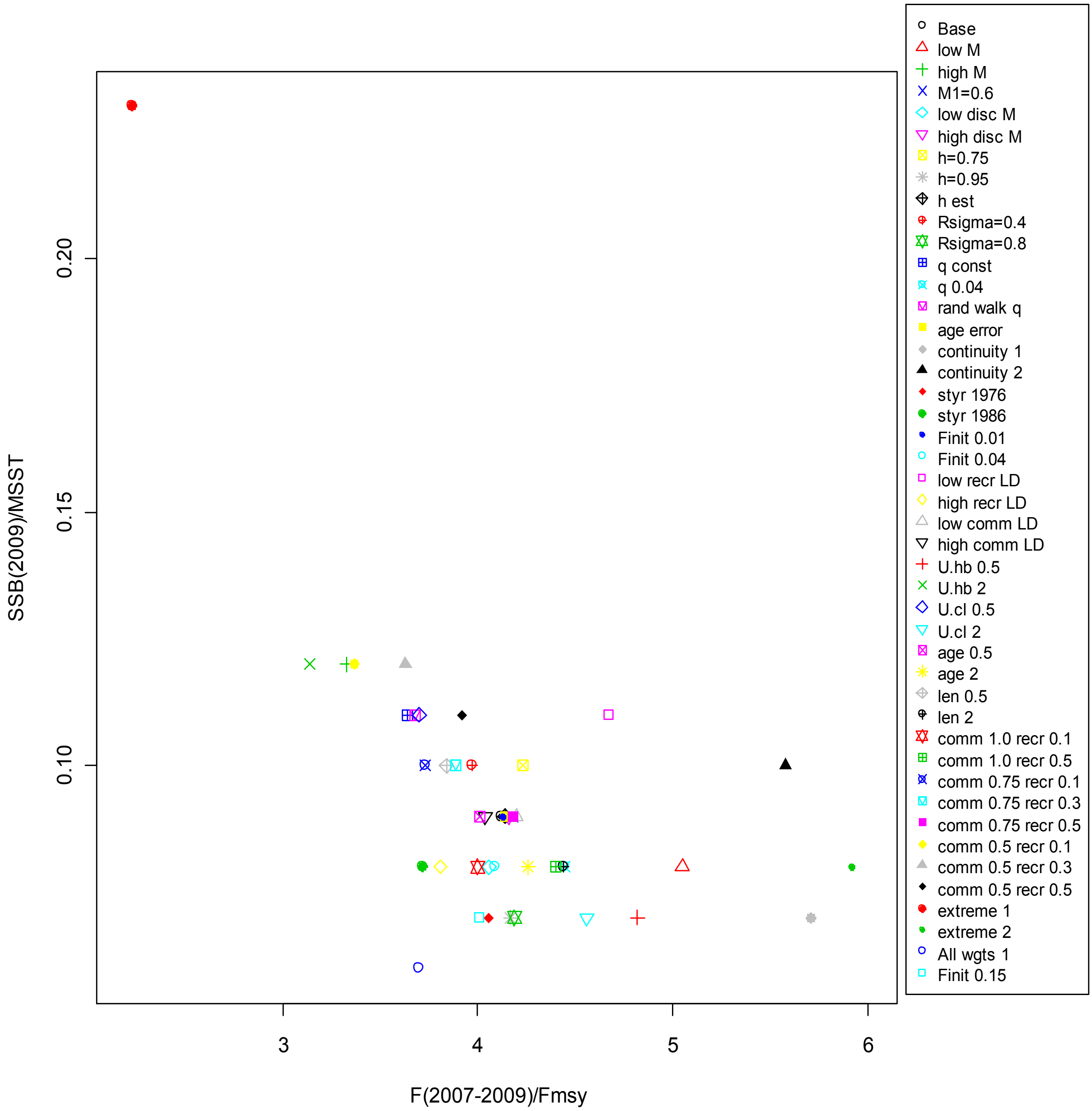
4) To help describe the differences between the ASPIC output and the base BAM results, the Review Panel requested that the analysts provide a plot of annual average fish weights in the landings by fleet from the BAM outputs and the equivalent average values for the recreational fleet from the ASPIC input data. The plots below show that the mean weights in the ASPIC data are much lower than the BAM model predictions for the commercial and recreational fleets, particularly during the early decades of the series. The Review Panel noted that an average weight of 9 lb had been used by the AW to convert commercial landings estimates into fish numbers for the 1955-80 period for use in predicting recreational catch numbers using the ratio method. If the model estimates of mean fish weight (14-18 lb) in the commercial landings are correct, this would imply a large overestimate of the historical recreational catch numbers using the 9-lb figure in the ratio method. The estimated age compositions of recreational catches, combined with the estimated selectivity parameters, lead to mean weights in historical recreational catches that are well above the 4.2 lb figure assumed for the ASPIC input data. The inconsistent treatment of weights in the BAM model appears responsible for the 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. The Review Panel recommends that the historical mean fish weights for the different fleets are thoroughly reviewed using additional evidence that may be available, and that the BAM model is adapted to ensure consistency in the way mean fish weights are estimated from input values.

Figure 2.2.3. Comparison of mean fish weights in landings, by fleet from BAM outputs and for recreational fleet from ASPIC input data.



5) The Review Panel also requested that the phase-plot (SSB/MSST by F/F_{msy}) figure be redone so the data points are more clearly visible. In addition, the analysts were requested to include results from runs using a higher F_{init} value (0.15) and all weights set to 1 (equally weighted). The analysts provided the following plot to address the Review Panel request.

Figure 2.2.4. Phase plot of terminal status estimates from sensitivity runs of the Beaufort Assessment Model, updated to more clearly show various points, and to include F_{init} = 0.15, and with all components equally weighted (from Figure 3.60 in AWR).



6) The Review Panel also requested that the analysts run BAM using the Ricker spawner-recruit model, fit internally to the model. They reported back to the Review Panel that this approach was tested but the model would not converge, and provided an implausibly large R_0.

2.3. Additional Comments (*if necessary, to address issues or discussions not encompassed above)*

The current configuration of BAM excluded length frequencies in years when adequate age compositions were also available, because these two data sources are not independent. A better approach would be to only exclude those length frequencies for which ages were obtained.

Presentation of sensitivity analyses would have been clearer if results were provided for both absolute stock size estimates and stock size estimates relative to reference points. This could occur in 2x2 panels of: (1,1) SSB with SSB_{ref} as a horizontal line; (1,2) SSB relative to SSB_{ref}; (2,1) F with F_{ref} as a horizontal line; and (2,2) F relative to F_{ref}.

The barplots of apical Fs by fleet should be better described in the figure caption.