Report on SEDAR 20:
Atlantic menhaden
and
Atlantic croaker
review

March 8 –12, 2010

North Charleston, South Carolina

P.L. Cordue

Fisheries consultant

New Zealand

***For* CIE Independent System for Peer Review**

18 April 2010

Executive summary

During 8-12 March 2010, a SEDAR Review Workshop was convened in North Charleston, South Carolina, to review two draft stock assessments: Atlantic croaker, and Atlantic menhaden. I was a member of the Review Panel which consisted of four independent reviewers (three CIE appointed), and a Chair.

The menhaden base-model assessment was technically defensible and the conclusions that the stock was not overfished and overfishing was not occurring, with regard to the current reference points, were robust to most defensible alternative assumptions. As such, the base model represents the best currently available science on which to base management decisions. However, the assessment could be improved with greater attention paid to the structuring of the fisheries within the model and the appropriate weighting of data sets. The "filling in" of gaps in time series because of missing observations should not be done again. Also, overfishing determination must not be made, as it has been in the past, using number-weighted F (full F should be used). Finally, the use of F_{MED} based reference points should be reconsidered. Alternative reference points which give better protection to the spawning stock biomass and fecundity should be considered.

The croaker assessment contained several errors which the Panel, working with the Assessment Team, tried to correct during the meeting. We were not successful in producing a fully defensible base model. However, the data are such that it is reasonable to conclude that overfishing was probably not occurring in 2008. The overfished status is not determinable at this stage. The croaker model needs to be re-thought with careful consideration given to how to incorporate the bycatch from the shrimp fishery into the assessment. The key point is that the croaker bycatch and the croaker landings from the shrimp fishery must be linked, in the model, through a common effort term (being the effort in the shrimp fishery). Also, careful thought needs to be given to the definition of reference points – e.g., F_{MSY} should probably be defined as the optimal F in the directed fisheries given a certain level of effort in the shrimp fishery.

Background

During 8-12 March 2010, a SEDAR Review Workshop was convened in North Charleston, South Carolina, to review two draft stock assessments: Atlantic croaker (ASMFC, 2010a), and Atlantic menhaden (ASMFC, 2010b). The Review Panel operated under the SEDAR review workshop guidelines although the Data and Assessment workshops had been convened under the ASMFC review process.

I was one of three CIE reviewers appointed to the five person Review Panel. The meeting had an independent Chair and a NEFSC reviewer was also on the Panel (see Appendix 3 for a list of participants). This report presents my review findings and recommendations in accordance with the Terms of Reference (TOR) for the review (Appendix 2, annex 2). My views are mainly consistent with those expressed in the Summary Report, which contains the agreed findings and recommendations of the Panel.

Review Activities

Pre-meeting

Meeting documents and materials were made available in electronic form in advance of the meeting (see Appendix 1). I familiarized myself with the background material and read the main assessment documents in detail prior to the meeting.

There was a conference call between some Panel members and some other meeting participants, a few days before the meeting. A number of issues were discussed. In particular, I raised the issue of the inappropriateness of some of the TORs. I was advised that the Data and Assessment workshops had been convened under the ASMFC review process and that that process used the same TOR in each workshop (i.e., the same TOR in the Data, Assessment, and Review workshops). We were advised that in the Review workshop we were to primarily consider whether the TOR had been met by the Assessment Team or not.

I raised this issue with CIE by email before the meeting. They consulted NMFS who advised that the CIE reviewers were to conduct the review with regard to the existing TOR and must not just consider whether the Assessment Team had met their TORs. Since some of the TORs actually required the CIE reviewers to run the stock assessment models (e.g., “Croaker: TOR 5. Perform retrospective analyses…) I requested of the Chair that further clarification be obtained from NMFS before the meeting. An email was received in which the CIE reviewers were assured that they were not required to adhere to the TORs in a literal sense where they called for inappropriate actions, such as performing an assessment.

At a lunch meeting prior to the opening of the formal review, the Panel members agreed to a division of tasks with regard to the first draft of the Summary report. I accepted the role of menhaden leader (responsible for compiling the first draft of the menhaden report); and I also agreed to write the first draft of the report for menhaden TOR 2-7.

Meeting

The meeting began on schedule and generally followed the agenda during the five days (Appendix 2, annex 3). The meeting convened at 1pm on the first day with introductions and a review of the TOR. The formal presentations of the assessments began with croaker. Before the adjournment of the meeting on the first day, the Panel formulated a written set of requests for the croaker Assessment Team to address over-night and the following morning.

On the next day, before the first menhaden presentation, I gave a brief presentation on a flip chart with regard to overfishing determination. I had noted that both Assessment Teams were making the over-fishing determination using number-weighted Fs at age (rather than the full F). I explained why this approach was invalid (see Appendix 4). The menhaden presentations then began. During the remainder of the meeting, presentations, formulation of requests, and responses to requests, were interspersed as appropriate. The meeting concluded on Friday morning with a response from the croaker Assessment Team to a Thursday evening request. The croaker assessment contained several errors which the Panel and the croaker Assessment Team had tried to correct during the meeting. As there were still issues with the croaker assessment it was agreed that a further request would be made, by email, after the meeting.

Post-meeting

The Panel formulated a final croaker request by email which the Chair forwarded to the croaker Assessment Team. I completed a first draft of my assigned sections of the Summary Report (menhaden TOR 2-7) which I circulated to Panel members before flying home. On my return to New Zealand, using section drafts that were available, I compiled a near to complete draft of the menhaden report which I forwarded to the Chair.

From time to time, over the following two weeks, as a member of the Panel, I contributed comments and editorial suggestions for drafts of various sections of the Summary Report. I also contributed comments on a draft of an appendix to the croaker Assessment Report which described the corrected assessment that the Panel had requested.

Summary of findings: Menhaden

The menhaden assessment was technically defensible and the stock status conclusions from the base model (not overfished, not overfishing), with regard to the current reference points, were robust to most defensible alternative assumptions.

That is not to say that the assessment was without fault. The Assessment Team had “filled in” gaps in a CPUE time series and a catch-at-age time series where direct observations were not available. This is common practice for catch-at-age data when using VPA models but is unnecessary and very bad practice when using a statistical catch-at-age model. Also, the weighting scheme used in the base model was inappropriate with far too much weight being given to catch-at-age data (see Discussion section below).

The Panel formulated a "reference model" which made better assumptions than those in the base model. However, it gave essentially the same results as the base model. This is one of the main reasons why the Panel were comfortable with the results from the base model.

Each of the TOR are specifically considered below.

1. *Evaluate precision and accuracy of fishery-dependent and fishery-independent data used in the assessment.*

   The Atlantic Menhaden fishery was modeled as one east coast stock. Data included commercial and recreational landings at age from Florida to Maine, a fishery-dependent adult index developed from Potomac River Fisheries Commission (PRFC) pound net survey, and a juvenile index (JAI) developed from coast-wide beach seine information. In addition, growth, weight, and maturity at age were estimated using fishery dependent and independent information, while age and time-variant natural mortality was estimated using a multi-species virtual population analysis (MSVPA-X, see TOR 2).

   Landings and biological data from the commercial purse-seine reduction fishery were well characterized. Reliable data from the commercial bait fishery have only been available since 1985. CVs were estimated for the landings data; the bait landings were considered less reliable and were given higher values – especially in the early years. Commercial discard is not documented but is assumed trivial compared to total landings, so it was not included. Recreational harvest and discards were estimated through the Marine Recreational Fisheries Statistical Survey (MRFSS). Recreational harvest is minimal and is believed to be caught primarily with cast nets for use as bait. Biological data were not available for the recreational fishery, so the recreational landings were included with the bait fishery.

   Data from biological sampling for length and age for the reduction fishery were available from 1955 through 2008. Biological samples for the bait fishery are available since 1988, and sampling improved in 1994 when a pilot study was initiated to increase the sampling intensity compared to the reduction fishery. Ages are determined using scales. Estimation of growth was complicated due to size dependent migration. This was accounted for by weighting mean fish weights by catch in numbers by year, season, and fishing area, which is a reasonable approach. The use of cohort-specific weight and length-at-age to account for apparent density-dependent growth is also justified.

Maturity was re-examined on the recommendation of the 2004 Peer Review Panel. New estimates were based on 2004 and 2008 collections. The results were similar to previous studies. However, there is a potential confounding of maturity estimates because the samples are mainly taken during a major spawning event – therefore, estimates are probably biased high.

Two alternative methods were used to estimate adult relative abundance indices from pound-net bait fishery landings collected by the PRFC. One was a simple total-catch over total-effort index (CPUE) and the other was developed using a generalized linear model (GLM). There were years where the data were not collected, producing data gaps. Values were estimated to fill the gaps in the CPUE index. The PRFC adult CPUE index, with gaps filled, was used in the base model. Making up data is always bad practice and was unnecessary in this case since the model can accommodate gaps within time series.

Data from state beach-seine surveys were used to construct two alternative sets of juvenile indices. In the first approach a single coastwide juvenile index was calculated (JAI). The Assessment Team used this coastwide juvenile index in the model, assuming that the number of samples from each survey would provide suitable weightings of the different regional trends. The second approach combined relative abundance data from groups of adjacent states according to the similarity of trends in the state-specific time series and fitted all of the regional time series within the model (letting the model decide on the relative weighting of each regional index to the total coastwide juvenile abundance). I favor the second approach since the number of samples in each survey is very unlikely to provide an appropriate weighting of different trends.

Specific questions specified in TOR 1. are addressed below.

a. *Discuss data strengths and weaknesses (e.g. temporal and spatial scale, gear selectivities, aging accuracy, sampling intensity).*

Strengths of the fishery-dependent and fishery-independent data:

- The reduction fishery landings and biological sampling information have been collected since the 1950s in a consistent manner.
- Daily logbooks (Captains Daily Fishing Reports) have been collected since 1985, and detail purse-seine set locations and estimated catch. Vessel compliance is 100%. This information is used to decrease "topping off" bias. Topping off is the practice of taking one more set to fill the hold at the end of a long trip. These added fish are typically smaller than fish in the rest of the hold.
- Scales have been used for ageing since the 1950s, and have been read by the same person since 1969. A re-aging program was conducted in 2009 to determine precision of aging. The standard deviations associated with age estimates were used to provide the error associated with the age composition data. It was assumed constant over time.

Data weaknesses:

- The Panel had some questions on the age structure and estimated selectivities of the commercial reduction and bait fisheries by area. After inspection, it was revealed that the age structure of the landings was a result of the area harvested rather than the type of fishery. I agree with the Panel recommendation that the commercial fishery be modeled by area (north vs south) rather than by fishery (reduction vs bait).
- A re-aging study was conducted to estimate precision (see above), but little age validation work has been conducted. An ongoing validation study at Old Dominion University has had good agreement between scale and otolith ages, but few age 2 and 3 fish have been processed, and fish >3 are not included.
- The PRFC pound-net index had a number of years with missing data, which the Assessment Team filled in with estimated values. Genuine data-gaps should be left alone.

b. *Report metrics of precision for data inputs and use them to inform the model as appropriate.*
   This is an instruction to the Assessment Team. I provide some comments on the assumed precision of data inputs in the Discussion section (below).

c. *Describe and justify index standardization methods.*
   Another instruction to the Assessment Team. The standardization methods were generally appropriate. The Assessment Team used standard GLM approaches to derive juvenile indices and one of the alternative adult relative time series.

d. *Justify weighting or elimination of available data sources.*
   The Assessment Team did not include adult effort-based indices from the reduction fishery. I agree with their judgment that these indices were unlikely to be tracking abundance given the numerous changes in the structure of the fleet and processing factories. For my comments on the weighting of data sources see the Discussion section (below).

2. *Evaluate models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points.*

The Beaufort Assessment Model (BAM) was the only model used to produce final assessment results. This is a statistical forward-projection model with separable selectivities using the Baranov catch equation. Catch histories, catch-at-age, juvenile and adult abundance indices were all fitted in the model assuming two fisheries (reduction and bait). Constant selectivities were estimated for the fisheries and the indices. Lognormal likelihoods were assumed for the catch histories and indices with multinomial likelihoods for the catch-at-age data.

The MSVPA-X model was used to estimate age and year specific natural mortality from 1982-2008. The estimates were then assumed known in the base BAM run in those years with the average at-age estimates applied to the years 1955-1981. The MSVPA-X model was peer reviewed in 2005 and recommended for use in estimating natural mortality but not as a full assessment model. The Review Panel did not revisit this recommendation. However, I am not convinced of the appropriateness of using these estimates in the base model. In reality, natural mortality is age and year specific. However, it is not clear that the current understanding of the eco-system and the available data are sufficient to provide reasonable estimates of menhaden natural mortality at this level of detail. Fortunately, the debate is academic, as it happens that the assessment results are not sensitive to the use of age-specific natural mortality or age-and-year specific natural mortality.

The base model has a number of strengths:

- well tested software, population dynamics equations, and likelihoods
- based on a good understanding of stock structure and migration patterns
- reasonable certainty in the catch history over an extended period
- extensive catch-at-age data from the main fishery sampled in a consistent manner over many years
- defensible recruitment indices and an adult abundance time series (pound-net CPUE)
- defensible estimates of age and year specific natural mortality.

However, there are also some potential weaknesses in the base model:

- the definition of the fisheries in the model is based on the product produced rather than the fishing method and time and location of fishing
- gaps in the catch-at-age data for the bait fishery and the pound-net CPUE indices were "filled in" with unobserved data
- the recruitment indices and adult abundance time series may not be representative of the whole population (the pound-net CPUE is very limited spatially)
- the input variance assumptions, especially with regard to effective sample sizes, are inconsistent with the model residuals (the effective sample sizes are too high, and therefore uncertainty in model outputs is underestimated)
- there are strong residual patterns for the reduction fishery catch-at-age.

Some of the problems with the base run were examined by the Panel in a number of sensitivity runs. A "reference run" was specified by the Panel:

- define two fisheries based on location (north – where the larger/older fish are typically caught; and south – where smaller/younger fish are typically caught; most catch in both areas is by purse-seine)
- reinstate the gaps in the data which were filled in with unobserved data
- use a maximum effective sample size of 200 for catch-at-age data
- allow the southern selectivity pattern to be domed

Sensitivities to this run included the use of time-blocked selectivities for the southern fishery based on known changes in the fishery (three time blocks were used). This run had the lowest AIC amongst comparable runs, suggesting that the use of the additional parameters was justified by the improvement in fit. A visual examination of the catch-at-age residuals also showed some reduction in the extent of the residual patterns. The determination of stock status for the reference run and sensitivities was the same as in the base model.

At my request, the Panel also evaluated the status of the stock relative to unfished fecundity. Two alternative “productivity periods” were considered. A “recent” period (1992-present) and the “full” period (1955-present); productivity in each period was determined by the mean and variance of the recruits, and the average natural mortality and mean weight-at-age over the period. Unfished fecundity for each period was determined by running the model forward, with stochastic recruitment, with no fishing until stochastic equilibrium was established (the mean fecundity is then, by definition, the unfished fecundity; representing the “carrying capacity” of the population under the assumed productivity regime).

For the base model, fecundity since 1998 was estimated at 5-10% of unfished fecundity for the full-regime and 10-20% for the recent-regime. The results for the best-AIC model were similar, but higher (10-15% for the full-regime since 1998 and 20-30% for the recent-regime), and also showed a slowly increasing trend since 1965 (the base model was fairly flat from 1965 to 2008).

Such an evaluation should have already been performed by the Assessment Team and included in the Assessment Report since it highlights a potentially serious problem with too little spawning stock biomass being available as a buffer for the stock should an extended period of poor recruitment occur.

Specific questions specified in TOR 2. are addressed below.

a. *Did the model have difficulty finding a stable solution?*

   The Panel requested that convergence be checked with some jittered starting values. Twenty five runs were performed and all runs converged to the same solution.

b. *Were sensitivity analyses for starting parameter values, priors, etc. and other model diagnostics performed?*

   An extensive set of sensitivity runs were performed for the base model including higher and lower *M*, alternative weights on data sets, alternative selectivities, and an alternative start year. The only result of note was that leaving out the juvenile index (JAI) resulted in an over-fishing status in 2008. On investigation it was found that the JAI supported higher recruitment in the last three years than other data sets. Removal of the JAI was sufficient to move the point estimate of 2008 F just above the overfishing threshold.

c. *Have the model strengths and limitations been clearly and thoroughly explained?*

   These were discussed in the assessment document and were also considered by the Panel (see above).

d. *Have the models been used in other peer reviewed assessments? If not, has new model code been verified with simulated data?*

   BAM has been used in several other peer-reviewed assessments.

e. *Compare and discuss differences among alternative models.*

   The Panel formulated an alternative BAM run which addressed the main problems identified with the base run. Given the other uncertainties, the differences in the assessment results between the two models are relatively minor (see above).

3. *Evaluate the potential for conducting assessments at a sub-regional level (e.g. Chesapeake Bay).*

All of the recent research results are consistent with a single Atlantic coast-wide menhaden stock. Although data are available to enable assessments at a sub-regional level, the results would be meaningless from a biological point of view (and would be of no use in making sensible management decisions). The issue of potential sub-regional quotas or fishing limits is outside the TOR for this review. However, I note that the implementation of such an approach could not sensibly be done by sub-regional assessment (and setting sub-regional quotas on the basis of the assessments).

4. *State and evaluate assumptions made for all models and explain the likely effects of assumption violations on model outputs.*

These were discussed under TOR 2. However, each point in the checklist is addressed below.

a. *Calculation of M.*
   Discussed under *b.*

b. *Choice to incorporate constant or time-varying M and catchability.*
   Year and age-specific M were estimated in the MSVPA-X model and assumed known in the base BAM run. Sensitivity runs with higher and

lower M and age-specific but time-invariant M did not change the status determination.

c. *Choice of selectivity patterns.*
These were estimated in the base run although domed-selectivities were not allowed. The potential impact of mis-specification was investigated by allowing domed-selectivities in some runs, but this did not change the status determination.

d. *Choice of time steps in models.*
The model had a simple annual cycle and assumed that all fisheries were operating year-round. This is a significant departure from reality but it is unlikely to have a major impact on assessment results. Nevertheless, it would be better to model the timing of the fisheries more accurately, particularly as the timing of some of them has changed in recent years.

e. *Error in the catch-at-age matrix.*
The catch-at-age data are assumed to follow a multinomial distribution in each year with effective sample sizes equal to the number of trips sampled. This is a mathematically convenient and commonly made assumption which is almost certainly violated. In this particular case, the effective sample sizes appear to be too high as the model residuals are much more variable than they should be given the assumed sample sizes. Also, there are obvious patterns in the residuals for the reduction fishery. Lower sample sizes and alternative splits of the fisheries, together with alternative selectivities alleviated these problems to some extent. Different point estimates were obtained but stock status determination was unaltered.

f. *Choice of a plus group for age-structured species.*
A plus group was used at an appropriate age.

g. *Constant ecosystem (abiotic and trophic) conditions.*
Ecosystem conditions are unlikely to have been constant over the period in which the stock was modeled. There are attempts in the model to deal with changing conditions in terms of year-specific natural mortality and cohort specific growth. The reference points used assume that the time period modeled is representative of a single constant regime. This is a reasonable approach as without a full understanding of the processes involved it is not possible to know how long a “regime shift” might last (or even if it has occurred). There is some evidence of a “regime shift” in 1992 to lower productivity. This was considered by the Panel when calculating unfished fecundity (two alternatives: 1992-present or 1955-present).

h. *Choice of stock-recruitment function.*

It is assumed that there is very little relationship between population fecundity and recruitment (i.e., steepness is close to 1). There is no evidence for a relationship in the model estimates of fecundity and recruitment. However, recruitment is quite variable and there could be a stock-recruit relationship which is not discernable for this reason. The current reference points are independent of steepness, so this assumption has no consequences for status determination.

i. *Choice of reference points (e.g. equilibrium assumptions).*
The use of F_{MED} based reference points is of concern. It appears that the stock has been at low levels of population fecundity for many years and yet the current reference points (and the F_{MED} reference points of previous years) provide a determination of "not overfishing" and "not overfished". I recommend that alternative reference points be considered and chosen on the basis of providing better protection for SSB or population fecundity relative to the unfished level.

5. *Evaluate uncertainty of model estimates and biological or empirical reference points.*

Sensitivity runs were discussed under TOR 2.b. Almost all sensitivity runs gave the same stock status determination as the base run. However, from the bootstrap analysis of the base run, it is clear that there is considerable uncertainty with regard to the overfishing status of the stock in 2008 (with 37% of the runs indicating that overfishing was occurring). I note that uncertainty is substantially underestimated in the bootstrap analysis as the assumed effective sample sizes are too high (see discussion of potential base-model weaknesses under TOR 2.)

Specific questions specified in TOR 5. are addressed below.

a. *Choice of weighting likelihood components.*
The likelihood components were un-weighted which is appropriate for maximum likelihood estimation. It allows standardized residuals to be calculated and AIC to be used to rank models.

6. *Perform retrospective analyses, assess magnitude and direction of retrospective patterns detected, and discuss implications of any observed retrospective pattern for uncertainty in population parameters (e.g., F, SSB), reference points, and/or management measures.*

A retrospective analysis was performed by the Assessment Team for the base model. There were no retrospective patterns of any consequence, nor were there likely to be. It is a common misconception that retrospective analysis can, of itself, provide useful information on estimator uncertainty or bias. Retrospective patterns, if they are particularly severe, can be an indication of problems with the model which would warrant further investigation. However, retrospective patterns are common for estimators and of themselves do not indicate a problem. It is easy to construct a very poor estimator which will have no retrospective pattern. It is also true that many good estimators will, purely by chance, often show a retrospective pattern.

7. *Recommend stock status as related to reference points.*

I support the recommendation of the Assessment Team that the stock status determination is "not overfished" and "not overfishing", relative to the current reference points. Further, I also agree with the Assessment Team that the uncertainties in the assessment are such that there could have been overfishing in 2008 (removal of the JAI from the base model gave that determination and many bootstrap runs also fell in the overfishing zone).

I also note that a strictly valid determination of the overfishing status requires comparison of full Fs and not number-weighted Fs. This is not a well-known result, but it is obvious once the problem is pointed out. When full F is constant at a reference level, the number-weighted version of F changes depending on recruitment strengths (see Appendix 4).

8. *Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made by next benchmark review.*

The Panel and the Assessment Team gave numerous recommendations in their reports. Below, I focus on what I consider to be the highest priority recommendations.

Short term (improvements for the next benchmark review)

a. Consider model specifications similar to the Panel's reference run for future assessments. This includes realistic effective sample sizes for catch-at-age data, allowing the gaps in the pound net index and bait fishery age composition where data are not available, modification of the reduction and bait fleets to northern and southern fleets (consider how the definition of these fleets might vary over time given the expansion and contraction of the stock and the fisheries), and time-varying domed selectivity for the southern fleet.
b. Overfishing determination should be done with full F. The number-weighted fishing mortalities relative to the number-weighted F-reference points do not always provide the correct overfishing determination.
c. Do not use F_{MED} based reference points. Change to reference points which provide more protection for SSB and fecundity.
d. Examine weighting of datasets in the model. As a starting point, determine the weights so that the input variance assumptions are consistent with the estimated variance of residuals. Deviations from this weighting pattern may be desirable but need to be justified.

e. Examine the timing of fisheries and indices in the model. Many of the fisheries are seasonal and need to be timed appropriately (i.e., if they are not year-round fisheries, do not model them as such). Incorrect timing may affect model fits.

Long Term

a. Develop a coast-wide adult menhaden survey. Possible methods include an air-spotter survey, an acoustic survey, or an industry-based survey with scientific observers on board collecting the data. In all cases, a sound statistical design is essential (involve statisticians in the development and review of the design; some pilot surveys may be necessary).

Summary of findings: croaker

The croaker assessment brought to the SEDAR review by the Assessment Team contained several errors which the Panel, working with the Assessment Team, tried to correct during the meeting. In the end, a fully defensible base model was not achieved. However, the available data are such that the overfishing determination from the corrected base model can be considered fairly robust (and I support the conclusion that overfishing was probably not occurring in 2008).

The primary problem with the corrected base model (and all of the models considered by the Assessment Team) is that the croaker bycatch from the shrimp fisheries is not properly incorporated into the model runs. There are two problems.

The method used to construct the bycatch estimates was inappropriate. A fixed ratio of croaker bycatch to shrimp landings was used in the calculations (within three periods, each of the three ratios based on some observations). This ignores the variation in croaker year class strengths and shrimp effort which undoubtedly occurred during each of the periods which were not directly sampled. Essentially, the method uses an extrapolation which cannot be justified. The Assessment Team were right to be wary of using the estimates, but they were wrong to ignore the issue of bycatch in the shrimp fisheries. It is possible to restructure the model and fit the measurements of bycatch within the model (see research recommendations below).

Also, the Assessment Team incorporated the croaker *landings* from the shrimp fishery into the commercial catches within the model. This creates a problem with the calculation of F_{MSY} (and hence B_{MSY}) because the effort in the shrimp fishery cannot be properly modelled. The link between the age-0 discards in the shrimp fishery and the older-age landings has to be maintained so that higher landings from the shrimp fishery necessarily result in higher discards from the shrimp fishery and vice versa; also, the effort from the shrimp fishery has to be independent of the effort in the directed fisheries. In the calculation of F_{MSY} done by the Assessment Team, a single F is optimized (even when shrimp discards are included) – this ignores the independence of the directed and bycatch fisheries. F_{MSY} needs to be determined conditional on the presence of some level of landings and discards from the shrimp fishery (see research recommendations).

Each of the TOR are specifically considered below.

1. *Evaluate precision and accuracy of fishery-dependent and fishery-independent data used in the assessment, including the following but not limited to:*

   The Atlantic croaker fishery was modeled as one east coast stock. Data included: commercial and recreational landings and discards at age; a fishery dependent index developed from the Marine Recreational Fishery Statistical Survey (MRFSS); and four fishery independent indices: National Marine Fisheries Service (NMFS) bottom trawl survey, Virginia Institute of Marine Science (VIMS) survey, SEAMAP-South Atlantic survey, and North Carolina 195 survey. In addition growth, weight, maturity, and natural mortality at age were estimated using both fishery dependent and independent information.

   Commercial landings data by gear were available from 1950 to 2008 from Florida (FL) to New Jersey (NJ) which spans the range of the stock. These data were collected by NMFS and state agencies at various reporting levels over the time series. The commercial landings data from 1981 through 2008 from FL to NJ were used in the assessment to conform to the years where recreational landings are available. Daily or trip-level data are currently collected in most states in the ASMFC management region. Data collection methods have changed over time for a number of states, and therefore data may not be comparable throughout the time series. The bulk of the landings come from Virginia (VA) and North Carolina (NC).

   The Assessment Team was concerned about how to assign CVs to the commercial landings data. This is generally an arbitrary decision, which is one reason why it is better to assume that a catch history is known and to explore implications of the uncertainty in catch history through sensitivity analysis (see the Discussion section below).

   The Panel had questions about the use of gillnets, which is a significant part of the fishery in recent years. There is a legitimate concern about potential changes in fishery selectivity but the current effort data by gear are not adequate to examine changes.

   There are three major types of commercial discard, from the scrap, finfish, and shrimp fisheries. Data on the quantity of discards by year and area is much more uncertain than for landings.

   The scrap fishery is one in which the fish species that are unmarketable as food, are sold unsorted, usually as bait. NC initiated a scrap-fishery sampling program in 1986, which was used to estimate the proportion of croaker in the unsorted landings. Atlantic croaker is a major component of the NC scrap fishery. There is concern that there are no data to estimate landings from scrap fisheries in other states. Different gears are used in the scrap-fisheries in other states, so it may not be appropriate to use the NC data to estimate bycatch in other states. Estimates of scrap landings have declined by an order of magnitude since the early part of the assessment time period. This decline may be due to the enactment of various gear related regulations along the coast.

A variety of gears used to catch finfish along the coast also have a bycatch of Atlantic croaker. NMFS observer data were used to estimate the bycatch in gillnets and otter trawls. The Assessment Team estimated croaker bycatch using the method recommended for scup during the 2009 data poor workshop. This method may not be reliable for croaker, due to the low number of trips which landed croaker.

Atlantic croaker is also a bycatch in the south-eastern Atlantic shrimp fishery. The Assessment Team developed rough bycatch estimates using the ratio of croaker catch to shrimp catch. These estimates suggested that in some years the bycatch was larger than the directed harvest. However, the method used is far from ideal and is more likely to track shrimp landings than croaker bycatch. (See discussion above.)

Recreational landings and discards were provided through MRFSS from 1981 through 2008. The majority of the harvest was in VA (62%), with FL, NC and Maryland (MD) next in importance. MRFSS harvest estimates for croaker were fairly reliable with low proportional standard errors. The Panel inquired about the use of 10% discard mortality for the recreational fishery. There are no discard mortality studies on Atlantic croaker; the 10% is based on rates used for red drum and weakfish (other sciaenids).

Data from biological sampling for length, weight and age for the commercial fishery were available from a number of states over differing time frames. NC (1979 to 2008) and VA (1989 to 2008) had the longest sampling programs, with NC being the only state that sampled over the assessment time series. NC initiated a biological sampling program for the scrap fishery in 1986, and is the only program along the coast. The information collected from the scrap fish sampling is used to estimate the proportion of croaker in the fishery and the size structure. There are no long-term programs for collecting biological data on the bycatch of croaker in the shrimp fishery, but historical work indicates that nearly all the discarded bycatch were age 0. Recreational length information was collected in the MRFSS intercept survey.

Croaker ageing was originally determined using scales, but switched to otoliths in 1996. NC's biological sampling collected paired-samples of scales and otoliths from 1996–1999 which were used to develop a scale-otolith transition matrix. The matrix was used to convert scale based age-length keys (ALK). The 2005

Peer Review Panel had concerns about ageing protocols, so an ASMFC ageing workshop was conducted in 2008. New ageing protocols were developed. The Panel had concerns that length, weight and maturity at age might be mismatched with cohort due to the new ageing protocol and the protracted spawning period. The Assessment Team reviewed the length and weight at age and found that they were cohort based. I continue to have concerns about the maturity at age, since new maturity estimates have a much higher percentage of mature age 0s compared to the past (over 40% mature at age 0 is very difficult to believe; it appears that the new ageing protocol assigns a lot of age 1 fish to the age-0 year class).

A fishery-dependent and four fishery-independent time series were developed. Two different methods were used for calculating recreational CPUE indices. The first method, "directed trips", used total catch divided by total effort for trips which targeted or caught croaker. This was the preferred method of the Assessment Team and was used in their initial base model. As an alternative, they used the method of Stephens and MacCall (2004). The Assessment Team was concerned that the Stephens-MacCall method resulted in some unrealistic species associations. The unrealistic species associations were probably due to use of the full data set without stratification. This method, or an alternative method of identifying trips which fish in "croaker habitat" should be explored in future analyses. The directed-trips method is not defensible as an abundance index without a detail analysis in support of the time series (for example, it must be demonstrated that the indices are not hyper-stable; this could occur if the distribution of the directed trips contracts and expands with the availability of croaker).

The NEFSC multi-species trawl survey was used to develop an index. The survey uses a stratified random design based on three depth strata. On examination, I found that the inshore strata were not consistently sampled. Also, the indices were calculated as numbers-per-tow rather than being worked up as an area-swept estimate of total-numbers or biomass (area-swept estimates enable the estimated trawl-survey proportionality constant to be used as a model diagnostic). The Panel recommended dropping the inshore depth strata, development of a depth-by-latitude based stratification using the mid and offshore depth strata and estimating the index using the area-swept approach. The Assessment Team also developed fishery-independent indices using data from the VIMS juvenile trawl survey, the SEAMAP South-Atlantic coastal survey, and the NC Survey-195 which catches young of the year (YOY) croaker.

Specific questions specified in TOR 1 are addressed below.

a. *Discuss the effects of data strengths and weaknesses (e.g. temporal and spatial scale, gear selectivities, aging accuracy, sample size, standardization of indices) on model inputs and outputs.*

Strengths of the fishery-dependent and fishery-independent data:

- Landings data were available from all states in the range of Atlantic croaker distribution, and biological samples were available from states with the major fisheries (88 to 99% of total catch).
- Paired scale/otolith collections were used to develop a scale-otolith transition matrix and applied to the scale-based age-length keys.
- The 2005 Peer Review Panel recommended the standardization of ageing procedures across states. The ASMFC held an ageing workshop in 2008 which developed standardized ageing protocols.

Data weaknesses:

- Very unreliable estimates of croaker bycatch in the shrimp fishery.
- Reporting methods for landings data have changed over time and estimates may not be comparable before and after the change.
- The method used to estimate the finfish-fishery bycatch using NMFS observer data may not be reliable.
- There are no reliable estimates of landings in the VA scrap fishery.
- Otolith ages have not been validated with known age samples.
- The proportion mature at age 0 is unreliably estimated.
- For catch-at-age estimation the use of only one ALK may not be adequate for some fisheries if they operate year round, or during a period of fast growth.
- Use of the directed trips method to estimate a recreational CPUE index is unlikely to be appropriate.
- The NMFS survey inshore depth stratum was sampled inconsistently. A new time series was derived during the meeting but this should be revisited with a more considered analysis.

b. *Report standard errors of inputs and use them to inform the model if possible.*

   This is an instruction to the Assessment Team. However, I do comment on the assumed sample sizes and precision of data inputs above and in the Discussion section (below).

c. *Justify weighting or elimination of available data sources.*

   Another instruction to the Assessment Team. However, in an attempt to correct the initial base model the Panel did recommend the exclusion of the CPUE time series (see discussion under TOR 1). Also, on the recommendation of the Panel, the assumed effective sample sizes for commercial catch-at-age data were reduced from the number of fish aged to the number of length-samples taken (this was just a temporary measure to get the sample sizes down to something more reasonable given the multinomial assumption).

2. *Evaluate models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points.*

The model used was described as a "hybrid model" by the Assessment Team because they had modified an age-structured production model to fit catch-at-age data. However, what they ended up with was simply a very standard statistical catch-at-age model. They could have performed the assessment in several available packages or even borrowed the BAM model used in the menhaden assessment.

Strengths of the croaker assessment are:

- The model was able to be changed at short notice since the Assessment Team had used their own code.
- The model is predominantly based on well sampled time series of catch data from commercial and recreational fisheries (including discards) derived with reasonable certainty, catch at age information from the two main fisheries and fishery independent survey data.
- The model was compared against alternative models that apply differing structural assumptions (an alternative ASPM structure and a biomass dynamic model - ASPIC), which gave similar trends in the biomass trajectories and exploitation rates.

With regard to the corrected base model (see Appendix A of the croaker stock assessment report, ASFMC, 2010a):

- Although based on known formulations, the code implementation was developed by the Assessment Team prior to and during the meeting and no evidence of code or model validation was presented.
- The corrected base model has not been fully reviewed (indeed the Panel helped construct the model, so who has reviewed it?)
- The level of croaker bycatch in the shrimp fishery is ignored in the base model (and all sensitivity runs which incorporate shrimp bycatch of some level use scaled versions of the *unreliable* shrimp-bycatch estimates).
- The model is not correctly formulated to account for the shrimp fishery effort (which catches a range of age classes of croaker). It is a mistake to split the shrimp fishery into commercial catch and age-0 discards. The link between these catches, in terms of effort, needs to be maintained to allow F_{MSY} and other reference points to be properly calculated (see discussion above).

Specific concerns, raised about model structure and coding issues, which were discussed and reviewed with the Assessment Team (and incorporated into the corrected base model) include:

- The coding of several parts of the model which did not follow standard formulations e.g. the multinomial assumption on proportions at age and the scaling of the selection pattern used within the estimation of F_{MSY}; and incorrect formulas for standardized residuals.

- The assumption of a population age structure at equilibrium in the first year when strong year class effects were apparent throughout the available catch-at-age data.
- In the initial base model, using the age data from the NFMS and SEAMAP surveys in a "first step" to estimate selectivities and then assuming the selectivities fixed in the second and final step (and excluding the age data).
- The inclusion of the "directed trips" recreational CPUE times series (which was not justified as an abundance index and could be hyper-stable; certainly it is contradictory to the NMFS survey time series).
- The use of the shrimp by-catch data which is based on a raising procedure which results in croaker by-catch being closely related to shrimp landings rather than the effort expended and incoming croaker year class strength.

Specific questions specified in TOR 2 are addressed below.

a. *Did the model have difficulty finding a stable solution? Were sensitivity analyses for starting parameter values, priors, etc. and other model diagnostics performed?*
   Sensitivity analyses were presented within the assessment report and during the review. The dominant sensitivities in model estimates are not dependent on the model structure or starting values but derive from the data sets to which the model is fitted and assumptions concerning the biological characteristics of the stock, specifically the shrimp by-catch and the maturity of the age-0 croaker.

b. *Have the model strengths and limitations been clearly and thoroughly explained?*
   Model strengths and weaknesses were reviewed with the Assessment Team and are discussed above in this section and below in TOR 3

c. *If using a new model, has it been tested using simulated data?*
   No, this is discussed above in this section.

d. *Has the model theory and framework been demonstrated and documented in the stock assessment literature?*
   Yes, discussed above in this section.

3. *State and evaluate assumptions made for all models and explain the likely effects of assumption violations on synthesis of input data and model outputs. Examples of assumptions may include (but are not limited to):*

   Maturity: Maturity for age 0 was initially modeled at 43% mature. The Panel considered this unlikely for a species that spawns primarily in the autumn and winter. A review of the species spawning and growth patterns established that there is potential for uncertainty as to which year class (as required by the assessment model) a fish counted as age 0 is derived from. Fish from the previous year class could potentially be included within the new maturity ogive applied in the assessment. It was established that this was unlikely to be the case for the catch-at-age data, for which the adjustment was made when reading and compiling the otolith data. Following a review with the Assessment Team the maturity ogive from the previous assessment, which assumes that age 0 fish are not mature, was used.

Specific questions specified in TOR 3 are addressed below.

a. *Calculation of M.*
   The assessment uses instantaneous natural mortality rates which decline with age and are constant over all years. The values are averaged across values derived from a series of methodologies applied to historical growth data and, although the analyses showed a range of values, the Panel agreed that the appropriate selections had been made and appropriate structure applied within the model.

b. *Choice to use (or estimate) constant, time-varying, or age-varying M and catchability.*
   See a. above.

c. *No error in the catch-at-age or catch-at-length matrix.*
   Catch-at-age data were assumed to have a multinomial distribution. As noted within TOR 1 the original formulation of the error model was incorrect; following a review with the Assessment Team this was corrected and the appropriate formulae were (hopefully) used.

d. *Choice of a plus group.*
   This was modeled appropriately.

e. *Population is at equilibrium.*
   This is also addressed in TOR 2, model structure. The assumption of a population age structure at equilibrium in the first year was considered inappropriate, when strong year class effects were apparent throughout the available catch-at-age data. Following discussions with the Assessment Team the starting numbers for each cohort present within the first year were estimated, improving the fit of the model.

f. *Constant ecosystem (abiotic and trophic) conditions.*
   There was no information on changes in ecosystem conditions, they are not considered within the assessment model. However, discussions did note anecdotal reports relating to environmental conditions in Chesapeake Bay that could impact on the population dynamics of this species.

g. *Choice of stock-recruitment function.*

A Beverton and Holt stock-recruitment relationship was assumed with a fixed steepness. I consider this appropriate given the lack of information in the data to estimate steepness within the model.

h. Choice of proxies for MSY-based reference points.
MSY-based reference points were used so no proxies were needed.

Determination of stock structure.
The assessment of croaker assumes a single population. Although alternative hypotheses of multiple stocks have been suggested, the information available for deriving separate assessments is too sparse and therefore the current level of aggregation is appropriate.

4. *Evaluate uncertainty of model estimates and biological or empirical reference points.*

Confidence intervals for the estimated stock metrics were provided and suggested that estimated trends in biomass and fishing mortality for the base model were well determined (given the model assumptions). Sensitivity runs gave similar trends in stock metrics as those from the base run apart from when shrimp by-catch estimates were included in the catch data. The uncertainty of model estimates and biological and empirical reference points is therefore dominated by the catch data set to which the model is fitted rather than the estimation procedure or model structure.

5. *Perform retrospective analyses, assess magnitude and direction of retrospective patterns detected, and discuss implications of any observed retrospective pattern for uncertainty in population parameters (e.g., F, SSB), reference points, and/or management measures.*

Retrospective analyses of the model were conducted and some retrospective patterns were apparent. However, as noted in response to the menhaden TORs this is probably of little consequence. During the meeting, with all the other issues that needed to be addressed for croaker there was no point in spending time looking for the cause of the retrospective pattern.

6. *Recommend stock status as related to reference points:*

In 2008 overfishing was probably not occurring. There has been an upward trend in biomass since the 1980s and a decreasing trend in F. There has also been an expansion in age classes in the catch and indices, which is consistent with increasing biomass and decreasing F. This recommendation is qualified because F_{MSY} has not been calculated in an entirely appropriate manner (see above).

With regard to the determination of overfished status, there is the major issue of croaker bycatch in the shrimp fisheries (in addition to the definition of F_{MSY} and B_{MSY}). It is not possible to make a confident determination of overfished status until the shrimp-fishery bycatch is properly incorporated into the croaker assessment and B_{MSY} is properly defined.

Specific questions specified in TOR 6 are addressed below.

a. *Biomass threshold and target.*
   The overfished status is very uncertain, because of the poor determination of croaker bycatch in the shrimp fisheries.

b. *F threshold and target.*
   The method used to calculate the biomass and fishing mortality thresholds and targets is not entirely appropriate. However, trends in biomass, directed catch, and bycatch from the shrimp fisheries are all such that overfishing was probably not occurring in 2008

7. *Compare trends in population parameters and reference points with current and proposed modeling approaches. If outcomes differ, discuss potential causes of observed discrepancies.*

This is a very poorly written TOR. Population parameters and reference points will only have trends if the modeler chooses a parameterization which allows a trend. I assume that this TOR is referring to trends in model outputs such as biomass and fishing mortality rates.

Comparisons were made with an alternative age-structured production model and with a biomass dynamic model. Both models gave similar perceptions of increasing biomass levels and decreasing mortality rates.

8. *If a minority [stock assessment] report has been filed, explain majority reasoning against adopting approach suggested in that report. The minority report should explain reasoning against adopting approach suggested by the majority.*

There was no minority report submitted to the review.

9. *Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made by next benchmark review.*

The Panel and the Assessment Team gave numerous recommendations in their reports. Below, I focus on what I consider to be the highest priority recommendations. It is a given that the shrimp-fishery croaker bycatch has to be incorporated into the base model.

Short Term (improvements for the next benchmark review)

a. Develop lower and upper bounds on the catch histories for the croaker fisheries together with a best estimate. Fit catch as known in the models. Explore the effects of alternative defensible catch histories through sensitivity runs.
b. Give very careful consideration to how the bycatch of croaker in the shrimp fishery can be properly incorporated into the model. In the absence of an effort time series for the shrimp fishery (which would potentially be useful for assessments of other bycatch species) the current estimates of bycatch (from the three time periods) could be fitted in the model. That is, structure the model to have a shrimp fishery which catches a range of age classes (including age 0 which are discarded). Fit to the shrimp fishery age-0 catches and older-age landings in the model by estimating the annual Fs for the shrimp fishery.
c. Give very careful consideration to how F_{MSY} (and hence B_{MSY}) should be defined for croaker given the presence of the shrimp fishery (again, the model must be structured to have a shrimp fishery catching a range of ages). It seems likely that F_{MSY} should be defined as the directed F which optimizes yield in the presence of the shrimp fishery (at some fixed level of shrimp-fishery effort).
d. Sort out the proportion of age-0 fish (43% mature at age 0 is not appropriate given the model structure which assumes a very discrete spawning period).
e. Analyze the recreational catch and effort data in detail to derive a defensible abundance index or to eliminate these data as a potential source of an abundance index.
f. Consider using alternative types of reference points that do not rely on a defined stock–recruitment relationship. SPR based reference points should be considered. An appropriate level of SPR can be determined for croaker by considering the trade-off between yield and SSB over a range of plausible levels of steepness. This evaluation can be done using models with deterministic recruitment or stochastic recruitment. The shrimp fishery must be modelled in all cases.

Long Term

a. Atlantic croaker otolith ageing methods need to be validated.
b. Develop and implement compatible and coordinated sampling programs for state-specific commercial scrap and shrimp fisheries in order to monitor the relative importance of Atlantic croaker in these fisheries.
c. Estimates of catch-at-age for a year-round fishery may not be reliably estimated from a length frequency and a *single* age-length key if some of the vulnerable fish are growing significantly during the fishing season (because age proportions at given length keep changing). If this is a problem for some of the croaker catch-at-age data, there are two alternative methods for avoiding the problem that should be investigated: development of separate age-length keys for different times of year; or directly sampling for age (otoliths) year round.

Discussion

I will limit this section to three areas where both Assessment Teams were in need of correction or advice. I also include a section on the review process.

Fitting to catch histories

In some parts of the world, it is common practice to fit to catch histories within the stock assessment model. This entails estimating annual Fs within the model for each fishery and also requires that a statistical error distribution be specified for the catch histories. The usual practice is to assume that the catch history is unbiased (in some sense) with a lognormal error distribution; a CV is specified by some means – normally just a subjective assignment (e.g., 5% if it is thought to be very accurate).

This is a reasonable approach if there really is a basis for assuming such a statistical distribution. However, catch data are normally very well known or alternatively very biased. Catch histories, by nature, are very different from other time series which are fitted in a model. They are normally largely made up of reported results. "Uncertainties" in catch histories are not driven by variance (observation error or measurement error) but by *bias*. And the bias is generally very hard to quantify. The methods used to correct for under-reporting are generally fairly ad hoc, making use of available data that in some sense should/might contain information on the missing catch. There will be a procedure, which will often be statistical in nature, but there is no certainty that the procedure will produce *unbiased* estimates of the catch.

A far better approach for dealing with major uncertainty in catch histories is to use sensitivity analysis to see if the stock assessment results are sensitive to the assumed catch history. Ad hoc methods, using whatever relevant data are available, can be used to bound the catch in each year, thus producing a lower bound and an upper bound on the catch trajectories. The "best guess" is still used as a base case, but it is assumed known exactly. The whole point is that the assessment results are *conditional* on the catch history. Within the bounds on the catch history, alternative defensible catch trajectories can be used in sensitivity runs. Thus, it can be seen whether different assumptions about catch history matter or not.

The alternative approach of fitting to the "best guess" and assuming that it is unbiased ignores any alternative catch histories. It is not the case, that by specifying CVs at different points in the catch history that the uncertainties "flow through" to the final assessment results (e.g., through a bootstrap). All that happens is that the confidence intervals become a bit wider depending on the CVs which were assigned. The crucial aspect of *bias* is completely ignored.

Number-weighted F for overfishing determination

For many years the menhaden overfishing determination has been made using number-weighted F. A simple mathematical argument shows that this is technically incorrect (Appendix 4). I note, that at least the menhaden Assessment Team were comparing the current number-weighted F with the number-weighted F reference point. The croaker team were making their overfishing determination by comparing the current number-weighted F with the full F from the reference point – clearly inappropriate.

If it is desired to compare "average" Fs in some sense, rather than full Fs (e.g., when there is a very domed selectivity) this can be done in a technically correct manner. It is just that the weightings at age need to be constant (e.g., use numbers at age in equilibrium at the reference point to weight the reference Fs and the current Fs).

Weighting of data
It is an open question as to how best to weight data in a stock assessment model. By this, I mean the CVs to use for abundance and biomass indices and the effective sample sizes to assume for catch-at-age or catch-at-length data (when a multinomial distribution is assumed). There is certainly no case for scaling likelihood components by an arbitrary factor (except, perhaps, as a crude way of doing data-weighting sensitivity runs).

While the question remains open, there are some weighting schemes which can be eliminated. Both Assessment Teams used inappropriate assumed effective sample sizes for the catch-at-age data in their models. The assumption that the catch-at-age data follow a multinomial distribution is a mathematical convenience which is unlikely to be satisfied. Nevertheless, it is a useful convention which is often followed. However, when doing so, it is important not to place too much weight on the data. As a rule of thumb the effective sample size (i.e., an appropriate n to assume for independent multinomial draws) is the number of samples taken rather than the number of fish aged. Also, even when a large number of fish are aged or a large number of samples are taken, it is generally not appropriate to assume effective sample sizes in the hundreds. The Panel suggested that sample sizes be truncated at a maximum of 200 – again, just a rule of thumb. Ideally, the process used to generate an estimated age frequency would be bootstrapped to determine an appropriate distribution (possibly not multinomial) and the corresponding precision.

One way to get defensible CVs and sample sizes is to iteratively re-weight data sets until the standard deviation of the standardized residuals is approximately equal to 1 (generally this is done by "adding" on process error to externally estimated observation error – which is always maintained – process error may be zero, but not negative). The resulting weights have variously been referred to as the "natural weights" or, somewhat cynically, as the "naive weights". I suggest that such weights should be used as a starting point. It may be desirable to deviate from them, but this should only be done with good cause (e.g., if a reliable abundance index is not being adequately fitted – then it should be up-weighted). The first step is to determine these weights; one of the reasons for calculating standardized residuals.

Critique of the NMFS review process
The TORs for this review were a major impediment to producing coherent review reports. The Summary report and this CIE report are compromised because of the need to respond to ambiguous and overly prescriptive TORs. I understand how we ended up with

these TORs because of the mixed nature of the process (ASMFC for Data and Assessment workshops, then SEDAR for the Review) and that the main fault lies with the ASMFC process. It is nonsense to have the *same* TORs for the Data, Assessment, and Review workshops.

On speaking to some meeting participants about why we had such prescriptive TORs I was told that this was because they had had poor review panels in the past that had not addressed the important issues. The high degree of prescription was an attempt to force the review panel to do a "good review". Unfortunately, the "book" on how to do a good review has not yet been written. When/if it is, then it will be written by an experienced reviewer rather than a council or a committee.

It is entirely appropriate that a council include in the TORs, for a review, specific questions that they would like to be addressed for the particular stock. It is counter-productive for them to specify TORs which purport to be a comprehensive technical checklist. They do not understand the technical terms and therefore mix up concepts and produce ambiguous or poorly written TORs.

The standard SEDAR TORs are a reasonable default to use for a review. However, these are also too prescriptive and require the review panel to tick far too many boxes. For example, it should not be required that the review panel determine whether the Data Workshop and the Assessment Workshop have fulfilled each of their TORs. A review panel should spend their time in determining whether the assessments are technically correct, whether the assessments represent the "best available science", and whether those assessments can be used as a basis for providing sound management advice.

Conclusions and Recommendations

The menhaden assessment was technically defensible and the stock status conclusions from the base model, with regard to the current reference points, were robust to most defensible alternative assumptions. However, the assessment could be improved with greater attention paid to the structuring of the fisheries within the model and the appropriate weighting of data sets. The "filling in" of gaps in time series because of missing observations is unnecessary and should not be done again. Also, overfishing determination must not be made, as it as been in the past, using number-weighted F (full F should be used). Use of F_{MED} based reference points should be reconsidered.

The croaker assessment contained several errors which the Panel, working with the Assessment Team, tried to correct during the meeting. We were not successful in producing a fully defensible base model. However, the data are such that, with reasonable confidence, we can conclude that overfishing was probably not occurring in 2008. The overfished status is not determinable at this stage. The croaker model needs to be re-thought with careful consideration given to how to incorporate the bycatch from the shrimp fishery into the assessment. The key point is that the croaker bycatch and the croaker landings from the shrimp fishery must be linked, in the model, by a common F.

Also, careful thought needs to be given to the definition of reference points – e.g., F_{MSY} should probably be defined as the optimal F in the directed fisheries given a certain level of effort in the shrimp fishery.

Detailed recommendations are given for menhaden under TOR 8 and for croaker under TOR 9.

References

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

Appendix 1: Bibliography of supplied material

Croaker Data Workshop

Anon. North Carolina Fishery Dependent Commercial Data Report for the ASMFC Atlantic Croaker Technical Committee Data Workshop, July 2009, Richmond VA. 51p.

Anon. NJ Delaware River Seine Survey (Fishery Independent). 2p.

Anon. Maryland Fishery-Dependent Programs Report for the ASMFC Atlantic Croaker Data Workshop, July 2009. 3p.

Anon. Georgia Department of Natural Resources Fishery-Independent Surveys. Report for the ASMFC Atlantic Croaker Data Workshop, July 2009. 15p.

Anon. Maryland Department of Natural Resources Fishery Independent Surveys. Report for the ASMFC Atlantic Croaker Data Workshop, July 2009. Chesapeake Bay Seine Survey 11p.

Anon. North Carolina Fishery Independent Data Report for the ASMFC Atlantic Croaker Technical Committee Data Workshop, July 2009, Richmond VA. 26p.

Anon. NJ Biological Characterization (Fishery Dependent). 1p.

Brown, K. 2009. Interstate Fisheries Management Program Implementation For North Carolina. Completion Report for NOAA award No. NA05NMF4741003. Study II: Documentation and Reduction of bycatch in North Carolina Fisheries. 29p.

Clark, J. The Commercial Atlantic Croaker Fishery in Delaware.1p.

Defilippi, J. Atlantic Croaker Commercial Landings for the Atlantic Coast from the Atlantic Coastal Cooperative Statistics Program Data Warehouse. 3p.

McDonough, C. Analysis of the NMFS Fall Ground-fish Survey for Atlantic croaker off the East Coast of the United States 1972 – 2008. 25p.

McDonough, C. Evaluation of SEAMAP survey data for Atlantic Croaker, *Micropogonias undulatus,* in the South Atlantic bight from 1990 to 2008.12p.

McDonough, C.; Arnott, S. Atlantic Croaker in the SCDNR Fishery Independent Surveys. 28p.

McDonough, C.J.; Charles A. Wenner. 2009. Estimating the size and age composition of the B–2 fish (caught and released alive) in the recreational fishery for red drum and spotted seatrout in South Carolina. Summary Report for Cooperative Research Program: Grant Num. 00750-83502. 20p.

Munyandorero, J. The Atlantic croaker fishery data on Florida's Atlantic coast, 1950-2008. A report to the ASMFC Atlantic Croaker Technical Committee. 61p.

Croaker Assessment Workshop

Munyandorero, J.; Eric Robillard. 2009. Fitting the non-equilibrium production model to Atlantic croaker (*Micropoginias undulatus*) fishery data. 13p.

Rickabaugh, H.; Laura M. Lee. Catch Curve Estimates of Total Mortality Rates for Atlantic Croaker. 6p.

Menhaden Data Workshop

Anon. Summary of SEDAR 20 Menhaden Data Workshop. 4p.

Anon. Summary of SEDAR 20 Menhaden Post-Data Workshop Call. 4p.

Anon. Atlantic menhaden habitat. 42p.
NMFS. 1995. Instructions for mehadedn sampling program. Revised May 1995 (from 1972, 1777, 1985 editions). 16p.
Dietrich, C.S. 1979 Fecundity of the Atlantic Menhaden, *Brevoortia Tyrannus*. Fishery Bulletin. 1979. 4 p.
Higham, J.R.; Nicholson, W.R. 1964. Sexual maturation and spawning of Atlantic menhaden. Fishery Bulletin. 1964. 17p
JHK. 1966. Verification of menhaden conversion factor. 4p.
Lewis, R.M.; Ahrenholz, D.W.; Epperly, S.P. 1987. Fecundity of Atlantic menhaden *Brevoortia Tyrannus*. Estuaries, 10: 347-350.
Lynch, Abigail J.; Jan R. McDowell; John E. Graves. 2010. A molecular genetic investigation of the population structure of Atlantic menhaden (*Brevoortia tyrannus*). Fishery Bulletin 108: 87-97.
Nelson, W.R.; Ingham, M.C.; Schaaf, W.E. 1977. Larval transport and year-class strength of Atlantic menhaden, *Brevoortia tyrannus*. Fishery Bulletin 1977. 19p.
Punt, Andre E., David C. Smith, Kyne Krusic-Golub, and Simon Robertson. 2008. Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's southern and eastern scalefish and shark fishery. Can. J. Fish. Aquat. Sci. 65: 1991–2005
Smith, J.W.; Douglas S. Vaughan; Jeffrey Brust. 2009. Commercial Harvest and Catch-at-Age for Atlantic Menhaden. 49p.
Vaughan, D.S. 2009. History of Assessments of the Atlantic Menhaden Stock along the U.S. Atlantic Coast. 17p.
Vaughan, D.S.; Matthew Cieri; Genevieve Nesslage. 2009. Life-History Based Estimates of Natural Mortality for Atlantic Menhaden. 22p.
Vaughan, D.S.; Joseph W. Smith. 2009. Reconstructing Historical Commercial Landings of Atlantic Menhaden. 27p.
Vaughan, D.S.; Joseph W. Smith; Erik H. Williams. 2009. Growth and Reproduction of Atlantic Menhaden. 36p.

Menhaden Assessment Workshop

ASMFS; Lance Garrison; Jason Link. 2004. An expanded Multispecies Virtual Population Analysis Approach (MSVPAX) to evaluate predator-prey interactions in Exploited Fish Ecosystems. VERSION 1.1, Users Manual and Model Description, 17 April 2004. 90p.
Various authors and dates. Directory: "old assessments", 12 pdf files.

Review Workshop

ASMFC. 2010a. Atlantic Croaker 2010 stock assessment report. 235p.
ASMFC Atlantic Menhaden Stock Assessment Subcommittee, 2010b. Draft Atlantic Menhaden Stock Assessment for Peer Review. 343p.
ASMFC Multispecies Technical Committee. 2008. Update of the Multispecies Virtual Population Analysis. 54p.
Northeast Fisheries Science Center. 2006. 42nd Northeast Regional Stock Assessment Workshop (42nd SAW) stock assessment report, part B: Expanded Multispecies

Virtual Population Analysis (MSVPA-X) stock assessment model. *U.S. Dep. Commer., Northeast Fish. Sci. Cent. Ref. Doc.* 06-09b; 308 p.

Appendix 2: Statement of Work for Patrick Cordue

External Independent Peer Review by the Center for Independent Experts

SEDAR 20 Atlantic Menhaden and Atlantic Croaker Review

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 with 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.com.

Project Description: SEDAR 20 will be peer reviews of assessments of Atlantic menhaden and Atlantic croaker stocks conducted by the respective stock assessment subcommittees of the Atlantic States Marine Fisheries Commission (ASMFC). The Southeast Data, Assessment and Review (SEDAR) process will coordinate the peer reviews during a single workshop. SEDAR peer reviews typically involve a panel composed of one NOAA/NMFS chair, one reviewer selected by the resource management agency, and three CIE reviewers. The lead assessment agency is the Atlantic States Marine Fisheries Commission with consultation by the Southeast Fisheries Science Center, NMFS. Peer reviews of the Atlantic menhaden and Atlantic croaker stock assessments are approved items of the SEDAR Steering Committee assessment schedule. Atlantic menhaden is an important industrial and bait fishery resource and contributes to commercial fisheries in portions of its range. It is also recognized as a vital ecological resource as a prey species. The most recent assessment of Atlantic menhaden was the 2006-update of a full assessment conducted in 2003. Atlantic croaker is an important recreational fishery resource and contributes significant commercial landings throughout its range on the Atlantic coast. The most recent assessment of Atlantic croaker status was in 2004 and presents stock status for the mid-Atlantic region. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment, statistics, fisheries science, and marine biology to complete their primary task of conducting an impartial and independent peer review report in accordance with the Terms of Reference to determine if the best available science is utilized for fisheries management decisions. 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 North Charleston, South Carolina during 8-12 March 2010.

Statement of Tasks: Each CIE reviewer 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 will forward this information to the NMFS Project Contact no later than 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>Foreign National Security Clearance</u>: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/sponsor.html).

<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.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

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

Other Tasks – Contribution to Review Panel Report: Each CIE reviewer shall assist the Chair of the panel review meeting with contributions to the Review Panel Report, based on the terms of reference of the review, and may assist the Chair in review and comment of an Assessment Summary Report. 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 in North Charleston, South Carolina during 8-12 March 2010.
3) In North Charleston, South Carolina during 8-12 March 2010 as specified herein, conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than 26 March 2010, 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 David Sampson, CIE Regional Coordinator via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and shall 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.

1 February 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
22 February 2010	NMFS Project Contact sends the CIE Reviewers the pre-review documents
8-12 March 2010	Each reviewer participates and conducts an independent peer review during the panel review meeting
26 March 2010	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
8 April 2010	CIE submits CIE independent peer review reports to the COTR
15 April 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Science 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 or 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 be 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 the regional Science Center Director.

Key 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

Dale Theiling, SEDAR 20 Coordinator, NMFS Project Contact
South Atlantic Fishery Management Council
4055 Faber Place, Suite 201
North Charleston, SC 29405
Dale.Theiling@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 20 Atlantic Menhaden and Atlantic Croaker Review

Atlantic Menhaden

1. Evaluate precision and accuracy of fishery-dependent and fishery-independent data used in the assessment:
    a. Discuss data strengths and weaknesses (e.g. temporal and spatial scale, gear selectivities, aging accuracy, sampling intensity).
    b. Report metrics of precision for data inputs and use them to inform the model as appropriate.
    c. Describe and justify index standardization methods.
    d. Justify weighting or elimination of available data sources.
2. Evaluate models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points.
    a. Did the model have difficulty finding a stable solution?
    b. Were sensitivity analyses for starting parameter values, priors, etc. and other model diagnostics performed?
    c. Have the model strengths and limitations been clearly and thoroughly explained?
    d. Have the models been used in other peer reviewed assessments? If not, has new model code been verified with simulated data?
    e. Compare and discuss differences among alternative models.
3. Evaluate the potential for conducting assessments at a sub-regional level (e.g. Chesapeake Bay).
4. State and evaluate assumptions made for all models and explain the likely effects of assumption violations on model outputs, including:
    a. Calculation of M.
    b. Choice to incorporate constant or time-varying M and catchability.
    c. Choice of selectivity patterns.
    d. Choice of time steps in models.
    e. Error in the catch-at-age matrix.
    f. Choice of a plus group for age-structured species.
    g. Constant ecosystem (abiotic and trophic) conditions.
    h. Choice of stock-recruitment function.
    i. Choice of reference points (e.g. equilibrium assumptions).
5. Evaluate uncertainty of model estimates and biological or empirical reference points.
    a. Choice of weighting likelihood components.
6. Perform retrospective analyses, assess magnitude and direction of retrospective patterns detected, and discuss implications of any observed retrospective pattern for uncertainty in population parameters (e.g., F, SSB), reference points, and/or management measures.
7. Recommend stock status as related to reference points.

8. Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made by next benchmark review.

Atlantic Croaker

1. Evaluate precision and accuracy of fishery-dependent and fishery-independent data used in the assessment, including the following but not limited to:
    a. Discuss the effects of data strengths and weaknesses (e.g. temporal and spatial scale, gear selectivities, aging accuracy, sample size, standardization of indices) on model inputs and outputs.
    b. Report standard errors of inputs and use them to inform the model if possible.
    c. Justify weighting or elimination of available data sources.
2. Evaluate models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points.
    a. Did the model have difficulty finding a stable solution? Were sensitivity analyses for starting parameter values, priors, etc. and other model diagnostics performed?
    b. Have the model strengths and limitations been clearly and thoroughly explained?
    c. If using a new model, has it been tested using simulated data?
    d. Has the model theory and framework been demonstrated and documented in the stock assessment literature?
3. State and evaluate assumptions made for all models and explain the likely effects of assumption violations on synthesis of input data and model outputs. Examples of assumptions may include (but are not limited to):
    a. Calculation of M.
    b. Choice to use (or estimate) constant, time-varying, or age-varying M and catchability.
    c. No error in the catch-at-age or catch-at-length matrix.
    d. Choice of a plus group.
    e. Population is at equilibrium.
    f. Constant ecosystem (abiotic and trophic) conditions.
    g. Choice of stock-recruitment function.
    h. Choice of proxies for MSY-based reference points.
    i. Determination of stock structure.
4. Evaluate uncertainty of model estimates and biological or empirical reference points.
5. Perform retrospective analyses, assess magnitude and direction of retrospective patterns detected, and discuss implications of any observed retrospective pattern for uncertainty in population parameters (e.g., F, SSB), reference points, and/or management measures.
6. Recommend stock status as related to reference points:
    a. Biomass threshold and target.
    b. F threshold and target.

7. Compare trends in population parameters and reference points with current and proposed modeling approaches. If outcomes differ, discuss potential causes of observed discrepancies.
8. If a minority [stock assessment] report has been filed, explain majority reasoning against adopting approach suggested in that report. The minority report should explain reasoning against adopting approach suggested by the majority.
9. Develop detailed short and long-term prioritized lists of recommendations for future research, data collection, and assessment methodology. Highlight improvements to be made by next benchmark review.

Annex 3: Tentative Agenda SEDAR 20 Atlantic Menhaden and Atlantic Croaker Review

SEDAR 20 REVIEW WORKSHOP

Atlantic Menhaden and Atlantic Croaker

Hilton Garden Inn – Charleston
5265 International Blvd., North Charleston, South Carolina

TENTATIVE AGENDA

TBN, Chair
Mr. Dale Theiling, SEDAR Coordinator

Monday, March 8, 2010

1:00pm – 5:30pm	**Afternoon Session**	
	Convene	Chair
	Introductions and Opening Remarks Coordinator	Chair and SEDAR
	Agenda Review	Chair
	TOR Review	Chair
	Task Assignments	Chair
	Croaker Data Presentation	Linda Barker
	Croaker Assessment Presentation	Laura Lee Katie Drew
	Croaker Assessment Discussion Analysts	Review Panel and

<u>Tuesday, March 9, 2010</u>

8:00am - 11:30am	**Morning Session**	
	Croaker Assessment Discussion	Review Panel
12:00nn Lunch		
2:00pm – 5:30pm	**Afternoon Session**	
	Menhaden Management History	Brad Spears
	Menhaden Data Presentation	Doug Vaughan (data) Rob Latour (indices) Matt Cieri (MSVPA and M)
	Menhaden Assessment Presentation	Doug Vaughan (model selection) Erik Williams (Beaufort Assessment Model) Behzad Mahmoudi (complementary model)

	Menhaden Assessment Discussion	Review Panel and Analysts

Wednesday, March 10, 2010

8:00am - 11:30am	**Morning Session**	
	Menhaden Assessment Discussion	Review Panel and Lead Analyst
2:00pm – 5:30pm	**Afternoon Session**	
	Stock Topical Discussions as needed	Review Panel

Thursday, March 11, 2010

8:00am - 11:30am	**Morning Session**	
	Complete Croaker Topical Discussions	Review Panel
	Croaker Review Workshop Report	Review Panel
12:00nn	**Lunch**	
2:00pm – 5:30pm	**Afternoon Session**	
	Complete Menhaden Topical Discussions	Review Panel
	Menhaden Review Workshop Report	Review Panel
	Croaker Assessment Summary Report	Panel, Stock Leader, Coordinator
	Menhaden Assessment Summary Report	Panel, Stock Leader, Coordinator

Friday, March 12, 2010

8:00am - 11:30am	**Morning Session**	
	Final Review of Panel Documents	Chair
12:00nn	Adjournment	Chair

Discussion Topics

Evaluation of data and their preparation and presentation
Choice and utilization of assessment models and methods
Continuity run from previous assessment(s)
Alternative assessment approaches
Identification of additional analyses, sensitivities, and corrections
Review of additional analyses and sensitivities
Initial Review Workshop recommendations and comments
Review of Data and Assessment Workshop research recommendations
Identify Review Panel research recommendations
Improvement of the SEDAR process
Assure all Terms of Reference are addressed
Develop and review draft Review Panel Report sections
Finalize workshop recommendations
Finalize Review Panel Report
Post-Review Workshop tasks and products due Chair and CIE

The timing of particular events is tentative, and the Chair may modify this schedule during the workshop as needed to complete stated tasks. However, to accommodate

travel planning the workshop will start as scheduled and will conclude no later than the stated time.
SEDAR is a public process, and the public is welcome to attend SEDAR workshops. Although no formal public comment period is scheduled, the workshop Chair will allow opportunity during the meeting for the public in attendance to comment on discussion items.

Appendix 3: Review workshop participant list

SEDAR 20
Atlantic Croaker and Atlantic Menhaden

Peer Review Workshop Participants
March 8-12, 2010
North Charleston, SC

Appointee	Function	Affiliation
Independent Review Panel		
Kim McKown	Chair and Reviewer	NY DMR
Dr. Tim Miller	Independent Reviewer	NMFS NEFSC
Patrick Cordue	Independent Reviewer	CIE
Dr. Chris Darby	Independent Reviewer	CIE
Dr. Geoff Tingley	Independent Reviewer	CIE
Rapporteurs		
Eric Robillard	Rapporteur – Atlantic Croaker	ASMFC AC SAS
Braddock Spear	Rapporteur – Atlantic Menhaden	ASMFC
Analytic Team and Presenters		
	ATLANTIC CROAKER	
Laura Lee	Lead Analyst	ASMFC AC SAS
Eric Robillard	Analyst	ASMFC AC SAS
Katie Drew	Analyst	ASMFC AC SAS
Linda Barker	Stock Leader	ASMFC AC SAS
	ATLANTIC MENHADEN	
Erik Williams	Lead Analyst (BAM)	ASMFC AM SAS
Matt Cieri	Model Presenter (MSVPA)	ASMFC AM SAS
Behzad Mahmoudi	Model Presenter (Stock Reduction)	ASMFC AM SAS
Genny Nesslage	Analyst	ASMFC AM SAS
Joe Smith	Analyst	ASMFC AM SAS
Amy Schueller	Analyst	NMFS SEFSC
Doug Vaughan	Stock Leader and Presenter (Data)	ASMFC AM SAS
Rob Latour	Presenter (Indices)	ASMFC AM SAS
Appointed Observers		
Robert Boyles	Commissioner	ASMFC
Nichola Meserve	Atlantic Croaker FMP Coordinator	ASMFC
Braddock Spear	Atlantic Menhaden FMP Coord.	ASMFC
Coordination		
Dale Theiling	Coordinator	SEDAR
Rachael Lindsay	Administrative Support	SEDAR

Acronyms
SEDAR 20 Participants List
Atlantic Croaker and Atlantic Menhaden

AC	Atlantic Croaker
AM	Atlantic Menhaden
ASMFC TC	Atlantic States Marine Fisheries Commission Technical Committee
BAM	Beaufort Assessment Model
CIE	Center for Independent Experts
FMP	Fishery Management Plan
MSVPA	Multispecies Virtual Population Analysis
NEFSC	Northeast Fishery Science Center, National Marine Fisheries Service
NMFS	National Marine Fisheries Service
NOAA	National Oceanic and Atmospheric Administration
NY DMR	New York Division of Marine Resources
SAS	Stock Assessment Subcommittee
SEFSC	Southeast Fisheries Science Center, National Marine Fisheries Service
SEDAR	Southeast Data, Assessment, and Review
TBD	To be determined
TBN	To be named

Appendix 4: Overfishing determination: why number-weighted F should not be used

Consider a population which is being fished at some reference level F_{REF} (e.g., F_{MSY}) with a constant selectivity pattern (e.g., logistic). Suppose that recruitment is constant and that the population is at equilibrium. The number-weighted version of F_{REF}, say F^*_{REF}, can be calculated from the equilibrium distribution of numbers-at-age and F_{REF}. Now, consider what happens to the number-weighted version of F, say F^*, when a large recruitment pulse is introduced into the population. As the pulse enters the first vulnerable age class (which is included in the calculation of F^* and F^*_{REF}), there is a large increased weight on a partially selected age class, and hence F^* is not equal to F^*_{REF} (it will probably be less than F^*_{REF}, but this depends on the particular selectivity and population parameters). As the pulse travels through each age class, the value of F^* changes, but it is unlikely to achieve equality with F^*_{REF} at any age. When it reaches the first fully-recruited age class, it is likely that $F^* > F^*_{REF}$. In any case, in this example, F remains constant at F_{REF}, but the number-weighted version of F varies – if used in an overfishing determination, it will give an incorrect status in most if not all years. This is an extreme example but, mathematically, it is sufficient to prove that the use of number-weighted F is not appropriate for overfishing status determination. In general, full F should be used.