Report on the 2003 Assessment of Yellowtail Snapper in the Southeast United States

NIWA Client Report: WLG2003-54
August 2003

NIWA Project: ERI04901

Report on the 2003 Assessment of Yellowtail Snapper in the Southeast United States

R.I.C.C. Francis

Prepared for

University of Miami

Independent System for Peer Review

NIWA Client Report: WLG2003-54
August 2003

NIWA Project: ERI04901

National Institute of Water & Atmospheric Research Ltd
301 Evans Bay Parade, Greta Point, Wellington
Private Bag 14901, Kilbirnie, Wellington, New Zealand
Phone +64-4-386 0300, Fax +64-4-386 0574
www.niwa.co.nz

Contents

Reviewed by:

Approved for release by:

Executive Summary

The 2003 assessment of yellowtail snapper in the southeast United States was reviewed as part of the SEDAR (South East Data, Assessment and Review) process. The Assessment Review Panel met 28-31 July 2003 at the Hilton Hotel in Tampa Florida. The assessment was presented to the Panel, additional analyses were requested and carried out, and the Panel discussed the assessment and wrote its two reports (one evaluating the assessment, and one on stock status).

The data used in the assessment appear to be the best available, and the assessment methods, and their presentation to the Panel, were of a high standard. I support the finding of the Panel that, according to the best available information, the stock is not overfished and not undergoing overfishing. However, I note that this conclusion is sensitive to assumptions about two key parameters (recruitment steepness and natural mortality) which are not well known for this stock.

Recommendations are presented which are intended to improve future assessments by improving

– the standardisation of CPUE

– the weighting applied to each data set

– the quality of age data

– the documentation of assumptions, and

– other minor matters.

Some suggestions are also made concerning the terms of reference of future Panels.

1. BACKGROUND

This report reviews the 2003 assessment of yellowtail snapper in the southeast United States, at the request of the University of Miami (see Appendix 1). The author was provided with a draft stock assessment report and web access to many associated files and documents (Appendix 2), and participated in the SEDAR (South East Data, Assessment, and Review) Assessment Review Panel Workshop that considered this assessment. This workshop constituted the last of the three phases of the SEDAR process, with the earlier phases being a data workshop (3-4 March) and an assessment workshop (9-13 June), both held in St Petersburg, Florida. The Panel also discussed data available for goliath grouper but that discussion is outside the scope of this report.

2. REVIEW ACTIVITIES

The Assessment Review Panel Workshop was held 28-31 July 2003 at the Hilton Hotel in Tampa Florida (see Appendix 3 for the Panel membership and a list of other attendees).

Bob Muller presented the draft assessment (see Appendix 2) which used two models: Integrated Catch at Age Analysis (ICA) and a fleet-specific model. He also presented some additional material, including more details of the CPUE (catch per unit effort) data and analyses. The panel discussed the assessment and requested some additional analyses. These were done and the results presented to the Panel (see below). The Panel drafted their two reports (one evaluating the assessment, and one on stock status) with input from others present.

Anne Marie Eklund presented a summary of available knowledge on goliath grouper. This was discussed, and a note added to the Panel's assessment-evaluation report. I make no further comment on goliath grouper because it is outside my terms of reference (Appendix 1).

2.1 Additional analyses

The additional analyses requested by the Panel, and the results from these, are described here very briefly. The reasons for requesting these analyses and the implications of their results are discussed more fully in Section 3.

A catch-curve analysis was done, and this produced an estimate of $Z = 0.54\ y^{-1}$ (Z is the total mortality expressed as an instantaneous rate).

Those GLMs (generalised linear models) for biomass indices that included terms for interactions with year were redone without these interactions. This made the commercial index slightly steeper and the MRFSS (Marine Recreational Fisheries Statistics Survey) index less variable and flatter. Many of the model runs (all which produced results for the Panel's stock-status report) were rerun with these new indices.

SDSRs (standard deviations of standardised residuals) were calculated for each data set for the main runs for both the ICA and fleet-specific models. These were all less than 1 and had higher values for the two visual-survey indices than for other indices (which suggests that the visual-survey indices were over-weighted).

The iterative reweighting facility of the ICA model was used to balance the weights assigned to each data set. This made little difference to the model outputs.

The ICA model was run with the early catch-at-age data (which is poorly known because of the lack of early otolith data) more strongly down-weighted. This produced unsatisfactory results because it denied the model information about total catches and the selectivity in the early years.

A retrospective analysis was done for the main run using the ICA model. This showed no strong retrospective trends.

The main ICA model run was rerun without the commercial CPUE biomass index. This made little difference to the model outputs.

The fleet-specific model was rerun with the total catch removed from the likelihood function. This made little difference to the model outputs.

3. FINDINGS

I was impressed by this assessment and the way it was presented to the Panel. The data used seem to be the best available and the approach to modelling was consistent with international best practice, with only relatively minor exceptions (see below). I agree with the assessment team's conclusion that production models were not useful for this assessment. I also support their decision to present results from two age-structured models (ICA and fleet-specific). There is little to choose between these models (each has its advantages and disadvantages) and the fact that different models produced similar conclusions (in terms of stock status) is reassuring and strengthens the overall conclusions of the assessment. The presentation of the assessment to the Panel was always clear and the assessment team was unfailingly helpful in response to requests for clarification or further analyses.

I support the general finding of the assessment team, and the Workshop, that, according to the best available information, the stock is not overfished and not undergoing overfishing. However, I note that this conclusion is sensitive to assumptions about two key parameters (recruitment steepness, *h*, and natural mortality, *M*) that are not well known for this stock.

My more detailed comments on the assessment, which are covered in the remainder of this section, fall into two parts: the first concerning the assessment; the other relating to the SEDAR terms of reference.

3.1 The assessment

3.1.1 Standardisation of CPUE

It is normal practice to standardise CPUE indices using GLMs, as was done in the assessment. This helps to ensure that these indices track abundance and are not affected by extraneous factors (such as changes in the region or season of fishing). There were three ways in which I thought the standardisations could be improved.

The first concerns model selection. By this I mean the decision as to which of the candidate predictor variables (and interactions between these variables) should be included in the GLM. In the yellowtail snapper assessment, this decision was based on statistical significance. In my experience this criterion is usually poor for CPUE data. It tends to include too many terms in the model because the assumptions necessary for this statistical test (independent and identically distributed errors) are not met. I suggest instead a two-step process of model selection. First, use a stepwise procedure, starting with a model with only year as a predictor, and then adding predictors according to how much additional deviance they explain. Use a sensible threshold for "additional percent deviance explained" (in New Zealand we usually use a threshold of 1% or 0.5%) and stop the stepwise addition of predictors when this threshold is no longer met. Second, reject predictors that are not "plausible", where plausibility is judged from a graph of the estimated effect. For example, if the factor month is to be included we should expect a plot of the estimated month effect to show a reasonably smooth trend. A plot which showed a more or less random fluctuation of predicted CPUE from month to month would be implausible. For a month-year interaction to be plausible we would expect that the predicted annual trends for adjacent months would be relatively similar. Note that the adjective "plausible" applies to the estimated coefficients, not to the model term itself. It is certainly plausible that catch rates might vary from month to month in a way that is consistent from year to year. However, in some CPUE analyses the data are inadequate to estimate this variation plausibly. My experience is that the test of plausibility often rejects interaction terms because CPUE data sets are commonly inadequate to estimate many interactions.

The second possible improvement concerns interactions with year. When such interactions are included the GLM produces more than one time trend in CPUE. For example, a year-region interaction produces a different time trend for each region; a year-month interaction produces a different trend for each month. Such interactions are not implausible, but they are problematic for stock assessments. If they arise we have three options. The first is to accept them and use the multiple time trends in our assessment. This means we must complicate the structure of our assessment model. For example, if we accept a year-region interaction we must include the associated regions in our assessment model in such a way that the model can calculate the biomass in each region at any time and allow for different biomass trends in different regions. This approach is acceptable only if there are sufficient data to justify this increase in model complexity. The second option is not to use any CPUE indices from this source on the grounds that they contain conflicting information. The third option is to drop year interaction terms from the GLMs. The effect of this option is to generate a single time trend that is a data-weighted average of the multiple trends that would have been produced had the year interactions been included. (In the draft assessment, although year interactions were included, only a single time trend was produced from each GLM. It was unclear what this time trend was; it may have been a simple average of the multiple trends, or the trend associated with the reference level for each factor in the interaction.) I believe that this last option was most appropriate in the present assessment, and that was what the Panel adopted (there was not sufficient time to check to see whether these year interaction terms would have passed the model selection criteria described above). Where this option is used it is sensible to describe the interaction as background information for the assessment. For example, with a year-region interaction it would be useful to present a graph showing the different time trends in each region (as output from a GLM with a year-region interaction) but then to drop the interaction when generating a CPUE index for the assessment model.

The third possible improvement concerns explanatory variables that are continuous (e.g., DEPTH and TIMEFISH in Table 4.1.2.1.1). In the assessment, each such variable contributed just one degree of freedom, which implies that CPUE was treated as a linear function of these variables. This is clearly inappropriate for some variables (e.g., we might expect CPUE to be maximum for some optimal depth and less for smaller and larger depths). Unless a linear trend is obviously required I would suggest using a higher order polynomial, perhaps a cubic. (The naive way of doing this — including DEPTH, DEPTH^2, and DEPTH^3 — is sometimes not satisfactory because of co-linearity. A better way is to use orthogonal polynomials.)

3.1.2 Data weighting

In assessments with multiple data sources, the weighting of individual data sets (and even parts of data sets) can have a profound effect on the assessment results. For this reason, it is important that data weightings should be the best possible, and that the model's sensitivity to alternative weightings be evaluated. I'd like to emphasize that this is not easy to do. There is sometimes no obvious "correct" weighting for a data set, and it is sometimes quite a lot of work to generate an appropriate weight, which may turn out to have little effect.

In approaching this problem I think it is better to express the objective function in the form used for the fleet-specific model, rather than that used for the ICA model. That is, the contribution of the ith datum should be expressed as $0.5(\ln(\text{obs}_i) - \ln(\text{pred}_i))^2/\sigma_i^2$ rather than $\lambda_i(\ln(\text{obs}_i) - \ln(\text{pred}_i))^2$. These two forms are mathematically equivalent (if we define $\lambda_i = 0.5/\sigma_i^2$) but the latter makes you think of weighting (which is an essentially arbitrary, and thus subjective, process) whereas the former makes you seek an appropriate variance for each datum (i.e., the variance of its error distribution). The standardised residual for the datum is $(\ln(\text{obs}_i) - \ln(\text{pred}_i))/\sigma_i$. You can tell if you've got approximately the right variances (or weightings) by calculating the SDSR (the standard deviation of the standardised residuals) for each data set. These should have a value near 1. When all SDSRs are less than 1 (as they were in the yellowtail snapper assessment) then the σ_i are too big and overall uncertainty (in the form of parameter c.v.s in Table 4.2.2.1.4 and 4.2.2.2.2, and scatter in Fig. 4.2.2.1.9) will be overestimated. The SDSRs for the visual-survey indices were higher than those for other data sets, indicating that the visual-survey indices were (relatively) over-weighted. It should be noted that the SDSR is only an estimated quantity, and that small data sets produce poor (imprecise) estimates.

One circumstance in which it is not reasonable to expect an SDSR to be equal to 1 is when a data set is considered to be biased. This was the case for the early catch-at-age data, which were down-weighted because they are likely to underestimate strong year classes and over-estimate weak ones. Here, there is little choice but to down-weight arbitrarily, and the SDSR should be greater than 1 (but how much greater is hard to say). Some people like to down-weight CPUE indices on the grounds that they are not strictly proportional to abundance (and thus are biased). My preference is to leave them out as a sensitivity analysis (as was done during the Workshop).

The SDSR can also be used to examine the weighting within a data set. My guess is that if this were done with the catch-at-age data it would show that the small proportions (the tails of the age distribution) are over-weighted.

The σ_i represent the combination of sampling error (the difference between observations and the real world) and process error (the difference between the real world and the model). The approach that I have used, where possible, is to estimate the size of the sampling error from the data and then add a process error that is either derived from a meta analysis (e.g., Francis et al. 2003 for trawl surveys) or estimated within the model. The sampling error will typically be different for every datum within a data set, but the process error should be the same across the data set (and even across data sets of similar types). For some data sets (e.g., trawl surveys and the visual surveys used in the assessment) estimation of sampling error is straightforward. For other more complicated data a simulation approach can be used (e.g., I did this for survey- and catch-at-age data in Francis 2003).

On a related topic, a likelihood profile is a useful tool to examine the contribution of different data sets to a parameter estimate. The overall likelihood profile (which will have a maximum at the parameter estimate) may be thought of as a sum of the contributions from each data set. That is, for a given parameter (e.g., F_{2001}) we can construct a separate profile for each data set, and value of F_{2001} at the maximum of each profile is the preferred value of this parameter for that data set. This also shows which data sets have a strong preference for a specific value for the parameter and which are relatively uninformative about that parameter.

3.1.3 Recruitment variability

The recruitment variability estimated for yellowtail snapper is very low compared to values estimated for other marine species. From Table 4.2.2.1.7 I estimated σ_R =

0.15 for 1-year old recruits (σ_R is the standard deviation of log recruitment) whereas typical values are much higher (quantiles of σ_R for the data sets analysed by Beddington and Cooke (1983) are approximately 0.4, 0.6, and 0.8; see Myers et al. (draft) for a more comprehensive data set). This may be an under-estimate, because the use of multi-year age-length keys will tend to produce low estimates of σ_R. However, a back-of-the-envelope calculation using data from the years with individual age-length keys produced σ_R = 0.19 (I took the last 5 rows of Table 4.2.2.1.1, ignoring the plus-group column; divided each column by its mean; calculated the mean for each diagonal vector representing a year class, which gives an approximate year-class strength for each of the cohorts that were age 0 in years 1987 to 1997; and calculated the standard deviation of the logs of these year-class strengths).

Now it may be that this species really does have unusually low recruitment variability. On the other hand the low estimates of σ_R could be caused by either large ageing errors or non-representative age-length keys. Further work on age validation would address the issues of ageing errors (the current validation is based solely on marginal increment analysis, which is not a strong method). Because the distributions of ages for fish of a given length is likely to vary by month, region, and (possibly) fishing method (recreational versus commercial), age-length keys will be unrepresentative if otolith sampling does not match the month-region-method mix in the catch. I agree with the suggestion (on p.15 of the assessment report) that direct age estimation is preferable to the use of age-length keys (though I did not understand the explanation given, in the same paragraph, for the direct age estimates of Table 3.6.2 being so different from those based on age-length keys). However, I note that there is still a need for representative sampling of otoliths.

3.1.4 Natural mortality and recruitment steepness

The estimated stock status depends strongly on the values of these parameters and there is very little information from which to estimate them. Although I had no reason to question the values used in the assessment I didn't feel that the rationale for their use was well documented. The only information in support of M = 0.2 was a brief section (2.7) mentioning two rules of thumb that provide estimate of Z (total mortality), not M. Other estimates from the literature (which were presented to the Panel, but not included in the assessment report) should have been mentioned, as should have the catch-curve estimate of Z = 0.54. My opinion is that there do not seem to be any reliable data-based estimates of M for this species so we must fall back on analogies with other species based on life-history considerations. With regard to estimates of Z (which provide little more than an upper limit on *M*) I believe the catch-curve estimate is more reliable than those that arbitrarily assign a percentile to the maximum observed age.

I have no problem with the assumed steepness of 0.8 (it is close to the default value of 0.75 I recommended in Francis (1993)) but I think the reasons underlying the Stock Assessment Workshop Panel's recommendation of this value (p. 25) should be spelled out more fully.

3.1.5 Retrospective analysis

I support the Panel's concern about retrospective bias but do not believe that we can learn much from the brief analysis that was carried out during the Workshop.

Retrospective bias occurs in an assessment model if there is a consistent trend in the model's sequential estimates of some parameter. For example, consider the current assessment's estimate of F_{2001}. We can think of this as being the first in a sequence of estimates of that parameter for this stock. If a new assessment were carried out next year, with a further year's data, we would get the second estimate of F_{2001} in our sequence. A third estimate would result from a further assessment in 2005, etc. There is a similar sequence of estimates for F_{2000}, and another for F_{1999}, etc. If when we plot these sequences they all show a trend in the same direction then we have an instance of retrospective bias (see, e.g., Sinclair et al. 1991; Parma 1993; fig. 1 in Francis & Shotton 1997). Unfortunately, we do not then know what caused this bias, and nor do we know whether the sequence tends towards or away from the true value.

With a relatively short data series, such as was available for yellowtail snapper (very short if we consider that only the last five years of catch-at-age data were really reliable) it is difficult to draw any strong conclusions from a retrospective analysis. Even if some trends occur this may not be serious. For example, a biomass index may, by chance, have several positive residuals (overestimates) in a row. This will mean that a biomass estimated near that time will be overestimated, but that as additional years' data are added that overestimate will gradually be corrected. In other words, estimates of that biomass will show a trend, but this is normal model behaviour and not indicative of the sort of data or model problem that a retrospective analysis is intended to find.

3.1.6 Other minor assessment matters

The ignoring of stage 7 gonads in Section 2.5 will probably produce bias in the maturity curves because it is unclear what proportion of such fish were mature. The possible extent of this bias could be evaluated by plotting two additional maturity curves on each panel of Figure 2.5. Both curves would be calculated using all the data (including stage 7), but in one we would assume that all stage 7 fish were mature, and in the other we would assume that they were all immature. These two curves would then bound the true maturity curve.

I think it would have been useful to include a plot, for each fishery sector, of mean length against time. This would have had no direct input to the assessment, but would have been useful background information for the Panel. It is more reliable information than the catch-at-age data (because of the problem of multi-year age-length keys). The lack of a strong trend in the stock biomass ought to mean no strong trend in mean lengths. Also, it would have been good to see whether fishers' perception of increasing mean length in recent years (as expressed during the Workshop) was evident in the data.

It was unclear to me why the commercial logbook hook-and-line index was used in the fleet-specific model but not in the ICA model.

It would have been useful to see more discussion about the reasons for the decline in effort in recent years in the three fishery sectors (commercial, recreational, and headboat). It is unusual to see effort decline in a fishery where biomass is stable or increasing. Unless we are given good reasons for the decline in effort there is a temptation to doubt the assessment and assume that effort had declined because there were fewer fish.

There were some model outputs that I felt were unnecessary. I now (after the Workshop) understand what the "phase plot" (Figure 4.2.2.1.9) represents and how it was generated (though this not described in the assessment report) but I still don't know why it was presented. What should we infer from it? The assessed stock status seems to have been based on point estimates from various model runs, not the uncertainty surrounding these point estimates. Most of Table 4.2.2.1.5 is unexplained, and some of the parts I think I do understand don't make sense to me. The statistics for skewness, kurtosis, and partial chi-square are neither explained nor interpreted for the reader. The assignment of degrees of freedom to each data set, the associated ANOVA table, and the significance of fit values make no sense to me. These imply that we can assign each estimated parameter to a particular data set, whereas I believe that many of the parameters affect the fits of several data sets. A similar comment applies to Table 4.2.2.2.1.

I have some difficulty with the inclusion of total catch in the likelihood function for the fleet-specific model (p. 29). Certainly, it looks like double counting to include both this and numbers at age in the catch. This double counting could be avoided if the catch-at-age data were used a proportions, rather than numbers, at age. Even then, I'm unsure of the advantage of treating total catch as a quantity observed with error (and thus to be included in the likelihood function) rather than being known exactly (as is the done in New Zealand assessments).

The term "asymptotic recruitment" is mistakenly used as being synonymous with R_0 (the mean recruitment when $F = 0$). Actually, the stock-recruitment curve used has an asymptote at $R = 1/\beta = 4hR_0/(h\text{–}1)$.

3.2 SEDAR Terms of Reference

3.2.1 Management Recommendations

The phrase "including management recommendations" in the sixth term of reference for the Assessment Review Panel caused some difficulty during the Workshop and in the drafting of the stock-status report. It was unclear to the Panel what this meant, particularly given that two Fishery Management Councils (whose requirements appeared to differ) were involved. The Panel's primary job was to review the stock assessment. This leads quite naturally to drawing conclusions about the status of the stock, and the Panel found no difficulty in doing so. However, it is a big step from conclusions about stock status to management recommendations, and this step requires a complete change of framework and a great deal of additional information. Fisheries management is an essentially political process that involves balancing the needs of various stakeholders in the context of complex legislative and administrative requirements. If management recommendations are to be required from future SEDAR panels then some very clear guidelines will be necessary.

3.2.2 Assessments and Assessment Reports

There is a distinction to be made between the assessment and the assessment report. My understanding is that the Panel's task was to review the former, not the latter. I felt free to comment on various parts of the assessment report (which was my main, but not only, source of information about the assessment) but I did not feel that I was reviewing this document. I raise this matter because I learnt, after the Workshop, that at least one Panel member had concerns (not expressed during the Workshop) about some aspects of the assessment report. My experience is that what is deemed necessary in an assessment report varies widely amongst institutions (and I understand that the Fisheries Management Councils for the Gulf of Mexico and the South Atlantic have quite different expectations). Thus the Panel could not have reviewed the assessment report without clear guidelines as to what was required in it. My view is that it is preferable to restrict the Panel to reviewing the assessment only. However, should a review of the assessment report be desired from the Panel then clear guidelines would be necessary, and these might vary according to which fisheries management council(s) was involved.

4. RECOMMENDATIONS

1. That consideration be given in future assessments to:

   – the issues of year interactions, polynomial terms, and model selection in the standardisation of CPUE (see Section 3.1.1);

   – the use of less arbitrary data weightings (see Section 3.1.2);

   – further validation of yellowtail snapper ageing, an examination of the “representativeness” of age-length keys, and more work on direct age estimation (see Section 3.1.3);

   – better documentation of the rationale for the assumed values of natural mortality and recruitment steepness (see Section 3.1.4); and

   – the various minor matters discussed in Section 3.1.6.

2. That consideration be given, in writing of terms of reference for future SEDAR Assessment Review Panels, to

   – either removing the phrase “including management recommendations” or giving clear guidance as to what sort of management recommendations are appropriate (see Section 3.2.1); and

– clarifying what is to be reviewed — the assessment or the assessment report — and, if the latter (not recommended), providing clear guidelines as to what is required in an assessment report (see Section 3.2.2).

5. REFERENCES

Beddington, J.R.; Cooke, J.G. (1983). The potential yield of fish stocks. *FAO Fisheries Technical Paper 242*. 50 p.

Francis, R.I.C.C. (1993). Monte Carlo evaluation of risks for biological reference points used in New Zealand fishery assessments. *Canadian Special Publication of Fisheries and Aquatic Sciences 120*: 221-230.

Francis, R.I.C.C.; Shotton, R. (1997). "Risk" in fisheries management: a review. Canadian Journal of Fisheries and Aquatic Sciences 54: 1699-1715.

Francis, R.I.C.C. (2003). Analyses supporting the 2002 stock assessment of hoki. *New Zealand Fisheries Assessment Report 2003/5*. 34 p.

Francis, R.I.C.C.; Hurst, R.J.; Renwick, J.A. (2003). Quantifying annual variation in catchability for commercial and research fishing. Fishery Bulletin 101: 293-304.

Myers, R.A.; Bowen, K.G.; Zouros, I.A. (draft). Recruitment variability of fish and aquatic invertebrates. Draft paper available in pdf form from http://fish.dal.ca/~myers/papers.html.

Parma, A.M. (1993). Retrospective catch-at-age analysis of Pacific halibut: implications on assessment of harvesting policies. *In*: Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. Alaska Sea Grant College Program Report 93-02. Kruse, G.; Eggers, D.M.; Marasco, R.J.; Pautzke, C.; Quinn II, T.J. (eds.). University of Alaska, Fairbanks. pp 247-265.

Sinclair, A.; Gascon, D.; O'Boyle, R.; Rivard, D.; Gavaris, S. (1991). Consistency of some northwest Atalantic groundfish stock assessments. *Northwest Atlantic Fisheries Organization Scientific Council Studies 16*: 59-77.

APPENDIX 1: Statement of Work

This appendix contains the Statement of Task that formed part of the consulting agreement between the University of Miami and the author.

STATEMENT OF TASK

Consulting Agreement between the University of Miami and Dr. Chris Francis

June 11, 2003

General

The **S**outh **E**ast **D**ata, **A**ssessment, and **R**eview (**SEDAR)** process for stock assessment and review is used in the NMFS- Southeast Fisheries Science Center's area of responsibility. This new program provides the framework for independent peer review of stock assessments undertaken jointly by NMFS-SEFSC, three Regional Fishery Management Councils, and two Interstate Fishery Commissions, and state fishery agencies. The SEDAR process uses a three phase approach: a data workshop, an assessment workshop, and a peer review panel workshop. The peer review panel is composed of stock assessment experts, other scientists, and representatives of the Council, the fishing industries, and non-governmental conservation organizations. The communication elements of SEDAR include a stock assessment report from the Assessment Workshop, a review panel report evaluating the assessment(s) (drafted during the Review Panel Workshop), presentation of the peer reviewed assessment results to the Council(s) and public, and publication of collected documents for stock assessments considered in that cycle of the SEDAR process.

The assessment to be reviewed by this SEDAR Peer Review Panel is yellowtail snapper in the area of jurisdiction of the South Atlantic and Gulf of Mexico Fishery Management Councils. The Review Panel will meet July 28-31, 2003 at the Tampa, Florida Airport Hilton Hotel. A data workshop was held March 3-4, 2003 in St. Petersburg, FL. The assessment workshop was held during week of June 9-13, 2003 in St. Petersburg, FL. The SEDAR Review Panel for the yellowtail snapper assessment may include 12+ members: 1 senior assessment scientist each from NMFS- NEFSC and -SEFSC, 2 Council Staff scientists (South Atlantic and Gulf of Mexico), 2 assessment scientist members of the Scientific and Statistical Committee of South Atlantic and Gulf of Mexico Fishery Management Councils, commercial/recreational fishermen from the Snapper-Grouper (SA) and Reeffish (GM) Advisory Panel with special experience with the species, 2 scientist representatives (SA and GM) from non-governmental organizations, and 2 members from the Center for Independent Experts (Chairperson and reviewer). Assessment scientists from Florida FWC and

NMFS-SEFSC will present the assessment and be available to provide supplemental information as requested by the review panel.

SEDAR Assessment Review Panel Tasks

The Panel will evaluate the yellowtail snapper assessment, the input data, assessment methods, and model results as put forward in the stock assessment workshop report.

Specifically, the review panel will:

1. Evaluate the adequacy and appropriateness of fishery-dependent and independent data used in the assessment (i.e. was the best available data used in the assessment).
2. Evaluate the adequacy, appropriateness and application of models used to assess these species and to estimate population benchmarks (MSY, Fmsy, Bmsy and MSST, i.e. Sustainable Fisheries Act items).
3. Evaluate the adequacy, appropriateness, and application of models used for rebuilding analyses.
4. Develop recommendations for future research for improving data collection and the assessment.
5. Prepare a report summarizing the peer review panel's evaluation of the yellowtail snapper stock assessment. (Drafted during the Review Workshop; Final report due two weeks after the workshop-August 15, 2003).
6. Prepare a summary stock status report including management recommendations. (Drafted during the Review Workshop, Final report due two weeks later -August 15, 2003).

It is emphasized that the panel's primary duty is to review the existing assessment. In the course of this review, the Chair may request a reasonable number of sensitivity runs, additional details of the existing assessment, or similar items from technical staff. However, the review panel is not authorized to conduct an alternative assessment, or to request an alternative assessment from the technical staff present. To do so would invalidate the transparency of the SEDAR process. If the review panel finds that the assessment does not meet the standards outlined in points 1 through 3, above, the panel shall outline in its report the remedial measures that the panel proposes to rectify those shortcomings.

The Review Panel Report is a product of the overall Review Panel, and is NOT a CIE product. The CIE will not review or comment on the Panel's report, but shall be provided a courtesy copy, as described below under "Specific Tasks." The CIE products to be generated are the Chair's report, also discussed under Specific Tasks.

Specific Tasks

Designee will serve as panelist of a SEDAR Stock Assessment Review Panel which is to convene in Tampa, FL at the Airport Hilton during 28-31 July 2003. The Panel meeting will begin mid-day on the 28^{th} and conclude early afternoon on the 31st. The Panel will review the stock assessment provided for yellowtail snapper in the area of jurisdiction of South Atlantic and Gulf of Mexico Fishery Management Councils. The SEFSC shall provide the Panelist with copies of the all background documents.

It is estimated that the Panelist's duties will occupy a total of 14 days - several days prior to the Review Panel meeting for document review; four days at the SEDAR meeting; several days following the meeting to ensure that the final documents are completed, and several days to complete a Chair's report for the CIE.

Roles and responsibilities:

1. Prior to the meeting panelists will be provided with the stock assessment workshop report and other associated documents on the yellowtailsnapper. All panelists shall read these documents to gain an in-depth understanding of the stock assessment and the resources and information considered in the assessment.
2. During the review panel meeting, participate, as a peer, in panel discussions on assessment validity, results, recommendations, and conclusions. Participate in the development of the Peer Review Panel Report and Summary Stock Status Report.
3. Review and provide comments to the Panel Chair on the Draft Peer Review Panel Report and Summary Stock Status Report.
4. No later than August 14, 2003, submit a written report[1] consisting of the findings, analysis, and conclusions, addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Sampson, via email to David.Sampson@oregonstate.edu, and to Mr. Manoj Shivlani, via email to mshivlani@rsmas.miami.edu.

Contact persons:
NMFS contact: Dr. John Merriner, Beaufort Laboratory, 101 Pivers Island Road, Beaufort, NC 28516. Phone 252-728-8708. FAX 252-728-8784. E-mail john.merriner@noaa.gov
SAFMC contact: Mr. Gregg Waugh, One Southpark Circle, Suite 306, Charleston, SC 29407, phone 843-571-4366, FAX 843-769-4520, E-mail gregg.waugh@safmc.net.

[1] The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

Annex I to Appendix 1: Report Generation And Procedural Items

1. *The report should be prefaced with an executive summary of findings and/or recommendations.*

2. *The main body of the report should consist of a background, description of review activities, summary of findings, conclusions/recommendations, and references.*

3. *The report should also include as separate appendices the bibliography of all materials provided and a copy of the statement of work.*

Please refer to the following website for additional information on report generation: http://www.rsmas.miami.edu/groups/cie.

APPENDIX 2: Materials Provided

All material provided to the author was web-based. The primary web site was ftp://ftp.floridamarine.org/users/assess/SEDAR_YT_Assessment, which, as well as holding many data and model-output files, included the draft assessment report:

Muller, R. G., Murphy, M. D., de Silva, J., and Barbieri, L. R. 2003. A stock assessment of yellowtail snapper, *Ocyurus chrysurus*, in the Southeast United States. Draft Report submitted to the National Marine Fisheries Service and the Gulf of Mexico Fishery Management Council as part of the Southeast Data, Assessment, and Review (SEDAR) III. St Petersburg, FL; Florida Marine Research Institute: 182 pp.

Two other sites described key data sources:

http://www.st.nmfs.gov/st1/recreational/survey for the Marine Recreational Fisheries Statistics Survey, and

http://www.sefsc.noaa.gov/alsprogram.jsp for commercial catch data.

APPENDIX 3: Attendees at SEDAR Assessment Review Panel Workshop

Panel

PANEL CHAIR:	Dr Andrew Payne
REVIEW PANELIST:	Mr Chris Francis
SAFMC:	Mr Gregg Waugh
GMFMC:	Mr Steve Atran
NMFS SEFSC:	Dr Joseph Powers
NMFS NEFSC:	Dr Larry Jacobson
FISHERS:	Mr William Kelly Mr Robert Zales Mr Peter Gladding
NGO REPRESENTATIVE:	Ms Nadiera Sukhraj
SSC REPRESENTATIVES:	Mr Doug Gregory Mr Billy Fuls Dr Al Jones Ms Carolyn Belcher Dr Robert Trumble Dr Rocky Ward

Non-Panel

PRESENTERS:

AW Coordinator	*Dr Luiz Barbieri*
Lead Analysts	*Dr Robert Muller* *Mr Michael Murphy*
Goliath grouper	*Dr Anne Marie Eklund*

AW/RPanel SUPPORT STAFF:

Dr John Merriner
Dr Janaka de Silva

MEETING SUPPORT STAFF & OTHER ATTENDEES

Mr Roy Williams
Dr Tom McIlwain
Dr Joe Kimmel
Mr Mark Robson
Mr Stu Kennedy
Dr Roy Crabtree
Dr Behzad Mahmoudi