Center for Independent Experts (CIE) Independent Peer Review Report for:

The Southeast Data, Assessment, and Review (SEDAR) 93 Atlantic Red Drum

Kotaro Ono[1]

Prepared for:
The Center of Independent Experts (CIE)

[1] Institute of Marine Research. P.O box 1870 Nordnes. NO-5817 Bergen. Norway. Email: kotaro.ono@hi.no

1 Executive summary

The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations and specify whether the science reviewed is adequate

The SEDAR 93 Review Workshop (RW) was held in Charleston, SC August 13-16, 2024. The RW was conducted in an open, friendly, yet professional manner and included two stocks: the northern and southern red drum. Both stocks were assessed following advice from the 2022 simulation work and were done rigorously. The northern stock was assessed using Stock Synthesis (SS), Traffic Light Analysis (TLA), and skate methods. The skate method was a new data poor assessment method that was presented in this study for the first time and was not evaluated in the 2022 simulation work. Results from the SS models developed for the northern stock were not reliable (contrasting from the simulation work), suggesting poor information in the input data. Consequently, the status of the northern stock was assessed using both the TLA (qualitative assessment) and skate methods (semi-quantitative) which suggested the stock was likely not overfished but that there has been an increasing trend in fishing mortality which warranted attention. That said, the review panel recommended revising some survey index standardization as they showed a problematic residual pattern which could affect the derived abundance indices, thus, the assessment outcomes as well. Part of the reason could be attributable to the changes in survey coverage in space and time that were not adequately captured by the model, and some calculated indices of abundance that did not properly reflect the influence of environmental effects that were included in the CPUE standardization model.

For the southern stock, the developed SS model was more stable in performance and was recommended as the method of choice for stock status determination. The model indicated that the stock is likely overfished, and that overfishing was likely happening in the terminal year of the assessment. That said, some reservations were made regarding the abundance indices currently used as inputs to the SS model as many were derived from a CPUE standardization model and showed a problematic residual pattern.

It is therefore recommended that the Stock Assessment Subcommittee (SAS) team revise all their survey CPUE standardization models and make sure none exhibit problematic residual patterns. The SAS team also needs to revise the approach taken to derive the abundance indices. This includes a proper treatment of space and time (i.e., if the final model includes a variable that changes its value over space and time, e.g., temperature or an area x time "interaction", the derived index needs to properly account for these effects), and proper propagation of uncertainty. Once done, the SS models need to be rerun to ensure that results are consistent/stable.

Despite some aforementioned setbacks, I believe the SS models merit further development for both stocks in the future. This is especially relevant since the models might perform differently with the updated indices, and SS is a flexible framework that allows integrating many different data sources (existing and any new ones, e.g., tagging). A few additional recommendations have been included in this report for the team to explore further before the next benchmark. These include further use of the simulation framework to evaluate reference points for each of the three assessment approaches, conducting a value of information analysis (in SS), exploring a seasonal model (in SS), and incorporating tagging data (in SS).

2 Background

*The report must contain a **background** section, description of the individual reviewers' roles 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 ToR*

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage the USA's marine living resources based upon the best scientific information available (BSIA). NMFS science products require scientific peer reviews that are strictly independent of any influences. Thus, these external reviews are essential to strengthening scientific quality assurance for fishery conservation and management actions. The Southeast Data, Assessment, and Review (SEDAR) is the cooperative process by which stock assessment projects are conducted in NOAA Fisheries Southeast Region. SEDAR was initiated to improve planning and coordination of stock assessment activities and to improve the quality and reliability of assessments. The focus of the SEDAR 93 was on the northern and southern Atlantic Red Drum stocks benchmark assessment evaluation.

Description of the individual reviewer's roles

*The report must contain a background section, **description of the individual reviewers' roles** 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 ToR*

The role of the reviewer is set out in the CIE Statement of Work, Attachment A, attached here in Appendix 2 and reads as follows:

- Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
- Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.
- Reviewers should elaborate on any points raised in the summary report they believe might require further clarification.

- Reviewers shall provide a critique of the NOAA Fisheries review process, including suggestions for improvements of both process and products.
- The 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 report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.

3 Summary of findings for each ToR

*The report must contain a background section, description of the individual reviewers' roles 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 ToR*

3.1 TOR 1. Evaluate responses to Simulation Assessment Peer Review Panel recommendations.

There were two recommendations from the Simulation Assessment Peer Review Panel:

a. Revise the grid search for deriving reference points for the TLA and only include data that would be available to a TLA model in the current assessment (i.e., pre-2023).

b. Demonstrate that the southern Stock Synthesis (SS) estimation model (EM) could produce unbiased estimates when fitted to data generated from the OM without observation error.

The revised grid search was used to calculate the TLA reference points used in the analyses presented in the assessment report.

The SAS team used the data generated from an iteration of the OM without observation error and used an EM that had fewer misspecifications (growth and M were fixed at the OM true values) to demonstrate that the model can indeed exhibit less relative error in derived quantities (with the relative error approaching zero as the degree of misspecification decreased). However, it would have been beneficial to evaluate whether the model can consistently produce unbiased estimates across multiple iterations (i.e., with different recruitment time series) to ensure its robustness, i.e., that the mean (or median) relative error across iterations is zero.

But overall, the SAS team appropriately responded to the Simulation Assessment Peer Review Panel recommendations.

4 TOR 2. 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:

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

The presentation of variance in the various data sources was generally good throughout the report. Values of standard error and/or confidence intervals were provided in tables or plots.

4.2 Justification for inclusion or elimination of available data sources.

Justification for inclusion or exclusion of available data sources was generally satisfactory except for the southern stopnet survey. For the latter, part of the justification provided from the SAS team seemed inappropriate (i.e., a post model run showed that the recruitment index that was not in conflict with other indices). In general, the choice of inclusion/exclusion of data sources should be decided pre-analysis, based on holistic thinking about 1. Whether the spatial coverage of the data source can be assumed to be representative of the whole stock for particular life stages (in this case, recruitment, subadult and adult stages). 2. Whether there are enough reliable samples to derive an abundance index. If so, 3. Determine whether any observed deviations from the original sampling plan/design could be properly addressed during a CPUE standardization.

Some age data derived from scales were available for some years, but this data source was shown to be only accurate for age 3 and below and increasingly biased for older fish. The SAS team therefore excluded all age information derived from scales in the current assessment model. During the RW, the panel asked for further detail on the available scale ageing data (age 4 and below), i.e., sample size, spatio-temporal coverage, but not all answers were provided during the workshop time. Therefore, it is not fully clear yet whether ageing data from scales might help expand the spatial coverage of ageing data for some years or not.

4.3 Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, aging accuracy, sample size).

The SAS team provided sufficient information in the stock assessment report and in RW presentations to describe the available datasets. Further clarifications were asked and provided during the RW to understand any sampling heterogeneity (type of gear used, bait type, tidal cycle, spatio-temporal coverage, etc.).

One novelty and a big strength in the current benchmark is the inclusion of the tagging data (size at release information) which provides very valuable information on the discard length composition.

However, challenges in modelling the adult longline surveys continued from the last benchmark assessment (2017) for both the northern and southern stocks. The indices derived from the longline survey are of great interest because they may provide historical and contemporary scaling of the adult abundance to the SS models. The value of information from the contemporary longline survey was expected to be higher than before, as the time series has more than doubled in length since the last benchmark assessment. However, the SS models were not able to fit to both the historical and contemporary longline indices. The review panel recommended some sensitivity analysis on the exclusion / replacement of the longline indices

and the analysis showed a minor effect (a bit more for the south when removing the historical indices).

This means that in its current state, the longline survey does not provide much information to the SS model and/or does not align with the other data sources. There could be several reasons for such an observation:

- The existing longline survey data and coverage may not accurately reflect the status of the southern adult stock. If neighboring states were to conduct similar surveys in their waters, it could potentially provide a clearer understanding of the situation.
- The issue could be related to the CPUE standardization underlying the longline survey indices. The panel requested residual diagnostics on all CPUE standardization models and all longline surveys for the southern stocks (in addition to other surveys) showed problematic residual patterns. A trial on improving the CPUE standardization model was further requested during the RW and the SAS team was able to find a model (for the SC_longline_contemporary) with no residual pattern. This model led to a few changes in the index trend compared to the original model but had minimal influence on the SS model output. That said, it has been shown in the literature that CPUE standardization of longline data could be tricky at times due to hook saturation effects, depredation effects (e.g., shark), predator/prey interactions, and tidal conditions affecting fish activity level, among others. It is thus recommended to further explore and evaluate the potential influence of these factors when standardizing the longline survey CPUE data.
- There might not be enough ageing data from the longline survey to inform on the cohort strength. Before expanding biological data collection, it is advisable to develop age-specific indices for selected age groups (using existing data and a modeling framework such as VAST or sdmTMB) and verify their consistency with one another. If such work is inconclusive, then some simulation testing on the value of longline ageing data could be investigated under different scenarios (e.g., assumptions about the representativeness of the longline data in indexing the adult portion of the stock).

4.4 Calculation and/or standardization of abundance indices.

Several surveys showed poor residual performance during the CPUE standardization process. As these models are used to create indices of abundance - which are used in all three assessment methods presented during the workshop – problematic residual patterns should be eliminated to ensure reliable indices. This belief was supported during the RW when the panel requested and was presented with the results of some sensitivity analysis on the southern SS model when using abundance indices derived from CPUE standardization models without any problematic residuals pattern. The test was conducted on the SC_trammel sub-adult index and on the SC_longline_contemporary adult index and the southern SS model was sensitive to the updated SC_trammel sub-adult index.

It is therefore advisable to consider the following recommendations when standardizing the survey CPUE data:

- Proper consideration of changes in the spatio-temporal coverage or sampling design for

each survey. This requires including some spatio-temporal effect in the model and/or any available variables that reflect those changes.

- Proper inclusion and treatment of variables that could define red drum "suitable habitat" (i.e., any static (e.g., depth) or dynamic (e.g., temperature, salinity) variables potentially affecting the underlying red drum abundance as opposed to "catchability" variables which only affects the effectiveness in catching red drum). If the habitat variable is dynamic, its effect should be properly included when deriving the abundance index as well.

Once all the abundance indices satisfy the above recommendations, the SAS team should check that the indices are consistent with each other for overlapping age classes / cohorts across surveys to identify when signals on abundance trends / year class strengths may be different or similar.

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

5.1 If modeling approaches differ from those recommended during the Simulation Assessment, were these differences warranted and appropriate?

During the workshop presentation, the SAS team provided a summary table highlighting the differences between the SS estimation models (for the northern and southern stocks) used during the Simulation Assessment and the ones used during the current benchmark. All except for the choice of steepness value were appropriate. The steepness value was fixed in the SS model for both stocks due to the lack of data to estimate the parameter. However, the SAS team decided to fix its value to 0.99 as opposed to 0.84 (as previously used in the Simulation Assessment and the value was based on the literature) and the decision felt arbitrary and ad-hoc. Moreover, this decision effectively means that there is no stock-recruit relationship for the red drum which is a strong assumption. It is therefore important to provide enough evidence to support such a strong assumption. Otherwise, the use of 0.84 (which is supported by the literature) would be preferable.

For the TLA approach, the SAS team followed the recommendation provided during the Simulation Assessment on the choice of reference period to perform the grid search for deriving reference points.

The Skate method (applied to both the northern and southern stocks) is a new data-poor assessment approach that was not presented in the 2022 Simulation Assessment. The approach seems reasonable and is used as a back-up approach to provide catch advice in case the SS models are unstable.

The Cormack-Jolly-Seber (CJS) tagging model is a new approach that was not included in the 2022 Simulation Assessment. While it is not an assessment method, the CJS provides a framework for analyzing temporal trends in total mortality. Furthermore, its development was recommended by past SAS teams and review panels.

Overall, the modifications made in the study were justified and appropriate, apart from the issues outlined in the preceding paragraphs.

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

Both the northern and southern stocks were assessed using SS, TLA, and skate methods.

For the northern stock, the SS models proved unreliable (unstable) leading to the decision to favor the TLA (a qualitative assessment method) and the skate method (a semi-quantitative assessment method). This choice was appropriate and justified. For the southern stock, the developed SS model was more stable in performance and was chosen as the preferred method. This decision was warranted.

5.3 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).

All 3 methods (SS, TLA and Skate) heavily depend on the use of reliable indices of abundance. This reliance is particularly important for the data poor methods (TLA and skate method) as they directly use these indices to determine stock status as opposed to SS models which can include additional data to weight the influence of indices. Not all survey data were used to create population indices for different life stages, i.e., recruitment, sub-adult, and adults. The decision to exclude certain survey data for index standardization was generally well-justified, but the rationale behind the inclusion of some data was less clear (see 2.2). All indices were equally weighted (in terms of contribution to the total likelihood) which was appropriate. However, problematic issues were identified during the index standardization process which would justify a thorough re-analysis. This re-evaluation is crucial to ensure the accuracy and reliability of the indices, which are fundamental to all three methods.

5.3.1 SS models

In general, the choices made by the SAS team when developing the SS model parameterizations and specifications were consistent with the available data and knowledge of the stock(s).

- Changes to using fishing year (September – August) instead of calendar year (January – December) are better aligned with the life history of the modeled species and the data collected.
- The justification for using an age-dependent natural mortality was well-documented and aligned with current common practices for stocks in the southeastern US.
- The choice of a length-based double normal selectivity and retention curve for the fishing fleet was well-justified. Furthermore, the retention curve was made time-varying (in blocks) to capture known changes in slot limits regulations. Such a decision was sensible. Finally,

the choice of fixing the width of the full-selectivity dome was well-justified and the sensitivity analysis conducted to test its influence was appropriate.

- Similarly, and more generally, the use of prior probabilities for parameters poorly informed by data were well-justified.
- On the other hand, the choice for fixing the Beverton-Holt stock-recruitment's steepness parameter to 0.99 was poorly justified. One of the justifications was that the parameter was hitting the upper bound when freely estimated but this is not appropriate. It is more appropriate to use existing information on the most likely steepness value for red drum, which is 0.84 (a value used as the basis in the 2022 simulation work).
- For the composition data (length and age), the SAS used multinomial distributions with the number of trips as sample size. The approach is reasonable, but other commonly used self-weighting distributions (e.g., Dirichlet) could have been worth exploring.
- Finally, there is a mismatch between the distributional assumption around the indices of abundance (which depends on the distribution used during CPUE standardization) and the choice offered within SS (normal, lognormal, Student's t). Such an issue is not specific to red drum but applies for any other species modelled using SS. At the moment, it remains unclear how to properly do this adjustment/correction.

5.3.2 TLA analysis

The TLA method relies heavily on the specification of the reference period, which is supposed to cover the time span when the stock was in a "good status". However, the choice of the reference period for both stocks was based on previous assessment results rather than the time frame of the available time series. It would be preferable (and more generalizable) if the reference period could be chosen based on the available time series using some objective criteria rather than past assessment outputs, which could be unreliable and/or not available in certain cases.

Moreover, the optimized value of the thresholds from the Simulation Assessment was arbitrarily adjusted for the adult abundance indicator. The rationale for changing this was well-expressed but the choice of the scalar was not justified.

5.3.3 Skate method

In general, model parametrization of the skate method was appropriate. The SAS ensured that both the index of abundance and the catch time series used were selecting the same age group of animals, meaning that both the numerator and denominator reflected the same unit and focused primarily on the young segment of the population that is being "harvested" by the fishery and recreational activities. Additionally, the use of a three-year moving average for both the catch and index was appropriate. This allowed the analysis to focus on the main trends while minimizing the noise effects. That said, a weighed moving average (where the weight is, for example, based on the inverse of the variance around the catch:index ratio, which could be calculated using a linear approximation of the formula, could have been used to consider the

uncertainty around the catch:index ratio, even though the moving average would have likely weighted down large variations.

The reference point in the skate method was static, which is appropriate for such an analysis. However, the choice of the "reference point" which is currently based on the median of the catch:index ratio over the chosen/available time series seems arbitrary and its choice can have large consequences on the catch advice.

However, there are some weaknesses in the presented data-poor approach that should be pointed out. The method is sensitive to variable recruitment (variability in year class strength). As the catch:index indicator transitions from strong year classes to weak year classes (or vice versa), the catch advice will be over and under-estimated respectively. And there isn't any proposed approach to adjust for this effect. Moreover, the method is prone to ratcheting of advice even though the three-year moving average reduces its impact.

5.3.4 The Cormack-Jolly-Seber (CJS) tag model

The CJS model used for analyzing the red drum tagging data was appropriate. That said, some key information (e.g., changes over time and space in the gear type used for capturing red drum which has potential implication on post-release mortality) was missing from the report but was provided later during discussion. The possible effect of gear type on post-release mortality can be easily incorporated into the CJS model (as an extra covariate for example) and it would be recommendable to test for its effect in the future. Furthermore, it would be relevant to have updated information on the tag shedding rate as well. Otherwise, there was no notable issue on the CJS model except maybe the overly smoothed curve that was fitted on top of the CJS model output. A less smooth curve would have been more appropriate.

6 TOR 4. Evaluate the diagnostic analyses performed, including but not limited to sensitivity analyses to determine model stability and potential consequences of major model assumptions and retrospective analysis.

The SS model diagnostics were done following the standard practices outlined in the Carvalho et al. 2021 "cookbook" paper. For each stock (northern and southern) the SAS team checked that the model was stable and converged by verifying low final gradient, hessian being positive definite, no highly correlated parameters nor very low correlations, no parameters at bounds, and performed some jitter analysis.

The Francis plot was appropriately utilized to summarize goodness-of-fit to composition data. The SS models' fit to abundance indices was also appropriately presented in addition to the corresponding residual plots. The latter suggested some misfit which could be partly caused by non-adequate CPUE standardization process (see section 2.4). The SAS team also performed a thorough retrospective analysis focusing on key reference point-related quantities, i.e., spawning stock biomass (SSB), relative SSB, Age-2 fishing mortality, and spawner-per-recruit (SPR) estimates. The analysis revealed (for the southern stock) a minor retrospective pattern related to the low 2019 index value, which further reinforces the importance of proper and

reliable index standardization. The team also conducted a historical retrospective analysis to compare current *vs.* past assessment model outputs and performed an R0 likelihood profiling as well to check match/mismatch on R0 information content between data sources.

The SAS team also performed many sensitivity runs to evaluate the impact of key model assumptions and data choices on model estimates. This revealed that most of the sensitivity runs fell within the 95% confidence bound of the base case estimates for SSB, SPR, and age 2 F except for the change in the assumed width of the full-selectivity of the fishing fleet, and the use of a later starting year for the model. Some additional sensitivity runs were requested by the panel during the RW related to some issues with index standardization (re-calculation of the South Carolina trammel index as well as SC longline contemporary index), the assumed steepness value, some removal of sub-adult length composition data, removal of some Florida haul seine data for initial years with lacking age-0 data, and other combined scenario with adjusted MRIP catch estimate and other model assumptions (M and discard mortality). This indicated that many fell again within the 95% confidence bounds but some were at/beyond the bounds.

For the TLA approach, the choice of different reference periods was tested (for both the north and southern stocks) and change in outputs was compared for the different choices. The results were largely in agreement. Other sensitivity analyses were discussed during the meeting but nothing more was presented nor evaluated. These additional sensitivity analyses include: the use of updated abundance indices (based on CPUE standardization without concerning residual pattern), changes in TLA thresholds due to changes in the assumed steepness, possible inclusion of a new scenario (e.g., hyperstability/bias in the adult longline index) when evaluating the TLA thresholds.

For the skate method, a sensitivity test based on the choice of terminal year was conducted. No other sensitivity testing was performed. The SAS could have also explored other sensitivity analysis, e.g., choosing a reference period to calculate the reference F value (instead of basing it off the entire time series - similar to the TLA approach), sensitivity to the number of years used to calculate the moving average.

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

To characterize uncertainty in the SS model outputs from both the northern and southern models, the SAS team performed some sensitivity analyses, calculated the likelihood profiles for R0, and estimated the asymptotic standard errors (as a common practice in all ADMB models). All were appropriately conducted and their implications clearly stated in the report. Additional sensitivity analyses were requested during the RW which highlighted some issues (some sensitivity runs led to changes in the stock status estimate) which would merit further investigation. These include, among others, a full reworking of the index standardization and the use of a steepness value of 0.84 as a base case.

The TLA uses a simulation framework developed in 2022 to determine the "reference values", i.e., the threshold values and number of years to trigger management action. These are based on many iterations and scenarios. In this sense, the determination of these "reference values" clearly considers the uncertainty included in the operating model around the TLA inputs as well as uncertainty around the population dynamics (i.e., different scenarios). Furthermore, the SAS team conducted some sensitivity analysis on the choice of reference period and determined that there was no major change in the stock status for the southern stock. Finally, the SAS team included some precautionary principles when defining the management reference points (i.e., overfishing and overfished) in terms of frequency of any indicator being red but the choice of these reference points has not been fully evaluated using a simulation or a management strategy evaluation approach, which is recommended to do before establishing such reference points.

For the skate method, the SAS team appropriately used a moving average to focus on changes in trend while reducing the effect of noise. However, the approach could have better incorporated the uncertainty around the catch:index ratio (as already mentioned in section 3.3.3). In that way, a sudden drop in index with large uncertainty will be down-weighted compared to a highly precise drop in index value.

8 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 filed.

9 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.

For the southern stock, SS is the recommended approach to provide best estimates of stock biomass, abundance, and exploitation. The current base model appears to be adequate as the sensitivity analyses (reported and requested) presented during the meeting suggested that the model is generally conservative, i.e., most of the sensitivity analyses fell within the confidence interval of the base model and produced similar stock status outcomes. The base model also suggested that the southern stock is likely overfished, and that overfishing is likely happening in the terminal year of the assessment. That said, some reservations were made on certain input data, especially, the abundance indices. Many were derived from models with problematic residual pattern and some sensitivity runs indicated that an update of these indices might lead to a notable change to the assessment outputs. It is therefore crucial to revise all CPUE standardization models to ensure they all show no problematic residual pattern (which could be evaluated via the R package DHARMa) (see also section 2.4) and rerun the SS model to

ensure that results are consistent/stable. Consequently, there aren't any usable estimates of stock biomass, abundance and exploitation that are available to date, except for the qualitative description of the stock status. Despite this setback, I believe the SS model should continue to be used and developed for application to the southern red drum stock in the future.

The northern stock does not have any acceptable/reliable SS models but has only the TLA and skate method – two semi-quantitative assessment approaches – to evaluate the status of the stock. Thus, there aren't any available and usable estimates of stock biomass, abundance and exploitation to date. Nonetheless, these methods suggested that the stock was likely not overfished but that there has been an increasing trend in fishing mortality which warranted attention. Similarly to the southern stock, I believe a working SS model for the northern stock should keep being investigated (same recommendation as in the 2022 simulation).

10 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.

There are a few reference points for SS models that are already established by the fishery management plan (FMP). This included F30% and SPR30% as thresholds and F40% and SPR40% as targets. The Fxx% values are calculated based on age-2 fish and are the level of F that achieves an SPRxx%. The SSBxx% represents the level of SSB associated with a stock fished at SPRxx%. SSB30% is the overfished limit and SSB40% is the target. However, the SSBxx% reference points are new and not part of the FMP yet. These reference points are seemingly acceptable on the basis of previous use for other stocks but have not been specifically evaluated for the red drum stocks. It is therefore recommendable that these reference points be evaluated specifically for red drum stocks using a simulation framework for example.

The TLA approach also defined some overfishing/overfished reference points based on expert-based precautionary judgment, but not enough information was provided during the RW to be able to fully evaluate their performance. It is therefore recommendable that these reference points be evaluated in future work using, for example, a simulation framework or a management strategy evaluation.

There are no management reference points defined in the Skate method yet. If the Skate method is to be used to provide quantitative catch advice, it is then important to conduct either a simulation analysis or a management strategy evaluation to evaluate the effectiveness of different harvest control rules and/or reference points.

11 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 diverse recommendations provided by the Technical Committee (TC) were generally appropriate. However, there are a few additional recommendations for the short/medium-term.

First and foremost, the SAS team needs to fix issues with the CPUE standardization as soon as possible. There are obvious problems with some of the CPUE analysis, and all indices of abundance need to be recomputed while making sure that the underlying CPUE standardization does not show any residual pattern. Moreover, the SAS team needs to make sure that the derived indices properly account for the effect of all "habitat" covariates (both static and dynamic variables believed to affect the underlying biomass of the species during the survey period which are included in the final CPUE standardization model). The SAS team then needs to perform a historical retrospective analysis examining changes in the assessment outputs and recommendation for all included models: SS, TLA, and skate method.

Secondly, the simulation framework developed in 2022 needs to be further utilized for testing/determining a variety of assessment relevant information. This includes:

- The determination of all red drum reference points. Instead of using some values taken from the literature, the simulation framework can be used to tune in these reference points to the red drum case study. This includes the use of SPR30%, SPR40% reference points for the SS models but also the definition of overfishing and overfished status for the TLA.
- A value of information analysis should be conducted to determine the value of each survey data source (both as an index of abundance but also as the source of the composition data) and prioritize data collection. The SS models had a hard time fitting to the longline survey indices of abundance in general. One can determine how much improvement in bias can be expected if one increases, for example, the age composition sample of the longline survey to its maximum capacity. Similarly, one can determine whether the use of conditional age at length is more informative than marginal age composition data from the longline survey data. Finally, one can examine whether deriving a few age-specific abundance indices from the longline survey might be more informative to the SS model than an aggregated index (which will also allow checking the consistency of information).

Furthermore, tagging studies should be continued in the future as they provide important information to the assessment models (i.e., discard size estimate) and effort should be spent to integrate these data into SS for further use. It is also recommendable to keep monitoring/tracking tagging mortality rate as it would be beneficial to separate it from the total mortality estimate. During the RW, it was shown that the gear types from which tagging data came from changed over time. Thus, there is a need to explore a possible "gear type" effect in

the tagging model, or possibly conduct a field experiment to measure differences in tagging mortality by gear type.

As many of the above recommendations point out, there is a lot of potential for further development and use of the SS models for both stocks. Another possible axis of exploration is to investigate potential seasonal models to deal with the growth misspecification (i.e., seasonal change in growth) in addition to including the tagging data into the SS model (as mentioned in the previous paragraph).

Finally, it is recommendable to do some research focusing on realistic and effective implementation plans for the measures that might come out from the advisory process, i.e., explore ways of delivering necessary reductions in F, e.g., effort/participation control, spatio-temporal closures. To do so, a good understanding of the drivers of the fishery dynamics is required.

12 TOR 10. Review the recommended timeframe for future assessments provided by the TC and recommend any necessary changes.

The decision to have the next Benchmark Assessment in five years while updating the TLA every two years seems appropriate. That said, given the identified issues in the SS assessment for the southern stock, it would be good to update this model in 2025 as well. Such an update should consider the following.

- Include the most recent data available (including catch, biological and abundance indices information).
- Have an updated abundance indices which are derived following the panel recommendations.
- Fix steepness to 0.84.
- Include the expected revision of the catch history derived from the MRIP, if available.

13 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.

Such document was prepared according to the TOR and was completed in the allocated time frame.

Appendix 1: Bibliography of materials provided for review

Document #	Title	Authors
Reference Documents		
SEDAR93-RD01	Red Drum Simulation Assessment and Peer Review Report	Atlantic States Marine Fisheries Commission
SEDAR93-RD02	Estimating the tag-reporting rate and length-based selectivity of red drum (*Sciaenops ocellatus*) in South Carolina using a long-term tag-recapture study	Lukas Ugland Troha
SEDAR93-RD03	Spatial synchrony and temporal dynamics of juvenile red drum *Sciaenops ocellatus* populations in South Carolina, USA	Stephen A. Arnott, William A Roumillat, John A. Archambault, Charles A. Wenner, Joy I. Gerhard, Tanya L. Darden, Michael R. Denson

Appendix 2: A copy of this Performance Work Statement

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
NOAA Fisheries
Center for Independent Experts (CIE) Program
External Independent Peer Review
Southeast Data, Assessment, and Review (SEDAR) 93 Atlantic Red Drum Assessment Review

Background

The NOAA Fisheries is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NOAA Fisheries science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one (1) or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination. Specifically, science products that the agency can reasonably determine that will have, when disseminated*, "a clear and substantial impact on important public policies or private sector decisions*." Additionally, peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[2].

Scope

[2] https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2005/m05-03.pdf

The SEDAR is the cooperative process by which stock assessment projects are conducted in NOAA Fisheries Southeast Region. SEDAR was initiated to improve planning and coordination of stock assessment activities and to improve the quality and reliability of assessments.
SEDAR 93 will be a CIE assessment review conducted for Atlantic Red Drum. There are two (2) models to be reviewed: Southern and Northern Stocks. 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 the scientific basis of the assessment through the SEDAR process. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (ToR) of the peer review are listed in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3.**

Requirements

NOAA Fisheries requires three (3) reviewers and a chairperson to conduct an impartial and independent peer review in accordance with the PWS, OMB guidelines, and the ToR below. The reviewers shall have a working knowledge 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 for the stock assessment. The chair, who is in addition to the three (3) reviewers, will not be provided by the CIE. Although the chair will be participating in this review, the chair's participation (e.g., labor and travel) is not covered by this contract.

Tasks for Reviewers

1. Two (2) weeks before the peer review, the Project Contacts will send (by electronic mail) the necessary background information to the CIE reviewers and reports for the peer review. In the case where the documents need to be mailed, the Project Contacts will consult with the contractor on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance with the PWS scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.
2. Attend and participate in an in-person review meeting. The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to answer any questions from the reviewers, and to provide any additional information required by the reviewers.
3. After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and ToR, in

adherence with the required formatting and content guidelines. Reviewers are not required to reach a consensus.

4. Each reviewer shall assist the Chair of the meeting with contributions to the summary report.
5. Deliver their reports to the Government according to the specified milestones dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NOAA Fisheries Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the 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 NOAA Fisheries Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Foreign National Guest website. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The places of performance shall be in Charleston, SC.

Period of Performance

The period of performance shall be from the time of award through **September 2024**. Each CIE reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Schedule	Milestones and Deliverables
Within two (2) weeks of award	Contractor selects and confirms reviewers
Two (2) weeks prior to the panel review	Contractor provides the pre-review documents to the reviewers
August 13 – 16, 2024	Panel review meeting
Approximately three (3) weeks later	Contractor receives draft reports
Within two (2) weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three (3) performance standards: (1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each ToR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $15,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

Project Contacts:

Larry Massey – NOAA Project Contact
150 Du Rhu Drive, Mobile, AL 36608
(386) 561-7080
larry.massey@noaa.gov
Julie Neer - SEDAR Program Manager
4055 Faber Place Drive, Suite 201 North Charleston, SC 29405
Julie.Neer@safmc.net

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations and specify whether the science reviewed is adequate.

2. The report must contain a background section, description of the individual reviewers' roles 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 ToR.

- Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
- Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.
- Reviewers should elaborate on any points raised in the summary report they believe might require further clarification.
- Reviewers shall provide a critique of the NOAA Fisheries review process, including suggestions for improvements of both process and products.
- The 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 report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of this Performance Work Statement
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review
SEDAR 93 Atlantic Red Drum Assessment
Review Workshop Terms of Reference

1. Evaluate responses to Simulation Assessment Peer Review Panel recommendations.
2. 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, ageing accuracy, sample size).
    d. Calculation and/or standardization of abundance indices.
3. Evaluate the methods and models used to estimate population parameters (e.g., F, abundance) and reference points, including but not limited to:
    a. If modeling approaches differ from those recommended during the Simulation Assessment, were these differences warranted and appropriate?
    b. 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?
    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 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.
10. Review the recommended timeframe for future assessments provided by the TC and recommend any necessary changes.
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.

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

Review Panel

Gavin Fay (Chair) .....
Kotaro Ono ..... CIE Reviewer
Geoff Tingley ..... CIE Reviewer
Katyana Vert-Pre ..... CIE Reviewer

Analytic Team

Joey Ballenger ..... SCDNR
Tracey Bauer ..... ASMFC
Jared Flowers ..... GADNR
Angela Giuliano ..... MADNR
Jimmy Kilfoil ..... SCDNR
Jeff Kipp ..... ASMFC
CJ Schlick ..... SCDNR

Staff

Julie A Neer ..... SEDAR
Emily Ott ..... SEDAR
Rachael Silvas ..... SAFMC Staff

Workshop Observers

Chip Collier ..... SAFMC Staff

Workshop Observers via Webinar

Alan Bianchi ..... NCDNR
Pat Campfield ..... ASMFC
Manuel Coffill-Rivera .....
Dawn Franco ..... GADNR
Ryan Harrell ..... GADNR
Matthew Jargowsky ..... MADNR
Chris Kalinowsky ..... GADNR
Cara Kowalchyk ..... NCDNR
Laura Lee ..... NCFWS
Rebecca Scott ..... FWC

Meeting agenda - SEDAR 93 Atlantic Red Drum Assessment Review
August 13 – 16, 2024

Tuesday

8:30 am – 9:00 am	**Introductions and Opening Remarks**	**Coordinator**
- Agenda Review, TOR, Task Assignments		
9:00 am – 12:00 pm	**Assessment Presentations**	**Analytic Team**
- *Background*		
- *Assessment Data & Methods*		
12:00 pm – 1:30 pm	**Lunch Break**	
1:30 pm – 4:30 pm	**Assessment Presentations (continued)**	**Analytic Team**
- *Assessment Data & Methods*		
- Identify additional analyses, sensitivities, corrections		
4:30 pm – 5:00 pm	**Wrap Up/Public Comment**	**Chair**
5:00 pm - 6:00 pm	**Panel Work Session**	**Chair**

***Tuesday Goals*:** Initial assessment presentations completed, sensitivities and modifications identified.

Wednesday

8:30 a.m. – 11:30 pm	**Assessment Presentations (continued)**	**Analytic Team**
- *Assessment Methods*		
- Identify additional analyses, sensitivities, corrections		
11:30 a.m. – 1:00 pm	**Lunch Break**	
1:00 pm – 5:30 pm	**Panel Discussion**	**Chair**
- Review additional analyses, sensitivities		
- Recommendations and comments		
5:30 pm - 6:00 pm	**Public Comment**	**Chair**

Wednesday Goals: Presentations completed, additional sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun

Thursday

8:30 a.m. – 11:30 pm	**Panel Discussion**	**Chair**
- Review additional analyses, sensitivities		
- Recommendations and comments		
11:30 a.m. – 1:00 pm	**Lunch Break**	
1:00 pm – 5:30 pm	**Panel Discussion**	**Chair**
- Final sensitivities reviewed.		
- Projections reviewed.		
5:30 pm - 6:00 pm	**Public Comment**	**Chair**

Thursday Goals: Review final sensitivities, complete assessment work, and finalize discussions.

Friday

8:30 a.m. – 12:00 pm **Panel Discussion or Work Session** **Chair**

- *Review Summary Reports*

Friday Goals: Final results available. Draft Summary Report reviewed.