44th Southeast Data, Assessment, and Review (SEDAR 44)

Independent Peer Review Report on the **SEDAR 44 ASMFC Red Drum Assessment Review Workshop held in Charleston, South Carolina August 25th-27th 2015**

Sven Kupschus

Prepared for

Center for Independent Experts

The Centre for Fisheries and Aquaculture Science
Lowestoft Laboratory
Pakefield Road
Lowestoft
Suffolk NR33 0HT
England, United Kingdom

Phone: +44 1502 524454
Email: sven.kupschus@cefas.co.uk
www.cefas.co.uk

Executive Summary:

SEDAR 44 was scheduled to review the status of the southern and northern Atlantic red drum stocks in order to provide the council with guidance on the management of these predominantly recreationally exploited stocks. Prior to the review it became apparent that the move to more integrated assessments, now using SS3, had delayed the assessment development process as the assessment team was having difficulty in attaining plausible estimates of the stock dynamics using the new approach. Following the presentations of the latest models in the development process, the review panel agreed that the assessments would not be suitable for the provision of the type of management advice required. The remainder of the meeting was spent in developing strategies for rapid model improvement by investigating diagnostics and conducting alternative runs to better understand the causes of the inconsistencies. Time, however, was limited at the workshop and many inconsistencies in model and data setup could not be examined in detail at the workshop.

A large number of suggestions were made during the workshop as to how to deal with specific issues, and how to diagnose the effectiveness of those measures both in an absolute and relative sense. The overarching and agreed prerequisite to making progress was to simplify the model, as the large number of possible correlations in parameters estimates made diagnosing the root cause of the problem difficult in the complex base models. Removing selectivity blocks and simplifying the selectivity functions in terms of the parameters needed made this possible for the northern stock. For the southern stock there were additional issues, but the types of data and their information content are similar for the two stocks so it is expected that the approach will also apply here. The suggested way to proceed from here to a model suitable for providing advice is to investigate the effects of adding in different suggested options for improving certain aspects of model performance one at a time and to compare the utility of the various model changes in terms of their ability to explain changes in the dynamics, requirements for management metrics, number of additional parameters required, the ability of the data to support model complexity (avoid over parameterization) and plausibility in terms of the resultant changes in stock dynamics to derive at a new base model, where upon the process is repeated.

Where there was less agreement amongst panel members during the report writing phase was on the relative importance of the different options, and the order in which to best resolve the various issues. In response to some requested analysis on the data at the workshop and further investigations of my own, following the meeting I found the data were much more informative on the cohort structure than was suggested by the simplified model. Because the structure appeared to be coherent across several data sources, to me it is fundamental to resolve the reasons for this inconsistency before moving on with other investigations such as altering selectivities, because without an accurate cohort structure on the relative scale at least there is little chance of developing appropriate selectivity estimates.

The other major uncertainty in the northern model was the scaling of the assessment. The raw data does not suggest that there has been a large scale change F or abundance over the shortened time frame. The lack of contrast means it is unlikely that any developed model will be able to accurately scale the assessment without the use of the tagging data. Investigations of the effect of the tagging data in the model suggest that the tag reporting rate is crucial to estimating the scale of recruitment. Some more detailed examination of how best to implement this (several options were provided), and a review of the tagging information for plausible absolute reporting rates should be able to achieve this in the near future.

Background:

The SEDAR44 review took place in Charleston, SC from the 25th-27th of August 2015. I participated in the review as an independent reviewer on behalf of the Center for Independent Experts (CIE). I contributed to the discussions on the evaluation of the use of data and model for both the southern and northern red drum stocks. I contributed sections to the review report as well as edits to the final version.

The original intent of the SEDAR44 process was to provide the scientific basis for management advice on southern and northern red drum stocks following the TORs provided. However, prior to the meeting and confirmed at the webinar on the 18th of August 2015, it became clear that the assessment team was not in a position to recommend either a base or alternate model suitable as the basis of advice for either stock. The revised aim of the meeting was to use the advice of the review panel to develop a clear path to take in the development of the both the southern and northern SS3 models. The deadline for altering the TORs to specifically reflect these revised aims for the meeting had passed (**Modifications to the SoW and ToRs cannot 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.**). However, it was agreed in consultation with the CIE that the existing TORs allowed sufficient leeway in the process for the review group to proceed with model development assistance. It was made clear at this point that it would not be possible to provide management advice on any model developed because the review would not be independent. If stock status estimation via indicators or other means were sought then this would also be possible based on the data; however, there was insufficient time to do both in the allocated time. AMSFC decided that the long-term view of developing suitable models was more important than the requirement for short term advice. It was agreed that the review would proceed in the direction of model development. The latest progress in model development by the assessment group were considered as base models in terms of the review although it was never expected that these models would be used in the evaluation of stock status or management projections.

The assessment group presented summaries of the data review workshop including data and life history descriptions followed by the introduction of the two SS3 models. Because of the shortness of time available, the continuity models in SCA were not presented, as they had been deemed unsuitable for advice because of changes in data collection as well as methodological concerns raised at the previous review (SEDAR18). The SCA models were occasionally discussed during the remainder of the review, where they had informed on SS3 model development and perceptions of stock dynamics.

Description by TOR:

1. *Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:*

   Sizeable portions of the data report refer to the development of data files for the SCA model, whereas the assessment report largely dealt with the SS3 models without specifying how these data for these were derived. It was not clear to the review panel till later in the meeting that there had been significant amounts of data sharing

between fleets and years. Although necessary for the more restrictive age based SCA approach at least theoretically there is a much reduced requirement to share data in the length-based likelihood based SS3 approach weighted by effective sample size. The idea behind the likelihood approach is that it should down weight the importance of data represented by few samples compared to those data representing the conglomerate of many independent samples.

The data workshop should not only aggregate data, but also characterize the trends in the data and compare those trends between different data sources where possible, see example of different survey indices or age analysis performed under TOR 4. A qualitative assessment of the utility of certain data sources, and the trends that they emphasize is very helpful in structuring model development, especially in SS3 where there are many different implementation options that place different emphasis on different aspects of the data.

a. *Presentation of data source variance (e.g., standard errors).*

Information from the commercial fisheries (historic in the south, current in the north) were considered largely appropriate to support age based assessments. It was not clear in all cases how the data used in the final model was developed, because the data workshop to a large degree was still focusing on the provision of data for the SCA model which required estimates of the numbers at age. As this is an additional step in the development from length based information required by the SS3 models, it was assumed that the data treatment was generally appropriate. There was some 'borrowing' of data across fleets and ages, because a minimum sample size threshold was used for both length and age samples. This may be a hang-over from the SCA model which required raised numbers at age to be estimated. SS3 does not have the same requirements, being able to deal with the variation introduced by small sample sizes or alternatively being able to deal with individual years where no age or length information is available.

The review panel requested an examination of internal consistency of indices (coherent estimation of cohort variability between years within a survey). The longline surveys (both north and south) indicated that even at the oldest ages cohort signals were still very apparent. In fact cohorts from the early 70's could be tracked. For the southern stock a continuous estimate of relative cohort strength (1970-2013) could be developed by overlaying all surveys. This suggests that the assessment model at least in theory should be able to reconstruct the population dynamics. Tracking cohorts in the assessment model was complicated because the information was entered length conditional (see TOR 3).

Generally MRFSS (now MRIP) data collection for the recreational fleet is considered highly variable, especially the length and age information. For red drum, a major recreational species the issue seemed less problematic than expected, although estimates for the released fish (B2) was poor. Data from the recreational tagging program was used to supplement these data, but it is not entirely clear to which degree the tagging volunteers are representative of

the wider recreational fleet, nor are there data available prior to 2004. This affects both the recreational discard fleet and the MRFSS-CPUE index in the models.

The appropriate effective sample size of length measurements taken from a landing is somewhere between one and the number of fish measured, but it is not possible from the sample itself to determine where in this range the effective sample size is located. The data workshop decided to use the lower end of the range (i.e. 1) for the effective samples size. This has little impact when sample sizes are similar over time or between fleets, but particularly with a declining commercial fishery and large differences in the size of length samples between commercial and recreational fleets in this model it has the potential to inappropriately weight the certainty in different data sources.

Variance estimates were developed for all fleets and indices. For a number of the indices the variability across all samples was used. However, from some of the survey descriptions it is clear that not all samples are equally likely to encounter red drum so that a stratified variance estimate would be more appropriate.

b. *Justification for inclusion or elimination of available data sources,*
c. *Calculation and/or standardization of abundance indices.*
d. *Estimation of discards and size composition of discards.*

Abundance data for 3-year old red drum was excluded from the gillnet survey on the basis that it did not capture a large number of 3-yearolds, had high variance estimates, and the cohort was represented in the MRFSS fleet. Closer examination of the data during the workshop, however, revealed that the available information, though sparse, tracked the cohort abundance of 1 and 2 year old red drum well. Excluding data on the basis of redundancy (available from other sources) is inappropriate when data is entirely independent.

Although maximum selectivity is assumed for large fish in both the longline index (LL) and the B2 fleet, it appeared that there was a greater preponderance of very large individuals in the latter. It is possible that this is associated with the tagging data targeting trophy fish, or bias in the length estimates of recreational catches.

B2 lengths were measured in inches, which led to problems with the length binning to 2cm lengths with intermittent length bins remaining empty because they do not correspond to any integer inch measurements. This problem needs to be fixed prior to further attempts to develop SS3 models. Although it is unlikely to directly interfere with the central tendencies of parameter estimates, it will affect the variance estimates, and as such may change the relative weighting between fleets. There was insufficient time to resolve the data issue at the review.

Accurate aging from otoliths is not thought to be a problem in red drum even at the oldest ages. Certainly the aging information from the longline surveys shows that 34 year old fish can be consistently aged. Aging from scales is only

applied to smaller individuals and below 60 cm individual can be aged based solely on length and the date of capture.

The development of historic parts of the timeseries, particularly commercial and recreational discard fleets is problematic. The data workshop extended these series in proportion to the rates of discarding observed prior to the introduction of the main management measures (1989-1991). While a more accurate reflection of the discard mortalities would be helpful, the method is commonly employed in cases where there are no other options.

Natural mortality is based on length according to Lorenzen rescaled M. However, estimate off length were derived by non-parametric means (SEDAR 18) inconsistent with the growth model applied in the assessment. The estimation of growth (based on SEDAR18) assumes that the age information is collected randomly which applies only to the longline surveys. Much of the remainder of the age-at-length information is collected length stratified. Treating it as random will underestimate the mean length at age of the younger ages, while overestimating it at older ages. While in principle the non-parametric method seems appropriate when VB growth is inapplicable. It seems unlikely that this would have a major impact on the estimation of stock status, but this should be investigated once a more advanced model has been developed.

Index data from Florida in contrast to other indices was modeled using strata, so that it removes the between strata variance from the uncertainty estimates. Because of this, the model will be more strongly influenced by the modeled index data given equal data quality. Also it seems likely that bootstrap variance estimates of the delta-lognormal approach are skewed or at least not normally distributed, consequently the CVs are poor descriptors of the uncertainty in the IR indices. Variance estimate for indices should be reviewed. It is unlikely that these results will have a major impact on the presented model, but more refined models may be more sensitive.

Tagging data after 2007 has been eliminated for the northern stock. Data exists, but the data workshop decided to cautiously exclude this data because the monitoring design was changed based on the recommendations of the last review to develop methodologies that can estimate reporting rates. Given the importance of the data in scaling the assessment, the change in design has been justified. Providing a consistent time series of tagging data should be investigated.

2. *Evaluate the definition of stock structure used in the assessment. Is the definition appropriate given the biology and management of red drum?*

The assessment group presented a review of the stock structure which is still somewhat unclear genetically. The differentiation between Atlantic and Gulf stock is well documented, but the Atlantic north and south stocks are more difficult to distinguish genetically. Even if this is reproductively one stock, from a management and exploitation perspective, the north and south should be treated as separate units

since tagging data suggests that red drum migration in the Atlantic is limited to a degree where the assumption of a fully mixed stock would be compromised. The Indian River population may be as isolated from the southern stock as it is from the Gulf stock, possibly because the width of the shelf in that area largely restricts the life history to the coastal lagoon and nearshore area. From an assessment perspective the stock definition used appeared to be appropriate pending a better understanding of the contribution of the Indian River fish to the southern stock.

3. *Evaluate the methods and models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points, including but not limited to:*

   Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of red drum?

   Both stocks have been progressed to SS3 implementations. Fundamentally, the approach is robust, quality controlled and state of the art. However, SS3 is not a model but an approach covering many types of fisheries models, and appropriate implementation is conditional on how the available information is used. In this case, neither model seemed to represent realistic stock dynamics in the view of the assessment team. This opinion was confirmed by the review panel and consequently the review focused on way of diagnosing the problems and suggesting avenues for improvement. As usual with SS3, there are different ways of treating the sources of information and the review panel provided suggestions. Complete evaluation of the efficacy of a number of these options was not possible, because time was only sufficient to deal with the major issue of scaling. The panel was therefore not able to provide a step-by-step path to model improvement, so the suggestions should be seen more as options to try from a new base model. Several investigations and model implementations were run by the assessment team during the workshop to try to identify the cause of poor performance. These model runs are not described in detail here beyond the generalities of what could be gleaned about the possible development of more appropriate assessment models.

   The review panel endorsed the transition from the previous SCA methodology to an SS3 implementation of the red drum assessment. It felt that the progress in the development of SS3 provided a quicker and more robust means of attaining stable assessment results. Although it is possible to implement similar methodologies in the previous SCA approach through further development of that model, many of the planned model improvements such as an integrated evaluation of the tagging data, length based selectivities, as well as other potentially useful implementations such as non-parametric selectivity functions were already available in SS3. The assessment team's approach had been to try to recreate the SCA dynamics in SS3. However, this wealth of additional options and the move from age based to length based selectivity resulted in a complex model with alternate and often implausible dynamics. The assessment team was finding it difficult to make progress in model improvements, because the complexity of the interactions between parameters made it difficult to diagnose the origin of the unrealistic population dynamics within the model. The review panel, having examined the development of the red drum assessment also found it difficult to diagnose the cause of unintuitive stability of the model to sizable changes in the model assumptions for the same reasons. As a remedy, the panel

suggested significant simplification of the model was necessary to determine the underlying cause. The suggestion is not to be interpreted as a recommendation to use a simplified model for the advisory process. To develop a base model, starting from a functional SS3 implementation with simplified dynamics (removal of selectivity blocks, rationalization of the selectivity functions, and the removal of the lead-in period where only catch data is available) provides a quicker and more structured refinement towards a final model than trying to diagnose the differences between highly complex models with poorly understood dynamics. Construction of a simplified model according to these principles provided an opportunity to examine some potential avenues to improve model dynamics suggested by the panel:

Length vs Age data contributions to the total likelihood function:

As described under TOR 1, a comparison of the relative cohort strength (description of data) indicated that recently initiated longline surveys not only provide a useful indications of the current biomass of older ages, but the age structure they contained also provided a description of historic recruitment strength at least back to the 1973 and 1978 cohorts suggested to be well above average in both the north and south stocks. The simplified northern assessment model did reflect the abundance of these cohorts better than the original model, but particularly the 1973 cohort was still smeared across adjacent cohorts (1971-1975), for which little evidence existed in the age information.

The reasons for this appears to be that at the larger sizes captured in the long line survey length provides little information on age, i.e. the length conditional age is imprecisely estimated. As noted in the assessment report as well as SEDAR18 reviews, finding appropriate models to describe red drum growth has been difficult. Von Bertalanffy growth appears to be a particularly poor fit, but this is what is implemented in SS3 (including an age variant k option). It appeared the relatively poor fitting of the growth curve was overestimating the uncertainty in ages so that the model was unable to focus in on the 1973 cohort.

Options for improvement:

1) **Down weighting of the length information.** The model diagnostics suggest that the effective sample size of the length information for most fleets and indices is too large given the information content as determined by the model. This is surprising, since particularly for the commercial fleets an extremely cautious estimate of only 1df for every length sample collected was used. It is not entirely clear why samples taken from different landings should be correlated unless the selectivity data is persistently finding deviations from the assumed selectivity. This is something that should be re-examined if and when selectivity blocks are re-introduced into the model as this may have an impact. For the B2 fleet (northern model, and possibly southern model) there may be additional difficulties, because the conversion from inches to cm has meant that there are empty length bins in the length frequency which results in an inappropriately large contribution to the LL.

   Reducing the effective sample size of all length information to 10% of the original value had little impact on the parameter estimation of the northern

model, and because the age information is implemented conditionally on the lengths, the relative importance of these two sources of information did not change. It is not clear whether increasing the effective sample size for age information would have the desired effect of changing the relative importance of the two sources of information in the model. However, it appears realistic that length is poorly informative of age at the larger sizes, and that it is not possible to overcome the problem by reweighting the length information.

2) **Improved estimation of growth model.** Because of the interactions between growth and selectivities, it was difficult to optimise the growth function based on model diagnostics, but comparing the raw age at length information with the predicted growth from the model indicated that a model continued to have difficulties in fitting the data predominantly for the intermediate ages. The simplified model was using 2 k-estimates, one below 5 and one above (based on maturity on the basis of the findings for the southern stock), but the data suggested that switching between the two k's at a higher age in the northern stock should improve the fit. A new run using a transition age of 8 years improved the fit, reducing the LL by 55 units. However, the uncertainty in age conditional upon length at the largest sizes remained unchanged. The 1973 cohort had not been observed at younger ages in the simplified model (longline data starts at 2007), so that the model was diluting the persistent information on age in the LL survey by transitioning through length. Despite the minor improvement in the log-likelihood, there was little change in the estimation of parameters other than in growth.

I have examined size at age external to the model in order to illustrate the information content of the data. Figure 3 shows the expected length frequency distributions for each age based on average size at age (no growth function with mean and SD calculated from age data). Up to age 9, length is most informative on age despite sizeable variability in the size at age with adjacent ages sharing around 50-80% of the length frequency variation (Figure 1).

The next 10-15 cohorts show much reduced growth and hence differentiation in length (only around 10% difference), but are by eye at least still somewhat distinguishable based on a much reduced variability. Yet, numerically, it seems there is very little signal remaining with close to 0 percent of the variability in length being attributable to a single age (Figure 1). For the 10 oldest ages, the shared component of the variability in length across all 10 cohorts is around 50%, and a substantial part of the apparently unique variability is made up by differences in the estimates in SD, rather than the mean. Given the noisy length data, it is clear that the signal to noise ratio for the lengths is so small that effectively length provides no information on age for older ages.

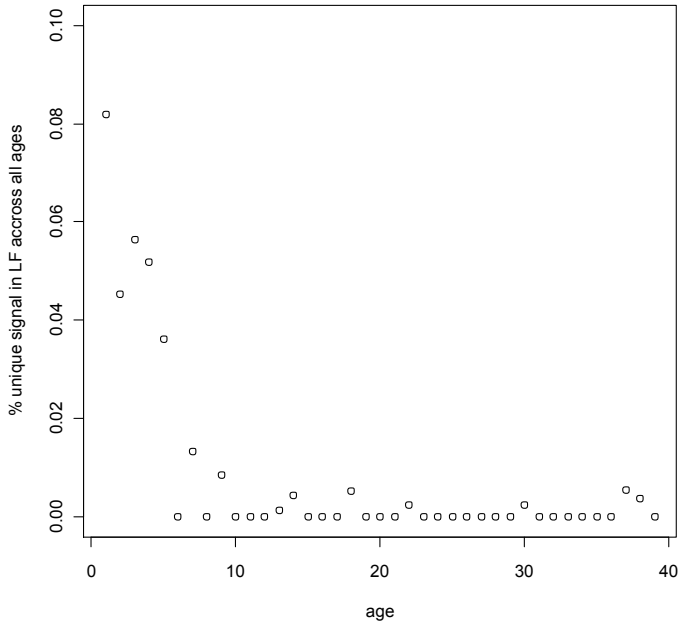

Figure 1: Plot of the unique variability in length frequency composition, indicating the ages that can be separated based on the relative length frequency information. P=1 means length is wholly informative on age, while 0 means length provides no information without accurate scaling of the relative abundance of aged individuals.

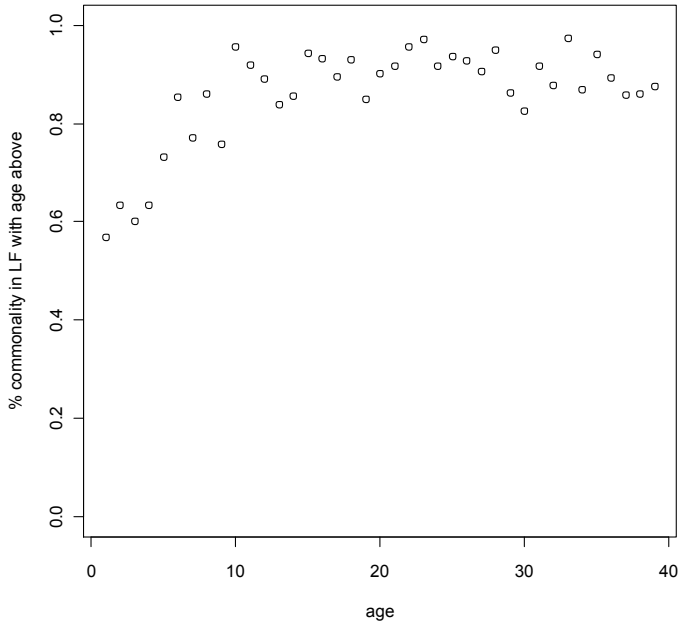

Figure 2: Proportion of age conditional variability in length frequency shared with the subsequent age, by age. This indicates that for ages > 10 separating adjacent cohorts based on the age conditional length will not be possible without an accurate reflection of the relative proportion of lengths.

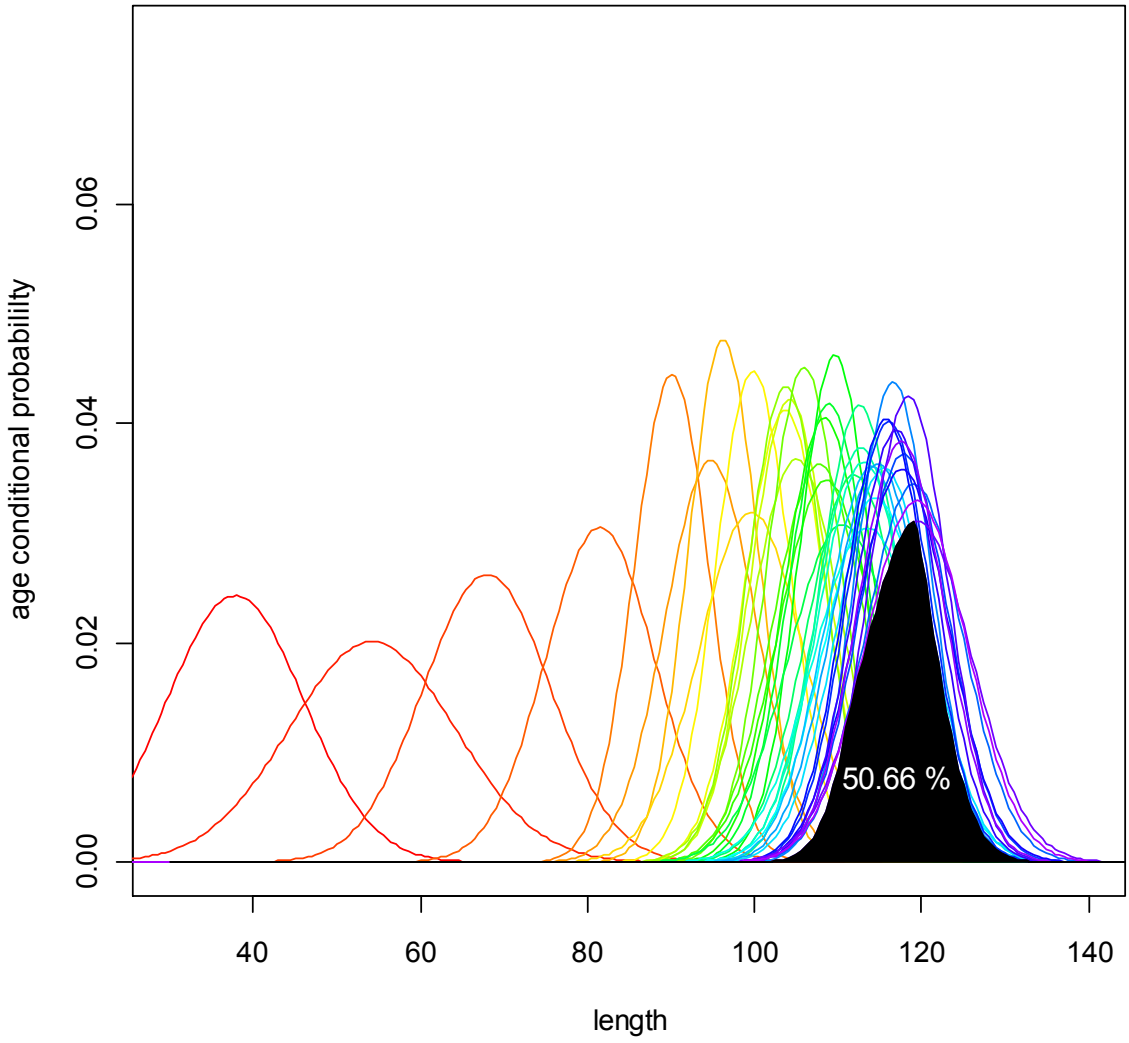

Figure 3: Age conditional probabilities at length for the combined age data from the northern assessment with increasing ages coloured red to violet. Marked in black is the overlap of the oldest 10 ages illustrating the minimal amount of information that lengths provide on ages.

3) **Age-based selectivities for longline fleet.** Discussions at the review were inconclusive as to whether treating age, conditional up on length would result in the same likelihood profile as treating ages multi-nominally. My view is that if the growth model is appropriate, the two will converge to the same thing; however, if there is process error particularly when the process error is for a specific cohort (faster or slower growing cohorts will be underrepresented) or when the estimated uncertainty does not match the true variability (contrast in cohort is reduced). Spline smoothers for cohorts with more than 4 different ages (
4) Figure 4) indicate that some long-term temporal bias in the growth estimate. Unfortunately, because different cohorts have different ages, the represented spline smoothers are not the best way to represent this

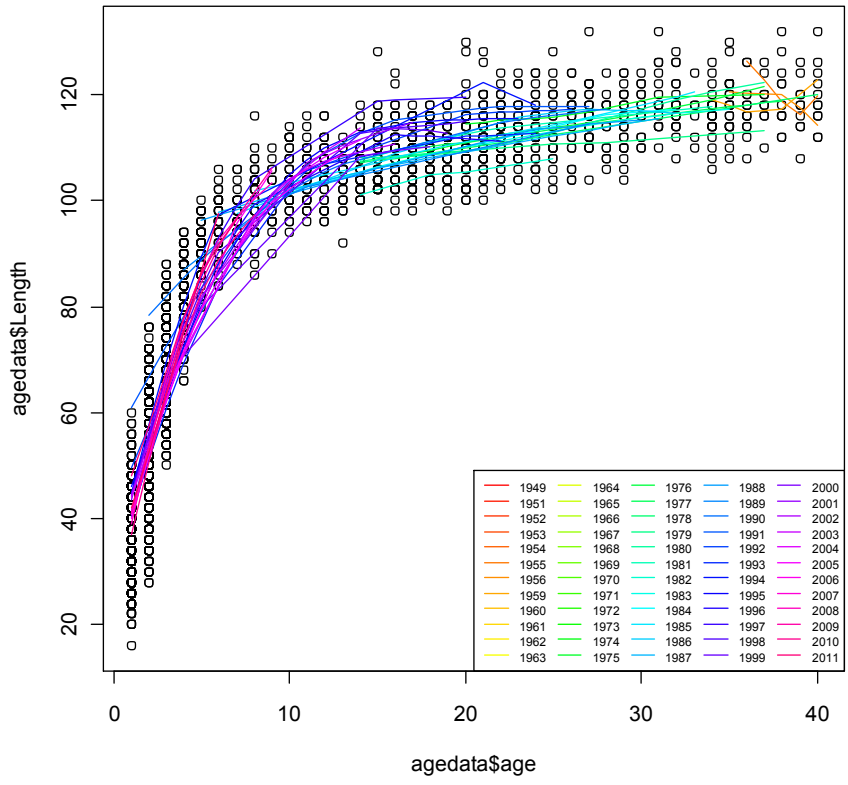

Figure 4: Spline based estimates of the growth of all cohorts in the data. Trends are not entirely comparable because the ages available are not the same for all cohorts; however, it does indicate differences between individual cohorts when examined over the same range of ages and it suggests some more systematic changes in growth over time. These differences will affect the age likelihood profile, because it is conditional on age.

information and more work along these lines would help. However, examining the data points for some individual cohorts clearly shows some cohorts are more poorly matches by a single growth function than others, even when examined over a similar data range (Figure 5). The result is that age data is not reconcilable with length data in the model. Therefore, the usually small penalty on recruitment deviates takes on a greater role in minimizing the likelihood. Reduced recruitment deviates means the contrast in the cohort signal is reduced. In most cases, there is little that can be done about this, because age data is usually collected length stratified and the issue has to be dealt with through sensitivity analyses. Here, however, random age samples from the LL survey exist that can reflect the actual ages much better than the lengths through a flawed growth model. Fixing the growth model would resolve the issue, but is a long-term goal. In the short term, the issue is best resolved by using the multinomial probability directly.

In my opinion it is likely that the longline survey will need to be implemented as age-based rather than length based selectivities, but certainly this option needs to be explored. At the review, there was a difference of opinion as to why the cohorts were being smeared. It was suggested that the catch data may contain other age information that was contradicting the view of a large 1973 cohort. Consequently, I performed the following analysis in discussions after the meeting:

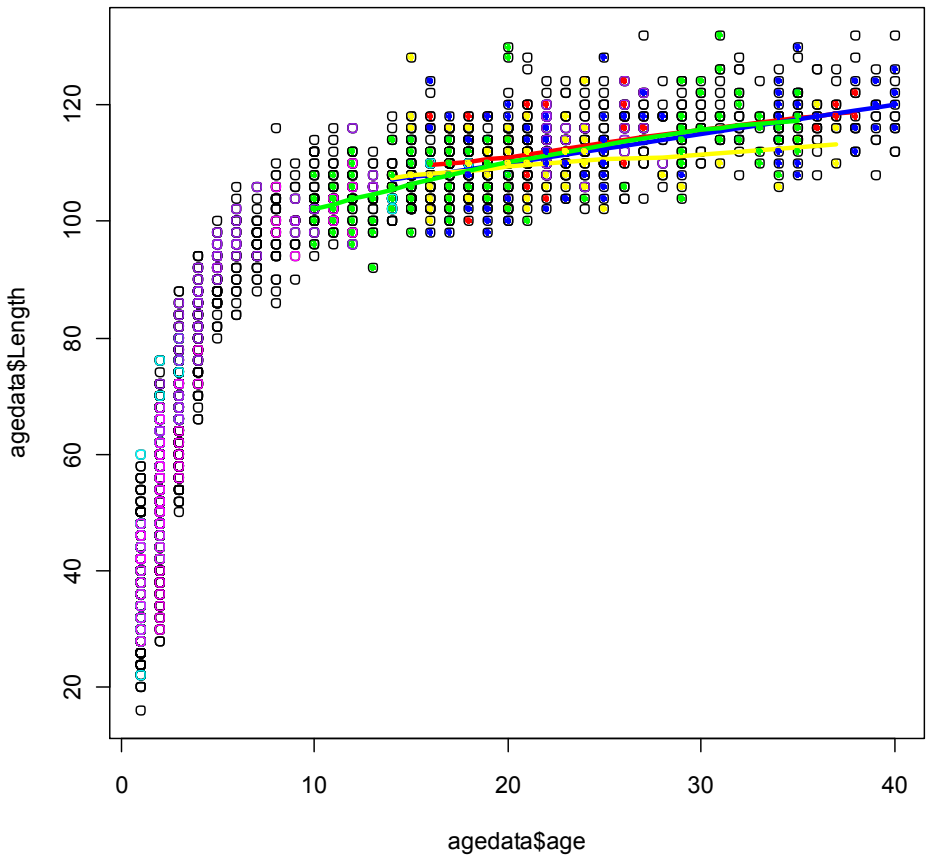

Figure 5: Aggregated age data treated unconditionally (ignoring lengths compositions). Although not statistically correct for some data sources, it shows some biased length at age information for some example cohorts. Younger cohorts are identified as circles (magenta = 2000, cyan = 1995, purple 1985). Older cohorts are marked in dots, but because of the over plotting of different cohorts, the central tendency is also shown as a spline smoother (red = 1972, blue = 1973, yellow = 1974 following the strong 1973 cohort, green = 1978 another stronger cohort).

i. Taking the length conditional information from the SS3 input file, I summed the ages across length as if they were length unconditional. My assumption is that for the lengths greater than the 27 inches that is probably appropriate (i.e. age samples are random with respect to lengths) which means roughly greater than age 5. In the data by year it is easy to track the cohort over years in the catch data until about 2000 (at age 27!) then it disappears for a few years reappearing as dominant in 2007 in the LL data and occasionally elsewhere. So the cohort is larger, but by how much?

ii. Taking the age data from fleets 1,2,3 (just the fleets) the numbers-at-age were expressed as a proportion of the total number of fish aged that year. Then, I summed the proportions of each cohort over all the years where they were > age 5. The proportion by year removes the effect of different annual sampling / catch levels, taking out ages < 6 should largely remove the age effect (catchability effect). The value of this sum is itself meaningless, but the relative scale is accurate. Plotting the proportions by cohorts in the fishery based on

the actual age information gives this plot. (It is not entirely clean, because different cohorts have different numbers of samples historically, and the +group is not dealt with neatly for speed (neither is a problem for comparing the 1973 and adjacent cohorts). But it clearly shows the 1973 cohort 300% as big as the adjacent cohorts (Figure 6) although these are generally larger than the average cohorts too. Comparison of more distant cohorts becomes increasingly inappropriate. Mostly, this is consistent with the LL data so I am not sure what information contradicts this.

iii. Interestingly, the 1953 cohort is also very strong (slightly over emphasised because it includes the +group, but it is clearly there and has far fewer years of data than the 1973 cohort. I appears first in the data at age 34, and then it is far bigger than the 1973 cohort at the equivalent age so it must have been very large or more likely have experienced much smaller Fs over it life. This is, in general, inconsistent with the impressions expressed by the assessment team that historic Fs were very high, but difficult to evaluate directly when exploitation is maximal at a very young age.

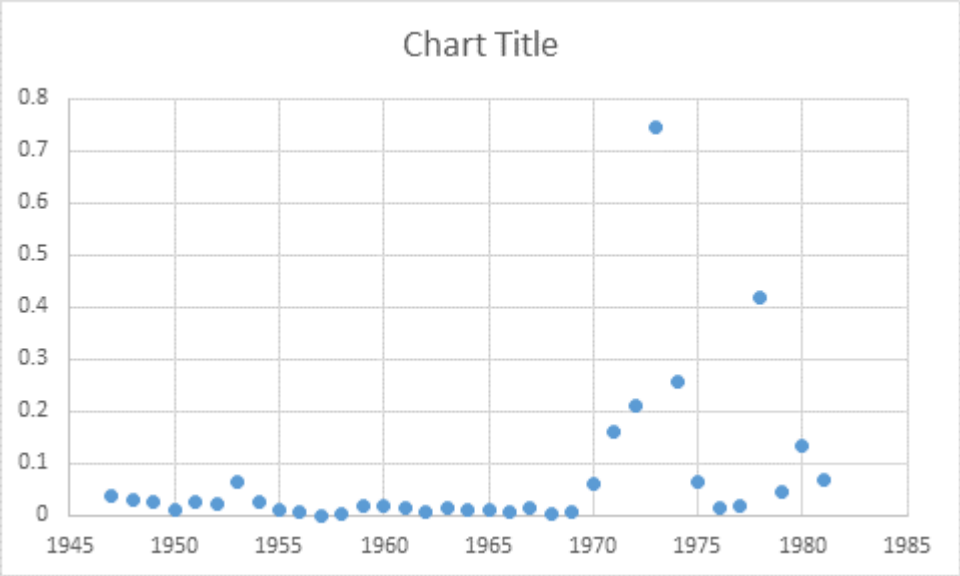

Figure 6: Relative cohort strength from fleet landings based assuming age collections are unconditional of length at length greater than 27 inches.

Developing a base model that more accurately reflects what we know about the cohort structure from various sources is an important next step in the development of functional model. I would suggest using the age information for the LL survey directly rather than length conditional. This removes the uncertainty in the growth function, and is the most appropriate way to model this data (in light of the uncertainty in growth function) since it is collected randomly, rather than length stratified. Given the consistency in cohort strength estimation across a number of abundance indices (southern stock) focusing on different ages, it is hoped that this will help the model connect F, which mainly occurs at younger ages with the resultant SSB signal which is predominantly affected by abundances at the older ages. One concern in this regard, however, is that the model may have difficulty fitting the vB

curve appropriately, because it is not clear how removal of length selectivity will affect the availability of length information to the growth model. If the age length pairs from the longline survey are unavailable to the growth function because of the move to age based selectivities, then this would be problematic, because the survey provides much of the information of length at the older ages. This should be investigated in detail after the implementation of age-based selectivities of the LL survey.

A prerequisite for age based selectivities for the LL surveys in the southern model is the removal of the age bins in the data. It is currently not clear, in terms of the implementation, if it is sufficient to remove these from this index or from all data. However, some clarification of why bins were implemented in the southern model might be helpful in understanding the implications, as it will be difficult to separate any changes in model parameterization between the un-binning and age-based selectivity effects.

R0 and tag reporting rates:

The northern model appeared to have significant problems scaling the stock size, suggesting the stock was highly depleted with unrealistically low values of stock size and high values of F (SPR averaging around 10%). SEDAR18 found similar problems with the SCA model in the absence of estimates of F from tagging. SS3, instead of using external estimates of F, is able to use tagging information directly. However, this appeared to have little effect in the original base model. Several model runs were conducted on a simplified version of the model demonstrating that the estimation of the reporting rate was very strongly correlated with the estimation of R0. Fixing either R0 or the tag reporting rate resolved the issue. Recent investigations on the reporting rate as determined by variable tag rewards suggested that the reporting rate for the recreational fleet were around 50% for the original reward value. Fixing only one of the reporting rates was sufficient to scale the model resulting in much higher R0 estimates and a much elevated SPR. Estimating reporting rates will therefore be crucial to the utility of the model in estimating stock status. Improved models are likely to remain sensitive to the assumption regarding reporting rates, but it should be possible to provide useful advice regarding the uncertainty around stock status estimates based sensitivity runs over a plausible range of reporting rates.

a. *If multiple models were considered, evaluate the analysts' explanation of any differences in results.*
b. *Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).*

As described under TOR 1 the effective sample size particularly of the commercial fleets appears to be underestimated by the decision to classify a length sample as having an effective sample size of 1 irrespective of the number individual measured in the sample. Similarly, there is some inconsistency in the estimation of variability in the indices which may have an effect on the parameter estimation.

Selectivities functions:

Selectivity has been implemented in the model based on lengths, fundamentally this is consistent with what we understand about the catch process of commercial gears. Where selectivity is based on spatially changing availability such as in many recreational, and some commercial, fisheries this becomes less appropriate. This happens when migration or ontogenetic movements are based on age due to the interaction with seasonal signals (e.g. age 1 fish move from the upper to the lower estuaries as temperature decreases in the fall). If growth is variable at age1, selectivity is separable only at age, not at length. For red drum these age-based changes in availability seem to be sufficiently small, or growth sufficiently consistent between years that generally speaking length based selectivity is appropriate. However, note the point about length being un-informative of age in the previous section.

The assessment team made use of a number of complex functions available in SS3 to model the likely selectivity patterns of fleets and indices for the base model. In addition, they attempted to account for a number of management changes that have occurred in both stocks by strategically developing selectivity blocks in line with the major management changes. The complexity of the functional parameterization along with the re-estimation of selectivities in each block meant that a large number of parameters in the model were dedicated to the estimation of selectivities. Some of the choice of functions were made appropriately given the characteristics of the gears, but in some instances the suitability of function, and where to delineate differences was based on the observed length frequency distributions which are not only a function of selectivity, but also of abundance so the latter should be avoided. The main guiding principle though is to reduce the parameter estimation to a manageable level to understand the dynamics of the stock as a whole, and then examine residual patterns to see where additional parameters may be warranted if they can be justified in what is understood about the management or gear characteristics.

In the simplified model, a single selectivity block made very little difference to the length residual pattern suggesting that the management changes had less of an impact on the behavior of the fleet than assumed. The model seems to suggest this is because the catch component is relatively small, and much of the mortality at higher ages comes from the release mortality by recreational anglers which has been unaffected by management. The changes in the mortality particularly at the younger ages (below slot limit, and changes in bag limits within the slot) occur over a small portion of the lengths, and affect only one or two ages. To save parameters, yet maintain the ability to assess the effectiveness of management changes, future models could maintain selectivity at larger sizes across all blocks and examine the effect of freeing up estimation at the shorter length. For the southern model, there were still some unintuitive patterns of selectivity suggested for some of the more complex selectivity functions. These need to be understood and corrected.

Recreational discards were modelled in the based case as separate fleets with their own complex selectivities. Fundamentally, this is inappropriate, as discarding in the recreational fleet is not separable, i.e. it is not orthogonal to length given management (bag / trip limits). Implementation of a logistic discard ogive is more appropriate for the portion of fish below the slot limit. However, because discarding increases again above the slot limit, it can currently not be implemented appropriately in SS3. Lastly, within the

slot limit the proportion of fish discarded is dependent on the abundance of those fish because of the bag limit. In years where abundance in the slot is low (or shorter trips) the proportion discarded will be low, while in years of high abundance (or longer mean trip length), the portion discarded will be high. Changes in recreational bag limits over time complicate the appropriate modeling of the discarded portion. In the short term there seem to be few options to resolve the issues without development of SS3. Consequently, it is unlikely that in the near future it will be possible to accurately demonstrate the effects of management on the exploitation of the younger individuals in detail so having fewer selectivity blocks is much less likely to have an impact and will save parameters. Some sensitivity of the effect should be conducted once a functional model has been developed to ensure that this does not negatively impact the long term management advice, though it seems the chance of this is small.

Although the review panel originally agreed that it was necessary to simplify the model to understand its behavior, and the information from different data sources, there were some suggestions that in order to fit the length data better additional model complexity should be returned to the selectivities. To me it seems that this is likely eventually necessary, but caution against doing so before the age structure is reasonably represented as it will once again become more difficult to assess model response. In my mind, the model is still acting very much as a biomass production model, without much contrast in apparent recruitment given the conflict between age and length information. Such models generally are unable to support complex selectivities (nor are they designed for that), and the parameters usually end up hiding residual variance rather than having a sound foundation in the data.

4. *Evaluate the diagnostic analyses performed, including but not limited to:*
    a. *Sensitivity analyses to determine model stability and potential consequences of major model assumptions.*

        Dithering was implemented to examine the risk of local likelihood minima. However, both models had problems with the results. The northern model largely resulted in a dichotomy of two sets of stock dynamics as if sitting on a knife edge. In contrast, the southern model was highly unstable with respect to the historic stock sizes providing a wide range in past stock dynamics while diagnostic plots produced in the SS3-package in R were provided to the review group for all model runs. Not all model diagnostics were sufficiently evaluated in the assessment report. Particularly, the residual patterns would have benefitted from further exploration when trying to improve model realism.
        Sensitivity runs / likelihood profiles were produced to explore model improvements through the identification of the specific likelihood components defining the convergence of different parameters. Attempts to improve model performance based on the profiles were initially unsuccessful, because the model complexity meant that profiles for most components were relatively flat over a wide range of settings.
        In principle, this range of model diagnostics should be sufficient to evaluate the suitability of a particular model specification as well as determining the best approaches for model development/ improvements. Following an initial

simplification of the base models allowed for progress to be made towards better assessment models using these diagnostics.

b. *Retrospective analysis*
There was insufficient time to investigate retrospective patterns in any of the experimental mode runs.

5. *Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.*

In the absence of an appropriate base model no characterization of uncertainty was or could be provided.

6. *If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.*

No minority report filed.

7. *Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods.*

No proposed estimates of stock dynamics were available due to the absence of a suitable assessment. There was insufficient time at the review to explore and evaluate alternative quantitative method to assess the dynamics. My own qualitative assessment of the state of the stock based on experience, rather than proven quantitative evaluation against management targets is that the stock is not in immediate danger despite the more pessimistic views of the proposed models (particularly the northern stock). I base this on: 1) the extended age structure with near maximum ages being seen in the longline survey, and 2) the lack of a large sudden response to what seems to be a major management intervention early in the timeseries, particularly the reduction of bag and trip limits, and the exclusion for the commercial fishery in the south. Assuming reasonable compliance, unaltered recruitment and a critically low stock, a significant effect should be noticeable. The first two assumptions are unknown, but if taken as read it suggests that F on the younger ages is not critically high.
This does not take into account the need for a more conservative approach to management given the characteristics of the stock and fisheries described under point TOR8.

8. *Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.*

No reference points were estimated during the review because of a lack of suitable assessment to develop these. SEDAR 18 used SPR methods which are appropriate given the lack of an obvious stock recruitment relationship. Given the available data, there is no indication that the stock dynamics have varied over a sufficient range in

terms of SSB or recruitment that this situation would change, so a continued use of SPR is likely irrespective of the choice of final model.

In general, it is possible to develop alternative methods suitable for advising management on red drum that are less detailed than a complete age-based model. There was insufficient time at the meeting to investigate this further in this instance. However, compared to other stocks this will be more difficult because the fishery is largely recruitment based (F is low for the majority of the population), whereas the management reference points are largely based on reproductive capacity. Trying to relate the effect of fishing on the SSB which has a long lag is always going to be associated with increased risks. The results of poor management will only become apparent in SSB way in the future, but once occurred it will take many years for the stock to recover even under appropriate management or 0 fishing. Lastly, there appears to be little or no effective management of a now substantial and often largest of mortality, recreational release mortality. Without some way of controlling the effort or reducing the release mortality rate, the options for reducing F are significantly limited. If all currently implemented legislative tools, i.e. bag limit, minimum size quotas, etc., were implemented to the extreme, F may still be above an imaginary F_{MSY} so it is not clear how F based reference points would necessarily be useful to management. Other management tools such as spatial and temporal closures may be more effective measures particularly as maximum yield appears not to be a motivator for the exploitation of red drum. In this case, the approach should be one based on constant escapement (some minimum number of animals from a cohort to make it through to the low fishing mortality part of the age structure) which would require management reference points more similar to those frequently used in salmonid management. If recruitment is as sporadic as suggested by the catch data, then protecting individual strong cohorts during their early life may be sufficient to maintain SSB long term, while maintaining fishing opportunities for most of the time as preferred for recreational fisheries.

9. *Review the research, data collection, and assessment methodology recommendations provided by the Technical Committee and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.*

In terms of model development, the review panel provided a sizable number of recommendations of things to "try" from changing selectivities, removing data or altering weighting and implementation methods. In the absence of a model to try these recommendations on, no conclusions can be reached as to whether such options would have the desired effects, so these are recommendations to try and subsequently assess their effects. Similarly, it is not possible to necessarily judge the weaknesses or the sensitivities of the assessment in terms of the data needs or where maximal gains can be made at minimal cost or effort. However, two general recommendations spring to mind for the red drum assessment.

a. Simplify the model to the absolute minimum. It must represent a realistic cohort structure, have qualitatively an appropriate F and biomass trends. It should also be consistent with the general pattern of

information that can be inferred with certainty from within individual sources (internally consistent YC estimation) or better still between sources. For this assessment, I would go so far as to say it does not even have to have realistic selectivity patterns, because much of the biomass and most of the cohorts are indistinguishable in behavior or size (>10 years) so are expected to have a more or less constant selectivity and mortality. To me, this means fixing the growth function or finding means to reduce its influence or eliminating it all together. Getting information to scale recruitment is probably very important to management, but for this initial stage doing this directly or through fixing the tagging reporting rate at some reasonable level is sufficient.

b. Such a simplified model, which will undoubtedly have inaccuracies in the magnitude of the change and the absolute value of estimates, but it should be sufficient to test the sensitivity of some of the major data assumptions (e.g. what is the effect of using recent discard data to hind cast historic discard rates), what are the likely effects of using recent tag release information to describe B2 length frequencies, are the *a priori* data choices setting the effective sample size or excluding source of data based on variability (age 3 trammel net index) still defensible.

c. Having examined the sensitivity and justified the choice of one particular method, while acknowledging its effects one can then start to introduce some more realism and complexity in the model, but one step at a time. The benefit of each alteration should not only examine the tradeoff between the necessary additional parameters (usually when the log likelihood or AIC is used as the only criterion in this process, one tends to end up with overly complex models), but also the benefit in terms of the improvement realism, reduction in model stability, retrospective bias and metrics relevant for management.

d. All the data in an assessment is important, not knowing what the assessment methodology is going to be, it is not possible to say where the emphasis on maintaining data should be. However, it is clear that the emphasis of the exploitation is on the youngest ages in the stock. Management is in part based on having good estimates of the spawning stock so any information that can improve the certainty in that aspect is especially valuable. The review panel showed that the longline surveys provide good data on the relative changes in SSB, but in addition they also demonstrate a high degree of internal cohort consistency that critically span the periods of management changes. The data series may be relatively short, but because they cover so many cohorts with a virtually identical selectivity, they provide very important stability to the assessment, particularly in the period where management changes have meant that F estimates and selectivity changes are correlated. I think it is important that this data source is maintained for future assessments, and future assessment development will progress much more rapidly because of the stability that this data offers. In the case of the southern stock it showed a high degree of consistency with other

fisheries independent indices, while in the north LL data was highly consistent with the catch data.

e. In past assessments and in the initial attempts to develop a basic SS3 model for the northern stock, it seems there is little information available on the scaling of recruitment. Part of the reason that the most recent tagging information could not be used in the current assessment is because the design of the tagging program had changed to incorporate a variable reward system in order to estimate the reporting rate within the program. This proved highly valuable in this review in helping to scale the model appropriately, and is likely to continue to do so in future assessments. I, therefore, consider this also a priority for maintaining the ability to manage red drum.

10. *Recommend timing of the next benchmark assessment and updates, if necessary, relative to the life history and current management of red drum.*

    Given that the process failed to provide assessments that can be used in management, the short term recommendation is as soon as an appropriate base model can be developed. However, even after this has been completed, I am of the opinion that regular (annual) examination of the data is necessary to evaluate if a full benchmark is necessary. As described under TOR8 the stock / fleet dynamic interactions are such that the risk to the stock are increased due to the severity of the impacts rather than the magnitude of the uncertainties. It certainly does not have to be a full assessment, or even an update at that frequency, but could be some simple indicators about recruitment.

11. *Prepare a peer review panel terms of reference and advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.*

    The report is currently at the second draft stage and on course to be delivered by the required deadline.

Appendix 1: Bibliography

Document #	Title	Authors
SEDAR44-DWReport	SEDAR 44 Atlantic Red Drum Data Workshop Report	
SEDAR44-AWReport	SEDAR 44 Atlantic Red Drum Assessment Workshop Report	
Documents Prepared for the Data Workshop		
SEDAR44-DW01	Adult Red Drum Genetic Diversity and Population Structure	Cushman, Jamison, and Darden 2014
SEDAR44-DW02	Red Drum Maturity Analysis	Arnott 2015 & South Carolina DNR
SEDAR44-DW03	Distance moved by red drum recaptured by recreational anglers	Arnott 2014
SEDAR44-DW04	Recreational Landings and Live Releases of Red drum (Sciaenops ocellatus) in the Southeast US using MRFSS-MRIP intercept data, 1981-2013.	Murphy 2014
SEDAR44-DW05	Sizes of tag recaptured red drum that were released alive by recreational anglers.	Arnott & Paramore 2015
SEDAR44-DW06	Estimating the age composition of the MRIP/MRFSS estimated landings and live-releases for red drum along the Atlantic coast, 1981-2013.	Murphy 2014
SEDAR44-DW07	Development of historical annual recreational landings of red drum from 1950 through 1980 for the Atlantic coast states from Florida through New Jersey.	Murphy 2015
SEDAR44-DW08	NC Biological Data Survey Descriptions and Background Information	Paramore 2014
SEDAR44-DW09	Fishery Independent Surveys of Sub-Adult Red Drum in South Carolina	Arnott 2014
SEDAR44-DW10	SCDNR adult red drum 1/3rd mile longline survey	Frazier and Shaw 2014
SEDAR44-DW11	Relative indices of abundance for Red drum (*Sciaenops ocellatus*) inhabiting estuarine waters along the Atlantic coast of Florida, 1997-2014.	Murphy 2014
SEDAR44-DW12	Relative indices of abundance for Red drum (*Sciaenops ocellatus*) inhabiting inland waters along the Atlantic coast based on 1991-2013 angler catch rate data.	Murphy 2014
Documents Prepared for the Review Workshop		

SEDAR44-RW01	Red Drum SEDAR 44 Stock Assessment Research Recommendations	Red Drum Technical Committee & Stock Assessment Sub-Committee
Final Assessment Reports		
SEDAR44-SAR1	Atlantic Red Drum Stock Assessment Report	To be prepared by SEDAR 44
Additional Supplementary Materials		
SEDAR44-RD01	SEDAR18-AW02: Nonparametric growth model for Atlantic red drum, and changes to natural mortality (M) estimates	Cadigan
SEDAR44-RD02	SEDAR 18 Atlantic Red Drum Review Workshop Report (excerpt from full Stock Assessment Report)	SEDAR 18 Review Panel
*The last assessment for Atlantic Red Drum was SEDAR 18. All SEDAR 18 documents (final assessment report, working papers, and reference documents) are available in a separate folder on the FTP site and on the SEDAR 18 web page (http://sedarweb.org/sedar-18). The two SEDAR 18 reference documents mentioned above were specifically suggested as supplementary materials for the SEDAR 44 Review Workshop.		

Appendix 2: SCOPE of WORK

Statement of Work

External Independent Peer Review by the Center for Independent Experts

SEDAR 44 ASMFC Red Drum Assessment Review Workshop

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from *www.ciereviews.org*.

Project Description:

SEDAR 44 will be a compilation of data, an assessment of the stock, and CIE assessment review conducted on ASMFC Red Drum. The review workshop provides an independent peer review of SEDAR stock assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and sensitivity runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. 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 should have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop Terms of Reference. 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 **Charleston, South Carolina** during **August 25-27, 2015.**

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

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: 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/

http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Pre-review Background Documents: 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 cannot 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 Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting in Charleston, South Carolina from August 25-27, 2015.
3) Conduct an independent peer review in accordance with the ToRs (Annex 2) in, Charleston, South Carolina, from August 25-27, 2015.
4) No later than **September 7, 2015** each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to *mshivlani@ntvifederal.com*, and Dr. 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 address each ToR in **Annex 2**.

Tentative Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

August 10, 2015	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
August 10, 2015	NMFS Project Contact sends the CIE Reviewers the pre-review documents
August 25-27, 2015	Each reviewer participates and conducts an independent peer review during the panel review meeting
September 7, 2015	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
September 18, 2015	CIE submits CIE independent peer review reports to the COTR
September 21, 2015	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (Allen Shimada, via Allen.shimada@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) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:

Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen.Shimada@noaa.gov Phone: 301-427-8174

William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

Manoj Shivlani, CIE Lead Coordinator
NTVI Communications
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136

Key Personnel:

NMFS Project Contact:

Julia Byrd
SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
(843) 571-4366
julia.byrd@safmc.net

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: Tentative Terms of Reference for the Peer Review

SEDAR 44 ASMFC Red Drum Assessment Review Workshop

12. Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:
    e. Presentation of data source variance (e.g., standard errors).
    f. Justification for inclusion or elimination of available data sources,
    g. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size),
    h. Calculation and/or standardization of abundance indices.
    i. Estimation of discards and size composition of discards.

13. Evaluate the definition of stock structure used in the assessment. Is the definition appropriate given the biology and management of red drum?

14. Evaluate the methods and models used to estimate population parameters (e.g., F, biomass, abundance) and biological reference points, including but not limited to:
    a. Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of red drum?
    b. If multiple models were considered, evaluate the analysts' explanation of any differences in results.
    c. Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).

15. Evaluate the diagnostic analyses performed, including but not limited to:
    a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions
    b. Retrospective analysis

16. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

17. If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.

18. Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods.

19. Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.

20. Review the research, data collection, and assessment methodology recommendations provided by the Technical Committee and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.

21. Recommend timing of the next benchmark assessment and updates, if necessary, relative to the life history and current management of red drum.

22. Prepare a peer review panel terms of reference and advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.

Annex 3: Tentative Agenda

SEDAR 44 ASMFC Atlantic Red Drum Review Workshop

Charleston, South Carolina, August 25-27, 2015

Tuesday

9:00 a.m.	**Convene**	
9:00 a.m. – 9:30 a.m	**Introductions and Opening Remarks**	
	Coordinator	
	- Agenda Review, TOR, Task Assignments	
9:30 a.m. – 12:00 a.m.	**Assessment Presentation**	**TBD**
12:00 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Continue Presentations / Panel Discussion**	**Chair**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
3:30 – 4:00	**Break**	
4:00 – 6:00	**Continue Discussion**	**Chair**

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

Wednesday

8:30 a.m. – 12:00 a.m.	**Panel Discussion**	**Chair**
	- *Continue deliberations*	
	- Review additional analyses	
12:00 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion**	**Chair**
	- Continue deliberations	
	- Review additional analyses	
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Recommendations and comments	

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

Thursday

8:30 a.m. – 10:30 a.m.	**Panel Discussion**	**Chair**
	- Final sensitivities reviewed.	
	- Projections reviewed.	**Chair**
10:30 a.m. – 11:00 a.m.	**Break**	
11:00 a.m. – 1:00 p.m.	**Panel Discussion or Work Session**	**Chair**
	- Review Reports	
1:00 p.m.	**ADJOURN**	

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

Appendix 3 List of Participants

Review Workshop Panelists

Jeff Brust	Review Panel Chair	ASMFC Appointee
Carmen Fernandez	Reviewer	CIE
Jaime Gibson	Reviewer	CIE
Sven Kupschus	Reviewer	CIE
Gavin Fay	Reviewer	ASMFC Appointee

Analytical Representatives

Jeff Kipp	Assessment Team	ASMFC
Mike Murphy	Assessment Team	FL FWCC
Steve Arnott	Assessment Team	SCDNR
Lee Paramore	Assessment Team	NCDMF

Observers

Pat Geer	South Atlantic Board Chair	ASMFC / GADNR

Council and Commission Staff

Julia Byrd	SEDAR Coordinator	SEDAR
Mike Collins	Admin.	SEDAR/SAFMC
Megan Ware	Red Drum Plan Coordinator	ASMFC
Pat Campfield	Science Program Director	ASMFC