CIE Independent Report
SEDAR 24 Review on South Atlantic Red Snapper Assessment
Savannah, Georgia during 12-14 October 2010

Dr. Noel Cadigan
Fisheries and Oceans Canada,
Northwest Atlantic Fisheries Center
St. John's, NL, Canada

Executive Summary

The red snapper stock is overfished and overfishing is occurring. Spawning stock biomass in 2009 (SSB2009) was estimated to be about 10% of MSST (SSB2009/MSST= 0.09), and the recent level of fishing was four times FMSY (F2007-2009/FMSY= 4.12). Numerous sensitivity analyses presented in the assessment all indicated very similar conclusions about stock status. The most significant sources of uncertainty include landings, the stock-recruitment relationship, CPUE catchability, and age compositions and fishery selectivity.

An important research recommendation is to better account for the uncertainty in the stock recruitment relationship when assessing red snapper stock status. More accurate estimates of historic catch at age, improved age sampling, and reliable fisheries independent indices of abundance should improve the statistical catch at age model.

Background

SEDAR 24 was 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. 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. SEDAR 24 was conducted for the South Atlantic Fishery Management Council (SAFMC). The lead assessment agency was the Southeast Fisheries Science Center of the US National Marine Fisheries Service (NMFS).

The CIE (Center for Independent Experts) reviewer was tasked with conducting an impartial and independent peer review in accordance with the Statement of Work (SoW) and Review Workshop (RW) Terms of Reference (ToRs; Appendix 2) for SEDAR 24, to determine if the best available science is utilized for fisheries management decisions, and to present the review in writing. SEDAR Review

Workshops provide independent peer review of stock assessments prepared through SEDAR data and assessment workshops. The goal of the review is to ensure that the assessment and results presented are scientifically sound and that managers are provided adequate advice regarding stock status, management benchmarks, and the general nature of appropriate future management actions.

The Review Panel (RP) was composed of a Chair, three CIE reviewers, and one reviewer representing the Scientific and Statistical Committee (SSC) of the SAFMC. The CIE reviewers were independent, meaning that they did not contribute to the assessment under review and did not have a role in any management actions that may stem from the assessment.

Role of reviewer

I attended the SEDAR 24 in Savannah, Georgia during 12-14 October 2010. I reviewed presentations and reports and participated in the discussion of these documents, in accordance with the SoW and ToRs (see Appendix 2). The review is structured according to my interpretation of the required format and content described in Annex 1 of Appendix 2.

Backgrounds documents, including the Data Workshop (DW) and Assessment Workshop (AW) reports were made available approximately one week before the RW. I spent three days reviewing these documents before the RW.

I would like to express my appreciation for the assessment team's thoughtful and respectful responses to all of my queries.

Summary of findings

In what follows I first provide the Review Panel conclusions, for context and completeness. I then provide my views on each ToR.

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

Panel conclusions

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 CPUE's 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.

<u>Additional reviewer views</u>

I did not completely agree with the RP that "commercial and recreational fisheries lack the adequate statistical design and spatial coverage that one would expect from a fishery-independent survey". Some CPUE's, and in particular the recreational CPUE used in this assessment, can have good spatial coverage.

Discard mortality is an important source of mortality for red snapper. Hooking mortality was thought to be substantial in the DW report, but not quantified. This mortality seems to be similar to tagging mortality, and there are numerous studies that describe methods to estimate tagging mortality. In any event, in the current assessment hooking mortality represents an unaccounted source of mortality and, as such, fishing mortality estimates may be too low. Unaccounted hooking mortality will be more important in 2010 because of the fishing moratorium on red snapper, in which case discards and discard mortality will increase.

CV's for landings were provided by the DW (see Table 3.8 in the report). Some CV's were 50%, which suggests that a 95% confidence interval for true landings was roughly 0 to double the reported value. It should be clarified if this was the intended interpretation of the DW. It would be useful for the DW to provide CV's that the AW could use in their Monte-Carlo bootstrap procedure for quantifying uncertainty (see ToR 6).

More thorough analyses of the internal consistency of age composition data would help when examining output from stock assessment models. A graphic that some stock assessment researchers, including myself, have found useful is a SPAY (standardized proportion at age) plot. They are simple to do in R. SPAY plots are also provided by the FLEDA component of the FLR (Fisheries Library in R) package. They are useful for tracking changes in annual age compositions, as strong or weak cohorts move through the gear selection age-range.

A SPAY plot of the recreational for-hire landings is given in Figure 1 in Appendix 3 of this report. The data were obtained from Appendix A of the RW working paper describing the application of BAM to red snapper (SEDAR 24 RW-01). The interpretation of this figure is complicated because of changes in management measures that affect the selectivity of this fishery. Nonetheless, the age composition data appear to be quite noisy and do not suggest cohorts have been consistently tracked. Note that no age 15 fish were recorded in the age compositions throughout the time-series, and the area of the bubbles for this age in Figure 1 are all zero's, which is why there are no bubbles for this age.

Contrast this with the age-compositions from the BAM output (i.e. Table 3.2 in AW report). The SPAY plot in Figure 2 (Appendix 3) clearly tracks strong and weak cohorts. This is expected of course because these age compositions were derived from a cohort model. Unfortunately there is little correspondence between Figures 1 and 2. This is discussed more for ToR3.

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

<u>Panel conclusions</u>

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/Fmsy values from ASPIC were at a lower scale compared to BAM, indicating a lower level of over-fishing. The values of B/Bmsy from ASPIC were below 1.0 over the entire assessment time frame (1955-2009), whereas BAM indicated biomass above Bmsy prior to 1970. BAM also indicated that current

(2009) biomass is much less than Bmsy (i.e. 10%), whereas ASPIC is somewhat more optimistic (B2009/Bmsy = 0.39; B2010/Bmsy = 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.

<u>Additional reviewer views</u>

The implementation of BAM in SEDAR 24 was improved compared to SEDAR 15, but the improvements may still require further investigation.

Although dome-shaped selectivity was suggested to be more plausible, this type of selectivity is usually confounded with the magnitude of fishing mortality (F). The primary source of information about F comes from the rate of decline in cohort catches as age increases, but length compositions and trends in age-aggregated indices also contribute some information. However, with dome shaped selectivity this decline can also be explained by reductions in selectivity with age – hence the confounding between selectivity and F. Fleets with flat-topped selectivity can be more informative about F. In this assessment only the commercial lines landings were assumed to have flat-topped selectivity, and there were limited age samples from this fleet available (1996-2000, 2004, 2007, and 2009). Hence, one might anticipate that the BAM model+data will not provide reliable estimates of F.

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. Many bias correction methods have been proposed for the purely likelihood setting (see Severini, 2000; Cox, 2006). However, such adjustments will probably be complicated for the types of data used by BAM.

The iterative re-weighting multiplied model weights and external CV's for input measurement error. I don't think this is the best way, and I am not sure of the statistical justification for it. I think of this problem as one of variance components. Let U denote a CPUE index. Assume U is unbiased for the population CPUE for the entire fleet, and U has a constant CV;

$$U = CPUE \times \varepsilon,\ E(\varepsilon) = 1\ and\ Var(\varepsilon) = \sigma^2_{cpue}. \tag{1}$$

Note that ε is the error in U for measuring CPUE. Now CPUE may also be an index of stock size (N); that is,

$$CPUE = qN \times \delta,\ E(\delta) = 1\ and\ Var(\delta) = \sigma^2_{\mathrm{mod}el}, \tag{2}$$

where δ is the model error in CPUE as an index of stock size. This model error cannot be estimated from samples of CPUE alone. For convenience let $\mu = qN$. Equations (1) and (2) can be combined to express U in terms of relative stock size (μ),

$$U = \mu \times \varepsilon \times \delta.$$

It is not difficult to show that

$$E(U) = \mu \text{ and } Var(U) = \mu^2 \left(\sigma_{cpue}^2 + \sigma_{model}^2 + \sigma_{cpue}^2 \sigma_{model}^2\right) \tag{3}$$

The CV of U is not just the product of the measurement error CV (σ_{cpue}) and the model CV (σ_{model}).

The measurement error variance (σ_{cpue}^2) is estimated external to the stock assessment model. If the estimates are the same for all indices then the model variance parameter is easy to estimate. In the lognormal setting, if MSE is the mean squared log error ($\sum_{i \in obs} \{\log(U_i / \mu_i)\}^2$), then the maximum likelihood estimate (MLE) of σ_{model}^2 is

$$\hat{\sigma}_{model}^2 = \frac{MSE - \sigma_{cpue}^2}{1 + \sigma_{cpue}^2}. \tag{4}$$

Note that if $\sigma_{cpue}^2 = 0$ then the MLE is the standard result, $\hat{\sigma}_{model}^2 = MSE$. Otherwise, the total variance in Equation (3) is simply the MSE. However, if σ_{cpue}^2 varies for different indices then the estimation of σ_{model}^2 is more complicated, but can be obtained via MLE.

In the re-weighting used in BAM, both CV's were simply multiplied. However, if the measurement error CV is really small then the total weight U gets may still be small even if the model CV is large, because they are multiplied together. In Equation (3), if the model CV is large then the CV for U is still large, and will basically be σ_{model} if σ_{cpue} is close to zero.

I propose that a better way to weight information in statistical catch at age is to use Equations (3) and (4). This would still have to be iterated, because each new choice of weights will yield new values for MSE, and a new estimates of σ_{model}^2.

References

Cox DR (2006) Principles of Statistical Inference, Cambridge University Press, Cambridge

Severini TA (2000) Likelihood Methods in Statistics, Oxford University Press, Oxford.

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

Panel conclusions

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/Fmsy 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.

Additional reviewer views

On the whole, I preferred the iteratively re-weighted base run compared to the equal weight run. Although the re-weighted run down-weighted the Headboat recreational index, which was thought by the DW to be the best indicator of stock size, this index displayed occasional large fluctuations that the cohort model could not explain well, and this is why it was down-weighted.

Historic estimates of stock abundance were very sensitive to the assumption about initial exploitation rate in 1955 (F_{init}). Initial 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; 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 (see top left corner of Figure 2 in Appendix 3), 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 (see top left corner of Figure 1). This was the only source of age-composition data for this period.

I feel that the estimate of virgin biomass (B0) was based on a highly uncertain stock recruitment relationship, and as a consequence the estimates of initial depletion were highly uncertain. Values of F_{init} that are more consistent with the for-hire recreational age compositions should be explored in the future, while at the same time ensuring that Pr(B1955<B0) is not too large. In this approach, the uncertainty in B0 is taken into account.

Subsequent to the RW I found some differences in tables and plots in the AW report. For example, in Figure 3 (Appendix 3) I plotted the recruitments shown in Table 3.2 of the AW report, and found differences compared to Figure 3.25 in the AW report.

*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.*

Panel conclusions

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 Fmsy, SSBmsy are less important than the ratios Fcurrent/Fmsy and SSBcurrent/SSBmsy. 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. F's much closer to Fmsy). 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 Fmsy 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/Fmsy 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.

<u>Additional reviewer views</u>

The declarations of stock status (over-fished and over-fishing) appeared to be robust to the assessment model formulation. However, I still can’t point to the main source of information (i.e. data) that leads to these conclusions. I expect that this makes the assessment difficult to explain to clients, who are probably skeptical of models (and BAM in particular), but less skeptical of the data. It would be helpful if runs of BAM could be provided in which F2009 was forced to equal Fmsy, or B2010 was forced to equal Bmsy. We could then look at the fits to the data to see why these runs are not plausible.

MSY benchmarks critically depend on the nature of the stock-recruit relationship. However, the assessment provided little information on the recruitment dynamics of the stock, such as spawning areas, and larval drift and settlement.

Potential changes in stock productivity (i.e. ecosystem effects) were not discussed.

I was surprised to read shortly after the assessment that Conn et al. (2010) concluded that external estimates of steepness were more reliable than internal estimates. I thought (or speculated) that the reverse would be true, because internal estimates could better account for uncertainty in SSB. External estimates of steepness were not presented in the red snapper AW report.

The Beverton-Holt stock-recruit steepness parameter was fixed at 0.85, which was the mode of a meta-analysis distribution. The mean of this distribution was 0.75, and it was unclear why the review group selected the mode, which is a less commonly used measure of central tendency. For example, only approximately 30% of the stocks in the meta-analysis (AW-07) had steepness greater than 0.85. This may not be an important concern however, because when steepness was estimated within BAM the value was close to one, but this did not appreciably affect conclusions regarding stock status,

The variance of the stock-recruit relationship (σ_R) was not estimated and was inferred from a meta-analysis. I am uncertain if σ_R is an exchangeable parameter among stocks, appropriate for meta-analysis. This would seem to depend, in part, on the environmental effects on recruitment. If the stock-recruit relationship was estimated externally then σ_R is easy to estimate.

The choice of σ_R also affects MSY benchmarks because of the bias correction applied to the estimated stock-recruit curve. The bias correction is an adjustment due to the lognormal assumption. Let *μ(s)* denote that stock (S) and recruitment (R) function. If we assume that $E\{\log(R) \mid S = s\} = \log\{\mu(s)\}$ and that *log(R)* is normally distributed with mean *log{ μ(s) }* and standard deviation σ_R, then it can be shown that $E(R \mid S = s) = \mu(s)\exp(\sigma_R^2/2)$. Hence, multiplying the estimated stock recruitment function *μ(s)* by an estimate of $\exp(\sigma_R^2/2)$ should reduce the transformation bias. However, if σ_R is poorly estimated then the bias correction will also be poorly estimated.

Note that the bias correction makes the estimated stock-recruitment function approximately mean unbiased, but medium biased. This is because $Median(R \mid S = s) = \mu(s)$. For skewed distributions some people prefer an estimate that is median unbiased rather than mean unbiased (e.g. Cabrera and Watson, 1997).

Another approach to estimating the stock-recruitment model that does not involve log-transformations and the associated bias is to use gamma MLE. The lognormal and gamma distributions are similar in that they both have CV's that are independent of the mean (i.e. constant CV models). Gamma MLE's are sometimes preferred to lognormal MLE's, even when the data are lognormally distributed (e.g. Cadigan and Myers, 2001). Cadigan (2006) found that gamma (i.e. Q2 in Cadigan, 2006) MLE's of the stock size corresponding to 50% of maximum recruitment, derived from stock-recruit models, were somewhat less sensitive than lognormal MLE's. However, the differences in gamma and lognormal estimates were not that large.

If one is willing to assume gamma errors then bias correction is not necessary. Usually it is difficult to differentiate in practice if data are gamma or lognormally distributed. Hence, gamma MLE's of stock-recruit curves are more straight-forward in that they do not involve bias correction. Note however that while the direct gamma MLE of the stock-recruit curve is approximately mean unbiased, it is not median unbiased.

To better understand differences in gamma and lognormal MLE's of stock-recruit curves, I applied these estimators to the 15 case studies considered by Cadigan (2009). These are ICES stocks I picked with evidence of density dependence in their estimated stock-recruit relationships. The results (Figure 4 in Appendix 3) demonstrate that gamma estimated stock-recruit curves are usually very similar to bias-corrected lognormal estimated curves. Differences between gamma and uncorrected lognormal curves were usually not large.

I estimated the stock-recruit curve for red snapper using gamma MLE's, and with steepness fixed at the same value as in the assessment (i.e. h=0.85). The data and stock-recruit model fits are shown in Figure 5 (see Appendix 3). The data seem very similar to Figure 3.37 of the AW report. All of these model fits seem plausible, but will probably lead to substantially different estimates of MSY benchmarks. This demonstrates that the benchmarks estimated by the assessment are highly uncertain.

Note that the differences between the gamma and lognormal MLE's in Figure 5 could be qualitatively predicted by the diagnostics in Cadigan (2006). He showed than gamma MLE's of the Beverton-Holt model are more sensitive, compared to lognormal MLE's, to anomalously large values of recruitment at low stock sizes, and less sensitive to anomalously small values of recruitment at large stock sizes.

References

Cabrera, J and Watson, GS (1997). Simulation methods for mean and median bias reduction in parametric estimation. J Stat Plann Inf, 57: 143-152.

Cadigan, N. G. and R. A. Myers. 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. Can. J. Fish. Aquat. Sci. 58: 560-567.

Cadigan, N. G. 2006. Local influence diagnostics for quasi-likelihood and lognormal estimates of a biological reference point from some fish stock and recruitment models. Biometrics. 62: 713-720

Cadigan, N. 2009. Sensitivity of common estimators of management parameters derived from stock-recruit relationships. Fish. Res. 96: 195–205

Conn, PB, Williams, EH, and Shertzer, KW (2010) When can we reliably estimate the productivity of fish stocks? Can J Fish Aquat Sci, 67: 511-523.

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).

Panel conclusions

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 Bmsy, 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 Bmsy. 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.

Additional reviewer views

For reasons outlined for ToR3, the panel found it difficult to identify a unique ‘best estimate’ model run. Consequently, there is no ‘best’ projection, and the deterministic projections should be considered as one of a range of plausible projections.

Hooking mortality may be more important in the projection, when all mortality is due to discarding of red snapper in the mixed-species fisheries (see Discussion for ToR1).

The adequacy of stochastic projections is addressed under ToR6.

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.

Panel conclusions

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 Finit 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, Finit, 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 Fmsy 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.

Additional reviewer views

Common problems when bootstrapping complex models with many parameters, some of which are poorly identified by the data, are: identifying convergence, local minima, and starting values. If the base-run parameter values were used as starting values then some of the bootstrapped parameter estimates may indicate local minima. A robust bootstrap methodology should check and correct for these problems.

A parametric bootstrap procedure was used, with additional uncertainty added to non-estimable parameters by Monte-Carlo sampling from assumed distributions for these other parameters. The range of σ_R was [0.4,0.8], although the external estimate I obtained was approximately one. The range used in the assessment should be defended.

A truncated beta distribution was used for Monte-Carlo sampling of steepness. The mean of this truncated distribution is approximately 0.8, which is more similar to the value assumed in the base run (i.e. 0.85) compared to the meta-analysis beta distribution which had a mean of 0.75. This is a good thing.

The DW provided CV's for historical catch estimates, and these could have been used in the Monte-Carlo bootstrap in the same was as m was re-sampled. However, it was not clear to me if the DW intended the CV's to be used this way, and this should be clarified for the next assessment.

The bootstrap procedure did not include enough uncertainty in the CPUE indices. Uncertainty described by Equation (3) in the section for ToR 2 should be included, and not just the measurement error cv, σ_{cpue}.

There was evidence of correlated residuals for many of the data series (i.e. length and age compositions, and indices). This lack of independence is not accounted for in the bootstrapping procedure, and results in under-estimates of uncertainty. Some statistical catch at age models attempt to account for these correlations, although it is not trivial to do. Nonetheless, this is another reason why the red snapper assessment may under-estimate uncertainty.

The parameter uncertainty was also carried forwarded in the projections, in addition to recruitment variation and some variation in projected fishing mortalities. Process error in population dynamics was otherwise not accounted for in the assessment, and was not carried forward into the projections. It is not possible to assess how important this process error could be.

Process error affects the values of MSY benchmarks (e.g. Bousquet, 2008). Generally Bmsy and Fmsy are lower with process error compared to deterministic results. There is evidence of this in Figure 3.43 for Fmsy, in which the mean of the stochastic Fmsy's is less than the deterministic result. However, this was not the case for Bmsy, and I am not sure why. In terms of Fmsy, this suggests that the status of the stock may be slightly worse than the assessment suggests.

References

Bousquet, N., Duchesne, T. and Rivest, L.-P. 2008. Redefining the maximum sustainable yield for the Schaefer population model including multiplicative environmental noise. J. Theor. Biol. 254 65-75

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.

Panel conclusions

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

Additional reviewer views

No additional comments.

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.

Panel conclusions

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 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.

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.

Additional reviewer views

An additional assessment team could have two tasks: (1) replicate assessment to check for programming errors, etc, and (2) conduct an alternative assessment.

A CIE reviewer at the AW could spend more time reviewing assessment input data, in conjunction with the stock assessment itself. Perhaps their focus should be on ensuring best practice in the assessment, rather than reviewing the final assessment – which is the task of the RW.

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.

Panel conclusions

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.

<u>Additional reviewer views</u>

1. Consider methods to estimate hooking mortality (e.g. cage experiments, etc). Unaccounted hooking mortality will be more important in 2010 because of the fishing moratorium on red snapper, in which case discards and discard mortality will increase.
2. Assess the internal consistency of age composition data, before modeling.
3. Use a variance components approach to iterative re-weighting.
4. Select values of Finit that are more consistent with the for-hire recreational age compositions, while at the same time ensuring that Pr(B1955<B0) is not too large.
5. It would be helpful if runs of BAM could be provided in which Fcurrent was forced to equal Fmsy, or Bcurrent was forced to equal Bmsy. We could then look at the fits to the data to see why these runs are not plausible.
6. Provide better information on the recruitment dynamics of the stock, such as spawning areas, and larval drift and settlement, and how this may have changed over time.

7. Should steepness be estimated internally or externally?
8. Provide bootstrap convergence diagnostics.
9. CV's for historical catch estimates should be used in the Monte-Carlo bootstrap.
10. Include model-residual uncertainty in the CPUE indices when bootstrapping.

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.

<u>Panel conclusions</u>

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.

<u>Additional reviewer views</u>

No additional comments.

Summary of conclusions and recommendations

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 (SSB2009) is estimated to be about 10% of MSST (SSB2009/MSST= 0.09), and the current level of fishing is four times FMSY (F2007-2009/FMSY= 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.

An important recommendation is to better account for the uncertainty in the stock recruitment relationship when assessing stock status. More accurate estimates of historic catch at age, improved age sampling, and reliable fisheries independent indices of abundance will improve the statistical catch at age model. More accurate age samples should improve estimation of fishing mortality, especially for ages that are fully selected by a fishing gear. The assessment should describe more clearly why conclusions about stock status that are substantially different from the assessment are not consistent with the available information.

Critique of the NMFS review process

See ToR8.

Appendix 1: Bibliography of materials provided for appointee's involvement

Documents Prepared for the Data Workshop		
SEDAR24-DW01	Discards of Red Snapper Calculated for Vessels with Federal Fishing Permits in the US South Atlantic	K. McCarthy 2010
SEDAR24-DW02	SEDAR 24 South Atlantic Red Snapper Management Summary	J. McGovern 2010
SEDAR24-DW03	Standardized catch rates of U.S. Atlantic red snapper (*Lutjanus campechanus*) from headboat data	Sustainable Fisheries Division, NMFS 2010
SEDAR24-DW04	Standardized catch rates of U.S. Atlantic red snapper (*Lutjanus campechanus*) from commercial logbook data	Sustainable Fisheries Division, NMFS 2010
SEDAR24-DW05	Red snapper standardized catch rates from the Marine Recreational Fisheries Statistics Survey for the U.S. Atlantic Ocean, 1981-2009	Indices Group MRFSS 2010
SEDAR24-DW06	Distribution of red snapper catches from headboats operating in the South Atlantic	Sustainable Fisheries Division, NMFS 2010
SEDAR24-DW07	Georgia Headboat Red Snapper Catch & Effort Data, 1983-2009	S. Amick, K. Knowlton 2010
SEDAR24-DW08	Sampling Procedures Used in the Trip Interview Program (TIP)	Sustainable Fisheries Division, NMFS 2010
SEDAR24-DW09	Pre-Data Workshop Development of Commercial Landings for the Red Snapper Fishery	D. Vaughan, D. Gloeckner 2010
SEDAR24-DW10	Age Workshop for Red Snapper	J. Potts, editor 2009
SEDAR24-DW11	Review and Analysis of Methods to Estimate Historic Recreational Red Snapper Landings in the South Atlantic	SEDAR24 Historic Rec Catch Group 2010
SEDAR24-DW12	Red Snapper Discard Mortality Working Paper	SEDAR24 Discard Mortality Group 2010
SEDAR24-DW13	South Atlantic Red Snapper Marine Recreational Fishery Landings: FHS-conversion of Historic MRFSS Charter Boat Catches	T. Sminkey 2010

SEDAR24-DW14	Marine Resources Monitoring, Assessment and Prediction Program: Report on Atlantic Red Snapper, *Lutjanus campechanus,* for the SEDAR 24 Data Workshop	MARMAP 2010
SEDAR24-DW15	Red Snapper Length Frequencies and Condition of Released Fish from At-Sea Headboat Observer Surveys, 2004 to 2009.	B. Sauls and C. Wilson 2010
	Documents Prepared for the Assessment Workshop	
SEDAR24-AW01	Assessment History of Red Snapper (*Lutjanus campechanus*) in the U.S. Atlantic	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-AW02	The Beaufort Assessment Model (BAM) with application to red grouper1: mathematical description, implementation details, and computer code	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-AW03	Standardized discard rates of U.S. Atlantic red snapper (*Lutjanus campechanus*) from headboat at sea observer data.	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-AW04	Additional age data of south Atlantic red snapper (*Lutjanus campechanus*) from Florida Fish and Wildlife's dependent monitoring program	J. Tunnell, 2010
SEDAR24-AW05	Selectivity of red snapper in the southeast U.S. Atlantic: dome-shaped or flat-topped?	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-AW06	Spawner-recruit relationships of demersal marine fishes: Prior distribution of steepness for possible use in SEDAR stock assessments	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-AW07	Red snapper: Regression and Chapman-Robson estimators of total mortality from catch curve data	Sustainable Fisheries Branch, NMFS 2010
SEDAR24-AW08	Overviews of NMFS fishery-dependent data source surveys referenced in the SEDAR 24 data workshop report	SEDAR 2010, Compiled by J. Carmichael
SEDAR24-AW09	Vulnerability to Capture of Red Snapper (*Lutjanus campechanus*) in the Fisheries of the Southeast United States - a Preliminary look	F. Hester and D. Nelson, 2010
SEDAR24-AW10	South Atlantic Red Snapper Fishery – A Fisherman's Perspective	D. Nelson, 2009
SEDAR24-AW11	Additional information for red snapper selectivity	F. Hester, 2010
SEDAR24-AW12	Selectivity of red snapper in the South Atlantic More than Just Depth	D. Nelson, 2010

SEDAR24-AW13	Pre-review draft of the Assessment Report, public comment version 8-26-10	SEDAR 2010
	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

Appendix 2: A copy of the CIE Statement of Work

Attachment A: Statement of Work for Dr. Noel CadiganN@DFO-MPO. CA

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 14 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

<u>Tuesday</u>

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

<u>Wednesday</u>

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

<u>Thursday</u>

8:00 a.m. – 11:30 a.m.	**Panel Discussion** *- Final sensitivities reviewed.*	**Chair**
	- Projections reviewed.	**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: CIE Report Figures.

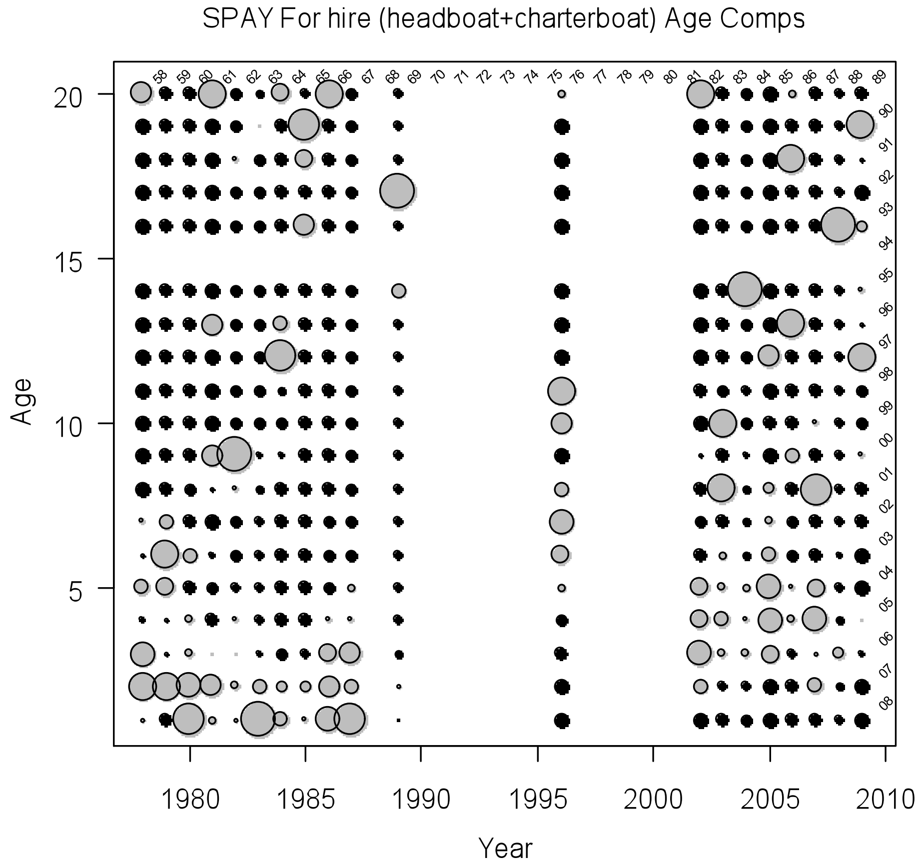

Figure 1. Standardized proportion at age for the recreational for-hire landings. A black circle means the proportion is less than the average for that age. Grey means greater than average. The area of a circle is proportional to the absolute value of the standardized proportion. A very small circle means the proportion is near average. Cohorts are indicated along the top and right sides.

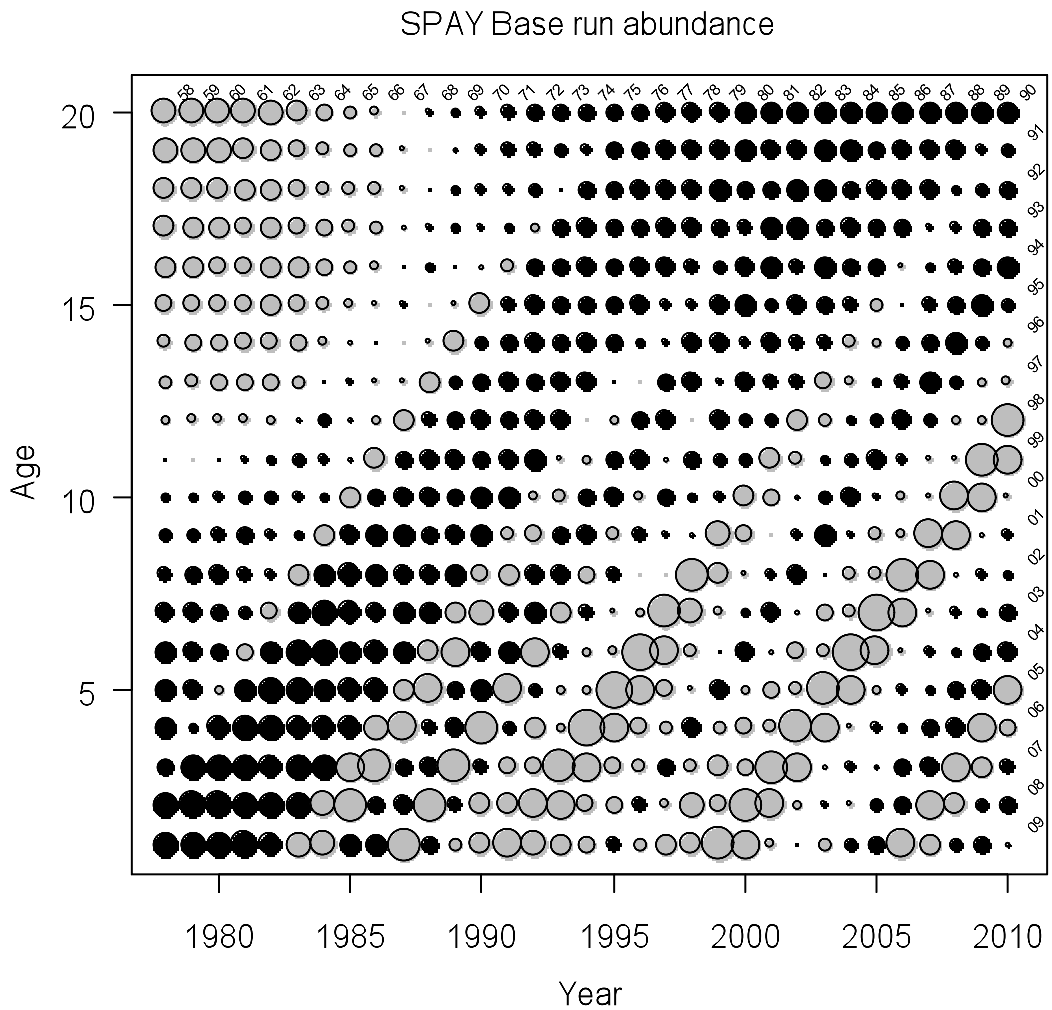

Figure 2. Standardized proportion at age for the BAM base run population abundance estimates. A black circle means the proportion is less than the average for that age. Grey means greater than average. The area of a circle is proportional to the absolute value of the standardized proportion. A very small circle means the proportion is near average. Cohorts are indicated along the top and right sides.

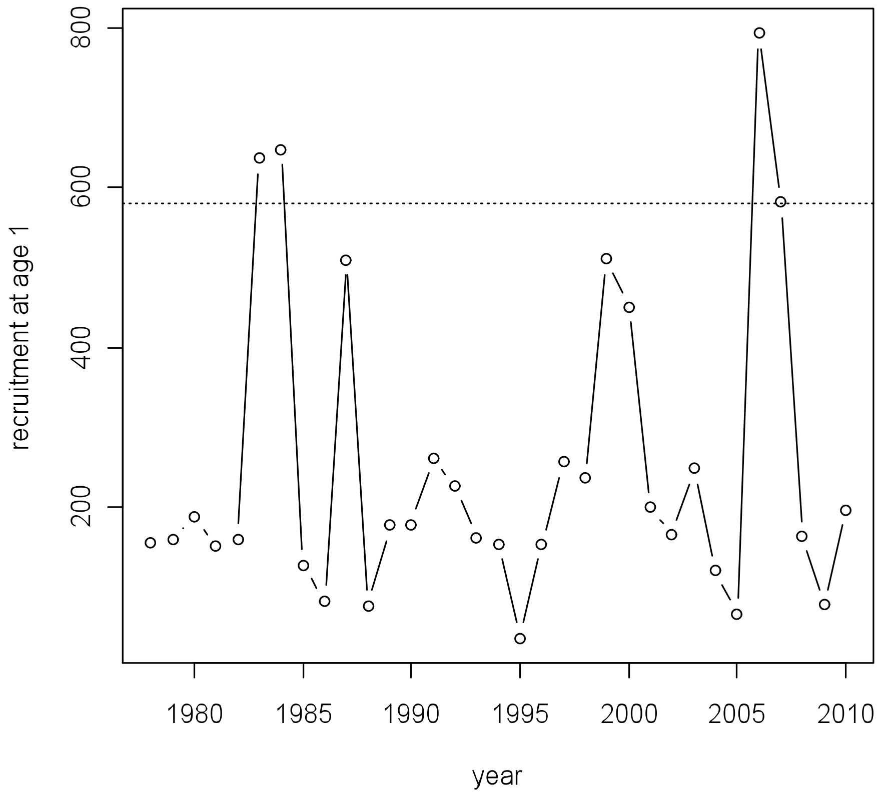

Figure 3 Time series of recruitment (since 1978) from Table 3.2 of the assessment workshop report.

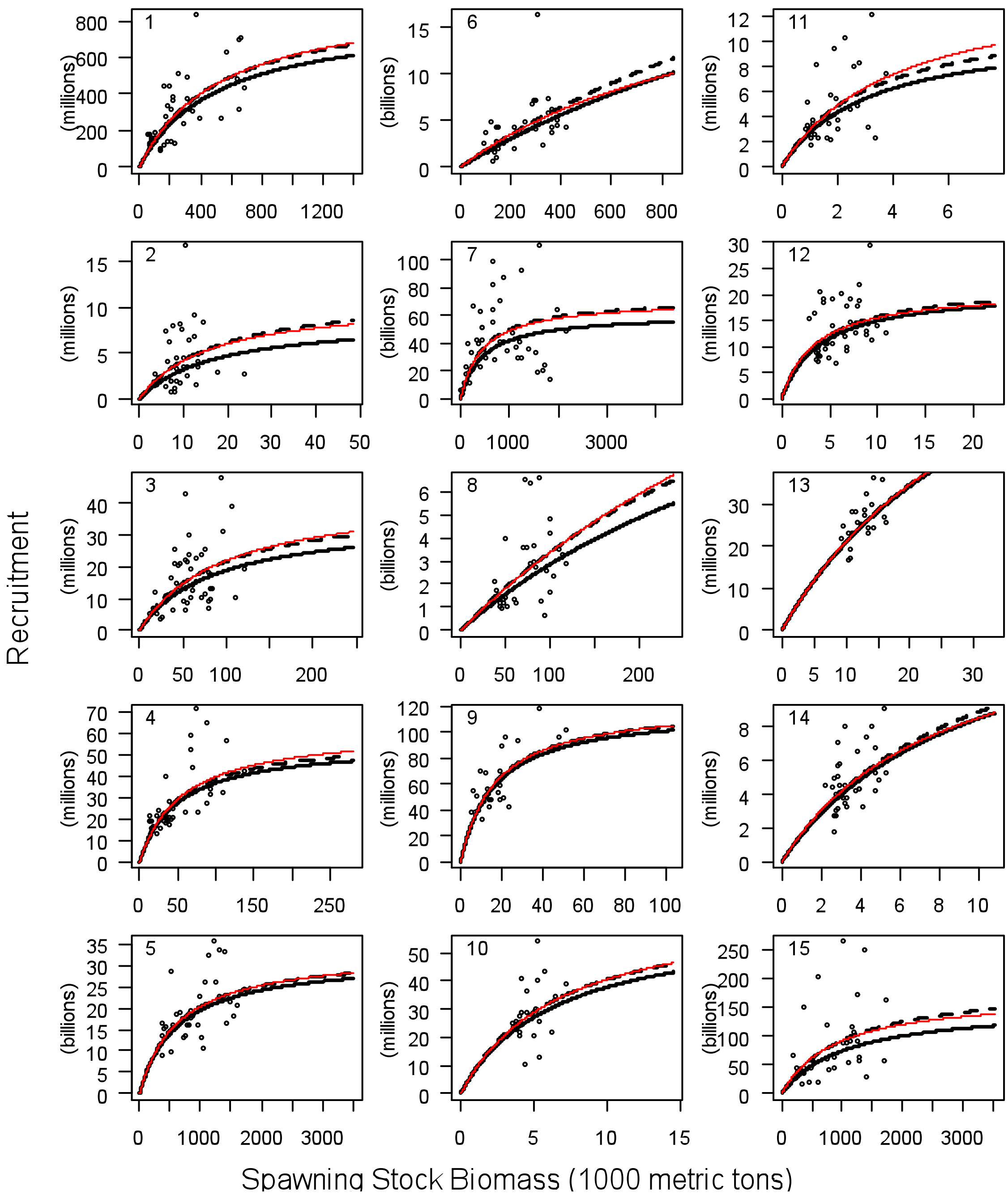

Figure 4. Lognormal (black solid lines), transformation bias-corrected lognormal (dashed lines), and gamma (red lines) maximum likelihood estimates of Beverton-Holt stock-recruitment curves. Each panel is for a stock (see Cadigan, 2009).

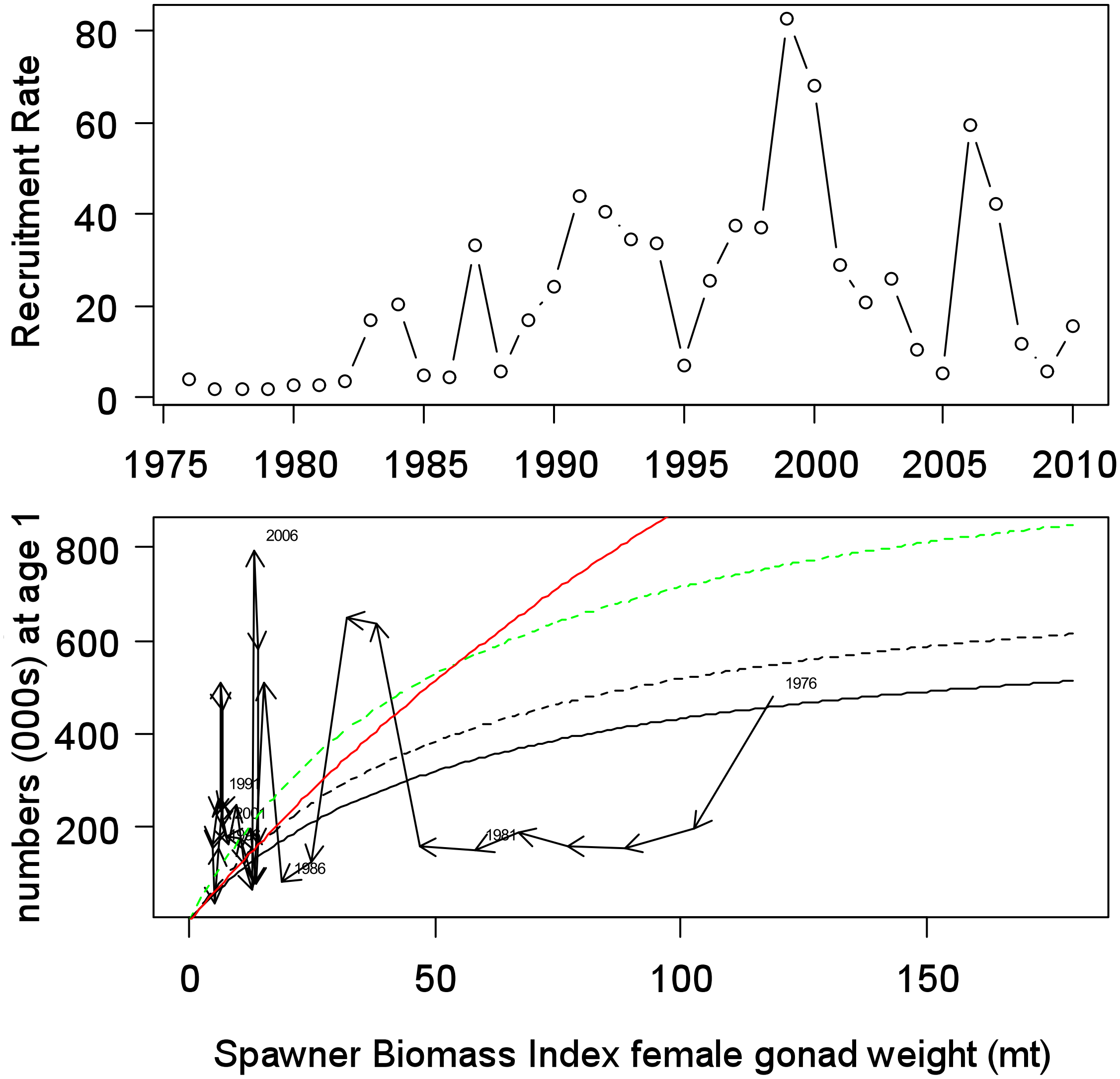

Figure 5. Lognormal (black solid lines), transformation bias-corrected lognormal (dashed lines), and gamma (red lines) maximum likelihood estimates of Beverton-Holt stock-recruitment curves for red snapper. The black dashed line is based on $\sigma_R = 0.6$, and the green dashed line is based on the external MLE of σ_R.