Center for Independent Experts (CIE) Independent Peer Review of the SEDAR 44 Atlantic Red Drum

Carmen Fernández

October 2015

Executive Summary:

Atlantic red drum is assessed as two separate stocks, with boundary between North and South Carolina. The stock assessments are in the process of switching from a statistical catch-at-age model (used in the previous benchmark assessment, SEDAR 18) to an integrated framework using Stock Synthesis (SS3). The latter allows dealing with some of the main recommendations made in SEDAR 18, such as incorporating the statistical analysis of tag-recapture data within the integrated stock assessment framework and combining age and length composition information, which is useful given the sparsity of age data for red drum. The Review Panel supported the efforts to switch to SS3 and I concur with that view.

A Review Workshop attended by the Assessment Team, a Review Panel (with some members appointed by the Center for Independent Experts (CIE) and others by Atlantic States Marine Fisheries Commission (ASMFC)), and other parties took place in Charleston, South Carolina, August 25-27, 2015. Prior to the Review Workshop the reviewers were informed that the Assessment Team had not managed to achieve a run they considered appropriate for use as the basis for management advice, due to a number of technical difficulties encountered during model development, as well as inconsistencies with SEDAR 18 output and plausibility of results. The Review Panel was therefore requested to review the Assessment Team's progress to date and to provide guidance on model inputs and parameterization rather than review completed models for their utility in management. This SEDAR review was therefore conducted with that approach in mind, and the responses that could be given to the ToRs were conditioned by that situation.

Progress was made with the development of both stock assessments during the Review Workshop. For the northern assessment, the explorations undertaken during the workshop led to an improved understanding of model behaviour. From my perspective, the biggest issue that remains to be resolved is that of overall scaling of model results, which has been found to be closely linked to the assumed values of tag reporting rates. The Review Panel has recommended conducting a comprehensive review of reporting rate values, which I strongly endorse. For the south, a problem with illogical selectivity estimates was encountered during the Review Workshop and this precluded progressing at the same rate as for the northern assessment. It was not possible to understand the causes for this behaviour during the Review Workshop and this issue needs to be understood and resolved in order to continue progress with this assessment. I offer some ideas (partly arising from discussion with other Review Panel members, and partly my own) in the body of this report, when discussing ToR 3, that can hopefully help resolve this issue. Once the problem with selectivity is resolved, the sensitivity of assessment results to the values of tag reporting rates should be explored, given the impact they have been found to have in the north. Much discussion took place during the review and many ideas were offered by Review Panel members with the intention of helping to make progress with these assessments. Most of the discussions took part during the Review Workshop, together with the Assessment Team, but a few aspects were further discussed among Review Panel members after the workshop (while in the process of preparing the Review Panel's Summary Report). The main body of my independent CIE peer review report explains the essence of these discussions, the main findings and conclusions, according to my understanding and perspective. I consider the science reviewed to be the best information available, but the assessments are not yet at a stage where they can be used as the basis for the provision of management advice.

Background:

Atlantic red drum is assessed as two separate stocks: The northern region covers North Carolina and north until New Jersey, whereas the southern region covers South Carolina, Georgia and the east coast of Florida. Commercial and recreational fisheries exist in the northern region, whereas fisheries in the southern region have been exclusively recreational since the late 1980s. The number of red drum caught and released alive in recreational fisheries has increased strongly in both regions over the last 2-3 decades and is at present much higher than the number of fish caught and kept in recreational fisheries. An 8% mortality has been assumed for these live releases since the previous benchmark assessment (SEDAR 18, in 2009).

A statistical catch-at-age model was developed for the previous benchmark assessment (SEDAR 18, 2009). In addition to the standard data used in statistical catch-at-age models (fishery catches, abundance indices, and age compositions), the SEDAR 18 assessment for the northern region used estimates of F-at-age, which were derived externally from tag-recapture data and incorporated as observations in the stock assessment model; these "observations" of F-at-age were seen to have a strong impact on the assessment results, which was interpreted in SEDAR 18 as conflicting information arising from different datasets. Another limitation of the SEDAR 18 assessments was the very high proportion of the population included in the plus group, for which very little information was available due to the shortage of fishery dependent or independent data for the adult fish. Largely as a consequence of this sparsity of information on the adult fish, SEDAR 18 was unable to determine biomass status relative to reference points and only determinations based on SPR (spawning biomass per recruit) could be made; even then, the uncertainty associated with SPR estimates for the southern region (for which no F-at-age "observations" from tag-recapture data were used) was very large.

The lack of information on the adult fish is related to the fact that red drum spend most time in offshore waters after maturity (which occurs at about age 4, whereas the fish can normally live up to about age 40 or even longer in the northern region), combined with fishery regulations that preclude harvesting larger fish. Important efforts have been made since SEDAR 18 to improve the information on the adult fish population, with fishery-independent long-line surveys targeting the adult fish being conducted in North Carolina and Georgia annually since 2007, and with the continuation and improvement of the South Carolina long-line survey, which already existed at the time of SEDAR 18.

The stock assessments for the northern and southern regions are in the process of moving from the statistical catch-at-age model developed for SEDAR 18 to an integrated modelling approach implemented in Stock Synthesis (SS3). The SS3 platform allows addressing some of the main recommendations made at SEDAR 18; in particular, it allows performing the statistical analysis of tag-recapture data within the stock assessment model and fitting index and fishery length composition data directly (by giving SS3 a growth model as an input or estimating such a model within the stock assessment), as well as age composition data, where available, which is helpful due to the sparsity of age sampling. SS3 is a very flexible and widely used platform that allows developing a wide variety of models for stock assessment. At the same time, its great flexibility also means it is not particularly easy to implement and, in my experience, setting up an assessment model requires time and experience.

For the SEDAR 44 review, the Assessment Team presented the Review Panel with a detailed description of their efforts to date to transition to SS3 (also documented in the SEDAR 44 Assessment Workshop Report, SEDAR44-AWReport). The Assessment Team explained that they had not managed to arrive to an assessment configuration that they considered appropriate to use as the basis for management advice. Model development had been hindered by model instabilities, and even though many alternative settings had been explored, the vast majority of them indicated very low SPR values for the northern region and very high SPR values for the southern region, which were inconsistent with SEDAR 18 results and with the general understanding of the biology and management of the two stocks. The Assessment Team also explained that their efforts had been focused on transitioning to Stock Synthesis, as this was considered the most promising platform for the present and future assessments, and that this is where they considered the main effort should continue to be focused during the present SEDAR review. Consequently, and in agreement with ASMFC and the SEDAR coordinator, the Review Panel was requested to approach this SEDAR review with an understanding that the aim was to review the Assessment Team's work to date and to help advance their efforts as much as possible towards achieving stock assessments in Stock Synthesis that could be used as a basis for management advice. The work in the Review Workshop (August 25-27) was, therefore, focused around this aim. Whereas progress was made during the Review Workshop, it was unfortunately not possible to reach by the end of the workshop a configuration that the Assessment Team or Review Panel considered appropriate as a basis for management advice; I concur with this view.

Description of review activities and reviewer's role:

The work was organized around a Review Workshop held in Charleston, South Carolina, August 25-27, 2015. The documents indicated in the Bibliography section of this report were provided to the reviewers in advance of the workshop. A slight difficulty I found was that one of the main documents, the Assessment Workshop report, was delayed by a few days relative to the expected timing (which had been 2 weeks in advance of the start of the Review Workshop). This is not in the least a criticism of the Assessment Team, which I am in no doubt was working as hard as possible to prepare all this wealth of material, but I did feel that those few extra days would have been given me a better chance of getting more understanding of the issues and of thinking through them more carefully in advance of the meeting. From my perspective, speaking also from experience in past reviews, receiving the review documents two weeks in advance of the meeting works well, but I find that less time than that makes the task of reviewing just a bit more difficult for me. Eleven ToRs were given for the review process, which are shown in Appendix 2 to this report.

The work during the Review Workshop broadly developed as follows:

Tuesday, August 25, 2015
Most of the day was spent on presentations given by the Assessment Team and in discussion with the Review Panel of their work to date. The following topics were presented:

Management regulations; Life history; Commercial and recreational fisheries; Abundance indices; Tagging data; History of red drum stock assessments; Progress with SS3 assessments to date.

By the end of the day, several topics had been identified for exploration and the Review Panel requested that assessment configurations with simplified settings were prepared for additional exploration on the following day.

Wednesday, August 26, 2015
The results from the assessment configurations with simplified settings were reviewed. A set of exploratory runs was then identified for both assessments and the rest of the day was spent mainly conducting those runs and reviewing results later in the day. For the northern stock, which had managed to progress further, an additional set of runs was requested for review on the following day.

The Review Panel also spent part of the afternoon going through each of the ToRs and reaching a common understanding of how each of them could be addressed in view of the issues seen in the assessments and the developments during the Review Workshop. Tasks for the joint write-up of the Summary Report were also agreed.

Thursday, August 27, 2015
Research recommendations were discussed at the start of the morning. The rest of the time was spent reviewing the results of the exploratory runs conducted since the previous day and discussing aspects that still needed to be resolved or further explored after this SEDAR review. The meeting adjourned at the end of the morning.

All reviewers fully participated in all aspects of the review. In order to provide this independent peer review report, I read carefully in advance the documents provided for the review, trying to understand the main issues and difficulties, as well as considering possibilities for ways forward. During the Review Workshop, I exchanged ideas and clarified questions that I had or arose during the workshop, and contributed to the overall discussion and progress as well as I could. Finally, I reviewed some of the documents and thought through some of the main issues once again (benefiting from the insights gained during the workshop) aiming to provide a (hopefully) clear and useful report.

The Review Workshop was very well organized and ran efficiently. Travel and other practicalities had all been taken care of and everything was very easy. The assessment scientists very helpfully and patiently clarified all questions that arose during the review workshop and I felt there was a very good exchange of knowledge and ideas among all participants. From the material presented, it is clear that a lot of work had taken place before this review, including a data workshop and an assessment workshop, and additional progress was achieved during the review workshop.

Summary of findings for each ToR:

This section presents the main points that arose during the review, according to my own perspective and understanding of the issues discussed and the progress made. A few additional thoughts I had from following up after the meeting on some aspects of the work are also included. This section is organized following the ToRs.

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:

a) Presentation of data source variance (e.g., standard errors).
b) Justification for inclusion or elimination of available data sources,
c) Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size),
d) Calculation and/or standardization of abundance indices.
e) Estimation of discards and size composition of discards.

The Data Workshop report (document SEDAR44-DWReport) provided a thorough examination of the available datasets and how they had been analyzed prior to inclusion (or non-inclusion) in the stock assessments. The Data Workshop report was organized starting with life history data, followed from data from commercial fisheries, recreational fisheries, and indices of population abundance. Concerning the aspects noted in parts (a)-(e) of this ToR, I was in general satisfied with the way they had been addressed. Below I go briefly through these points:

Measures of variance (be it standard errors, CVs, sample sizes for composition data, etc.) were presented in the DW report and in many cases were taken forward into the stock assessments. One comment is that the externally estimated CVs of the abundance indices have been used by the Assessment Team directly in the stock assessments without taking into account any additional components of the variance that would be related to how well each given index represents the total fish stock (e.g. related to survey coverage...); however, I do not regard this as an essential feature that should immediately be incorporated in the stock assessments, but as something to be explored in the future. Regarding sample sizes for the multinomial likelihood of length frequency data, the Assessment Team decided to treat each sampling event (e.g. a trip) as if it were a single measurement in the inputs to the SS3 models. The Review Panel noted that this has the potential to inappropriately weight the certainty in different data sources when sampling levels differ substantially between fleets or over time. Weights could be adjusted as part of an iterative reweighting process when fitting the stock assessment model if this is found to be necessary on examination of model diagnostics.

Reasonable justification for inclusion or exclusion of the different data sets from the stock assessments was provided, either in the Data Workshop or the Assessment Workshop report, with a particularly thorough job done for the choice of survey abundance indices. Data strengths and weaknesses were also discussed and taken into account as part of the decision process for inclusion or exclusion of datasets from the assessments (e.g. quality of biological sampling coverage in different fisheries, details of the surveys' coverage).

Concerning the calculation / standardization of the abundance indices, I did not have the opportunity to get into this in full detail and it was not discussed in detail during the Review Workshop. My general comment on this is that any survey index standardization process should, of course, take care to ensure that the variables used to standardize the survey index are solely representative of catchability (i.e. not confounded with abundance). Deriving relative abundance indices from the recreational fishery CPUE

requires dealing with the non-trivial problem of finding a way of selecting all trips that could have caught red drum, even if not all of them did, in the multispecies fishery. A clustering technique was chosen by the Assessment Team to select such trips. I am not familiar with the technique, but I take it as a good sign that the issue of zero-catch trips was addressed in the derivation of the index. Once the set of potential trips was identified, the corresponding CPUE was standardized using a set of explanatory variables and my previous comment about these variables not being confounded with abundance of course applies.

Regarding the estimation of discards and size composition of discards, commercial and recreational fisheries need to be considered. The only available data on commercial discards is from an observer programme in the North Carolina estuarine gillnet fishery (this fishery accounts for most of the commercial red drum harvest) during recent years. The Assessment Team decided to extrapolate the landings/discard ratio from the years with observer coverage to provide estimates of discards in all years without coverage since 1989. Length frequency distributions observed during the years with observer coverage were also extrapolated back into the past, but making appropriate modifications when regulations changed based on an assumption that no discards would have occurred within the legal slot limit. Despite the obvious uncertainties associated with this estimation procedure, it is a pragmatic approach which I support given the absence of other data. No additional work has been done since SEDAR 18 on mortality rates for red drum discarded alive and a 5% mortality, consistent with SEDAR 18, has been assumed by the Assessment Team. Recreational fisheries also produce live releases, of which 8% were assumed to die (again, as in SEDAR 18, given the absence of further work). Live releases in recreational fisheries (the so-called B2 component of these fisheries) have increased very significantly over the last 2 decades and the number of red drum dead from the B2 component constitutes a growing part of the total fishing mortality. The numbers caught and released are estimated by the recreational survey (MRFSS/MRIP), but a major limitation is the absence of directly obtained length composition data. Although various data sources are used to provide length composition samples, the major one is the lengths at the time of recapture of red drum that had been previously tagged in North Carolina and South Carolina by either volunteer angler programmes or fishery-independent surveys. The lengths of the fish at recapture have been assumed by the Assessment Team to be representative of the overall B2 component of the recreational catches. The Review Panel raised concerns that this tagging-recapture procedure will likely lead to biased length composition data relative to the lengths of the overall B2 catches, because it is unlikely that the population of tagged fish (the recaptures from which are used to provide length-composition data) will have the same length frequency distribution as the entire red drum population; for example, often only fish above a certain length are tagged, leading to an under-representation of small fish in the tagged population and in the recaptures. The Assessment Team agreed with this concern, which they had also identified; a recommendation from the Technical Committee, clearly supported by the Review Panel, relates to finding ways to improve the length composition data of the B2 component of the recreational fishery (see ToR 9 later in this report).

The Data Workshop report explains how length frequency distributions were derived for the recreational fisheries, based on stratum-specific (state/year/wave/fishing mode) data if a sufficiently large sample size was available (at least 20 sampling events for Type A catches or 5 for Type B2) and pooling data incrementally across strata when the sample size was below the set threshold. By looking at the length frequencies of the B2 components of the assessment data, I get the impression that, within each region (north or south), this procedure has resulted in strong commonalities between all the data that have a

strong B2 component (i.e. recreational B2 removals and recreational CPUE abundance index for the north, and FL B2, GA-SC B2 and recreational CPUE abundance index for the south), which could potentially lead to excessive weighting of this information in the model results. If the resulting length frequency distributions are very similar, then not entering all of them as data into the stock assessment, and instead assuming that selectivity of some of the fleets mirrors that of some other fleet could produce more appropriate weights in the stock assessment; however, if the length frequency distributions are sufficiently dissimilar, then I think it is best keeping all of them as different data streams in the assessment, but considering a possible reduction of their weights.

The Assessment Team estimated natural mortality at age using the Lorenzen method (scaled in agreement with Hoenig's constant M) externally using a non-parametric growth curve, due to the poor fit of the Von Bertalanffy growth model to red drum data. On the other hand, a Von Bertalanffy growth curve, with 2 values of K, is used in the SS3 red drum assessments, so this could create some inconsistency between M and growth in the stock assessment. A catch-curve analysis conducted on the basis of the South Carolina long-line survey during the Review Workshop produced a similar estimate of M to the one used in the southern assessment, suggesting consistency between this estimate of M and the age structure seen in the survey. Given the long-lived nature of the species and the fact that it is fished mostly only at young ages, assumptions about M could potentially have significant impact on some of the assessment results. I suggest sensitivity runs be conducted with alternative values of M (e.g. rescaling the M-at-age vector to higher or lower average values); I am aware the Assessment Team already conducted this type of sensitivity runs in their earlier work, but I suggest this be repeated as the new stock assessments develop. Specifying M-at-age as Lorenzen-based internally in the stock assessment is also possible in SS3 and a sensitivity run exploring the results from such an approach could provide additional insights.

As a final general comment, it would be good if in future occasions some additional emphasis could be placed on highlighting at the Data Workshop stage the main signals that can be seen in the data and which are relevant for the stock assessment, particularly cohort strength, and on examining the coherence of those signals across different datasets. An exercise along those lines was conducted during the Review Workshop, which complemented the earlier work by the Assessment Team; this exercise was found to be very useful for understanding the information content of the long-line surveys in terms of cohort strength, and how this related to other surveys for younger ages. In integrated models, such as SS3, where data are typically entered in the assessment in more raw form instead of being highly processed and aggregated, one often ends up with a large collection of somewhat patchy and disparate datasets, and it can be difficult to get a feel for what kind of signal the different data provide and how they relate to each other. However, achieving such an understanding can be very helpful for guiding modelling choices and for interpreting model results. There was consensus among all members of the Review Panel on this point.

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

The Data Workshop report provides information on stock structure, which was also presented by the Assessment Team during the Review Workshop. Atlantic red drum has been considered as two stocks for assessment purposes since the mid-1990s, a northern one covering from North Carolina northwards approximately until New Jersey, and a southern one covering South Carolina, Georgia, and the east coast

of Florida. The division is based on different life history characteristics and limited movement found from tagging studies. Recent genetics work by South Carolina DNR has detected significant differentiation between North Carolina and more southern areas at spawning time, but not during the non-spawning season, suggesting some mixing of adults occurs during the non-spawning season; however, because of fishery regulations, no significant mixed-stock fishery removals are expected. The Review Panel supported the Assessment's Team decision to continue treating Atlantic red drum as two stocks for assessment purposes, and I agree with that.

TOR 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:

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?

The model platform chosen by the Assessment Team is Stock Synthesis (SS3), and the Assessment Team provided an appropriate justification for this choice. This is a flexible platform that allows developing a wide range of stock assessment models. It allows dealing with some of the main recommendations made in SEDAR 18. In particular, the statistical analysis of tag-recapture data can be conducted in SS3 as part of the stock assessment itself (rather than externally), which allows a more appropriate integration of the tag-recapture data with the other data used in the stock assessment, leading (in principle) to more realistic weighting and propagation of uncertainty into stock assessment outputs. SS3 also allows combining length and age composition information by means of a growth model, which seems useful given the sparsity of age data for red drum; this avoids having to convert all observed length frequency distributions to age frequencies (as would happen in a standard statistical catch-at-age model, such as the one used in SEDAR 18), which often requires pooling age data across different strata, potentially leading to biases and to excessive weighting of some datasets in the model. The SS3 modelling approach is within the class of "integrated models", whose philosophy is to include the data in a form much closer to what was actually observed rather than in a highly processed format. If one sets up a realistic population dynamics model and is able to assign realistic weights to the different input datasets, then such a modelling framework should be able to provide more realistic stock assessment outputs and a better characterization of the uncertainty associated with those outputs. All this was explained by the Assessment Team and I support their choice of modelling platform and their way of structuring the model overall.

Another advantage of SS3 is that once a basic, but realistic, model configuration is obtained, the flexibility of the platform allows for future extensions that can make it even more realistic. For example, improvements can be made to the selectivity forms and how these change over time, aspects of the tag-recapture modelling, inclusion of other data types that SS3 can assimilate, or moving to a seasonal time step (e.g. quarterly time steps would allow more realistically handling spawning time, which would be at the start of quarter 4, the dates of tagging and recapture events, given that most tagging seems to take place in the autumn, and the estimation of growth, which is fast for the young fish and might be better inferred using time steps shorter than a year). Several of these extensions are, however, what I would call model refinements, which I think would be more appropriate for a future time than for now.

b) If multiple models were considered, evaluate the analysts' explanation of any differences in results.

Different model configurations in SS3 had been explored by the Assessment Team before the Review Workshop (see SEDAR44-AWReport) and more explorations took place during the Review Workshop. The results from the different run configurations conducted at the Review Workshop were examined and compared to each other in order to get a better understanding of the models' behaviour and the impact of assumptions on assessment results.

A continuity run based on the statistical catch-at-age model for SEDAR 18 (although with some changes to the modelling framework, so not strictly a continuity run) was also presented at the start of the Review Workshop. The Assessment Team presented this run as a means to facilitate comparisons between the previous model and the new SS3 developments, rather than as a potential alternative for a base run, given the limitations found in the SEDAR 18 review and the fact that SS3 is a more powerful tool to address those limitations. I agree with that conclusion.

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).

The Assessment Team had explored a wide set of alternative SS3 configurations in advance of the Review Workshop and presented one of those configurations in detail at the start of the workshop; this was not intended as a base run that could be considered for providing management advice, but rather as their best example of their progress thus far. Key elements of the configuration presented were:

- Stock assessment period extended back to 1950 (the SEDAR 18 assessment started in 1989, which can be considered the start of the "data-rich" period); it should be noted that the only data available before 1981 is catches, which for the recreational fisheries were reconstructed based on recreational fishing effort estimates and an assumed constant CPUE through time.
- Catches not fitted perfectly; instead assumed to be log-Normal with CVs given as inputs to SS3.
- Annual time step in the population dynamics model.
- Growth estimated within SS3 using a Von Bertalanffy model with a change in growth rate at a pre-specified age (i.e. a value of K is considered up until a pre-specified age, after which the fish continue growing with a different value of K).
- M assumed to be age-dependent but constant over time, following a Lorenzen specification based on a non-parametric growth model fitted externally in SEDAR 18; these M-at-age estimates were treated as fixed inputs in SS3. I have commented on the choice of M under ToR 1.
- Weight-at-length based on allometric relationship with parameters estimated externally and treated as fixed inputs in SS3.
- Maturity treated as age-specific and constant over time; estimated externally (based on an analysis specifically conducted for SEDAR 44) and treated as fixed inputs in SS3.
- No stock-recruitment relationship, which the Assessment Team implemented in SS3 as annual deviations around a Beverton-Holt curve with steepness fixed at 0.99. Recruitment deviations were allowed before the first assessment year (for the southern stock; they did not seem necessary for the

northern stock when the start year of the assessment was 1950) in order to allow a non-equilibrium age distribution for the population in the first assessment year.

- Fishing fleets coherent with the choices made in SEDAR 18: 4 fishing "fleets" in the northern region (Commercial gillnet and beach seine GNBS; Commercial Other; Recreational AB1; Recreational B2); 6 fishing "fleets" in the southern region (FL commercial; FL AB1; FL B2; GA AB1; SC AB1; GA-SC B2), where FL, GA and SC denote Florida, Georgia and South Carolina, respectively. In line with SEDAR 18, 8% of the fish released alive by recreational fisheries were assumed to die and, therefore, the removals of the B2 "fleets" were taken to be 8% of the fish released alive by the corresponding recreational fisheries.
- Selectivities assumed to be fleet-specific and length-based for all fishing fleets. The SS3 double-normal selectivity form was selected for all fishing fleets; this selectivity function is defined by 6 parameters, all of which were estimated except for the parameter defining the selectivity at the first length bin (which was fixed so that selectivity at that length equals 0). Time blocks of constant selectivity were assumed based on the regulations expected to affect selectivity of each of the fleets, with all selectivity parameters considered to be block-specific. Four blocks were chosen for all fleets in the north (same blocks for all fleets). In the south, no blocks were chosen for Florida commercial (the fishery ended in 1988), whereas two blocks were chosen for Florida recreational fleets, four blocks for Georgia AB1 and GA-SC B2, and five blocks for South Carolina AB1.
- Five and eleven abundance indices were used for the north and south, respectively. For each region, all abundance indices are fishery-independent except for one which is based on CPUE from the recreational fisheries. Many of the indices are considered to represent ages 0, 1 or 2, and an age-based selectivity concentrated on the age they are considered to represent was selected for those indices. Length-based selectivities were selected for the indices that are considered to represent a conglomerate of ages: double-normal forms were chosen for the recreational CPUE indices, and logistic forms for the long-line indices that represent the adult fish in the population. Abundance indices rarely start before the mid-1990s and some are much more recent (e.g. North Carolina and Georgia long-line indices start in 2007). Log-Normal distributions were assumed for all abundance indices, with CVs estimated externally and treated as fixed inputs in SS3. I have commented on the choice of CV under ToR 1.
- Length composition data were available for the fishing fleets since about the mid-1980s and for the abundance indices generally for the entire duration of the indices, whereas age composition data were sparser and entirely lacking for some of the fleets and abundance indices. The Assessment Team explained that age data arose from random sampling in some cases and length-stratified sampling in other cases; all age composition data were input in SS3 as conditional on length. The length and age composition data give rise to multinomial likelihoods in SS3; the number of samples was taken to be the number of sampling events (e.g. trips for commercial fisheries) for the length composition data and (I think) the actual number of fish sampled for the age conditional on length composition data. I have commented on the choice of sample size for the length frequency data under ToR 1.
- Tagging data available for both the north and the south were incorporated in SS3. SS3 requires that each tag release group be associated with an age and assumes that all tag releases occur at the start of the time step (the start of the year in the red drum models) whereas peaks in tagging activity occur in the autumn. To deal with this, the Assessment Team gave 0.75 as a latency period in the SS3 input file and a problem with that approach has been detected after the Review Workshop; I will comment on it below. Most parameters associated with the tagging sub-model (initial tag loss, chronic tag loss,

tag reporting decay rate, overdispersion of the negative binomial distribution for total recaptures) were treated as known and fixed; only the tag reporting rates were estimated for some fleets while they were treated as fixed inputs (i.e. as known) for other fleets.

The above settings were discussed in much detail between the Assessment Team and reviewers during the presentation at the start of the Review Workshop. In general, the reviewers considered the Assessment Team had done a good job (I concur with that), but also raised a number of issues as well as suggested many ideas in an effort to try and help progress these assessments towards the goal of achieving a configuration that can be considered appropriate to provide management advice. Many of the issues raised were similar for both assessments and I note the main ones below.

Discussion of results presented by the Assessment Team for the north:

Although the fits to commercial catch were good, this was not the case for recreational catches (model overestimated AB1 and underestimated B2 from about 1990 to 2005). Whereas the continuity run (which, as already explained, included some modifications relative to the SEDAR 18 settings) was seen to produce very similar SPR values to those estimated in SEDAR 18, the SS3 run produced much lower SPR values. The Assessment Team explained that all SS3 configurations that had been tried for the northern assessment, except for one where the catch data were forced to be fit perfectly (using the F hybrid SS3 method), produced similarly low SPR values (averaging less than 20% since 1990, which is not in agreement with the scientists' expectations according to the biology and management in place for this stock). The Review Panel's and Assessment Team's interpretation was that this is a scaling issue: the SS3 run seemed to be estimating overall lower average recruitment and abundance than the continuity run and, in parallel with this, higher Fs and lower SPRs. Understanding the causes of this effect, relative to the results of the continuity run and SEDAR 18, was considered a priority by the Review Panel.

The fits to length composition data seemed reasonable in broad terms. It was noted that a unimodal selectivity-at-length function, such as the double normal, is likely not the most appropriate form for the B2 component of the recreational catches, as one might expect increased selectivity just outside the edges of the size-box in which retention has been allowed in the last 2 decades (i.e. 18-27 inches or, approximately, 46-69 cm). This could be expected to result in local modes in the B2 selectivity-at-length somewhere around 46 and 69 cm and investigating more flexible selectivity forms seems appropriate (for both north and south assessments); this could be done via e.g. the SS3 non-parametric selectivity pattern, using a sufficient number of waypoints (perhaps 6 or 8) and judiciously choosing the locations of the end waypoints. Another issue for the north is that the B2 length composition was seen often displaying positive residuals around 120-125 cm, which must be related to the fact that these very large fish that seem to appear in the B2 catches are almost completely absent from the North Carolina long-line survey, for which asymptotic selectivity-at-length is assumed. The Assessment Team considered that the lengths of the long-line survey are correct, whereas the lengths of the largest fish are probably overestimated in the B2 catches (i.e. the fish reported as 120-125 cm are probably a bit smaller in reality, in line with the largest lengths observed in the long-line survey; see also my discussion in ToR 1 on the likely over-representation of large fish in the B2 length frequency data). Although this issue was not investigated during the Review Workshop, I think it should be examined by the Assessment Team, as the discrepancy between the B2 length frequency distributions and those of the long-line survey means that they may be

giving conflicting signals about population structure, with unknown effect on model results. Two possible ways of exploring this are: (1) manually changing the B2 length data for the large sizes so as to make those lengths smaller and more consistent with the long-line survey observed lengths (just as a sensitivity test, to help understanding potential impacts on model results), or (2) considering a more flexible shape for the B2 selectivity, permitting it to have a local mode around those large lengths. These explorations are relevant for both B2 removals and recreational CPUE index, as the latter also has a strong B2 component.

Modelling the AB1 and B2 catches as a single fleet with a discard pattern would be likely better than the current approach of treating them as two separate fleets. However, the pattern of discarding in the recreational fishery, with retention prohibited outside the 46-69 cm box, can currently not be implemented in SS3, because only logistic discarding patterns are currently allowed in SS3. In view of this limitation, I support the approach taken by the Assessment Team of splitting the AB1 (harvest) and B2 (live releases) components of the recreational fishery into two different fleets in the SS3 implementation.

There was substantial discussion around the tagging data. For the north, the reporting rate of the B2 tag recaptures had been assumed to be 100% but, upon discussion with the Assessment Team, it was concluded that the reporting rate was unknown so that assuming it was 100% was not appropriate. As there was originally some uncertainty among Review Panel members about how the B2 tag recaptures should best be handled in SS3 (later clarified and agreed that the approach followed by the Assessment Team of assuming 8% of the released recaptures as the actual reported recaptures was appropriate), the tag recaptures corresponding to B2 were left out of the analysis during the Review Workshop explorations and a zero tag reporting rate was assumed for the B2 "fleet". The results of the Review Workshop explorations (described below), indicated that the fleets' tag reporting rates were key parameters directly associated with the scaling of the assessment.

Another issue, common to north and south, discussed was whether the assessment should best start in 1950 or at some later point in time, closer to the start of the "data-rich" period. The year 1950 had been selected by the Assessment Team as the start of the assessment period to provide some contrast in the removals and to try and inform the model about stock productivity, particularly since both north and south stocks are believed to have been relatively depleted when the MRFSS recreational removal estimates start in 1981. However, the only data available for 1950-1980 are catches (without any associated length or age compositions or abundance indices) and the recreational catches for those years had to be estimated based on effort estimates and an assumed constant CPUE, so that there is also considerable uncertainty around the catch estimates. Given this, I find it difficult to see the gain of extending the assessment period back to 1950, and I think it is likely more appropriate to start the assessment towards the beginning of the "data-rich" period, i.e. at some point during the 1980s. That said, an examination of the age composition data of the long-line surveys conducted during the Review Workshop revealed that even now they were tracking some strong year classes from the 1970s, so there could be some basis for extending the assessments back in time as had been the original idea of the Assessment Team. In any case, the Review Panel agreed that model explorations conducted during the Review Workshop would use the "data-rich" period only.

For both the north and south, some concern was raised that using the recreational CPUE, together with length composition data, as an abundance index might be redundant with the AB1 and B2 components of the catch, and might give excessive weight to the datasets associated with recreational catches in the model fit. There was no complete agreement originally among Review Panel members on this point, and it was therefore decided to conduct sensitivity runs excluding this abundance index to gain understanding of any potential effects. Both the length compositions and their residual patterns are very similar for the recreational CPUE and the B2 component of the removals. In line with a comment I have made under ToR 1, I think one could consider excluding the length composition data of the recreational CPUE from the analysis and modelling its selectivity by mirroring the selectivity of the B2 fishing "fleet", but since the recreational CPUE also includes the AB1 component of the catch this does not seem an ideal solution (although it could be worth exploring this as a sensitivity run).

Although most of the discussion was focused on the results being shown for the north, it was agreed that many of the above issues, or similar ones, also applied to the south.

The Review Panel considered that the best way forward for explorations during the Review Workshop was to simplify the models, in order to facilitate model exploration and understanding. Therefore, runs with the following specifications were requested for both north and south regions:

Assessment to start in 1989; No time varying selectivity; Fix B2 fleet tag reporting rate near zero and delete observed recapture data for that fleet; Hybrid method for F (i.e. all fleets' catches to be fit exactly); Keep the stock-recruitment part simple and estimate recruitment deviations before the start of the assessment period (to allow for a realistic population age structure in the initial stock assessment year).

Developments during the Review Workshop for the north:

A run with the requested specifications was conducted and results discussed in detail. Much of the discussion focused on residuals and on whether the model was able to reflect main signals seen in the data. Several things were noticed: The strong 1973 year class, which could be clearly detected in the age structure of the North Carolina long-line survey in recent years, was being spread by the model among several adjacent years (as could be seen from the estimated recruitment deviations and population abundance at age over time), which could be a consequence of the uncertainty around the estimated growth model or, possibly, to conflict with other datasets, such as the B2 length composition data; the estimated selectivities for the recreational CPUE and B2 removals were very similar, and so were the residuals of their length compositions; the input sample sizes for the length composition data were too high according to the effective sample sizes calculated as model outputs by SS3; estimated tag reporting rates were not intuitive, in particular, the fact that reporting in recreational fisheries was estimated to be lower than in commercial fisheries; the model continued to estimate very low SPR values.

Four additional runs were undertaken in order to explore some of the main issues noticed. They included the following modifications with respect to the first run conducted during the Review Workshop:

- A run with non-parametric selectivity form for the recreational B2 removals (aiming to achieve a more realistic selectivity shape),
- A run changing the age at which K changes to age 9 (aiming to improve growth estimation),

- A run decreasing all length composition data sample sizes by a factor of 10 (in line with model diagnostics on effective sample sizes and aiming to allow other signals from the data, in particular the age data, to increase their weight relative to that of the length data), and
- A run fixing the tag reporting rates to values that might be considered reasonable (as a first guess they were fixed at 30% for the 2 fishing fleets and at 50% for the recreational AB1 harvest).

Convergence of some of the runs was poor and the first run did not converge at all. In general terms, the first 3 runs produced results that were not majorly different from the first run conducted during the Review Workshop, although I note that the run that decreased the sample size of length composition data by a factor of 10 altered the recreational CPUE estimated selectivity curve. The run with the change in K at age 9 led to some improvement in the overall model fit. All 4 runs continued to estimate the strength of the 1973 year class, relative to the surrounding year classes similarly to the first run conducted during the Review Workshop. However, the fourth run produced a significant change in overall scaling of assessment results: average recruitment and SSB were strongly scaled upwards and F downwards; in line with this, SPR estimates were scaled upwards. This suggested that the tag reporting rates are key parameters in the scaling of this stock assessment. Therefore, an attempt was made to try and gain additional understanding of this feature in the next set of explorations.

Four additional runs were conducted, where each included the following modifications with respect to the first run conducted during the Review Workshop:

- A run fixing R0 at the estimated value in the run that fixed the tag reporting rates; all tag reporting rates were estimated in this run.
- A run fixing the tag reporting rate of the recreational harvest fleet (at 50%) and estimating the reporting rates of the commercial fleets.
- A run without tagging data.
- A run strongly downweighting the recreational CPUE index and its length composition (to address the concern raised earlier about the use of this index in the stock assessment).
-

All runs continued to estimate the strength of the 1973 year class, relative to the surrounding year classes similarly to all other runs conducted during the Review Workshop. Excluding the recreational CPUE index (fourth run) did not have much effect compared to the run with the same settings but including the index, but it did alter the selectivity estimated for the large lengths of the B2 removals "fleet". A major difference found between the results of the four runs is that the first two produced a similar scaling to the run that had fixed all tag reporting rates, whereas the last two runs produced a similar scaling to the first run conducted during the Review Workshop, in which all tag reporting rates were estimated. This lends additional support to the understanding that the tag reporting rates play a key role in scaling this stock assessment. In view of these results, it would seem very important to get a reliable external estimate of the tag reporting rate for at least one of the fleets, which would then be plugged in the stock assessment (and would largely drive the overall scale of the assessment results); the results also suggest that as long as this externally derived estimate is available for one of the fleets, the tag reporting rates for the other fleets will scale themselves within the stock assessment. There was no time to undertake any additional explorations on the assessment for the north during the Review Workshop.

Developments during the Review Workshop for the south:

A run with the following specifications was conducted:

- Assessment starts in 1989;
- No time varying selectivity;
- Estimates tag reporting rates for all fishing fleets (some reporting rates had been previously fixed at values for which there was no real basis, so in this first run all were estimated); Hybrid method for F (i.e. all fleets' catches to be fit exactly);
- No estimation of recruitment deviations before the start of the assessment period (note: this simplification was most likely not warranted, given that recruitment deviations before the start of the assessment period are instrumental in achieving a realistic population age structure in the initial stock assessment year; all runs conducted in the Review Workshop after this first one incorporated recruitment deviations before the start of the assessment period).

A main problem observed in the results of this first workshop run for the south was that some selectivity shapes were not logical, indicating very high selectivity at very small fish lengths; they resulted in very obvious misfits to the length composition data. This was particularly problematic for the South Carolina AB1 component, although a misfit also occurred for the GA-SC B2 component at small lengths. For the length composition data, the SS3 output corresponded to higher effective sample sizes than those given as inputs for many of the fishing fleets and abundance indices. The age conditional on length composition data were fitted relatively poorly at the larger lengths. For the southern assessment, age bins grouping 5 ages had been used for ages ≥ 10 in the data, because of the sparsity of age data. However, as the age-length data are a major source of information about growth, which is a key component of these stock assessments, the Review Panel requested that the age data not be grouped into bins in subsequent model explorations. As was the case in the northern assessment, the observed length frequencies of the B2 components and recreational CPUE abundance index seemed very similar, and so seemed also their residual patterns.

Additional runs were requested removing the 5-age bin grouping and with recruitment deviations starting before the first assessment year. It was found that unbinning the age data led to very high estimates of the CV of the estimated growth function for the old fish and a run was also conducted fixing the value of this parameter instead of estimating it. In an effort to try and resolve the selectivity problem for the SC AB1 component, another run mirrored the selectivity of SC AB1 to that of the FL AB1 component. Finally, another run was conducted with two selectivity blocks for the B2 components, with a changepoint in 1992. The Review Panel had also requested a run excluding the recreational CPUE index, but there was no time to conduct this during the Review Workshop.

The illogical selectivity pattern (with very high selectivity at small lengths) in the SC AB1 component was present in all the runs conducted during the Review Workshop; the feature also appeared in other fleets depending on run configuration. It was unfortunately not possible to understand during the Review Workshop the causes of this behaviour, which had only occurred to a much lesser extent in the run presented by the Assessment Team at the start of the Review Workshop. This needs to be understood and resolved in order to be able to progress towards a stock assessment that can serve as the basis for management advice. I offer some suggestions here with the aim of trying to help resolve this issue. The

first obvious one is to double-check (once again) the input files for possible errors; the runs conducted during the Review Workshop used one fleet less than the earlier runs (because of the shortened time period for assessment) so non-trivial changes were required in the input files. The run for the south presented by the Assessment Team at the start of the Review Workshop had made extensive use of symmetric beta priors in order to get around the problem of illogical selectivity estimates, but (on checking an SS3 control file provided by the Assessment Team for the first run conducted during the Review Workshop) priors do not appear to have been used during the runs conducted in the Review Workshop; it may be that this feature is needed to get sensible selectivity estimates, just as the Assessment Team had noticed in their earlier work. Another possibility is that the lower bound of the first selectivity parameter (the parameter that determines the lower end of the double-normal selectivity peak) needs to be increased to more realistic values for the different fleets (e.g. the current value in the control file appears to be 9 cm for all fleets, which seems far too low for most fleets and, in particular, for the harvest fleets). Additionally, the runs conducted before the workshop seemed to have fixed the selectivity to be 0 at the lowest length bin, but this parameter appears to have instead been estimated during the Review Workshop runs; I would recommend using the previous setting. A final possibility that I recommend be explored is to increase the weight of the length composition data (i.e. the sample size), in order to direct the model towards more closely fitting those data (one would then, of course, need to check for other aspects of the model fit that might deteriorate with the increased weight on the length composition data). I note that on checking SS3 input files available on the SEDAR 44 FTP site after the workshop, I was left with the impression that the runs conducted during the Review Workshop used a lower input sample size for the length composition data than had been the case in the runs conducted before the workshop, and I wonder whether this was part of the explanation for the illogical selectivity estimates found during the workshop for the SC AB1 component (assuming I looked in the correct files). Hopefully, some of the ideas listed here can help resolve the problem encountered with the selectivity estimates.

Once the selectivity problem is resolved, I think explorations in relation to the values of the tag reporting rates, as were done for the north model, should also be undertaken for the south.

Issue with tagging sub-model noted after the Review Workshop:

SS3 assumes that all tag releases occur at the start of the time step, which is the start of the year in the red drum models, whereas peaks in tagging activity occur in the autumn. To deal with this, the Assessment Team introduced 0.75 as a latency period in the SS3 input file. Unfortunately, it was realised after the Review Workshop that SS3 only allows integer values for the latency period, and that it had internally reset the value to 0. This means that SS3 will interpret recaptures for the first year as corresponding to the entire year and not only to the last quarter of the year. The implications of this are not entirely clear to me, but my first interpretation would be that this could lead to some underestimation of F for the first year (because fewer recaptures are observed than expected by the model), although there are tag releases in several consecutive years and it is not easy for me to have a clear understanding of how various model parameters will interplay. In any case, I feel this finding gives rise to some concern, given how sensitive the scaling of the assessment for the north has been found to be to tag reporting rates. I think the best solution in the longer run could be to move to a model with quarterly time step, but I also feel this would probably require too much additional development at this stage, so I see it more as something for the

future than for now. For now, my suggestion would be that some sensitivity analyses be conducted to investigate the impact of this assumption on the results: since SS3 seems to have reset the latency period to 0, I suggest performing a sensitivity run setting the latency period to 1 (also suggested by the Review Panel). Sensitivity runs based on the alternatives the Assessment Team had initially considered to deal with this problem (section 2.1.4 of the Assessment Workshop report, SEDAR44-AW) could also be considered.

ToR 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

b) Retrospective analysis

The diagnostic methods employed by the Assessment Team (residual plots of all input data, automatically produced when running the r4ss R package, minimum likelihood values, likelihood profiles displaying the contribution of each dataset to the overall likelihood) are the standard tools and were used appropriately by the Assessment Team. The echoinput.sso file automatically produced in SS3 runs is a copy of the input files as have been interpreted by SS3, and is therefore a very useful tool to check that SS3 has interpreted all inputs as intended by the analyst and to debug the input files when this has not been the case (examination of this file uncovered the problem with the latency period value for tagging data discussed above). A considerable amount of time in the Review Workshop was spent conducting sensitivity runs in order to gain understanding of the implications of different model assumptions. This is a key step in the model development process and the Review Panel emphasized that this should be done with care, making only one modification each time so that the implications of each change can be more easily isolated and understood. Retrospective analysis should be conducted, particularly once configurations that are close to what may be considered a final model are achieved. In addition to conducting retrospective analysis on the usual quantities (recruitment, SSB, F or SPR), examining the retrospective behaviour of estimates of other key parameters can sometimes provide useful insights on model behaviour.

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

The method used so far by the Assessment Team to characterize parameter uncertainty given a specific model choice is the asymptotic standard errors, automatically produced by SS3, which I find appropriate. Markov Chain Monte Carlo (MCMC) methods could be considered at a later stage.

In addition to this, sensitivity analyses are useful to examine the impact of different structural assumptions, and I have covered this under ToR 4.

ToR 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 was submitted.

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

It was not possible to address this ToR during this SEDAR review. As explained already, the Assessment Team had not managed to achieve an assessment configuration that they considered appropriate as a basis for management advice in advance of the Review Workshop, and the Review Panel was requested to review their progress to date and to focus the workshop on progressing towards the aim of achieving final model configurations. The work during the Review Workshop was very intense and progress was made, but final model configurations were not achieved by the end meeting. As the efforts were focused on progressing with the SS3 work, there was no opportunity to consider alternative estimation methods during the Review Workshop. The Review Panel does support that efforts to switch to the SS3 modelling framework be continued, as SS3 can address many of the concerns raised at the SEDAR 18 review and should, therefore, be capable of providing more relevant management advice than the previous modelling framework; I agree with that.

I hesitate to provide possible ideas for alternative estimation methods, not having had sufficient time to think about it (and, as I said above, I agree that the SS3 work should be continued). But following from some discussion had with the Review Panel after the workshop, I would say that any population monitoring should consider both the older and the younger individuals in the population. As red drum is long-lived but most fishing occurs at young ages (mainly ages 0-2), it means that overfishing could go undetected for many years if only adult biomass was monitored. Therefore, in addition to adult biomass, it is important to monitor abundance-at-age for each of the ages 0-2, as well as to monitor fishing pressure (including mortality of B2 releases).

ToR 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.

Stock status determination could not be achieved during this SEDAR review, for the reasons explained above, and the Review Workshop did not get to the stage of considering reference points. Given this, the Review Panel did not discuss reference points in depth, although some thoughts were exchanged between Panel members while preparing the Panel's Summary Report. Since it is unlikely that there is sufficient contrast in SSB and recruitment to be able to estimate a stock-recruitment relationship, a reference point based on SPR, as has been used for these stocks in the past, still seems appropriate. An SPR target of 0.4 seems to me not out of line with choices made in other regions for sustainable exploitation levels (or F_{MSY} proxies), although this should of course depend on the perceived productivity of the stock. My suggestion would be that simulation-testing approaches be considered (in the future) to ascertain the potential long-term impacts of exploitation at alternative values of F on population dynamics. I note that, for the north assessment, the overall estimated scale (which directly impacts on the overall estimated SPR level) was found to be very sensitive to the fleets' tag reporting rates. There was no opportunity to explore this feature for the south during the Review Workshop.

ToR 9: Review the research, data collection, and assessment methodology recommendations provided by the TC 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.

The Review Panel reviewed the Technical Committee's research recommendations, which it generally supported with agreement from all panel members, and provided its perspective in terms of their prioritization. The recommendations were viewed as pertaining to two groups: one which refers to "technical" issues directly entering into the current stock assessment model, and another one relating to gaps in life history, biological and ecological knowledge relevant for red drum. Whereas both lines of work are obviously important and should be supported, the Panel recommended that higher priority be given to addressing model concerns at this stage. Rather than repeat the full list of the recommendations here, which are listed in document SEDAR44-RW01, I highlight some of the key ones as perceived by the Review Panel:

There was clear support for finding ways to improve length composition data of the B2 component of the recreational fishery, given its increasing contribution to the overall stock removals and the issues of potential bias that have been discussed by the Review Panel (see my mention of this under ToRs 1 and 3), as well as support for finding more appropriate ways of modelling the B2 selectivity in the assessment model. Gaining a better understanding of mortality rates of discards and live releases, and the covariates that influence them, was also considered high priority. Investigating methods for appropriately weighting the model inputs in SS3, including iterative reweighting, was viewed as an aspect to be addressed in the short rather than long term, and the cooperation / verification of aging practices between labs was considered very important. The panel also supported all recommendations pertaining to achieving appropriate sampling coverage, although it cautioned against sampling just "for the sake of increasing sampling size" and suggested that the adequacy of current sampling levels relative to the importance of a particular dataset for the stock assessment and the cost of data collection should be evaluated when considering the need for increased sampling. As I noted above, all other recommendations were also generally supported, though not all were considered to be at the same level of priority.

The Review Panel identified several additional items that could improve the current model and future assessments. One was the need to conduct a comprehensive review of available information on tag reporting rates, given the apparent impact they have on the scale of model results (as found out during the Review Workshop for the northern stock; still to be investigated for the south, but I expect similar issues will likely occur there). Another recommendation refers to some length frequency distributions which were seen during the Review Workshop to have systematically empty length bins at certain lengths, likely an artefact of the conversion from inch to 2-cm length bins; this has an unknown effect on model behaviour and should be rectified before progressing further. The Review Panel also recommended that the effect of sample size cutoffs (e.g. a minimum of 300 fish for tagged groups to be included in the model, minimum sample size before length composition data borrowing starts) be explored once a solid base model is available. Two additional recommendations by the Panel were to start simple when developing new assessment models and explore the effect of modifications incorporating only one at a time (see my comment under ToR 4), and to spend additional time evaluating data sources (e.g. cohort strength in different data sets and the coherence of signals between them) prior to the stock assessment modelling part. I concur with these recommendations, while noting that I consider the first of them (reviewing information on tag reporting rates) to be particularly important.

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

Stock assessments that could serve as the basis for management advice could not be achieved by the end of the Review Workshop, and therefore I think the immediate priority should be to finalize these assessments.

Once these assessments are achieved, and peer reviewed (in a process that may or may not be called a "benchmark"), the issue of when the next assessment should occur depends on life history characteristics, data availability, stock status (unknown at present in the absence of a final assessment at this point in time), management priorities and availability of human resources. Given the long-lived nature of the species and the available data, I think that update assessments every 2-3 years and benchmark assessments (more in-depth, with external review) every 5-6 years would be appropriate.

ToR 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.

This is currently being done together with the panel's chair and the other three reviewers. The Summary Report has a later deadline and has not yet been finalized; as such, there are still a few aspects currently being discussed by the Review Panel.

Conclusions and recommendations (in accordance with each ToR):

My conclusions, suggestions and recommendations were largely incorporated in the detailed discussions provided above for each of the ToRs. In this section I highlight main aspects.

ToR 1: Data evaluation

I consider the ToR was generally well addressed. A concern is that the current method to derive length frequency distributions for the B2 component of the recreational fishery is likely leading to biased length compositions, with unknown effect on the stock assessment results. Within each region (north and south), there also seems to be significant sharing of length data between different datasets that have a strong B2 component, including also the recreational CPUE abundance index, which may be overemphasizing the influence of these length compositions in the stock assessment results. There was clear support in the Review Panel for finding ways to improve length composition data for the B2 component of the recreational fishery, and I strongly endorse that recommendation. I also endorse the Review Panel's recommendation to spend additional time evaluating the signals in different datasets, and the coherence between them, prior to the stock assessment modelling part.

ToR 2: Stock structure

The information presented by the Assessment Team was sufficient and informative and supports continuing to assess Atlantic red drum as a northern and a southern stock.

ToR 3: Stock assessment methods and models

The model platform chosen by the Assessment Team for these assessments, Stock Synthesis (SS3), is appropriate and allows dealing with some of the main recommendations made at the previous benchmark assessment (SEDAR 18). I support this choice of modelling platform and the efforts that are being made to transition the red drum assessments from the previous statistical catch-at-age model to SS3. Once a basic, but realistic, model configuration is obtained, the flexibility of this platform will permit the incorporation of future extensions that can make the assessments even more realistic (e.g. selectivity blocks or, later in the future, quarterly time steps, that would more realistically handle spawning time, times of tagging and recapture events, and growth estimation).

The Review Workshop focused on reviewing the progress made to date by the Assessment Team in transitioning to SS3 stock assessments and in trying to progress these assessments towards a configuration that could be considered appropriate for providing management advice. The modelling decisions previously made by the Assessment Team were reviewed and considered sensible on the whole, although a number of issues were raised and ideas offered in an effort to help advance these assessments. The Review Panel considered that the best way forward for model explorations during the Review Workshop was to start with simple models, and explore issues one at a time in order to facilitate understanding; this is a recommendation made by the Review Panel for this and future assessment work, and I agree with it. Although it was not possible to achieve stock assessments that could be considered appropriate for providing management advice by the end of the workshop, progress was made, particularly for the north.

For the northern assessment, the explorations during the workshop led to an improved understanding of model behaviour. From my perspective, the biggest issue that remains to be solved is that of overall scaling of the assessment results, which has been found to be closely linked to the assumed values of tag reporting rates. When the stock assessment was allowed to internally estimate the tag reporting rates of all fleets, it produced unintuitive results (very low estimates of tag reporting rates, with lower tag reporting estimates for recreational than commercial fleets, and very low SPR values, contrary to SEDAR 18 and the continuity run). This suggests that the stock assessment may not be able to estimate all tag reporting rates internally. However, the explorations suggested that if a reliable, externally estimated, tag reporting rate can be provided for one of the fleets, the others could be adequately estimated in the stock assessment. As the tag reporting rates were found to play a key role in scaling the stock assessment, the Review Panel has recommended conducting a comprehensive review of reporting rate values, which I endorse. I also think it is important to investigate the issue of the large B2 lengths, due to the possible conflict of signals about population structure with the information provided by the long-line survey.

For the south, a problem was encountered during the Review Workshop explorations that prevented progressing at the same rate as for the northern assessment. In essence, an illogical selectivity pattern,

with very high selectivity at small lengths, was estimated for some fleets. It was not possible to understand during the Review Workshop the causes for this behaviour, which had only occurred to a much lesser extent in the run presented by the Assessment Team at the start of the workshop. This needs to be understood and resolved in order to make further progress with this assessment. My recommendation would be first of all to contrast the data and control SS3 input files from the run conducted before the workshop (i.e. the run whose results are shown on the Assessment Report document that was sent to the reviewers, SEDAR44-AWReport) with those from the runs conducted during the workshop. On checking the data input files available on the SEDAR 44 FTP site, I was left with the impression that the sample size values of the length composition data had changed, and were lower in the runs conducted during the workshop, which might partly explain the illogical results found during the workshop. Just as a check, I would also suggest increasing the weight given to the length frequency data in order to direct the model to fit those data more closely. In addition, I understood that symmetric beta priors had been used in the run conducted before the workshop, and that the selectivity had been fixed to be 0 at the lowest length bin; these features appeared to be absent from the control file I saw on the FTP site for the runs conducted during the workshop. Additionally, the runs conducted during the Review Workshop used one fleet less than the earlier runs (because of the shortened time period considered in the assessment), so non-trivial changes were required in the input files. I recommend double-checking all these aspects of the input files for consistency and, hopefully, some of the above may be able to resolve the problem encountered. Once the selectivity problem has been resolved, my recommendation would be to explore sensitivity of the assessment results to the values of tag reporting rates, as has been done for the north.

A problem concerning the 0.75 latency period used in the tag-recapture sub-model has been detected after the Review Workshop. Upon checking the echoinput.sso SS3 file it seems that the value 0.75 has been interpreted by SS3 as 0, and it seems likely that this is because only integer values are currently allowed in SS3 for the latency period. This probably means that SS3 has interpreted that the tag recaptures during the year of release correspond to the entire year, and not just to the last quarter of the year. The implications of this are not entirely clear to me, but I expect it can create bias in the estimates of fishing mortality and, hence, in the assessment results. The best solution in the longer run could be to move to a quarterly model. For the time being, I recommend that a run with a latency period of 1 be attempted and results contrasted with runs using 0.75 (i.e. 0 in practice) to understand whether or not it makes a big difference to the results. Other sensitivity runs based on the alternatives the Assessment Team had initially considered to deal with the latency period problem (explained in Section 2.1.4 of the Assessment Workshop report) could also be considered.

Once a basic assessment is achieved, my recommendation (in line with the Review Panel's) would be that the Assessment Team explores whether some of the features incorporated in their pre-workshop assessments (and removed during the Review Workshop exploratory phase) still seem necessary, i.e. if they help improve model diagnostics. My main suggestions go in the direction of selectivity: explore the effect of incorporating a more realistic shape for the selectivity of the B2 removals and recreational CPUE abundance index, as well as the possible need to incorporate changes in selectivity to reflect changes in the fisheries over time. Specifically for the north, I also suggest to consider replacing the exponential-logistic back to double-normal, which has a more flexible shape (I am just slightly concerned that the less flexible exponential-logistic shape could end up inadvertently driving some unwanted model result).

Another direction of exploration that has been discussed within the Review Panel (with part of the discussion taking place by correspondence after the Review Workshop) is to implement age-based selectivity for the long-line surveys for which age composition data are available, and to fit to their age composition data unconditionally on length. A number of explorations relating to the most appropriate treatment of the age and length composition data, as well as selectivity specifications, of the longline surveys are being suggested by the Review Panel and will be presented in the Panel's Summary Report. The rationale for them would be to see if a better translation of the cohort signals observed in the long-line surveys into stock assessment outputs can be achieved; this is because an approach based on age-selectivities fitting directly to age composition data (unconditional on length) would presumably provide a more direct linkage between model estimates of abundance at age and the age composition observed in the long-line surveys, hence avoiding going through the growth model (which is known not to fit the red drum age-length data particularly well) when fitting to the age composition data of the long-line surveys. It is a little unclear how the length composition data of the long-line surveys would best be used in this context, but I think it could still be kept as length composition data, be incorporated as observations of mean length at age, or potentially be excluded from the model fit. I am a little less convinced than some other Review Panel members that this is the best way forward. My sense is that this will lead to some loss of information with respect to that provided by the original survey dataset which contains, as I understand it, randomly sampled (age, length) pairs, plus additional length samples with no ages. I think the way the Assessment Team has treated the composition data (as a multinomial likelihood for the lengths, and also multinomial likelihoods for the ages conditional on length) is the most correct and will likely provide optimal information for the stock assessment overall. Nevertheless, I am happy for the suggested explorations to take place as I think they will help in gaining additional understanding. An additional exploration with the present model configuration could be to increase the weight of the age (conditional on length) data from the long-line surveys, but not for the other fleets, again with the intention of providing a closer fit to the age compositions observed in those surveys. It is clear that the whole assessment model is very reliant on being able to estimate an appropriate growth model to ensure an appropriate interpretation of the conglomerate of length and age data signals provided by the different datasets. Therefore, I think it could be worth doing some additional exploration of potential improvements to the growth model used in SS3, mainly by considering multiple ages at which K may change, instead of just a single age for the change in K. Another aspect that I think should be considered in this context, but in the future rather than now, is if there are any evidences of growth changes over time. There was no opportunity to explore any of this during the Review Workshop, so some of these are my own ideas arising from further thinking through the issues after the Review Workshop and from exchanges with other Review Panel members.

Given the long-lived nature of the species, and the fact that M is the main source of mortality for the adult fish, there is some potential for the choice of M to impact on the assessment results and it would be helpful to understand the magnitude of the potential effect. Therefore, I recommend sensitivity runs be conducted with respect to the choice of M.

ToR 4: Diagnostic analyses

The Assessment Team has been employing appropriate diagnostics (residual plots, examination of minimum likelihood values, likelihood profiles, sensitivity runs to understand the impact of different modelling assumptions). Two additional ideas specific to the workings of SS3 are: using the echoinput.sso file to ensure the inputs have been interpreted by SS3 as intended by the analyst, and perhaps also to include the age composition data (unconditional on length) for the long-line surveys as an additional input to SS3, but with a negative value in the fleet (e.g. for fleet 8 enter -8 in the input file), so that the data are not part of the likelihood calculation, but model predictions of age compositions are produced and can be checked against the age composition data (details in SS3 manual). Retrospective analysis should also be undertaken, particularly once configurations that are close to what may be considered a final model are achieved.

ToR 5: Methods to characterize parameter uncertainty

The Assessment Team has been employing asymptotic standard errors (obtained directly from SS3), which I find appropriate. MCMC methods could also be considered at a later stage. Sensitivity analyses are useful to examine the impacts of structural assumptions, and this is covered under ToR 4.

ToR 6: There was no minority report

ToR 7: Best estimates of stock biomass, abundance and exploitation status, if possible, or alternative estimation methods

It was not possible to address this ToR during this SEDAR review, as already explained earlier in this report. The work during the Review Workshop was very intensely focused on trying to progress the SS3 assessments, but no assessments that the Assessment Team or Review Panel considered appropriate to provide management advice had been achieved by the end of the Review Workshop. Of course, alternative methods could be considered (see my comment about monitoring aspects under ToR 7 in the previous section of this report), but there has been no time to address this during this SEDAR review. The Review Panel supports that the efforts to switch to the SS3 modelling framework be continued as SS3 can address many of the concerns from SEDAR 18, and provide more relevant management advice than the previous modelling framework, and I concur with that view.

ToR 8: Choice of reference points and methods to estimate them. Stock status determination or alternative methods.

The previous section of this report contains my conclusions and recommendations for this ToR (under ToR 8 heading), so I do not repeat them here.

ToR 9: Review research recommendations.

This ToR was addressed during the Review Workshop. I presented the conclusions and recommendations in the previous section of this report, so I do not repeat them here.

ToR 10: Timing of next benchmark.

The previous section of this report contains my conclusions and recommendations for this ToR (under ToR 10 heading), so I do not repeat them here.

ToR 11: A Summary Report is currently being prepared by the Review Panel; it has a later deadline and has not yet been finalized. Several aspects of the Summary Report are still being discussed by the Review Panel.

Appendix 1: Bibliography of materials provided for review

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: Copy of CIE 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

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:
    a. Presentation of data source variance (e.g., standard errors).
    b. Justification for inclusion or elimination of available data sources,
    c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size),
    d. Calculation and/or standardization of abundance indices.
    e. Estimation of discards and size composition of discards.

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

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:
    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).

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
    b. Retrospective analysis

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

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.

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

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.

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.

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

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.

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** *- Agenda Review, TOR, Task Assignments*	**Coordinator**
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** **-** *Assessment Data & Methods* *- Identify additional analyses, sensitivities, corrections*	**Chair**
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** **-** *Continue deliberations* *- Review additional analyses*	**Chair**
12:00 a.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion** *- Continue deliberations* *- Review additional analyses*	**Chair**
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session** *- Recommendations and comments*	**Chair**

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

Thursday

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

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

Appendix 3: Panel membership or other pertinent information from the panel review meeting

Panel members (alphabetical):

- Jeff Brust, USA (Panel Chair, appointed by ASMFC)
- Gavin Fay, USA (appointed by ASMFC)
- Carmen Fernández, Spain (CIE)
- Jamie Gibson, Canada (CIE)
- Sven Kupschus, UK (CIE)