SEDAR

Southeast Data, Assessment, and Review

Center for Independent Experts (CIE) Independent Peer Review of SEDAR 93

ASMFC Red Drum

Review Workshop Report

Katyana A. Vert-Pre

Prepared for

Center for Independent Experts

October 2024

Table of Contents

1. REVIEW PANEL REPORT

1.1 EXECUTIVE SUMMARY

The review workshop was attended by the stock assessment subcommittee (SAS), a review panel (with members appointed by both the Center for Independent Experts (CIE) and the Atlantic States Marine Fisheries Commission (ASMFC)), and other relevant parties.

This document serves as my independent reviewer report for the 93rd Southeast Data, Assessment, and Review (SEDAR 93), which focused on the updated assessment of the Atlantic red drum, divided into northern and southern stocks, consistent with the structure of previous assessments. To facilitate this review, a wide array of information was provided, including four comprehensive documents: the "SEDAR93-RD01-Red Drum Simulation Assessment and Peer Review Report", the "SEDAR93-ASMFC-2024 Red Drum Benchmark Stock Assessment Report", The “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” and the “SEDAR93-RD04-Spatial synchrony and temporal dynamics of juvenile red drum *Sciaenops ocellatus* populations in South Carolina, USA”. These documents laid the foundation for the review, offering essential background and details on the models used, methods applied, and data considered in the assessment.

Given that the previous 44th Southeast Data, Assessment, and Review (SEDAR 44) could not be fully completed due to insufficient model development, and no other benchmark assessments were conducted over the subsequent ten years, with only a simulation assessment being carried out during that time, the assessment team requested a preliminary review webinar on August 5th, 2024. This one-hour session provided an opportunity for the review panel to ask initial questions and request further diagnostic analyses, which allowed the assessment team to prepare additional materials for the main review workshop two weeks later. This preliminary meeting proved to be extremely beneficial, as it enabled the panel to identify gaps in the available diagnostics and request specific analyses that were not yet covered in the written reports.

The four-day review workshop was comprehensive, featuring 13 initial presentations covering various aspects of the assessment. These presentations delved into the data inputs, parameterizations, modeling methods, results, and diagnostics for both the northern and southern red drum stocks. Additionally, they covered the life history of the red drum, past assessments, and current management practices. Three further presentations were delivered to address special requests from the review panel. Overall, the information presented was thorough and represented the best available scientific information for assessing the status of the Atlantic red drum stocks.

The modeling framework selected for the assessment was based on Stock Synthesis 3 (SS3), a well-established model used in fisheries stock assessments. In addition to SS3, three alternative methods were presented: the Traffic Light Analysis (TLA), the Skate method, and the Cormack-Jolly-Seber (CJS) model. The presentations for each of these models and methods were detailed, providing sufficient information for the review panel to make informed recommendations about the assessment of the Atlantic red drum stocks.

The panel was largely in agreement on the key aspects of this review. In my independent CIE peer review, I provide a thorough and detailed account of the key points discussed during the SEDAR 93 review workshop, as well as additional observations that I believe are critical for strengthening the Atlantic red drum stock assessment. My review specifically addresses several

core aspects of the assessment process, which were brought forward during the workshop, including model parameterization, data handling, diagnostic methods, and uncertainty assessment.

One of the key issues I raised in my review was related to the parameterization of the Stock Synthesis (SS) models for both the northern and southern stocks of Atlantic red drum. While the SS model for the southern stock was generally considered robust, several issues required further scrutiny. In particular, the assumption regarding steepness, which drives the stock-recruitment relationship, was not adequately justified and could lead to overestimation of the stock's capacity to withstand high exploitation rates. Similarly, the northern SS model faced non-convergence issues, stemming from limitations in the data series and model inputs. I recommended a stepwise diagnostic approach to address these issues and suggested adding five more years of data to improve the model's stability and accuracy.

Another critical concern I emphasized was the need for improved index standardization across all models. The panel's discussions frequently highlighted inconsistencies in the standardization of survey indices, particularly for fishery-independent data. In some cases, survey designs did not follow best practices, leading to biased indices that could undermine the reliability of the stock assessment. Recalculating and re-standardizing these indices would help reduce bias. Additionally, I addressed the criteria for data exclusion and inclusion, proposing a more rigorous simulation-based evaluation to determine which indices and data sources are genuinely informative for the stock assessment.

In terms of uncertainty assessment, I strongly advocated for a more comprehensive approach. Although the Stock Synthesis model included diagnostic tools such as likelihood profiles and asymptotic standard errors, there were gaps in fully addressing uncertainty. The absence of sensitivity analyses on key parameters, such as steepness and catch histories, left room for potential biases. For alternative methods like the Traffic Light Analysis (TLA) and Skate models, the assessment of uncertainty was similarly limited. I recommended applying simulation frameworks to explore the effects of uncertainty on model outputs and management recommendations.

The alternative methods, including the TLA, Skate, and Cormack-Jolly-Seber (CJS) models, were also evaluated in my review. While these methods provided supplemental insights into the stock status, they lacked the depth and precision of the SS model.

In summary, my independent review highlights both the strengths and areas for improvement in the current assessment process for Atlantic red drum. The Stock Synthesis model for the southern stock represents significant progress and offers the best available science for management recommendations. However, critical areas such as model parameterization, index standardization, and uncertainty characterization still require attention. For the northern stock, the SS model holds potential, but non-convergence and data limitations must be resolved. Alternative methods, while informative, should not be used as standalone tools for management decisions.

To enhance the accuracy and reliability of future assessments, my report emphasizes the need for improved diagnostic methods and sensitivity analyses across all models, more sophisticated index standardization techniques, a more comprehensive approach to uncertainty assessment, and a refined approach to the inclusion and exclusion of key data sources. These steps are essential for producing more robust stock status estimates and sound management advice for the sustainable management of the Atlantic red drum fishery.

1.2 BACKGROUND

Atlantic red drum stocks are assessed separately for two regions: the northern stock, spanning from North Carolina to New Jersey, and the southern stock, covering South Carolina, Georgia, and the eastern coast of Florida. The northern region supports both commercial and recreational fisheries, while the southern region has been exclusively recreational since the late 1980s. Over the last few decades, the recreational fisheries have become the main source of catch for the species in both regions, with release numbers far exceeding those of fish kept due to the slots and bag limits regulations in place.

In 2015, the SEDAR 44 review was originally intended to provide scientific input for managing red drum stocks. However, the assessment team was unable to recommend a suitable base or alternate model for either stock. As a result, the review workshop shifted its focus to helping the team develop Stock Synthesis 3 (SS3) models for both stocks, rather than providing immediate management advice. It was understood that no management advice would be based on the discussed models, as they were not independently reviewed.

In May 2022, a simulation workshop was held to propose improvements to the SS3 models and explore alternative methods like Traffic Light Analysis and the Skate model in case the SS models were not viable. During this workshop, the review panel recommended dropping the SCA model previously used in SEDAR18 benchmark assessment. Progress was made in addressing the issues identified during SEDAR 44. It is notable that recommendations by the SEDAR 93 panel review differed from the simulation assessment recommendations. While the workshop deemed the northern model suitable and the southern model in need of further refinement, our panel concluded the opposite, finding the northern model had not converged while the southern model had.

During the SEDAR93 review workshop. The panel's primary responsibility was to evaluate the Assessment Team’s work on the Stock Synthesis models developed for the Atlantic red drum, providing expert guidance on model inputs, parameterization, and overall model structure. Additionally, the panel reviewed newly developed alternative assessment methods to ensure they aligned with recommendations from the Red Drum Simulation Assessment and Peer Review and assessed their utility for management decisions.

The SEDAR 93 review workshop took place in Charleston, South Carolina, from August 12th-16th, 2024. As an independent reviewer for the CIE, I reviewed all the reference documents and additional past benchmarks provided pre-meeting and during the meeting and participated in discussions on data usage and the modeling approaches for both the southern and northern Atlantic red drum stocks. My role included drafting key sections of the panel review report, providing edits and feedback on the final version, and submitting my independent reviewer report.

All reviewers were actively involved in every aspect of the review process, both during and after the plenary sessions. The main conclusions for the report were collaboratively drafted during side meetings, after which the report was divided equally among the reviewers for further development.

1.3 DESCRIPTION BY TOR

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

These were two main recommendations from the Simulation Assessment Peer Review Panel: 1) Perform a revised grid search for deriving reference points for the Traffic Light Analysis (TLA) that only includes data that would be available to a TLA when it is applied in practice; 2) Continue the work showcasing that the southern Stock Synthesis (SS) estimation model (EM) could produce unbiased estimates when fit to data with no observation error. The Simulation Assessment Subcommittee (SAS) appropriately addressed the initial request from the simulation assessment review panel report. The revised grid search, using only the most recent data available for TLA, was employed to optimize the TLA reference points. However, the second recommendation was not fully completed as per the panel's report. The SAS presented results from the estimation model run with data generated from a single iteration of the operating model without observation error. In the first estimation model run, growth and natural mortality were mis-specified, meaning they remained as in the initial simulation. These were compared against a second estimation where growth and natural mortality were set to the true values of the operating model. This comparison was intended to evaluate the performance of the southern Stock Synthesis (SS) model when run without observation error. While this scenario showed that EM models fit to data without observation error had reduced relative error in derived quantities compared to the original simulations, it was only conducted with a single iteration. Ideally, this analysis should be repeated across multiple iterations to confirm that the estimates consistently remain unbiased.

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:

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

The SAS made it a priority to present a clear representation of the variance in the data used for the stock assessment inputs. In instances where understanding variance was essential, they provided graphics with standard errors or confidence intervals, along with tabulated data. In cases where variance was not depicted, the panel emphasized the importance of including it.

b. *Justification for inclusion or elimination of available data sources.*
c. *Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivity, ageing accuracy, sample size).*

The assessment report and SAS presentations thoroughly detailed the complex array of data used or considered for the Stock Synthesis model, Traffic Light Analysis, and Skate analysis of the northern and southern red drum stocks. Generally, valid justifications were provided when data or indices were excluded.

For instance, the otolith data used to assess age composition of all age class was identified as a reliable source with no significant bias. The scale-based aging process was effective for ages 1 through 3, however, the aging process for older fish showed increased bias, leading the SAS to exclude this data source entirely. The panel requested further analysis to determine whether the scale-based age data provided additional information. Although the SAS presented monthly comparisons between scale-based and otolith-based age data (Figure 1) as a rationale for exclusion, there was insufficient evidence regarding the overall time series, shorter periods, and spatial coverage. I believe the justification for excluding scale-based age data was inadequate and incorporating the data for ages 1-3 would provide valuable additional information for the recruitment portion of the population. On the other hand, the inclusion of length composition data from angler tag releases was well-reasoned and clearly presented, effectively filling a gap.

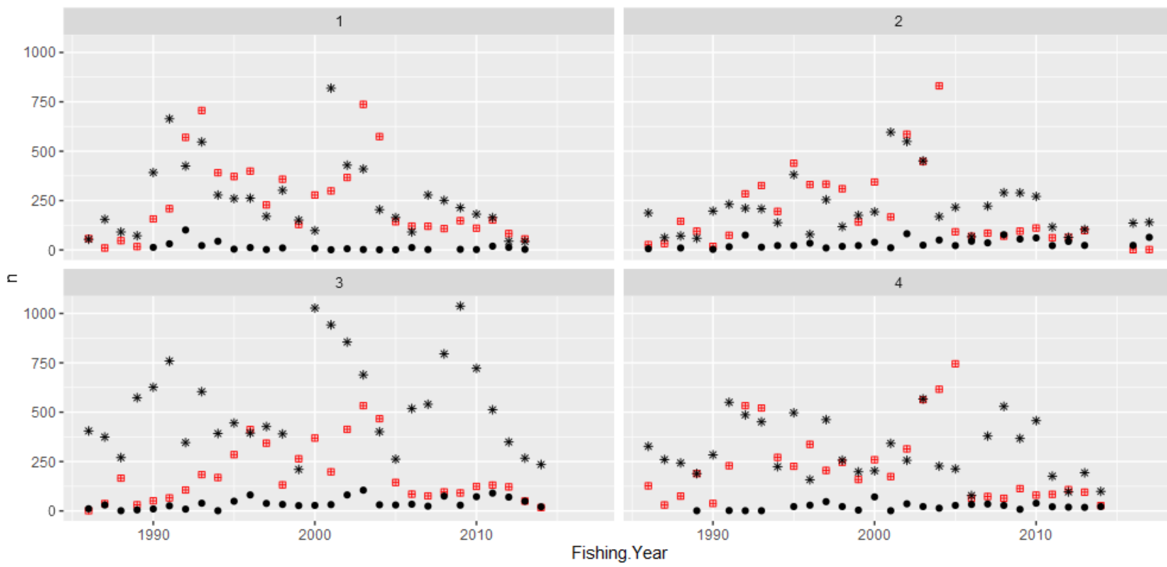

Figure 1: Monthly comparison between scale base (red squares), length based (black asterisks) and otolith based (Black circle) age determinations.

Some indices were excluded from both the northern and southern models due to their lack of representativeness of the population, such as the historical longline and stopnet indices. The stopnet index was excluded based on its limited geographical scope, though this justification was provided after the fact. Additionally, the exclusion of the abundance index derived from MRIP CPUE data was deemed reasonable due to concerns about potential hyperstability.

Furthermore, there are questions regarding the validity of including the contemporary longline index in the southern model, as it demonstrated poor informational value for the northern model. The SAS should systematically reassess the inclusion and exclusion of various indices, possibly using a simulation framework to test these decisions more robustly.

d. Calculation and/or standardization of abundance indices.

Given the transboundary nature of this fishery, the complex spatial distribution of the red drum's life history, and the multispecies nature of the surveys from which the data originates, the panel paid close attention to the standardization processes for survey indices and the spatio-temporal standardization of survey designs. Upon request and presentation of additional analysis (see Figure 2), the panel identified a lack of adherence to survey design best practices. For instance, some survey sampling was not conducted consistently in time or space, resulting in a lack of standardization. This issue was not initially apparent but became evident during further analysis, which highlighted these discrepancies. Specifically, it was revealed that changes in the FL 183m haul seine survey collection process (from 1997 to 2022) resulted in the absence of age-0 fish from marginal age compositions until 2010, a discrepancy that does not accurately represent early age compositions. Based on simulation results and the model's response to excluding these early years, the panel recommended removing them from the dataset used to generate indices. I acknowledge that achieving survey standardization can be challenging in this context; however, the SAS needs to maintain transparency regarding potential sources of bias in the data. Moreover, efforts should be made to address these biases during the standardization process, potentially utilizing tools like the DHARMa package to better account for these inconsistencies.

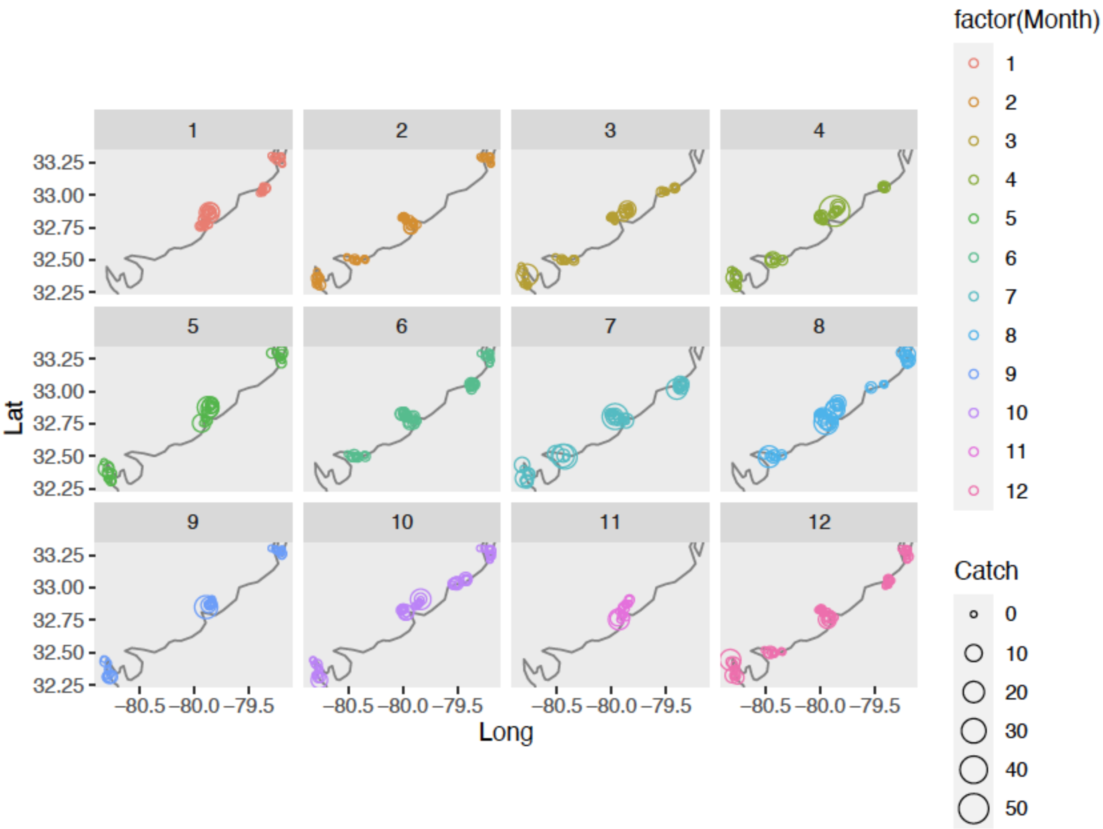

Figure 2: Trammel survey data plotted according to total catch by month for the year 2022.

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?*

The analytical team followed the recommendations from the peer review by proceeding with three performance models: the Traffic Light Analysis (TLA) of model-free stock indicators, the Statistical Catch at Age models used in the most recent red drum benchmark assessment (2017), and two Stock Synthesis (SS) models along with a simulation model based on the SEDAR 44 platform. While these modeling approaches aligned with the recommendations of the simulation assessment, there were some differences in parameterization. The SAS explored the potential use of an alternative tagging model for the southern stock. Several tagging models could be appropriate for these data, but the Cormack-Jolly-Seber (CJS) model stands out as a simple yet suitable choice. It aligns well with the data structure and the specific goals of the red drum tagging program.

For the northern stock, the SAS provided a clear summary table highlighting the parameterization differences between the SS estimation model and the model used in the current benchmark assessment. The differences within the approaches were presented more succinctly for the southern stock. Most parameterizations appeared consistent between the estimation model and the SS models, except for the choice of steepness. In the SS models steepness was fixed at 0.99, implying no stock-recruitment relationship due to insufficient data to inform the parameter estimation. Although direct estimation might not be feasible, it seems plausible to set the steepness closer to values typical of the species family. This issue was discussed more in depth in section 3c. This significant assumption requires further justification, which was not adequately addressed during the meeting.

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?*

The staff presented both methods and models, providing justifications for their selections in the report. While the TLA analysis and Skate methods are not considered models, the preferred models are the base models developed for the northern and southern stocks using the Stock Synthesis 3 platform. This platform is widely regarded for its robustness and flexibility in implementing stock assessment models, justifying its use in this context. However, the effectiveness of its implementation heavily depends on the appropriate selection of parameterizations and the sources of data inputs for the base models.

Throughout this report, I have discussed the justification, and the gaps therein, regarding the data inputs for the base models. In the following section, I will address the parameterization. However, it is already apparent that there is a discrepancy between the northern and southern models. While the southern model appears to fit the data reasonably well, the northern model continues to present issues. Given the similarities in model parameterization and data sources, this raises questions about the underlying reasons for these differences and highlights the need for a stepwise diagnostic evaluation of the northern model.

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 model parameterization and specifications were also described in detail. Overall, the SAS made consistent choices in developing the model parameterization and specifications, aligning them with the available data and the current understanding of the stocks' biology and life history traits.

c.1 Data input and sample sizes

The TLA and Skate analysis and SS models are performed using long time series of indices of abundance, good length, and age-at-length data sample sizes with the highest available quality.

For the Skate and TLA analysis, the time series were similar to those used in the final base model in Stock Synthesis (SS). Due to the data-limited nature of the analysis, these indices were directly used to infer stock status, which increased the importance of thoroughly scrutinizing the standardization process for these indices. I believe there was inconsistency in how these indices were standardized and in the diagnostics of those standardizations. For instance, QQ plots were only presented during the meeting upon request and were not included or analyzed in the report. Most of these diagnostics plots revealed some potential misspecification of error types along with misspecification of standardization assumptions. This is discussed in further detail in section 4.

For the SS models, when transitioning from simulations, it was crucial to exclude time series data that exhibited bias, were unrepresentative of the population, or lacked information. However, after reviewing the information presented at the meeting, it became clear that additional indices need further examination to determine whether they are suitable in their current form or require revisions.

c.2 Likelihood weighting schemes

In SS, likelihood weighting schemes are crucial for fitting the model to the data, ensuring balance among the different data sources influencing model output. The various likelihood components (such as abundance data, length composition data, recruitment deviations, and tagging data) contribute to the overall likelihood, which SS minimizes to find the best fit for the stock. The base model for both the southern and northern stocks is using empirical weighting. Sensitivity analyses showed that while the northern model was relatively insensitive to data weighting choices, the southern stock exhibited sensitivity to these decisions. For instance, the report assumes multinomial distributions for the length and age composition data from landings, discards, and indices. While the team explored the Francis reweighting method for composition data, they did not examine the Dirichlet-Multinomial weighting approach, which is often used to handle overdispersion in age and length composition data. Given the sensitivities in the northern model, I would recommend that the SAS explore additional weighting schemes, including Dirichlet-Multinomial, to ensure the model adequately accounts for the different data sources.

c.3 Plus group treatment

A detailed description of these indices can be found throughout this report. Compared to the previous assessment review there was a modification made as to how the model treated plus groups. The surveys conducted over the past decade have provided valuable new information. Specifically, the North and South Carolina longline surveys and the North Carolina gillnet survey have contributed conditional age-at-length composition and length composition data, respectively. This data is essential for informing the growth of larger fish and the relative strength of cohorts in older year classes within the estimation models. Consequently, the age structure of the estimation model could be expanded by replacing the 7+ age group with a full age structure, allowing for a maximum age of 62 years for the northern stock and 41 years for the southern stock.

c.4 Calculation/specification of natural mortality

The time-varying mortality calculation was conducted following best practices for stock assessment models. In addition, the SAS evaluated multiple methods to identify the most suitable approach for the red drum species' biology. As a long-lived species, red drum exhibits significant mortality variation across different age groups (with M being only significant for individuals under age 5). All estimation processes were thoroughly documented. The SAS opted for an age-based Lorenzen estimation, rather than a size-based one, utilizing a generalized length-inverse mortality framework fitted externally to the SS model. The resulting estimates were then used as inputs to inform M in the stock assessment model.

c.5 Retention rate

The SAS specified time-varying retention for the recreational fishing fleet to reflect shifts in fishing practices and regulations over time. The SS3 modelling framework allowed the SAS to model changes in retained versus discarded fish by incorporating regulatory shifts, such as slot and bag limits, using time blocks. Additionally, length-based retention curves and tuned maximum retention parameters could be applied. These specifications and tuning processes were well described in the report and seem relevant as they provide a method to estimate time-varying discard rates, especially given the limited availability of direct discard composition data from the recreational fishery.

c.6 Stock recruitment relationship

The panel raised several questions regarding the stock-recruitment relationship. In the simulation model, the SAS estimated steepness, but in the base model, steepness was fixed at the upper bound of 0.99. This would suggest that recruitment is almost entirely independent of the size of the spawning stock, implying that recruitment deviations are more influenced by randomness or environmental factors than by the stock biomass. Consequently, this assumption would suggest that the stock could sustain higher fishing rates without significantly affecting recruitment. Moreover, using SPR-based reference points would become less meaningful, as SPR is directly tied to the reproductive output of the stock. Fixing steepness at 0.99 inflates this output, potentially leading to delayed management interventions. Ultimately, a steepness of 0.99 does not seem like a realistic assumption, particularly for a long-lived species like red drum, where typical values usually range between 0.5 and 0.85. Given this, we requested the SAS to perform sensitivity analyses for the southern model and rerun SS for the northern model using steepness values recommended in the literature for the Sciaenid family, such as 0.84.

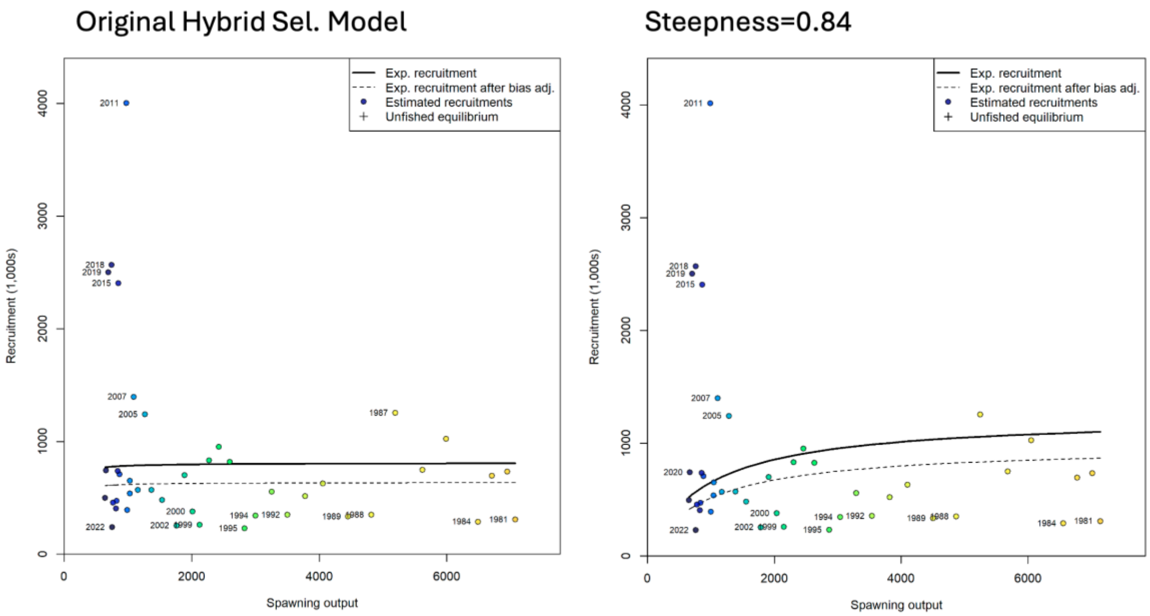

Figure 3: Stock-recruitment relationship fit for the northern Atlantic red drum stock when assuming a horizontal stock recruitment relationship (steepness of 0.99) on the left panel or when assuming a relationship (steepness of 0.84) on the right panel.

For the northern stock, assuming a stock-recruitment relationship appeared to better fit the low spawning outputs compared to assuming no relationship (Figure 3). However, this improvement did not seem to influence the estimation of SPR or enhance the overall model fit.

I believe the model should eventually be rerun for both models. Although, for the northern stock the SPR outputs didn't show significant deviation from the original model, the SAS should explore alternative values to provide a more accurate representation of stock dynamics. This should take into account the full timeseries, as well as scenarios where the 2011 recruitment outlier is excluded. The sensitivity analysis results and recommendations for the southern stock are discussed in ToR5.

c.7 Tagging-model specific parametrization

The parametrization of the tagging model was presented during the meeting, but it was not included in the report. During discussions, it became evident that much of the tagging data collected over the years was excluded from this analysis. This exclusion was not justified, making it difficult to assess whether the best available data had been incorporated into the model.

Furthermore, while the discussion addressed concerns regarding tag loss due to tag shedding, it lacked key information on several important biases. These included post-release mortality by gear type, age-dependent biases, spatio-temporal biases, and unequal recapture effort across space and time—factors particularly relevant to the red drum recreational fishery. Addressing these biases is crucial to ensure tagging models produce accurate estimates of key population parameters like survival, migration, and recapture rates. Unfortunately, the SAS did not provide sufficient discussion or presentation on these matters. Moreover, the current model used survival based on the age at release rather than the actual age of the fish at recapture. Implementing appropriate statistical methods, such as multi-state models or the inclusion of covariates, would significantly improve the robustness of the tagging study results. For example, a multi-state model would better address this issue by accounting for different states (e.g., age classes) of fish throughout the tagging study, resulting in more precise survival estimates.

c.8 Skate specific parametrization

The decision to use a three-year average and a static reference point for skate management was deemed appropriate, but several concerns were raised regarding potential sensitivities and mismatches in the method. The method appears to be sensitive to the presence of strong year classes and the management measures that were implemented. Although management actions are included in the model outputs, the method itself lacks a mechanism to adjust indicators based on those interventions, which could present a structural limitation. It may be beneficial to explore other data-poor methods that can explicitly address these structural mismatches and provide a way to adjust indicators accordingly, for example using a depletion-based stock reduction analysis. Moreover, the method seems prone to "ratcheting" advice, where overly conservative recommendations accumulate over time without built-in mechanisms for adjustment. Ratcheting in skate models can lead to overly conservative management that fails to respond to improvements in stock status, potentially hindering the optimal use of the resource. Addressing this issue requires more flexible management frameworks, dynamic reference points, regular reassessment of precautionary measures, and adaptive approaches that consider both biological and environmental changes. By mitigating the effects of ratcheting, fisheries managers can ensure that management advice is appropriately responsive to stock conditions while maintaining sustainability.

c.9 TLA specific parametrization

The correct parameterization of the TLA method relies heavily on the selection of the reference period, which is generally expected to be succinct. However, the reference periods chosen for both the northern and southern stocks were relatively broad in comparison to the available time series, with no clear justification provided. The selection was based on previous assessments rather than aligning with the actual length of the time series. For the northern and southern stocks, reference periods spanning 17 and 22 years, respectively, introduce the risk of including times when the stock was in poor or transitional conditions, potentially leading to a skewed baseline for stock status assessment. Additionally, another key parameter in the TLA method is the determination of thresholds, which were optimized using a grid search procedure. While this process was clearly communicated, the SAS confirmed that the threshold for the adult abundance index was reduced due to uncertainties surrounding this index. Although the rationale for setting the threshold to 50% of the estimated value was provided, the choice of this scalar still seems somewhat arbitrary.

4. *Evaluate the diagnostic analyses performed, including but not limited to sensitivity analyses and retrospective analysis.*

In general, the diagnostic analyses conducted for the skate, TLA, and tagging models were somewhat limited and could have been extended further to provide deeper insights. These analyses were not as comprehensive as they could have been, lacking in areas such as sensitivity testing. On the other hand, most of the diagnostic work done for the base model, within the Stock Synthesis framework, followed recognized standards and was carried out in accordance with established best practices for model evaluation.

4.1 Skate analysis

a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions.

The sensitivity analysis presented for the Skate model during the meeting focused only on the choice of the terminal year for estimating relative fishing mortality (F). Results revealed that removing the year 2022 did not significantly change the relative F estimate. However, further discussion revealed that sensitivities related to F trends had not been sufficiently tested, particularly regarding the reference period and the number of years used to calculate the moving average. Additionally, concerns were raised about the input data, such as the abundance index, which exhibited problematic residual patterns due to improper standardization, potentially affecting the analysis's reliability.

4.2 Tagging-model

The SAS did not provide any diagnostic analysis for the Cormack-Jolly-Seber (CJS) model apart from a basic goodness-of-fit test. Without additional diagnostics, such as performance evaluation or convergence checks, it becomes difficult to fully assess the model's reliability and accuracy. These diagnostics are essential for validating the CJS model, particularly in tagging studies, as they help ensure the robustness of the findings. This omission suggests that further diagnostic analysis is needed to confidently interpret the tagging data and model outputs.

4.3 TLA analysis

a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions.

The TLA analysis tested the model's outputs and management responses to changes in reference periods by adjusting the endpoints and lengths in 3-year increments, resulting in 11 and 8 scenarios for the southern and northern models, respectively. Although these analyses were presented during the meeting, they were not included in the report. The southern stock showed mostly consistent results, with some discrepancies in the adult abundance management actions, while the northern stock indicated management action was needed for recruitment and fishery performance, but not for adult abundance. While this was a solid initial sensitivity approach, it did not account for a reference period of consistent length that shifts over time. Furthermore, no sensitivity analysis was performed on the arbitrary 50% threshold used to account for the bias in the adult abundance index. This omission is particularly important given the results from the reference period sensitivity, as it could significantly impact the robustness and reliability of the analysis.

b. Retrospective analysis.

There was no retrospective analysis presented for the TLA, though performing a historical retrospective analysis could yield valuable insights. This analysis, while constrained by the length of the input dataset's time series, could still be conducted by using a shorter reference period. Incorporating retrospective analysis would allow for the evaluation of how past management decisions and model outputs compare with observed trends, potentially highlighting any inconsistencies or biases in the model, which could inform more accurate future assessments.

4.4 Stock synthesis 3 base model

The Stock Synthesis (SS) base model diagnostic analyses were conducted in accordance with standard practices for Stock Synthesis models. The SAS team delivered a comprehensive presentation of the model diagnostics, including assessments of convergence, goodness-of-fit, sources of information and structure, and sensitivity analysis.

The SAS team provided multiple elements to demonstrate the model's convergence. The model structure was deemed robust, with no parameters reaching their bounds. Additionally, the final gradient was minimal (5.76014e-05), and the Hessian matrix was positive definite, both indicators of successful convergence. These results were further supported by jitter analysis, which confirmed the model converged to a global solution. Additional convergence diagnostics, including the parameter correlation matrix, were presented, further validating the conclusions regarding the model's convergence.

Residual analysis was employed, in an appropriate way, to assess the goodness-of-fit across indices of recruitment, sub-adult, adult, and composition data. The Francis plot was utilized to summarize goodness-of-fit to composition data, which was deemed an appropriate choice. Although most residuals appeared random, the index residuals plot lacked the three residual standard deviation areas necessary for confirmation. Some residuals displayed biases and skewness, indicative of potential model misspecification. The panel identified possible sources for these issues, particularly concerning index standardization. It was noted that while diagnostics for index standardization were discussed during presentations, upon request from an earlier meeting, they were not included in the report. Once presented, the indices revealed poor diagnostics, characterized by residual patterns and skewed QQ plot distributions. The panel recommended a more comprehensive diagnostic evaluation of residuals during index standardization. Additionally, potential data issues were identified in section 2 of the report.

The retrospective analysis of information sources and model structure was thorough, employing a six-year peel to monitor key reference point-related quantities, including spawning stock biomass (SSB), relative SSB, Age-2 fishing mortality, and Spawning Potential Ratio (SPR) estimates. The detection of a minor retrospective pattern with a three-year peel divergence indicates that there may be inconsistencies in the model's ability to predict stock trends over time. The divergence, attributed to low 2019 index values, implies that the indices values for that year may not accurately reflect the stock's true status. This discrepancy suggests potential issues with the data collection in 2019 or the standardization processes. If left unaddressed, these inaccuracies could skew the overall stock assessment and lead to misinformed management decisions. Therefore, further scrutiny of these indices is crucial to ensure the reliability of the model.

The historical retrospective analysis evaluated the performance of the SS model by comparing current predictions to the previous benchmark assessment. This process helped assess the model's accuracy over time and highlighted any recurring biases. For the southern model, SPR was systematically overestimated but remained slightly over the upper bounds of the current model's estimation. In contrast, the northern model showed an underestimation trend. The analysis identified potential shortcomings, particularly in the northern model's stock predictions. Furthermore, discrepancies arose due to previous assessments using calendar years, while the current one used fishing years, the implications of which were unclear. Adjustments may be necessary to address these issues specifically in the northern model.

A detailed sensitivity analysis was conducted for the SS southern model, as outlined in section 5. Despite the thoroughness of the diagnostics, the panel suggested the inclusion of additional diagnostics from the SS cookbook. For instance, the SAS team could have considered Hindcasting cross-validation for indices, which would provide insights into the model's capacity to predict future catches. Additionally, while acceptable in this context, the absence of bridge runs from previous assessments was noted.

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

   5.1 Skate analysis

The SAS effectively employed a moving average in the skate analysis to highlight long-term trends while reducing the impact of short-term noise in the data, aiding in a clearer understanding of the fishery's performance. However, the evaluation of uncertainty in the estimated parameters was incomplete. Although the use of a simple moving average helped smooth out variability, the panel discussed the potential value of applying a weighted moving average, where the weights would be based on the inverse variance around the estimated annual index-to-catch ratio. This approach could have provided a more precise representation of the uncertainty in the estimates. Unfortunately, this method was not evaluated or presented during the meeting, leaving an important gap in the characterization of uncertainty.

Additionally, while some sensitivity analyses were conducted regarding the choice of reference period, particularly concerning the inclusion or exclusion of the 2022 fishing year, this analysis was ad hoc and lacked a systematic evaluation of the sensitivity of the results. A deeper exploration of the sensitivity to different reference periods, especially those used to calculate "relative F" (fishing mortality), was discussed but not fully presented or evaluated. This incomplete characterization of uncertainty limits the robustness of the conclusions drawn from the skate analysis.

The absence of management reference points for the Skate method further complicates the evaluation of uncertainty. Without established reference points, the implications of uncertainty in the estimated parameters remain unclear. The panel strongly recommended conducting a simulation analysis or management strategy evaluation (MSE) to assess how different harvest control rules and reference points would perform if the Skate method were to be used for quantitative catch advice.

The implications of not thoroughly evaluating uncertainty in these parameters are significant. Without a clear understanding of how sensitive the estimates are to key assumptions (such as the choice of reference period), technical conclusions about stock status and management recommendations may lack reliability. To ensure the Skate method can provide accurate and actionable catch advice, a more comprehensive approach to uncertainty characterization is essential, as it directly affects the robustness and credibility of the model's conclusions.

5.2 Tagging-model

For the Cormack-Jolly-Seber (CJS) model, the methods used to characterize uncertainty in the estimates were not thoroughly detailed. Asymptotic standard errors were employed to quantify the precision of parameter estimates, such as annual apparent survival and recapture probabilities. However, only the asymptotic standard errors (ase) for survival estimates were presented. While ase values provide some insight, they are not as robust as profile likelihood confidence intervals or bootstrap methods, both of which could easily be applied to this analysis. In particular, the bootstrap method offers a more reliable approach by generating empirical distributions of the parameter estimates, allowing for the calculation of standard errors, bias, and confidence intervals. Given the sample sizes and the complexity of the data structure in this tagging analysis, bootstrap methods would likely yield more dependable results.

5.3 TLA analysis

The uncertainty in the Traffic Light Analysis (TLA) framework is well-characterized through its use of a simulation-based approach, developed in 2022, which integrates multiple iterations and scenarios to account for both input variability and population dynamics. It establishes "reference values," which are essentially the threshold values and the number of years required to trigger management action. These reference values are derived from simulations, allowing for a robust determination that incorporates the inherent uncertainties within the operating model, as well as those related to the Atlantic red drum population dynamics. By doing so, the TLA framework ensures that the range of uncertainties, whether from input variability or ecosystem changes, are considered when establishing these thresholds.

However, a key source of uncertainty lies in the choice of the reference period. Thus, the SAS team conducted sensitivity analyses focused on the reference period to examine how different timeframes might affect the assessment outcomes. The results showed that, for the southern stock, changes in the reference period did not significantly alter the overall stock status. This suggests that the selected period was relatively stable across different potential management scenarios. Moreover, uncertainty also arises from the choice of length for the reference period itself. The current reference period used is relatively long, over 10 years, which may include years when the fishery was underperforming or experiencing suboptimal conditions. This could inadvertently increase the level of uncertainty in the model's outputs and its predictions for stock status. Reducing this uncertainty may be achieved by selecting a shorter, well-justified reference period that focuses on more stable years or a period when the fishery was functioning closer to optimal conditions. Such a shift could provide a more precise and actionable framework for stock assessment and management decisions.

Moreover, when defining management reference points, such as those used to indicate overfishing or an overfished status, the SAS incorporated precautionary principles. For instance, the decision-making process was tied to the frequency of an indicator turning "red" in the TLA framework. While this approach adds an element of caution, it is important to acknowledge that these management reference points have not been fully evaluated through more rigorous methodologies, such as a simulation-based or management strategy evaluation (MSE) approach.

Both methods are highly recommended before finalizing such critical management reference points, as they offer a more detailed analysis of how the stock might respond under various conditions.

Technically, this characterization of uncertainty underscores the need for further refinement in both the selection of the reference period and the full evaluation of management reference points using more advanced methodologies. Until these uncertainties are better addressed, management actions based on the TLA framework should be approached with caution, particularly when making decisions that could significantly impact the stock's future sustainability.

5.4 Stock Synthesis 3 base model

To address uncertainty in the southern Stock Synthesis model, the SAS conducted several diagnostic assessments, including sensitivity analyses, likelihood profiles for the parameter R0 (recruitment at unfished levels), and calculations of asymptotic standard errors. These methods were used to explore how variations in key assumptions and input data affect the model's outcomes, thereby providing a better understanding of the model's robustness.

The selection of model elements for the sensitivity runs was deemed appropriate and was consistent with recommendations from previous reviews. These elements included modifying assumptions on parameters such as natural mortality, recruitment selectivity sizes, discard mortality and timeseries like composition data, and catch histories. By testing the model's sensitivity to different assumptions, the SAS aimed to identify potential sources of bias and uncertainty in stock assessments, ensuring that the model reflected a range of plausible scenarios.

Although the sensitivity analysis conducted by the SAS was comprehensive, the panel identified necessary additional sensitivity analysis. A significant omission was a test for steepness. In the original model, steepness was fixed at 0.99, implying there is no meaningful stock-recruitment relationship. This assumption suggests that recruitment levels remain stable regardless of changes in spawning stock biomass, which may not align with the biological realities of the Atlantic red drum. Biological evidence and previous studies suggest that the stock-recruitment relationship for this species is likely stronger, but an assumed steepness value of 0.99 is unlikely. Given this discrepancy, the panel raised concerns about the robustness of the model in the absence of a more biologically informed sensitivity test for steepness. To address this, the panel requested an additional analysis during the meeting, where steepness would be set to 0.84. This alternative value was chosen based on previous biological analyses and literature on the species' recruitment dynamics, where a more moderate stock-recruitment relationship is expected. Without this analysis, there is a risk that the model could overestimate the stock's resilience to fishing pressure or environmental changes.

We requested these additional analyses to assess the impact of several key modifications like 1) The removal of the first 10 years (2000-2010) of age-0 data from the Florida haul index. This adjustment was deemed necessary due to a baseline shift in how age-0 data were recorded, as these data were only consistently included in the index starting in 2010. 2) The exclusion of sub-adult (SA) lengths. The analysis aimed to determine how the removal of SA lengths would influence model outcomes. 3) Switching to the updated standardized South Carolina trammel index. The panel identified that the standardization method used for many indices was inappropriate. Testing the effect of this misspecification on the model's performance was important to understand the implications of the incorrect standardization approach. 4) Testing the combined adjustments to MRIP catch estimates. With a 4% discard mortality or altering natural mortality by increasing or decreasing it by 20%.

These analyses were necessary to evaluate how changes in data and assumptions might affect the overall model results and management recommendations.

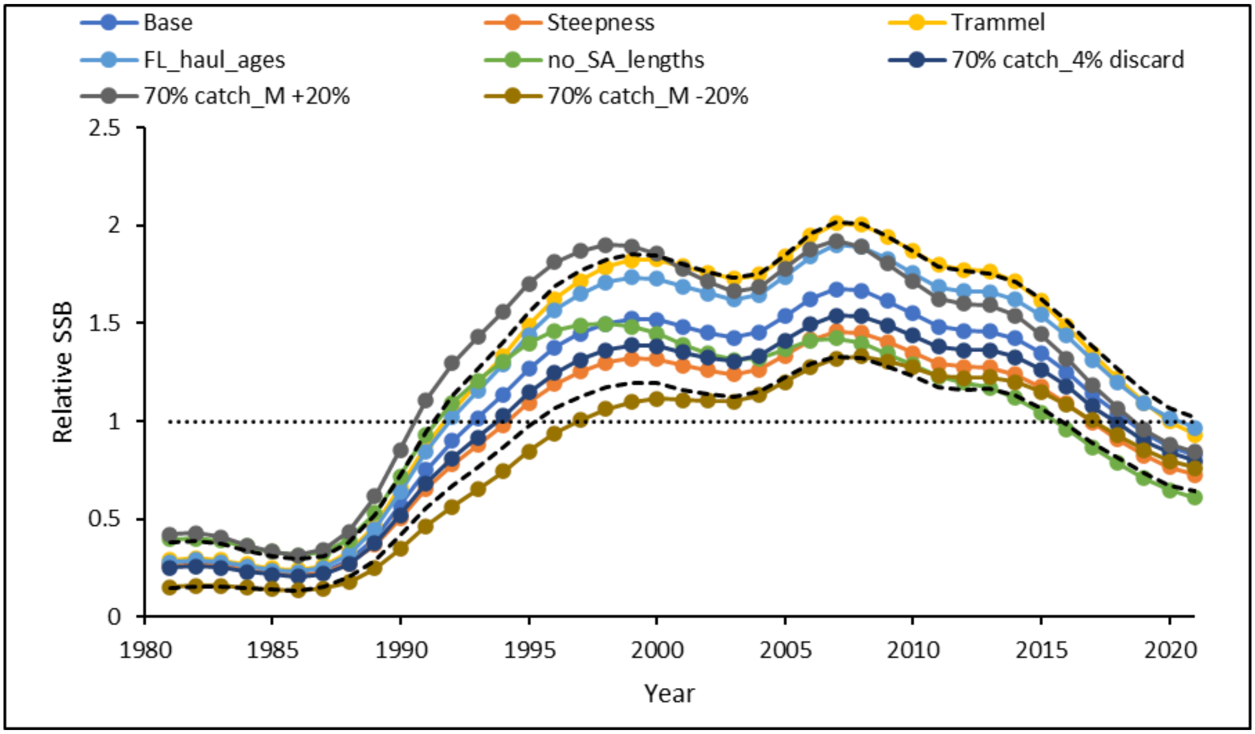

Figure 4: Timeseries response to additional sensitivity analyses requested by the panel.

Plots like Figure 4 for spawner per recruit, spawning stock biomass, and relative spawning stock biomass showed that while most analyses resulted in proportional shifts, only the removal of the Florida haul index data and the update of the South Carolina Trammel index led to a change in the stock status. Additionally, the exclusion of sub-adult (SA) lengths produced a noticeable shift in the patterns of SSB and relative SSB estimates. This suggests that the uncertainty associated with these datasets warrants further investigation.

The log-likelihood profiles for R0 showcased the contributions of both the total likelihood and the component likelihoods for each dataset. This analysis demonstrated that the model was primarily informed by the recruitment deviates, lengths, and discards, which had the strongest influence on the log-likelihood profile. However, the total log-likelihood appeared to reflect a trade-off, as the model struggled to simultaneously fit the age composition data and the index, highlighting potential conflicts between these two data sources.

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.

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

7.1 Northern stock

The northern Stock Synthesis (SS) model was not considered suitable, by both the SAS and the panel, to serve as the primary framework for providing reliable estimates of stock biomass, abundance, and exploitation rates for the Atlantic red drum northern stock. Although the model is not yet fully developed, it shows promise, especially given that the SS model has been accepted for the southern stock, and both models share a similar configuration. More effort should be focused on refining the parameterization and diagnosing the northern model. The inclusion of an additional five years of data will likely provide more informative indices and help address the non-convergence issue.

In the meantime, the SAS recommended using the Traffic Light Analysis (TLA) as a qualitative tool to provide a general indicator of the northern Atlantic red drum stock status. The results from the TLA suggest that while recruitment and fishery performance require moderate management intervention, the adult abundance does not currently warrant any management action. However, this assessment is not without its limitations. While there is relatively strong confidence in the recruitment data, which supports the status determination, the data concerning adult abundance is less reliable. This is compounded by uncertainties surrounding the scalar used for setting the management threshold, which may not adequately capture the nuances of the adult stock status. Also, an unexplained anomaly was identified in the data used to inform fishery performance for the 2011-year class, indicating an artifact in the data that could not be rationalized. Additional to index reliability the TLA analysis need more scrutiny on the reference period as the reference period sensitivity was limited. This raises concerns about the robustness of the TLA outputs for informing long-term management strategies.

Furthermore, the results of the TLA could be bolstered by incorporating a weight-of-evidence approach that draws on insights from other analyses, such as the Skate model and Stock Synthesis (SS) model, despite the imperfections in these methods. By integrating information from these various models, managers could derive a more comprehensive picture of stock status. However, it is important to note that in their current state, none of these methods (TLA, SS, and Skate) are sufficiently reliable to be used independently for formal stock assessments. Significant refinements would be needed before any of these models can be considered robust tools for decision-making in the management of the northern Atlantic red drum stock.

7.2 Southern stock

For the southern stock, the SS model is recommended as the primary tool to provide best estimates for metrics such as stock biomass, abundance, and exploitation rates. Based on the analyses conducted, the base model appears robust and provides a solid foundation for management decisions. Additional sensitivity analyses during the meeting suggested that the model is relatively conservative compared to other approaches, such as the Traffic Light Analysis (TLA) and the Skate model. Most of the sensitivity analyses fell within the confidence interval of the base model, resulting in similar stock status outcomes, further supporting the reliability of the SS model.

However, several areas require further attention to enhance the model's accuracy and reliability. These areas include input data quality, the index standardization process, and certain sensitivity analyses. The following updates are strongly recommended for the upcoming year :1) Revise the Index Standardization. I would suggest using the DHARMa package for index standardization. This tool will help identify and correct any potential spatio-temporal autocorrelation, which could affect the accuracy of the model's predictions. A robust standardization process is crucial to ensuring the model accurately reflects the fishery's underlying dynamics. 2) Update the Catch History. The catch history should be updated with the most recent available data. Incorporating the latest information will improve the precision of the model's estimates and ensure its relevance to current fishery conditions. Updating the catch history will also better equip the model to predict future stock trends and inform management decisions. 3) Reevaluate the Use of the Contemporary Longline Survey. Consider revisiting the inclusion of the contemporary longline survey in the model. The survey may not provide sufficient or reliable information for certain regions or time periods, and its continued use could introduce bias. Dropping the survey or reexamining its contributions could improve the overall performance and reliability of the SS model.

Addressing these concerns and refining the SS model will lead to more accurate and reliable guidance for managing the Atlantic red drum fishery. Implementing these updates will ensure the model is aligned with the best available science and supports well-informed, sustainable management decisions.

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.*

The selection of reference points for the SS models and the TLA method was presented in detail. While the approach was mostly relevant and aligned with the methods used for these models and the Atlantic red drum fishery, there are certain areas where improvements could have been made.

8.1 Northern stock

The reference points selected for the SS models are well-recognized standards and are commonly used in fishery management. These reference points include SPR30% (Spawning Potential Ratio at 30%), F30% (Fishing Mortality Rate calculated at age 2 fish and is the level of F at SPR30%), and SSB30% (Represents the level of Spawning Stock Biomass associated with a stock fished at SPR30%) as thresholds, as well as SPR40%, F40%, and SSB40% as target reference points. These metrics serve as critical indicators for assessing the health and sustainability of the fishery. SPR30% and F30% thresholds are widely accepted and have been applied previously in the fishery management plan (FMP) of the Atlantic red drum to maintain sustainable stock levels by ensuring that fishing pressure does not exceed the fishery's capacity to regenerate.

The introduction of the SSB30% threshold and the SSB40% target, however, represents a newer development. These reference points have not yet been incorporated into the fishery management plan (FMP) but are an option as a more precautionary management strategy. The SSB30% threshold sets a limit below which stock biomass should not fall to avoid recruitment overfishing, while the SSB40% target is considered an aspirational goal, encouraging the rebuilding of stock biomass to a more sustainable level. The inclusion of these new reference points marks a shift toward more robust, science-based management frameworks aimed at ensuring the long-term sustainability of the Atlantic red drum fishery. However, these reference points have not yet been thoroughly evaluated and may need to be assessed using the simulation framework, especially given their relationship to steepness. As previously demonstrated, the steepness value used in the model was inconsistent. For these reference points to be effectively integrated into the management plan, it is essential to carefully consider their potential impact on the fishery and ensure they are tailored to the specific dynamics of the Atlantic red drum population. Additionally, it is important to note that the northern SS model cannot be reliably used to assess stock status due to poor diagnostics, which must be addressed before any accurate evaluation can be made.

Traffic Light Analysis (TLA) is a tool used in fisheries management to visually represent the status of fish stocks and related indicators by assigning different colors red, yellow, and green to key metrics. Green indicates that the stock or indicator is within healthy limits and that management targets are being met. Yellow signals caution, suggesting that the stock or indicator is nearing a point where management intervention may be necessary. Red signifies that the stock or indicator is in poor condition, requiring immediate management action to prevent further depletion or negative impacts.

For both the northern and southern stock the reference points used in TLA are thresholds calculated for indicators such as Spawning Stock Biomass (SSB), Fishing Mortality (F), and recruitment. These reference points are crucial as they enable managers to visually assess the fishery's status, monitor trends over time, and determine when corrective measures are needed. However, as noted in ToR 3c, the SSB threshold was reduced by 50% without clear justification for selecting this specific scalar. I believe additional simulation work is necessary to evaluate the appropriateness of this 50% reduction of this reference point. The Skate analysis did not present management reference points though there are opportunities to develop these if the method is used to provide quantitative catch advice.

8.2 Southern stock

The southern stock presented the same thresholds and targets and did not present an evaluation of the new SSB threshold and targets (see comment for the northern model). However, one of the graphs incorrectly referred to SSB30% as Spawning Stock Biomass at 30% virgin biomass, when it should represent SSB at F30%. These two quantities are only equivalent under the assumption of no stock-recruitment relationship, with a steepness value of 0.99, as specified in the base model. If, as we suggested, the steepness is adjusted to a value different from 0.99, this equivalence no longer holds, and the distinction between SSB at 30% of virgin biomass and SSB at F30% becomes significant.

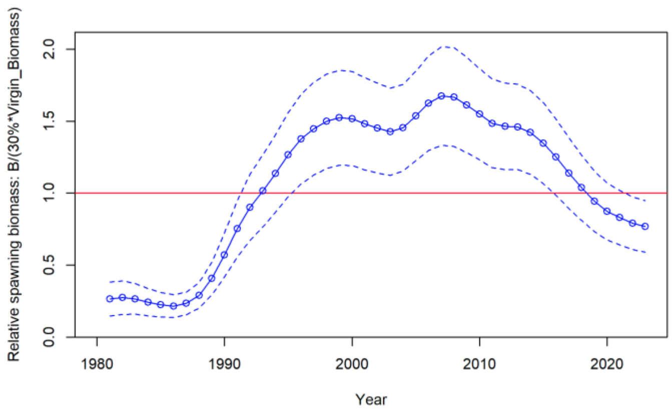

Figure 5: Relative spawning biomass over time.

Moreover, the reference points were calculated using a three-year window from 2020 to 2022. The reasoning for omitting 2023 was incomplete datasets for that year. However, Stock Synthesis (SS) can effectively handle missing data, and therefore, the reference point window should ideally cover 2021-2023 using the most recent data available. Additionally, instead of using an arithmetic average, a weighted average could be considered based on the reliability of each year's estimates, ensuring a more accurate reflection of the data's quality.

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 research recommendations made by the TC for the next benchmark assessment, and both short and long-term timeframes, seem appropriate. The prioritized list of improvements aligns with many of the needs identified throughout this review. While I support the overall recommendations, I suggest reconsidering the priority level of some research items and would recommend adding other essential topics to ensure a more comprehensive and reliable assessment in the future.

First, the SAS must prioritize addressing the issues with indices standardization. The request for showing diagnostics for standardization of abundance and CPUE indices revealed clear problems. Thus, it is crucial for the SAS to rerun the CPUE analyses, and that all indices of abundance are recalculated to ensure accuracy. In doing so, the SAS team must verify that the standardization process eliminates any residual patterns that could indicate underlying bias or errors in the model. This step is critical to ensure the robustness of the abundance indices, which form the foundation for subsequent stock assessments.

Second, the SAS team must ensure that all relevant "habitat" covariates that are believed to influence the species' biomass during the survey period are appropriately incorporated into the final index standardization model. These covariates could include environmental factors such as temperature, rainfall and other features that fluctuate over time and space and are relevant to the Atlantic red drum. Accurately accounting for these covariates is essential for deriving meaningful abundance indices that reflect the true distribution and availability of the species. Tools like the Vector Autoregressive Spatio-Temporal (VAST) framework (Thorson and

Barnett, 2017)[1] are particularly effective for identifying spatial and temporal variations that traditional models may overlook. By incorporating species distribution, abundance, and environmental factors, such frameworks enhance understanding of stock dynamics.

Last, the simulation framework should be used as a key tool for further exploring model configurations, especially considering the disparities in convergence between the southern and northern Stock Synthesis (SS) models. This would help distinguish relevant from irrelevant information for each model, ultimately improving the base SS model for a better fit. Through simulations, various scenarios can be tested to ensure that critical drivers of stock dynamics are accurately captured, reducing uncertainties, and enhancing the overall performance of the assessment models.

The investigation of reference points for red drum management should be prioritized as high because these points are essential for accurately identifying the stock status. Reference points serve as the benchmarks that guide management decisions, ensuring sustainable fishing practices. The review of Term of Reference (TOR) 8 revealed some shortcomings in how these points were handled, highlighting gaps that could impact the effectiveness of stock assessments and management strategies. Addressing these deficiencies would improve the reliability and precision of future red drum stock evaluations. Thus, instead of using some empirical values from the literature, the simulation framework can be used to tune in the Atlantic red drum specific reference points. This includes the use of SPR30%, 40% reference points for the SS models. Moreover, this is relevant for the definition of overfishing and overfished status for the TLA, especially the adult abundance threshold dampening coefficient.

The simulation framework could also be used to assess the informativeness of each dataset. For example, the model struggled to fit the longline survey abundance data, which was later deemed irrelevant for the southern model after additional sensitivity analysis during the review. More so, by increasing the age composition sample from the longline survey, potential improvements in bias could be determined. Similarly, the framework could assess whether using conditional age-at-length data would be more informative than marginal age composition data. This type of systematic analysis should be applied to all data sources using the simulation framework.

[1]Thorson, J.T., Barnett, L.A.K., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74, 1311–1321. https://doi.org/10.1093/icesjms/fsw193

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

The panel considered it appropriate to schedule the next Benchmark Assessment in five years. However, due to various issues identified in the Stock Synthesis assessment for the southern red drum stock, an update is recommended for 2025. This update should incorporate the most recent catch, biological, and abundance indices data, and include revisions based on the panel's recommendations, particularly regarding indices standardization. Regarding the recent news of a bias in MRIP-derived catches, if MRIP-derived catches show significant differences (around a 30% reduction), the southern assessment should be rerun within the inter-benchmark period. Additionally, the TLA for the northern stock should also be updated in 2025.

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

The advisory report summarizing the panel's evaluation of the stock assessment and addressing the terms of reference was collaboratively drafted by all panel members.

1.4 CONCLUSION AND RECOMMENDATIONS

My conclusions, and recommendations have been carefully integrated into the comprehensive discussions provided for each of the Terms of Reference (ToRs) in the sections above. These detailed evaluations address a wide range of aspects related to the stock assessment methodologies, data inputs, model parameterization, and the overall robustness of the analytical approaches. In this section, I will underscore the key takeaways and provide an overarching summary of the main points that emerged during the review process.

These highlights will focus on areas where the assessment either met or fell short of the expected standards, emphasizing where improvements are most urgently needed. Specifically, I will draw attention to issues such as the adequacy of model parameterization, the need for improved index standardization processes, and the importance of incorporating more rigorous methods to assess uncertainty. Additionally, I will revisit recommendations for refining the Stock Synthesis (SS) models and Traffic Light Analysis (TLA), and how these adjustments can significantly improve the accuracy and reliability of stock status estimates and management advice.

This section aims to provide a condensed overview of the critical insights gained throughout the evaluation and serve as a roadmap for future actions to enhance the reliability and effectiveness of the Atlantic red drum stock assessment process.

ToR.1: Response to Simulation Assessment Peer Review Panel Recommendations

SAS has effectively addressed the first recommendation from the Simulation Assessment Peer Review Panel by performing a revised grid search to optimize the Traffic Light Analysis (TLA) reference points. However, the second recommendation, to demonstrate that the southern Stock Synthesis (SS) estimation model can consistently produce unbiased estimates when fit to data without observation error, was not fully completed. The presented analysis was based on a single iteration of the operating model, which showed reduced relative error, but multiple iterations are needed to confirm the robustness of these estimates. Thus, I recommend that SAS to repeat the analysis with multiple iterations to ensure that the SS model provides unbiased estimates.

ToR.2: Data Collection and Presentation

The data collection and presentation of both fishery-dependent and fishery-independent data in the stock assessment were generally thorough and well-documented, with a clear emphasis on providing transparency and accuracy. The inclusion of variance measures, such as standard errors and confidence intervals, for key data sources was a positive aspect, although this was not consistently applied across all datasets. There were valid justifications for excluding certain data sources, particularly those with potential biases or representativeness issues, but in some cases, such as the scale-based aging data, these justifications could have been strengthened. Moreover, while the assessment effectively handled various data complexities, some inconsistencies in survey design and the standardization of abundance indices remain areas of concern. The panel's request for further analysis on excluded data and recommendations for addressing survey standardization highlight the need for greater transparency and methodological consistency.

Thus, I recommend the following steps:

1. Improve variance presentation across all data sources: While the assessment included variance measures for many data sources, this should be consistently applied across all datasets. Including standard errors or confidence intervals for every data source will ensure a clearer understanding of the uncertainties involved and improve the robustness of the assessment.
2. Reassess exclusion of scale-based age data: The exclusion of scale-based age data for older fish due to perceived biases was justified but incorporating this data source for ages 1-3 could enhance the assessment of recruitment. A more thorough investigation should be conducted to determine the potential value of this data source for younger age classes.
3. Standardize survey design and index calculation: The inconsistencies identified in the survey design, particularly the lack of consistent sampling in time and space, need to be addressed. Moving forward, the SAS should adhere to best practices in survey design and standardization to ensure more reliable abundance indices. Tools like the DHARMa package and VAST can be employed to analyze and better account for these biases and inconsistencies.
4. Systematically reevaluate inclusion of indices: Some indices, such as the contemporary longline index, provided limited informational value and should be reevaluated for their inclusion in the models. A simulation framework should be used to test these decisions more rigorously to ensure that only the most reliable indices are included in future assessments.
5. Address biases in fishery-dependent data: For indices derived from data sources such as MRIP CPUE, which showed potential hyperstability, further scrutiny is necessary. Applying a more rigorous standardization process could reduce these biases and improve the overall quality of the assessment.

ToR.3: Evaluation of Methods and Model used for estimating Population Parameters

The methods and models used to estimate population parameters and reference points were largely appropriate, particularly in the case of the southern stock. The analytical team successfully implemented multiple approaches, including the Traffic Light Analysis (TLA), Skate model, and Stock Synthesis (SS) models, in combination with a simulation model. However, several challenges and inconsistencies were identified, where key assumptions (e.g., steepness fixed at 0.99) were not fully justified.

The SS model for the southern stock showed robustness in its parameterization and use of available data, but additional scrutiny and refinements, particularly in the treatment of abundance indices and assumptions around the stock-recruitment relationship, are necessary. The alternative tagging model, Cormack-Jolly-Seber (CJS), appears suitable, but the exclusion of tagging data and key biases (e.g., post-release mortality) remains an unresolved issue, affecting the overall reliability of the tagging model outputs.

Furthermore, the Skate and TLA analyses exhibited sensitivity to reference periods and threshold choices, which were not fully justified, raising concerns about the robustness of the outputs.

Following are my recommendations:

1. Revise steepness assumptions: In both the northern and southern stock models, the assumption of steepness at 0.99 should be revisited. A sensitivity analysis using biologically plausible values (e.g., 0.84) should be conducted to provide more accurate estimates of stock dynamics and ensure that the stock-recruitment relationship is appropriately modeled.
2. Improve tagging model and data inclusion: The SAS should incorporate all available tagging data into the CJS analysis and address key biases such as tag shedding, post-release mortality, and age-dependent biases. Implementing multi-state models or incorporating covariates could significantly enhance the accuracy of survival and migration estimates, improving the reliability of the tagging study outputs.
3. Address data-weighting and likelihood schemes: For the SS models, particularly the northern stock, the SAS should explore additional data-weighting schemes, such as the Dirichlet-Multinomial approach, to better handle overdispersion in age and length composition data. Sensitivity to likelihood weighting was noted in the southern stock, and addressing this will ensure a more balanced model fit across all data sources.
4. Review reference periods and thresholds in TLA: The choice of reference periods and thresholds in the TLA method needs to be more succinctly justified. The current periods are overly broad and risk skewing baseline assessments. Shorter, well-justified periods should be selected, and the 50% reduction in the adult abundance threshold should be reconsidered with further simulation work to evaluate the impact of this arbitrary scalar.
5. Skate model adjustments: The Skate model's ratcheting tendency should be addressed by implementing more flexible management frameworks that allow for dynamic reference points and regular reassessments of precautionary measures. This will ensure that the model remains responsive to stock improvements and does not lead to overly conservative management that may hinder resource use.
6. Conduct stepwise diagnostics for northern model: The northern SS model, despite its challenges, has potential. A stepwise diagnostic evaluation, including re-running the model with additional years of data and more appropriate steepness values, is recommended to identify the underlying issues and improve model performance.
7. Further simulation-based evaluations: The simulation framework should be used more extensively to test the robustness of different model assumptions and parameterizations. This approach will help clarify which configurations best reflect the actual dynamics of the red drum stocks and improve the overall performance of the assessment models.

ToR.4: Evaluation of Model Diagnostics

The diagnostic analyses conducted across the different models (Skate, TLA, Tagging, and Stock Synthesis) varied significantly in depth and thoroughness. While the Stock Synthesis (SS) model was evaluated using recognized best practices, with detailed sensitivity analyses, convergence diagnostics, and retrospective analyses, the diagnostic work for the Skate, TLA, and Tagging models was notably limited. These omissions hinder a complete understanding of model reliability and uncertainty, particularly for the northern stock. For instance, key diagnostic tools like sensitivity analysis were either minimally applied or absent in these models, particularly in the case of the Tagging model, which lacked convergence checks or performance evaluation.

For the Skate and TLA models, the failure to explore important sensitivities, such as the effect of reference periods and the 50% threshold scalar for adult abundance, raises concerns about the robustness of these models. Additionally, retrospective analyses, especially for the TLA model, were not performed, limiting the ability to assess long-term consistency in stock assessment predictions.

The Stock Synthesis (SS) model, in contrast, demonstrated a more thorough approach, with comprehensive diagnostics that supported the model's reliability, though certain improvements, such as better residual analysis and additional sensitivity tests, are still recommended. The retrospective analysis for the SS model identified minor discrepancies in the 2019 index values, which should be investigated further to prevent misleading conclusions in future assessments.

My recommendations for this ToR are as followed:

1. Expand diagnostic analyses for Skate, TLA, and Tagging models: The diagnostic work for these models should be significantly extended. Sensitivity analyses must be performed on key assumptions, including reference periods, input data, and threshold scalars. For the TLA model, a retrospective analysis should be conducted to compare past management decisions with observed trends and evaluate model biases over time.
2. Improvement in residual and sensitivity analyses for SS model: The SS model is robust, but improvements can be made in evaluating residuals and sensitivity to various assumptions, particularly the standardization of indices. The inclusion of additional diagnostics, such as hindcasting cross-validation, would strengthen the model's capacity to predict future trends.
3. Scrutinize input data and index standardization: For all models, concerns about data input, particularly in the abundance indices, must be addressed. Misspecifications in the standardization process can lead to problematic residual patterns and skew results. A comprehensive evaluation of index standardization, including improved diagnostic tools like QQ plots, should be prioritized.
4. Conduct sensitivity analysis on the 50% scalar used in TLA: The arbitrary selection of a 50% threshold for adult abundance in the TLA method warrants additional analysis. The SAS should use a simulation-based approach to determine if this scalar is appropriate or if an alternative value provides a better fit for stock dynamics.
5. Incorporate bridge runs and historical retrospective analyses: To provide continuity between assessments, the inclusion of bridge runs from previous models is recommended. This will allow for a comparison between past and current assessments and ensure consistency in model predictions over time.

ToR.5: Evaluation of Methods to characterize uncertainty in estimated parameters

The assessment methods used to characterize uncertainty in estimated parameters across the various models (Skate, Tagging, TLA, and Stock Synthesis) reveal significant gaps, particularly in the less comprehensive models. The uncertainty in the Skate and Tagging models was not fully explored, with limited diagnostic evaluations and missing key methodologies like weighted moving averages and bootstrap methods. These omissions reduce confidence in the precision of parameter estimates such as fishing mortality and survival rates. The TLA model did better in addressing uncertainty through simulations, but it still faces challenges due to potential biases introduced by the long reference period. The Stock Synthesis (SS) model showed a more robust approach to uncertainty, with multiple sensitivity analyses and likelihood profiles, though the absence of a steepness sensitivity analysis leaves room for improvement.

The incomplete handling of uncertainty in these models limits the confidence in technical conclusions and management recommendations, potentially leading to unreliable stock status assessments. Accurate quantification of uncertainty is critical, as it directly influences the reliability of key outputs, such as stock biomass and exploitation rates, and ultimately the effectiveness of the management strategies informed by these models.

Below are my final recommendations in regard to ToR 5 for each model or methods.

1. Skate Model:
   Implement weighted moving averages: To better characterize uncertainty, apply a weighted moving average where the weights are based on the inverse variance of the estimated index-to-catch ratios. This would provide a more precise measure of variability and improve the reliability of the Skate model's output.
   Expand sensitivity analyses: Conduct systematic sensitivity analyses on key assumptions, particularly regarding the choice of reference period and the calculation of fishing mortality. This will help identify the impact of different assumptions on model stability and stock status estimates.
2. Tagging Model (Cormack-Jolly-Seber):
   Use additional diagnostics methods: Either use bootstrap methods to replace or complement asymptotic standard errors to generate empirical distributions of parameter estimates. This approach will provide more robust estimates of standard errors, bias, and confidence intervals, especially given the complexity of the data. Or, explore profile likelihood confidence intervals: Apply profile likelihood methods for survival and recapture probabilities, offering a more rigorous way to assess uncertainty than asymptotic estimates.
3. Traffic Light Analysis (TLA):
   Shorten the reference period: per description in previous ToR recommendations.
4. Stock Synthesis (SS) Model:
   Perform steepness sensitivity analysis: per explanation in previous ToR recommendations.
   Expand sensitivity testing on data inputs: Further sensitivity testing on data inputs (e.g., catch histories, natural mortality) is recommended, particularly on key datasets like the Florida haul index and sub-adult lengths. This will help assess the model's robustness to various assumptions and highlight any potential biases.

ToR.6: Minority report

There was no minority report.

ToR.7: Recommend best estimates of stock biomass, abundance and exploitation.

For the northern stock of Atlantic red drum, the Stock Synthesis (SS) model was deemed unsuitable by both the SAS and the panel to provide reliable estimates of key population metrics such as stock biomass, abundance, and exploitation rates. Despite its current limitations, the model shows promise, particularly since the SS model has been accepted for the southern stock, with which it shares a similar configuration. However, the northern SS model still requires refinement, particularly in parameterization and diagnostic testing. An additional five years of data is expected to improve the model's indices and help address the issue of non-convergence. In the interim, the Traffic Light Analysis (TLA) has been recommended by the SAS as a qualitative indicator for northern stock status, although it also has limitations. However, concerns remain around the reliability of adult abundance data, unexplained anomalies (e.g., the 2011 year class), and sensitivity to the choice of the reference period. The TLA results could be enhanced by adopting a weight-of-evidence approach, integrating insights from other analyses such as the Skate and SS models. However, none of these models, in their current form, are robust enough to be used independently for formal stock assessments.

For the southern stock, the SS model has been recommended as the primary tool to estimate stock biomass, abundance, and exploitation rates. The base model is considered reliable, with sensitivity analyses indicating that it is relatively conservative compared to other approaches like TLA and Skate models. Despite this, improvements are needed in input data quality, index standardization, and specific sensitivity analyses to ensure the model's accuracy. Addressing these issues will make the SS model a more effective tool for guiding the management of the southern Atlantic red drum stock.

I would recommend as:

1. Primary model: Use the Stock Synthesis (SS) model for the southern stock and, once converging with additional years in the times series, for the northern stock.
2. Supplementary model: Use Traffic Light Analysis (TLA) and Skate models as supplementary tools, but not as the primary methods for stock assessment.
3. Tagging model: The Cormack-Jolly-Seber (CJS) model can be used with improvements, particularly in uncertainty characterization and diagnostics.

<u>ToR.8: Evaluate the choice of reference points and the methods used to estimate them</u>

The reference points selected for the Stock Synthesis (SS) models for both the northern and southern stocks are well-established metrics commonly used in fisheries management, such as SPR30% and F30%. These are crucial in determining the sustainability of the fishery. However, the introduction of newer reference points, such as SSB30% and SSB40%, while promising, needs further evaluation through a simulation framework, especially given concerns about the inconsistency of the steepness value used in the model. This is critical to ensure these reference points are tailored to the Atlantic red drum's biological and population dynamics.

For the Traffic Light Analysis (TLA), the reference points allow managers to assess stock status visually and take necessary actions. However, the 50% reduction of the SSB threshold used in the TLA method was applied without sufficient justification, warranting additional simulation work to evaluate its appropriateness.

In the southern stock**,** the reference points mirror those for the northern stock but were misrepresented in certain analyses. The equivalence of SSB30% and SSB at F30% is valid only under specific assumptions of steepness, which were not properly addressed. Moreover, the use of a three-year window (2020-2022) for calculating reference points omitted data from year 2023, which could have been incorporated, as SS can handle missing data. Additionally, a weighted average should be considered for calculating reference points to account for the varying reliability of the data.

I would recommend exploring the following steps:

1. Establish a simulation framework for reference points:

Conduct simulation work to evaluate the appropriateness of the newer reference points (SSB30%, SSB40%), especially considering their relationship to steepness. This will ensure that these reference points are robust and well-suited to the specific dynamics of the Atlantic red drum fishery.

2. Evaluate the 50% scalar for TLA SSB threshold:
   Further simulation analysis should be conducted to justify or revise the 50% scalar reduction in the SSB threshold used in the TLA method. Without this, the current threshold may not accurately reflect the stock's status.
3. Improved calculation of reference points for southern stock:
   Use the most recent data (2021-2023) to calculate reference points, as SS can handle incomplete datasets. A weighted average should also be applied instead of an arithmetic average, allowing for more reliable assessments that reflect the uncertainty in each year's data.
4. Clarify and address misrepresentations:
   Ensure that reference points, such as SSB30% and SSB at F30%, are clearly differentiated and correctly applied in all analyses, where a misrepresentation was identified.

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

The recommendations made by the Technical Committee (TC) for the next benchmark assessment are generally appropriate and aligned with the needs highlighted throughout this review. However, some research priorities should be re-evaluated, and additional topics should be considered to ensure a more comprehensive and reliable stock assessment for Atlantic red drum. Key areas requiring attention include the standardization of indices, the incorporation of relevant environmental covariates, and the use of a simulation framework to refine model configurations and address issues in reference point estimation.

I would recommend:

1. Prioritize index standardization:
   Recalculate all abundance and CPUE indices to ensure proper standardization and remove residual patterns indicating bias. This should be a top priority to provide accurate and reliable inputs for stock assessments.
2. Incorporate Environmental Covariates:
   Include habitat-related covariates (e.g., temperature, rainfall) in the final index standardization models. This will allow for a more accurate reflection of environmental influences on stock biomass and abundance. Utilize tools like the Vector Autoregressive Spatio-Temporal (VAST) framework to identify spatio-temporal variations that may affect fish distribution and abundance.
3. Use Simulation Framework for Model Refinement:
   The simulation framework should be employed to resolve model convergence issues, particularly between the northern and southern SS models. This will improve the fit of the base models and ensure that the most relevant drivers of stock dynamics are identified and addressed. It should also be used to fine-tune reference points and evaluate dataset informativeness.

ToR.10: Timeframe for future assessments

The panel supports the recommendation for the next benchmark assessment in five years, with an interim update for the southern stock in 2025. If significant changes in MRIP-derived catches are observed, the assessment should be rerun within the inter-benchmark period. However, I would add to this recommendation to proceed with the scheduled updates in 2025, incorporating new data and revising the indices and models based on the panel's recommendations.

ToR.11: Prepare a peer review panel ToR and summary advisory report

The advisory report summarizing the panel's evaluation of the stock assessment and addressing the terms of reference was collaboratively drafted by all panel members and should be delivered given the allocated time.

By addressing these ToR recommendations, future stock assessments for the Atlantic red drum will become more accurate, reliable, and better suited to inform sustainable management decisions.

2. APPENDICES

2.1 APPENDIX 1: LIST OF REVIEW WORKSHOP DOCUMENTS

Document #	Title	Authors
Reference Documents		
SEDAR93-RD01	Red Drum Simulation Assessment and Peer Review Report	Atlantic States Marine Fisheries Commission
SEDAR93-ASMFC	Red drum stock assessment report for review	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

2.2 APPENDIX 2: 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[1].

Scope

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

---

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

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.

   a. 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.

   b. 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.

   c. Reviewers should elaborate on any points raised in the summary report they believe might require further clarification.

   d. Reviewers shall provide a critique of the NOAA Fisheries review process, including suggestions for improvements of both process and products.

   e. 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

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

These were two main recommendations from the Simulation Assessment Peer Review Panel:

   a. Performed a revised grid search for deriving reference points for the Traffic Light Analysis (TLA) that only includes data that would be available to a TLA when it is applied in practice.
   b. Continue the work showcasing that the southern Stock Synthesis (SS) estimation model (EM) could produce unbiased estimates when fit to data with no observation error.

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 selectivity, 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.

Annex 3: Tentative Agenda

TENTATATIVE 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.

2.3 APPENDIX 3: REVIEW PANEL

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